Modeling multidimensional spectral lineshapes from first principles: Application to water-solvated adenine

Journal Name

ARTICLE TYPE

Modeling Multidimensional Spectral Lineshapes from First Principles: Application to Water-Solvated Adenine†

Javier Segarra-Martí,*[a‡] Francesco Segatta,*[b‡] Tristan A. Mackenzie,[a] Artur Nenov,*[b] Ivan Rivalta,[b,c] Michael J. Bearpark[a] and Marco Garavelli[b]

Received Date
Accepted Date

DOI: 00.0000/xxxxxxxxxx

In this discussion we present a methodology to describe spectral lineshape from first principles, providing insight into the solvent-solute molecular interactions in terms of static and dynamic disorder and how these shape the signals recorded experimentally in linear and nonlinear optical spectroscopies, including two-dimensional electronic spectroscopy (2DES). Two different strategies for simulating the lineshape are compared: ~~in which both the coupling to the intra-molecular vibrations and the influence exerted by the different water distributions attained along a molecular dynamics (MD) simulation are accurately described.~~both rely on the same evaluation of the coupling between the electronic states and the intra-molecular vibrations, while they differ in describing the influence exerted by the diverse water configurations attained along a molecular dynamics (MD) simulation. The first method accounts for such water arrangements as first order perturbations on the adenine energies computed for a single reference (gas phase) quantum calculation. The second method requires computation of the manifold of excited states explicitly at each simulation snapshot, employing a hybrid quantum mechanics/molecular mechanics (QM/MM) scheme. Both approaches are applied to a large number of states of the adenine singlet excited manifold (chosen because of its biological role), and compared with available experimental data. They give comparable results but the first approach is two orders of magnitude faster. We show how the various contributions (static/dynamic disorder, intra-/inter-molecular interactions) sum up to build the total broadening observed in experiments.

1 Introduction

Spectral lineshape is a fundamental characteristic of any electronic spectrum.[1] Its structure and dynamical evolution contains information about the electronic structure of a system, vibrational and electronic dynamics, coupling strength to the environment and solute-solvent correlation time scales. Different levels of sophistication can be applied to the theoretical description of broadening effects:[2–10] from purely phenomenological models to more elaborate methodologies, capable of accounting for the various contributions (intra-molecular vibrations, fluctuations of the system's environment, etc.) that add together to produce the observed spectral shape. From being able to realistically model lineshapes we expect to gain a better understanding of the ~~underlying physical effects, which may for example differ for different electronic excited states.~~ physical origin of their many different underlying contributions.

The most common strategy employed for simulating spectral lineshapes from first principles involves performing ground state classical molecular dynamics (MD) simulations and subsequently extracting the autocorrelation of the excitation energy fluctuations, usually computed at the time-dependent density functional theory (TD-DFT) level.[17] The Fourier transform of the autocorrelation function, known as spectral density, is the key feature in describing the coupling of the electronic states of the system to the classical bath, accounting for both intra- and intermolecular degrees of freedom. This approach suffers from the fact that there is no guarantee that the MM potential faithfully re-

sembles the QM ground-state surface, therefore affecting the accuracy of the high frequency region of the spectral density. Coker et al. have suggested to decompose the spectral density into intra-molecular (high-frequency) vibrational contributions and an inter-molecular (low-frequency) environment component.[18] They have shown that with a quantum mechanical description of the intra-molecular vibrations and a classical treatment of collective solute-solvent motions, which accounts for both dynamic and static disorder contributions, a reliable representation of the spectral density is obtained. This strategy has been recently successfully applied to simulate linear absorption (LA) in light-harvesting complexes.[19]

The above strategy can be extended to the simulation of transient nonlinear spectroscopies, such as transient absorption (TA) and two-dimensional electronic spectroscopy (2DES), which allows the study of coherent excited state dynamics. ~~Concretely, the strategies devised by Coker et al. suffice to simulate the ground state bleach (GSB) and stimulated emission (SE) signals of the nonlinear spectra.~~ Recently, some of us have demonstrated through state-of-the-art electronic structure computations that the incorporation of absorption signals from higher lying states, the so-called excited state absorptions (ESA), are of paramount importance for the proper simulation of transient spectra:[3,20–24] ESA signals may overlap with GSB and SE, making the analysis of the recorded spectra convoluted. We have further demonstrated that multi-configurational wavefunction techniques are capable of resolving the high-frequency part of the spectral density of higher lying states.[2,3,25,26] We have also looked at the effect of static disorder on the spectral lineshapes of 2DES.[5,27]

In this article we move beyond previous approaches by introducing the effects of both static and dynamic disorder to spectral broadening by means of high-level multireference quantum chemistry calculations of the system's full manifold of states, accessing the lineshape broadening of a larger number of states and enabling its applicability for the simulation of (multi-dimensional) transient spectroscopy. Intra-molecular contributions are taken into account. In order to describe both static and dynamical broadening effects of the environment surrounding the molecule, two different models are formulated:

- **Perturbative model**: first order perturbation theory is applied to capture the effect of the Coulomb interaction between the gas phase densities of the solute and different solvent arrangements extracted from a MD simulation, requiring a single electronic structure evaluation (i.e. a reference) in absence of the solvent and computing only the solute-solvent interaction term along the MD trajectory;

- **Brute force model**: a full QM/MM approach is employed, whereby an explicit electronic structure computation (thus considering exchange interactions and orbital relaxation) is performed at each snapshot extracted from the MD.

Both protocols are benchmarked against one another and against available experimental data through the simulation of LA, TA and 2DES in adenine, one of the DNA/RNA nucleobases.[31,32] Non-reactive ultrafast deactivations in DNA/RNA nucleobases and their derivatives are crucial in order to better understand to which extent they are linked to DNA's overall photostability.[33,34] A wide range of efforts have been carried out in the literature to study these intricate events. DNA nucleobases are traditionally among the first systems to be measured in novel experimental set-ups due to their unique photophysical properties. This has been recently witnessed in the the fields of X-ray[35,36] assisted by free electron lasers (XFELs)[37] and in vacuum-UV[38] time-resolved spectroscopies. This is also the case for non-linear optical spectroscopic methods, the main workhorse for studying ultrafast processes, which range from the seminal work of Kohler, Markovitsi and others employing pump-probe[39,40] and fluorescence up-conversion[41,42] set-ups to the recent 2DES measurements of Miller and co-workers.[43]

In the following, we first outline a general protocol for incorporating static and dynamic disorder into spectral lineshapes in linear and nonlinear spectroscopy simulations for an arbitrary number of electronic excited states. Then we provide the technical details regarding the implementation of the protocol to adenine. Finally, we assess the performance of the perturbation and brute force methods and the overall agreement of the simulated lineshapes with the experiment.

2 Theory

2.1 Origin of the spectral line width

The lineshape of experimentally recorded spectra contains contributions of different origin. The ability to disentangle them and correctly compute their effect on the overall broadening allows one to both faithfully reproduce the experiment and to gain unique information about the intra-molecular dynamics of the system and its interaction with the environment.

Figure 1 shows a schematic picture of the different broadening sources and of their impact on both linear and non-linear spectra. The stick spectrum of Figure 1(a) is informative about the vertical transition energy of electronic transitions (in the order of tens of thousand cm^{-1}) and the probability of such a transition to occur (peak position along the x axis and peak height along the y axis, respectively). The description is improved by considering the coupling of the electronic structure to (intra-molecular) nuclear degrees of freedom, which allows for harmonic fluctuations of a few thousand cm^{-1} (Figure 1(b)), which dress the electronic transition with its underlying vibrational structure. Three other sources of broadening are considered next, namely the state lifetime, the homogeneous and the inhomogeneous broadening (Figure 1(c)-(e)):

- The lifetime broadening comes from the finite lifetime of the considered excited state, that once populated can return to the ground state or transfer its energy to another excited state by either radiative (long time scale) or non-radiative (short time scale) decays. Excited state dynamics not only affect spectra by controlling the appearance and disappearance of peaks in time resolved techniques (as e.g., the 2DES), but it also gives a symmetric (Lorenzian) contribution to the lineshape, with a significant impact on the spectra only for ultrashort state lifetimes (i.e., in this spectral window, in the sub 100 fs timescale).

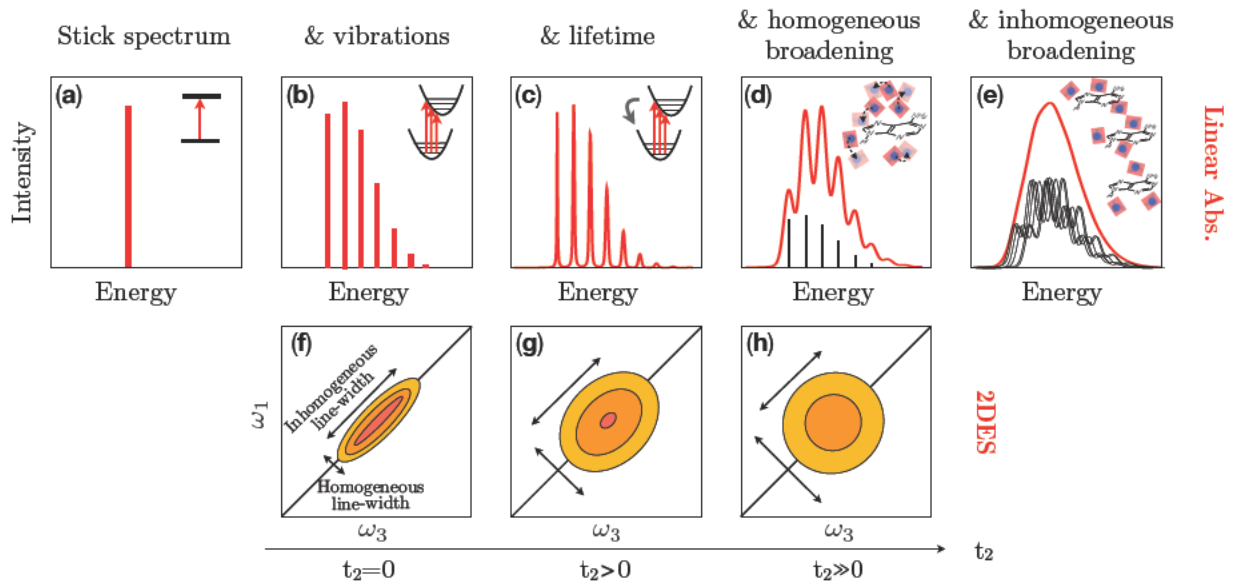

Fig. 1 Schematic representation of the different contributions that add up to produce the experimentally observed spectral lineshape. (a) Stick spectrum. (b) Stick spectrum plus vibrations. (c) Addition of the lifetime broadening, (d) homogeneous broadening and (e) inhomogeneous broadening. The insets of each spectrum pictorially represent the physical origin of the various contributions. (f) shows that 2DES is able to disentangle static and dynamic disorder,[28–30] which instead appear superimposed in the linear absorption. (g-h) show the evolution of the lineshape due to spectral diffusion (i.e. erasure of the system memory due to interaction with the solvent). The time scale at which the spectral diffusion occurs gives information about the time scale of the system-bath interaction (and is encoded in the homogeneous broadening term).

- The homogeneous broadening (dynamic disorder) originates from the dynamic interaction of the excited molecule with its moving environment, which produces low frequency fluctuations ranging from hundred to thousand cm^{-1} in the energy gap over time. The time scale of these fluctuations is shorter than or comparable to the state lifetime, so that the explicit time-dependence of such a energy gap modulation needs to be considered. This is captured by the energy gap fluctuation autocorrelation function, whose Fourier transform is the so called *spectral density* function $J(\omega)$,[44] the key quantity to describe the frequency dependent coupling ~~to both the intra-molecular vibrations and this additional source of spectral broadening~~ between the electronic states and both the intra-molecular vibrations and the moving solvent. Dynamic disorder is also responsible for spectral diffusion in 2DES (see Figure 1(f-h)), by controlling the time scale on which the system looses memory of its initial excitation wavelength.

- The inhomogeneous broadening (static disorder) accounts for the different excitation energy of each of the large number of molecules contained in the probed samples, due, at a given instant of time, to the different instantaneous solvent configuration surrounding them. Such different local environments (locally) tune the absorption energy, producing a distribution of possible transitions, as opposed to a single sharp transition energy. The ratio between homogeneous and inhomogeneous broadening can be extracted from 2DES experiments at very early times (Figure 1(f)). We note that energy gap differences due to intra-molecular distortions of different molecules in the sample, do not contribute to static disorder. These contributions are in fact already accounted for in the explicit calculation of the vibronic coupling between the electronic state and the intra-molecular modes.

All these different contributions (intra-molecular vibrations, state lifetime, static and dynamic disorder) are usually accounted for by phenomenological broadening models, bringing the focus of the computational chemist towards the accurate evaluation of transition energies and probabilities.[16,45–47] Phenomenological models range from very simple approaches, which add Gaussian or Lorentzian linewidths to the stick transitions, to more sophisticated ones that use model spectral densities to capture the solute-solvent interaction like the widely employed Overdamped Brownian Oscillator (OBO).[2–5] These are usually fitted to obtain linewidths comparable to linear absorption experiments, which however make it impossible to separately determine both static and dynamic disorder contributions.

Several methods exist for obtaining the coupling of the electronic structure to the intra-molecular modes. These range from classical (Boltzmann) and quantum (Wigner) sampling,[6,7] time-dependent and time-independent quantum-mechanical formulations,[8–10] with varying degree of sophistication. In a nutshell, sampling techniques do not require any preliminary knowledge regarding the potential energy surface (PES) of the electronic states involved in the transition intrinsically, unlike quantum-mechanical formulations which rely on Taylor expansion of the PES at key points in coordinate space. Despite this complication, quantum approaches have been shown to reproduce the vibrational structure of the absorption spectra with remarkable accuracy,[9] whereas sampling techniques correctly reproduce the

overall broadening of the absorption bands,[10,14–16] but are fundamentally limited by the impossibility to resolve vibronic structures.

In this work we not only compute *ab initio* energies and intensities of the transitions of interest, but also intra-molecular vibronic contributions as well as homogeneous and inhomogeneous broadening from first principles, thus leaving the lifetime broadening as the only contribution that was not explicitly computed: although it is an important component towards the symmetric contribution of the spectral broadening, its computation from first principles is costly and heavily dependent on the electronic structure method used for its determination. In order to avoid including additional complexity we preferred to use the lifetimes recorded experimentally in the literature associated to its bright L_a state, making use of the available experimental lifetimes of Kwok et al. estimated at ∼130 fs.[48] We checked a-posteriori that the addition of such lifetime contribution does not affect linear absorption lineshapes, still remaining important to describe appearance and disappearance of 2DES signals.

The asymmetry of the lineshape, as well as the static to dynamic disorder ratio, are generally measurable in different spectroscopic techniques, and can therefore be compared with their theoretical counterpart. In what follows we introduce the computational strategies outlined to account for each of the broadening contributions shown in Figure 1 (but for the state lifetime), namely: coupling to intra-molecular vibrations, coupling to dynamical solvent/environment fluctuations and static disorder.

2.2 Simulation of linear and non-linear spectra

The quantities that enter in the simulation of spectra are: energy levels (ω_{ga}, where the subscript g denotes the GS and a run over the manifold of excited states), transition probabilities (transition dipole moments (TDMs), μ_{ga}), static disorder, and a number of spectral densities $J(\omega)$, containing both intra-molecular vibrations and the dynamic effect of environment fluctuations (i.e. the homogeneous broadening). As detailed below, intra-molecular vibrations contribute to the high-frequency window of the spectral density, $J^H(\omega)$, while the dynamic effect of environment fluctuations (i.e. the homogeneous broadening) enters in the spectral density low-frequency window $J^L(\omega)$. The total spectral density is then obtained as $J(\omega) = J^H(\omega) + J^L(\omega)$.

2.2.1 Linear spectra

The theoretical expression of the linear absorption spectrum, that connects all these quantities in a single equation, is given by:

$$A(\omega) = \frac{\omega}{N_{shots}} \sum_{j=1}^{N_{shots}} |\mu_{ga}^2| \mathscr{R} \int_0^{\infty} \exp\left[i(\omega - \omega_{ga}^j) - g_{aa}(t) - t/\tau\right] dt \quad (1)$$

The static disorder is included by adding together a large number N_{shots} of independent spectra, each of which has transition energies sampled from a given distribution (usually a Gaussian centered at the average transition energy $\bar{\varepsilon}_a^*$, and with standard deviation σ_a, see Section 2.3).

The $g_{aa}(t)$ term, is the so called lineshape function for state a, which, by employing the cumulant expansion of Gaussian fluctuations (CGF) approach,[4] can be evaluated from the spectral densities as:

$$g_{aa}(t) = -\frac{1}{2\pi} \int_0^{+\infty} \frac{J_{aa}(\omega)}{\omega^2} \left[\coth\left(\frac{\beta\omega}{2}\right) (\cos(\omega t) - 1) - i(\sin(\omega t) - \omega t) \right] \quad (2)$$

J_{aa} is the total spectral density of state a, i.e. the sum of intra-molecular contributions $J_{aa}^H(\omega)$ and dynamic solvent rearrangement contributions $J_{aa}^L(\omega)$. The lifetime τ determines the exponential decay of the integrated function, which translates in a symmetric broadening contribution in $A(\omega)$.

2.2.2 Non linear spectra

In non-linear third-order techniques, contributions of different physical origin are subdivided in three groups: ground state bleaching (GSB), stimulated emission (SE) and excited state absorption (ESA), which add up to build the actual measured signal. The evaluation of these different contributions involves the computation of the four-point lineshape functions $\phi_{dcba}(\tau_4, \tau_3, \tau_2, \tau_1)$. These rule the shape of the various contributions, and can be expressed in terms of the $g_{ab}(t)$ functions of equation 2, as:[4]

$$\begin{aligned} \phi_{dcba}(\tau_4, \tau_3, \tau_2, \tau_1) = & -g_{cc}(\tau_{43}) - g_{bb}(\tau_{32}) - g_{aa}(\tau_{21}) - g_{cb}(\tau_{42}) + \\ & + g_{cb}(\tau_{43}) + g_{cb}(\tau_{32}) - g_{ca}(\tau_{41}) + g_{ca}(\tau_{42}) + g_{ca}(\tau_{31}) - \\ & - g_{ca}(\tau_{32}) - g_{ba}(\tau_{31}) + g_{ba}(\tau_{32}) + g_{ba}(\tau_{21}) \end{aligned} \quad (3)$$

Here a, b, c and d represent state indices, and $\tau_1, \ldots \tau_4$, with $\tau_{12} = \tau_1 - \tau_2$, are different time values, related to the instants at which the different pulses of the non-linear techniques interact with the sample.

The computation of the $g_{ab}(t)$ lineshape functions (with $a \neq b$) requires $J_{ab}(\omega)$ and thus the evaluation of energy gap cross-correlation functions $C_{ab}(t)$, which we explicitly compute. In the results section we show that considering or neglecting these cross terms does significantly change the spectral lineshape.

2.3 Modelling of signal broadening: how to compute the different contributions from first principles

1. **Intra-molecular vibrations.** Following previous work,[49] we describe the coupling between the electronic states and the intra-molecular vibrations with a different strategy with respect to the coupling between the electronic states and their fluctuating environment. Both terms are accounted for in spectral densities functions.

   Intra-molecular contributions were described in the framework of the displaced harmonic oscillator (DHO) model, which assumes ground and excited states PESs to have the same curvature along the same coordinates (i.e. the normal mode coordinates) but a shifted equilibrium position (see inset of Figure 1(b)). This is an assumption that has been shown to be reasonable for a number of related systems in the past and that we expect to be valid here as well.[2,3] From a computational point of view, all parameters of the model can be readily computed at the initial geometry: these are normal modes and frequencies (obtained via a frequency computation at the ground state minimum) and excited state PES energy gradient

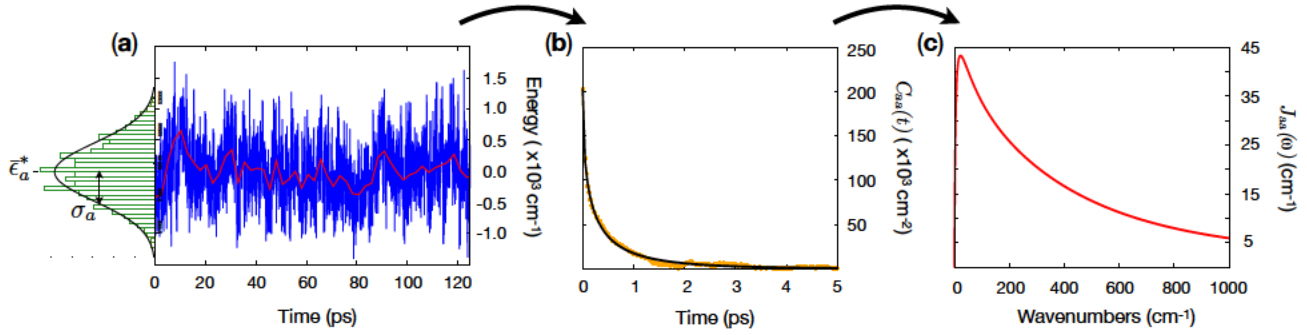

Fig. 2 From the fluctuating energy gap (a), through the gap fluctuation autocorrelation function $C_{aa}(t)$ (b) and the static disorder energy distribution (on the side of (a)), to the spectral density $J_{aa}(\omega)$ (c). $C_{aa}(t)$ is fitted by a stretched exponential (black line in (b)), while the static disorder is fitted by a Gaussian function ($\bar{\epsilon}_a^*$, σ_a).

at the Franck Condon point, which furnishes the displacement between the two PESs minima. This information is (necessary and) sufficient to build the high frequency sector of the spectral density, responsible for the vibronic peaks in the spectra, which can be expressed as:

$$J_{ab}^{H}(\omega) = \pi\omega \sum_{k=1}^{M} \frac{d_{ak} d_{bk}}{2\hbar} \omega_k^2 \delta(\omega - \omega_k) \qquad (4)$$

where the index k runs over the M molecular normal modes, ω_k are the normal mode frequencies, while d_{ak} and d_{bk} are the computed displacements of, respectively, a and b excited state potential wells with respect to the GS equilibrium position. The evaluation of the intra-molecular contributions should in principle be repeated at different environment configurations, to capture the environment influence on frequencies and excited state gradients. This is important especially for structured environments (as e.g. the protein scaffold of light harvesting complexes), where the system may remain staked for some time in different local minima. As we do not expect different solvent configurations to strongly affect frequencies and gradients, we here performed a single evaluation of the intra-molecular contributions at a given solvent configuration.

2. **Homogeneous broadening.** The low frequency window of the spectral density is instead responsible for the homogeneous broadening, and comes from the interaction between the "low frequency" modes of the environment with the electronic transition. This is usually captured by following the evolution in time of the manifold of state energies along a MD trajectory, and computing the so called energy gap fluctuation correlation function $C_{ab}(t)$ (where a and b run over the system states), defined as:

$$C_{ab}(t) = \left\langle (\delta\varepsilon_{ga}(t) - \delta\bar{\varepsilon}_{ga})(\delta\varepsilon_{gb}(0) - \delta\bar{\varepsilon}_{gb}) \right\rangle \qquad (5)$$

where the symbol $\langle \ldots \rangle$ denotes the average over the trajectory, $\delta\varepsilon_{ga}(t) = \varepsilon_a(t) - \varepsilon_g(t)$ is the energy gap between the state a and the ground state g at time t, and $\delta\bar{\varepsilon}_{ga} = \langle\, \delta\varepsilon_{ga}(t)\, \rangle$ is the average energy gap. The subscript g denotes that all gap fluctuations are referred to the ground state. Usually only the diagonal terms of this correlation functions are evaluated, i.e. the so called autocorrelation functions $C_{aa}(t)$; here we go one step further by computing also cross-correlation terms $C_{ab}(t)$ with $a \neq b$, which, as we show later, are required for a proper simulation of 2DES signals. The Fourier transform of $C_{ab}(t)$ is linked to the spectral density, as:

$$J_{ab}(\omega) = f(\omega,\beta) \int_{-\infty}^{+\infty} C_{ab}(t) \exp(i\omega t) dt \qquad (6)$$

The function $f(\omega,\beta)$, which depends on both frequency and temperature ($\beta = 1/K_b T$), is a quantum correction term needed to translate classical correlation functions into spectral densities.[50] An harmonic correction is usually employed, with the quantum correction term taking the form $f(\omega,\beta) = \beta\omega/2$;[50] we here applied the so called standard correction, given by $f(\omega,\beta) = \tanh(\beta\omega/2)$.[50] The two corrections are similar (for fixed β) at small ω values, and deviate at large ω values (with the former growing linearly with ω and the latter saturating at the constant value 1). An example of correlation function $C_{aa}(t)$ (for the La adenine state) is shown in Figure 2(b), while its corresponding spectral density is shown in Figure 2(c).

As we already accounted for energy gap fluctuations due to intra-molecular vibrations within a different approach (introduced in point 1.), the system of interest should be kept frozen in the MD to avoid double counting of these intra-molecular motions. This choice is motivated by recent works demonstrating how classical force fields can fail to reproduce these high frequency modes reliably.[49]

3. **Inhomogeneous broadening.** The MD simulation naturally samples the excitation energies produced by the different instantaneous configurations of the solvent around the molecular system of interest. These configurations describe the fact that, at a given instant of time, each molecule of the probed sample has its own local environment, which tunes its excitation energy. This allows one to build a distribution of excitation energies for each state a, i.e. the state static disorder, which can usually be fitted by a Gaussian function: the center of the Gaussian is the average transition energy along the

trajectory ($\bar{\varepsilon}_a^*$), while its width, σ_a, describes how broad is the excitation energy distribution (see Figure 2(a)).

If the system were to visit different local minima of the solute-solvent system along the MD, the distribution of energies would no more be suitable to be fitted with a Gaussian function.

The path that brings one from energy gap fluctuations to spectral densities and static disorder distributions is summarised in Figure 2.

2.4 Models for the computation of the environment induced energy fluctuations

The key aspect for the computation of the terms described at points 2. and 3. of the previous list, is the evaluation of the effect of the surrounding solvent (e.g. water) on the system's (e.g. adenine) excitation energy(ies). This can be accounted for with different approaches. The first consists in performing, for each snapshot, a single QM computation of all the states of interest, obtaining both energy values and transition dipole moments. This procedure directly furnishes the required flow of energy from which to compute both static and dynamic broadening contributions. We call this the *brute-force* approach.

The other approach consists in accounting for the presence of the solvent molecules as a first order *perturbation* of the state energies, on top of a reference set of states (e.g. obtained from a gas-phase computation) by computing the Coulomb coupling term due to the environment in a classical way, thus avoiding explicit QM calculations. This will be called the *perturbative model*.

2.4.1 The brute force approach

As a first approach we have performed high-level CASSCF computations (see Section 3.2) along the 5,000 extracted snapshots from the MD (Section 3.1), which provide the required accurate energies to build the fluctuating energy gaps as well as the transition dipole moments (including both $GS \rightarrow S_x$ and $S_x \rightarrow S_y$), which give the intensities of the different linear and nonlinear signals. We have also computed CASPT2 energy corrections on top of the CASSCF states to obtain more accurate reference energies and also to assess the fluctuating energy gaps at this level of theory.

Given the high number of excited states studied, an issue one may encounter is that of the swapping of the states along the trajectory. The state swapping problem, together with a number of possible strategies to cope with it, is discussed in Section 3.3.

2.4.2 The perturbative approach

In the perturbative approach one performs a single high level quantum chemistry computation on e.g. the gas phase molecule, which is taken as the un-perturbed reference for all subsequent calculations. The system is described by the gas-phase Hamiltonian H_0, while its eigenvalues and eigenstates are denoted ε_a and ψ_a, respectively, with a running over the states of interest. The presence of the surrounding solvent is included by a perturbation term H', different for every snapshot, which describes the interaction between charge densities of the solute and its environment as a classical electrostatic coupling term. The total Hamiltonian describing system and solvent at a given snapshot n can be written as:

$$H(n) = H_0 + H'(n) \tag{7}$$

In the framework of perturbation theory, the first order correction to states energies can be computed employing the wavefunctions of the reference computation, as prescribed by the following equation:

$$\varepsilon_a^*(n) = \varepsilon_a + \langle \psi_a | H'(n) | \psi_a \rangle \tag{8}$$

ψ_a being the unperturbed states (i.e. the eigenstates of the gas phase Hamiltonian H_0).

The perturbative correction to the energy of state a at snapshot n can be rewritten explicitly as the Coulomb interaction between the charge density of the solute (adenine) molecule and the field of the solvent (water) point charges surrounding it, as:

$$\langle \psi_a | H'(n) | \psi_a \rangle = -\sum_j \int \frac{\rho_{aa}(r) q_j^S(n)}{|r - r_j(n)|} dr + \sum_{ij} \frac{Z_i q_j^S(n)}{|r_i - r_j(n)|} \tag{9}$$

where $\rho_{aa}(r)$ is the electron density, Z_i are the atomic numbers of the various atoms in the system of interest (adenine) and $q_j^S(n)$ are the solvent (water) charges at snapshot n. The indexes j and i run over the environment charges and the nuclei of the (adenine) molecule, respectively. From the computational point of view, the Coulomb coupling between the state densities and the point charges surrounding the molecule was computed for all the states of interest by numerical integration of the charge densities $\rho_{aa}(r)$ (see Section 3.4).

The introduced perturbative approach is an extension of the charge density coupling (CDC) method presented elsewhere.[51,52] The method was originally applied to describe linear absorption techniques, and as such it considered a limited number of states. Here we combine it with multiconfigurational electronic structure calculations, making it possible to i) extend it to a large number (30) of transitions in the considered system and ii) compute accurate $\rho_{aa}(r)$ charge densities for all the states of interest. Moreover, since we want to simulate also non-linear spectroscopic techniques, the method has been adapted to describe not only energy gap fluctuation auto-correlation functions, but also energy gap fluctuations cross-correlation functions, that capture the correlation in between energy gap fluctuations of different states, as the general formula given in equation 5 shows.

3 Computational Details

3.1 MD simulations

MD simulations of water-solvated adenine were performed employing molecular mechanics force field (FF) as implemented in the Amber 11 package.[53,54] The system comprises the 9H-adenine molecule solvated by a 12 x 12 x 12 Å cubic box of TIP3P[55] water molecules. For the non-standard 9H-adenine we generated FF parameters based on the generalised Amber FF (GAFF)[56] and the Restrained Electrostatic Potential (RESP) charges computed at the Hartree-Fock 6-31G* level of theory.[57] As our aim is to isolate the role played by the solvent reorganisation dynamics towards spectral broadening, the adenine moiety was positionally restrained at the optimised geometry in vacuum

Table 1 Electronic excited states of interest ($L_{a/b}$ and HL^2) for the diverse CASSCF levels of theory considered in the present work. The different states are described by their main contributing configuration state functions (CSFs) and their relative weight, the norm of their transition dipole moment (TDM), the state number in their respective CASSCF procedure given to highlight state swapping, and the average norm of the TDM ($\langle|TDM|\rangle$).

	Reference[21]			(16,12)-2nN				(16,12)-3nN[b]				
State	CSF	Weight	|TDM|	CSF	Weight	|TDM|	St. n°	CSF	Weight	|TDM|	St. n°	$\langle\|TDM\|\rangle$
L_b	H → L+1	0.40	0.08	H → L+1	0.30	0.12	2	H→ L+1	0.26	0.13	2	0.11
	H-1 → L	0.29		H-1 → L	0.29			H-1 → L	0.36			
L_a	H→ L	0.62	1.35	H → L	0.59	1.34	5	H→ L	0.52	1.42	5	1.33
	H-1→ L+1	0.09										
HL^2	H =>L	0.22	1.15[a]	H =>L	0.17	1.26[a]	16	H =>L	0.17	1.13[a]	19	0.93[a]
				H→ L+2	0.15			H → L+2	0.13			

[a] |TDM| values are referred to the $L_a \rightarrow HL^2$ transition. [b] The values here reported for weight, |TDM| and state number, refer to the computation in the first MD snapshot.

(at the MP2/ANO-L-VDZP level) during all molecular dynamics simulations.

The entire system was pre-equilibrated by heating from 0 to 300 K with steps of 100 K (of 100 ps each) at constant volume and temperature using the Berendsen thermostat and a time step of 2 fs (using the SHAKE algorithm for bonds involving hydrogens), followed by volume equilibration in the NPT ensemble, maintaining constant pressure of 1 atm with the Bresendsen barostat and using isotropic position scaling. Then, a production run at constant volume and temperature (with the Berendsen thermostat) was performed for 1 ns using a time step of 0.5 fs, with a nonbonded cutoff of 10 Å. The coordinates of the system were extracted every 25 fs, i.e. accounting for 40,000 configurations.

3.2 CASSCF/CASPT2 computations

All multiconfigurational electronic structure calculations were carried out with the OpenMOLCAS package.[58] An atomic natural orbital basis set (ANO-L) was used throughout in its valence double-ζ polarised contraction.[59,60] The resolution of identity based on the Cholesky decomposition (CD) was used to speed up the calculation of the electron repulsion integrals.[61–63] An initial MP2 optimization was carried out employing CD-accelerated analytic gradients[64] and using numerical differentiation for obtaining the second-derivatives (Hessian) externally through GAUSSIAN[65] as implemented in the COBRAMM interface.[66]

Complete active space self-consistent field (CASSCF) computations on adenine contained the full π valence occupied and virtual orbitals plus two occupied n lone pair (n_N) orbitals to include the contribution of ^1n_N states, totalling in 16 electrons in 12 orbitals. This level of theory, called "(16,12)-2nN" hereafter, was used in the perturbative model calculations as they compare favourably with reference computations (see Table 1).[21] A second level of theory, called "(16,12)-3nN" hereafter, was employed for the 5,000 snapshots, which contained all 3 n_N lone pairs and all π occupied and virtual valence orbitals except the least contributing π fully bonding occupied orbital. The reason we employed two different active spaces was due to the need to include all 3 n_N lone pairs to mitigate orbital rotations and enable convergence in different solvent arrangements.

CASSCF wave functions were averaged over the lowest-lying 30 states and were subsequently used for single-point perturbation theory (CASPT2) energy corrections.[67–69] An imaginary level shift of 0.2 a.u. was employed in the perturbative step to avoid the presence of intruder states[70] and an IPEA shift[71] of 0.0 was used throughout. CASPT2 computations were performed in its single-state formulation.[67–69]

CASSCF analytic gradients on top of the starting MP2-optimized geometry in the gas phase was computed as required for modelling the high-frequency spectral density window for ~~both~~ the (16,12)-2nN ~~and (16,12)-3nN active spaces~~ active space, employing these data for the computation of the molecule's vibronic structure in both the brute force and the perturbative approaches. An ideal displaced harmonic oscillator was here employed, assuming negligible Dushinsky and anharmonic effects.

The CAS state interaction method[72] was employed to evaluate the transition dipole moments between the ground state and excited states, as well as between excited states.

A hybrid QM/MM scheme was employed on top of the first 5,000 extracted snapshots, where the solvent boxes were cropped to contain a 10 Å radius droplet of waters centred around the adenine moiety in order to isotropically account for solvent effects. The adenine moiety was treated quantum mechanically employing the CASSCF/CASPT2 protocol within the (16,12)-3nN active space mentioned above whereas the water distributions were introduced in the QM computation through an external field as detailed elsewhere.[66]

Table 1 displays the reference values obtained by Nenov *et al.*[21] for the adenine molecule employing the highly accurate SA-25-RASSCF(0,0|12,10|2,12)/SS-RASPT2 level of theory, and a comparison with those employed in the present study. As can be seen, the main configurations describing the different electronic transitions, namely $L_{a/b}$ and the doubly excited HL^2, are adequately described with the different approaches considered here, featuring differences of ~10% in the overall weight of their main contributing CSFs (see $H \rightarrow L+1$ for L_b or $H \rightarrow L$ for L_a in Table 1) that do not heavily alter the properties of interest for modelling the spectra like the |TDM|. The L_a state, mainly described by the $HOMO \rightarrow LUMO$ transition, displays a lesser weight in the (16,12)-2nN and (16,12)-3nN active spaces~~, particularly in the latter, which might explain the worse results obtained for this state due to state mixing~~. TDMs seem to be in qualitative agreement with the reference, (16,12)-2nN displaying much closer values than (16,12)-3nN and which highlights the importance of including the full valence occupied π orbitals in the active space whenever possible. It is worth noting that non-negligible differences arise when considering the norm of the averaged transition

dipole moment ($\langle|TDM|\rangle$, see Table 1), which highlights the additional difficulties faced due to state mixing along the trajectory and that will be discussed in the next section.

3.3 State swapping

Figure 3 displays the norm of the transition dipole moments during the first 12.5 ps of the dynamics in order to illustrate the issue of state swapping along the MD. The first 12.5 ps (500 snapshots) were selected as they are representative of the overall trajectory in describing this problem.

Adenine features two different low-lying excited states in the mid-UV energy window, namely L_b and L_a. The L_a state can be easily identified through its high oscillator strength from the ground state (see Fig. 3a-b). The use of a non-dynamically correlated method like CASSCF leads to the appearance of additional dark states within this window which, depending on the specific water arrangement can lead to an effective state swapping. This means a single CASSCF root (state number) does not correspond to the same electronic configuration on all snapshots. This can be seen in Fig. 3a, where S_3 and S_4 representing the bright L_a state swap in order along the trajectory, which is due to the specific position of the water molecules around the adenine moiety. Much lesser oscillations are observed for the HL^2 state. A detailed discussion on the specific water coordination prompting state swapping is outside the scope of the present manuscript and will be reported elsewhere.

Following the right electronic configuration is of paramount importance, as it is required for spectral simulations through the computation of the energy gap fluctuation autocorrelation functions $C_{ab}(t)$. To do so we have devised a protocol which (in a diabatic fashion) follows the nature of the excited state by computing the state overlap matrix throughout the dynamics, which relates the state ordering of a reference state (the gas-phase (16,12)-3nN reference) with all subsequent snapshots considered, and that allows us to revert the state swapping as shown in Fig. 3b. Only when the bright L_a state is identified along the MD can one assign the $L_a \rightarrow HL^2$ transition (Fig. 3c).

The state swapping issue does not appear in the perturbative model, where only the gas phase reference energies are perturbed, and not the states, which therefore preserves their order throughout the trajectory.

The swapping is, however, not the only issue arising along the MD: by looking at the S_3/S_4 |TDM| as plotted in Fig. 3a or the reordered $S_3(L_a)$ in Fig. 3b, one can notice not only the oscillations between states of the bright signal, but also the oscillatory differences in intensity displayed along the whole trajectory, which is directly related to state mixing: indeed, the changes in the wave function due to solvent reorganization along the trajectory, can lead to effective state mixing. This means the description of a given excited state deviates from the reference, which complicates its monitoring. Moreover, state mixing can lead to a single reference state mapping onto two (or more) adjacent states due to their similar state overlaps, which then makes the assignment ambiguous. This can be particularly seen when monitoring the HL^2 state (lower panel in Fig. 3), as the intensity (|TDM|) vanishes at time=∼2 ps despite the state being described by the same wave function according to the overlap criterion. This is associated to a mixed state character and the intensity of the HL^2 state is not transferred to any of the neighbouring states.

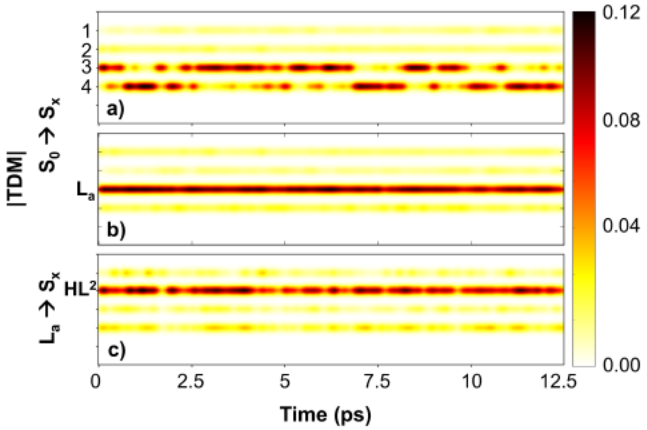

Fig. 3 Norm of the transition dipole moment (|TDM|) over the first 12.5 ps of the molecular dynamics for the CASSCF (16,12)-3nN level of theory. The first block of transitions (a) displays the ground to excited state ($S_0 \rightarrow S_x$) mapping of the L_a state, the second panel (b) is the reordered states from (a) according to the overlap criterion, and (c) displays the reordered $L_a \rightarrow HL^2$ transition and neighbouring states. The z axis denotes the values in arbitrary units of |TDM|, convolved with a 20-pixel wide Gaussian kernel for clarity.

We are currently contemplating four different ways in which these problems might be mitigated: i) to use a reference calculation with a fully solvated molecule instead of a gas phase reference, ii) to consider CASPT2 corrections that not only improve the energies of the states but also the description of the states themselves (i.e. MS- or XMS-CASPT2), iii) by designing more adequate measures to map states by combining state overlaps with the monitoring of the full transition dipole moment vector, and iv) by employing diabatization procedures to follow the diabatic character of the excited state. As stated above, a diabatization scheme based on the maximal overlap of the states along the trajectory with chosen reference states was employed here, while ongoing work is in progress on all four fronts. ~~which we think will lead to a more robust and systematic protocol to map the states and thus enable its use for simulating spectral lineshapes.~~

3.4 Numerical integration of the Coulomb coupling

The integration of the Coulomb interaction term between the electronic charge density $\rho_{aa}(r)$ and the water charges (equation 9) was performed numerically. Each water was considered to be made of three point charges (two q_H charges centered on the two hydrogen atoms, and one q_O charge on the oxygen), whose values were taken from the TIP-3P force field. $\rho_{aa}(r)$ was then computed by expressing it in terms of the (spatial) orbitals $\varphi_i(r)$, as:

$$\rho_{aa}(r) = 2\sum_{i}^{N_{occ}} \varphi_i^2(r) + \sum_{ij}^{N_{aa}} T_{ij}^{a} \varphi_i(r)\varphi_j(r) \qquad (10)$$

Two terms of the system's electronic charge density are here considered: the electron density of the doubly occupied (i.e. inactive)

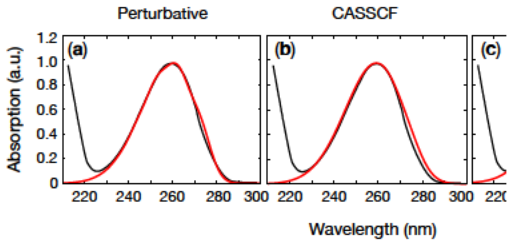

Fig. 4 Comparison between experimental (Ref. 74) simulated models for the linear absorption spectra of in the L_a state absorption window: (a) Perturbative m model and (c) CASPT2 model. The L_a state computec has been centered on the experimental one, applying a cm^{-1} (a), 1200 cm^{-1} (b) and 1100 cm^{-1} (c). Spect match the experimental La intensity. No lifetime bro added, as the use of experimental fitted L_a lifetime va negligible modifications to the lineshape.

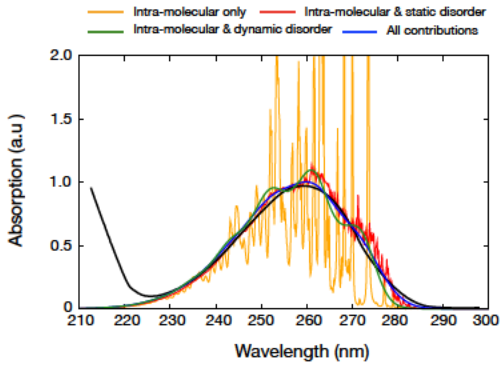

Fig. 5 Simulation of the linear absorption spectrum (perturbative model) in the L_a adenine state absorption window, accounting only for selected disorder contributions: static+dynamic disorder (blue line), only dynamic disorder (green line), only static disorder (red line), no disorder (orange line). The experimental spectrum (Ref. 74) is also shown (black line).

orbitals, in the number of N_{occ}, and the electron density of the active orbitals, in the number of N_{act}. The first contribution, as well as the nuclear contribution to the full charge density, is equal for all the states, while the latter changes from state to state, as prescribed by the transition density matrix T^a_{ij}. This summarizes the contribution with which each configuration, involving a pair of active orbitals φ_i and φ_j, participate to the state of interest (a) (i.e. the different way in which every state fills up its active orbitals with its active electrons).

From a practical point of view, the orbitals $\varphi_i(r)$ were transformed in cube files by means of an in-house modified version of the *Orbkit* toolkit,[73] which can be integrated numerically. The discrete densities are now represented as a collection of charges on a 3D grid ($\rho_{aa}(r) = \sum_i q^a_i \delta(r - r_i)$), and the integral of equation 9 is here substituted by a double sum over the state a density charges and the water charges:

$$\sum_{ij} \frac{q^a_i q^S_j(n)}{|r_i - r_j(n)|} \tag{11}$$

The evaluation of the double sum can be significantly sped up with a screening procedure that filters out very small charge values: the molecule is in fact characterized by a non-uniform electron density in the grid volume, so that points close to the edge of the grid boundaries carry negligible contributions. By discarding all charges whose absolute value is smaller than a threshold value we significantly reduced the number of interaction terms

	Perturbative			CASSCF		
	σ	λ_H	λ_L	σ	λ_H	λ_L
L_a	450	2530	410	630	~~2620~~2530	800
HL^2	490	8980	425	576	~~145008~~8980	575
$L_a - HL^2$	-	4024	276	-	~~5035~~4024	295

Table 2 Summary of the disorder parameters employed in the simulation of linear and non-linear spectra. Static disorder σ_a, reorganisation energies for intra-molecular modes (λ_H) and for the fluctuating waters (λ_L) are expressed in cm^{-1}.

that have to be computed. The accuracy of this procedure was checked by comparing numerical dipoles, computed by means of the discrete and screened set of charges q^a_i, with the analytic ones provided by the complete active space state interaction (CASSI)[72] method available in the OpenMOLCAS package.[58]

The extraction of the transition density matrix and of the orbital cube files from CASSCF computations (performed with OpenMOLCAS) has been already detailed elsewhere.[75,76]

3.5 Computation of the spectral densities

The spectral densities $J_{aa}(\omega)$ were composed by summing the intra-molecular vibration contribution, $J^H_{ab}(\omega)$, and the contribution coming from the water dynamics around the frozen adenine, $J^L_{ab}(\omega)$. $J^H_{ab}(\omega)$ were computed ~~for both considered active spaces, (16,12) 2nN and (16,12) 3nN~~ employing the (16,12)-2nN active space. A Gaussian broadening of 5 cm^{-1} was applied to the discrete spectrum of frequencies, which physically translates in accounting for the vibrational dephasing of the modes. We also neglected all contributions coming from modes below 450 cm^{-1}, as the gradient projection on these low frequency modes can be remarkably inaccurate.[9]

To ensure that the auto- (cross-) correlation functions $C_{ab}(t)$ decay to zero at the end of the time window, a Gaussian damping function with $\sigma = 3$ ps was used.[50,77] Moreover, instead of numerically Fourier transforming the $C_{ab}(t)$ functions, we fit these with a stretched exponential, i.e. a two parameters function given by:

$$f_{\tau,\beta}(t) = \exp\left[-(t/\tau)^\beta\right] \tag{12}$$

The Fourier transform of the stretched exponential was then obtained employing an open-source library.[78]

3.6 Computations of spectra

All the spectra simulations (both linear and non-linear) were carried out employing an in-house modified version of the *Spectron* software.[4]

Non linear spectra were convolved with a Gaussian pulse shape

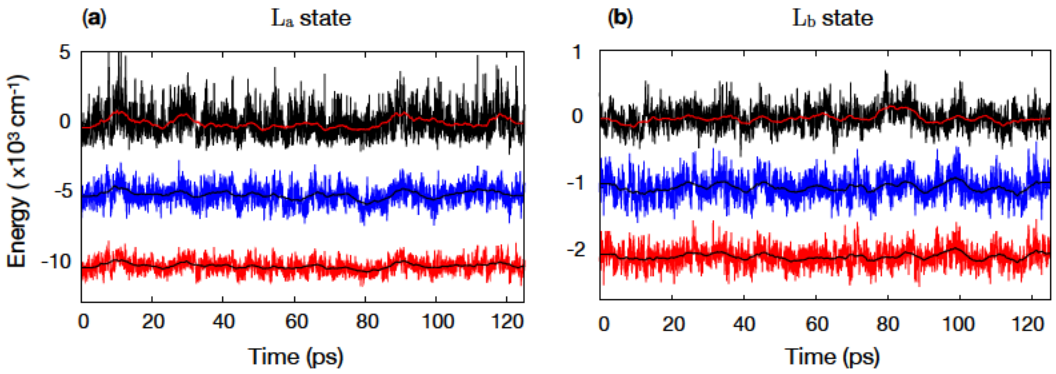

Fig. 6 L_a and L_b state energy flow along the MD trajectory. From bottom to top, the levels of theory are: Perturbative (red curve), CASSCF (blue curve) and CASPT2 (black curve). The curves have been displaced (by steps of 5000 cm^{-1} and 1000 cm^{-1} for L_a and L_b states, respectively) to allow for comparison. The states were reordered in all the models (except the perturbative one) according to the maximal overlap criterion with respect to the initial reference state (see Section 3.3). L_b flows are much closer to their average energy gap value and look nearly identical for all models, while the much wider L_a flows do not. For the L_a state it is clearly visible that the CASPT2 model has the highest number of outliers.

envelope of $\sigma = 7$ fs, in order to faithfully reproduce the experimental setup of Ref. [79], in which a pulse of width 16 fs was employed.

4 Results

In the following we apply the presented protocol to adenine. As a proof of concept we simulate linear absorption (LA), as well transient absorption (TA) and 2DES

Nenov et al.[21] reported the ele in gas-phase up to 10 eV, thereby e yond the lowest far-UV accessible st band). They showed that a multi CASSCF/CASPT2 is required to acc ergy spectrum. They identified a cl a double HOMO to LUMO (H => I intense "fingerprint" ESA for L_a in n of outlying the practical application comparing different approaches (br computing static and dynamic diso linear absorption and on both the (labeled HL^2) for TA and 2DES.

We start by benchmarking the va perimental linear absorption spectr perturbative and the QM/MM brute sient absorption and 2DES signals.

4.1 Linear absorption

In Figure 4 a comparison of linear s sition is presented. Three levels of t vative, CASSCF (in which the energ at the CASSCF level) and CASPT2 tions are applied on the top of the For the pertubative method we rely and SS-CASPT2 excitation energies phase structure, while for the othe trajectory averaged TDMs (CASSCF level for both models) and excitation energies (SS-CASPT2 level for both models).

The L_a linewidth is described remarkably well by most of the considered methods, CASPT2 being the only one that presents a slightly broader linewidth. This is caused by a small mixing with other states, which produces large deviations in the energy gap at the CASPT2 level as it degrades the CASSCF reference wave function in which the correction is applied. CASPT2 therefore exasperates the issues coming from states mixing, and we thus

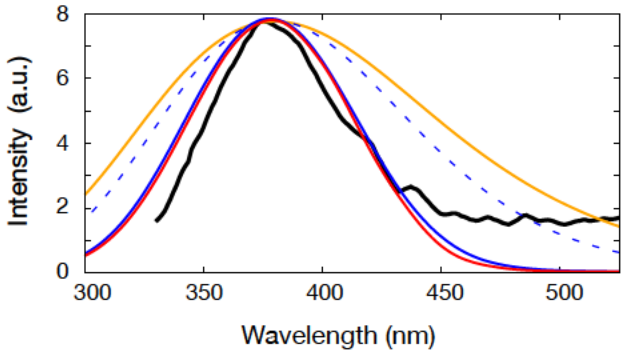

Fig. 7 Comparison of experimental[79] (black curve) and theoretical (red and blue curves, for perturbative and CASSCF models, respectively) transient absorption spectra at time $t_2 = 100$ fs. In the theoretical maps only the ESA contribution was considered. Simulation of the ESA signal that does not consider inter-state spectral densities $J_{L_aHL^2}(\omega)$ (orange line) and employing $J^H(\omega)$ computed with the (16,12)-3nN active space (blue dashed line) are also shown. ~~a is significantly less accurate with respect to the complete description (red line). The signal intensity above ~425 nm in the experimental spectrum is likely attributed to additional $L_a \rightarrow S_n$ ESA signals that were not considered in the model.~~ A Gaussian pulse envelope of ~16 fs width was applied to the simulated spectra, in order to mimic the effect of experimentally employed pulses. Simulated curves were shifted and normalized to the experimental spectrum, in order to allow for a comparison of the lineshapes.

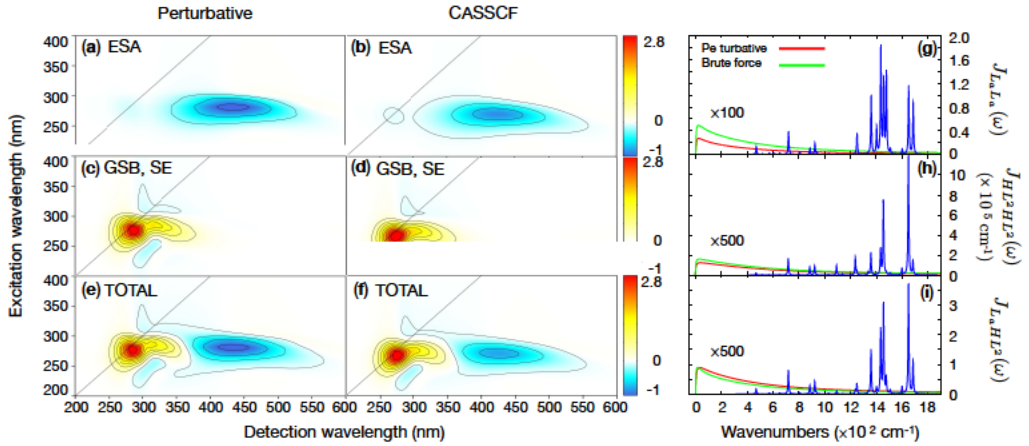

Fig. 8 Comparison of theoretical (perturbative model (a,c,e), CASSCF model (b,d,f)) 2DES maps at time $t_2 = 0$ fs: (a-b) ESA contribution, (c-d) GSB and SE contributions, (e-f) total 2DES spectrum. The intensity range has been fixed for all six maps to 90% of the min and max of the perturbative total map. Convolution of the simulated maps with a Gaussian pulse envelope of ∼16 fs width was applied. (g-i) Spectral densities used in the calculation of linear and non-linear spectra. A scaling factor was applied to $J_L(\omega)$ to make them visible in the range of $J_H(\omega)$. ~~The insets in (a) and (b) show that the computed ESA lineshape for the perturbative and CASSCF models have a different shape.~~

ing a large amount of states. We nonetheless note that, for states that do not swap or mix along the trajectory (as it happens for the dark L_b state), CASSCF and CASPT2 energy fluctuations (and therefore the spectral densities and eventually the broadening of the spectra) are comparable (Figure 6(b)).

In Figure 5 we show how the various disorder contributions are all necessary to describe a correct lineshape. The orange curve represents the spectra simulated by neglecting both static and dynamic contributions, and therefore including only intra-molecular vibrations (high-frequency window of the spectral density). One notices that the origin to the asymmetric lineshape of the L_a absorption, with an intense tail extending in the blue is provided by the intra-molecular vibrations. Still, it is the addition of homogeneous broadening (green curve) that smooths the spectral shape. Finally, the inclusion of static disorder (blue line, which is the same spectrum of Figure 4(a)) washes out the vibronic bands that would otherwise appear.

It is worth highlighting that the perturbative approach works as well as the higher level CASSCF approach. This is visible in the spectra but also when comparing the energy fluctuations (Figure 6), which follow the same dynamics (even if the CASSCF flow displays a larger variability, which is reflected in the higher intensity of the spectral density and the larger static disorder σ and reorganization energy λ_L, as detailed in Table 2).

4.2 Non-linear spectra

The simulation of non-linear spectroscopy was here limited to the GSB and SE contributions coming from the L_a bright state, and ESA signals from the $L_a \rightarrow HL^2$ transition. Not only is this transition the brightest from the initially accessed bright L_a state, but it also remains clearly detectable in gas phase and in solution. We have collected data and are planning to perform a more complete study, by including additional bright states in different windows.

First of all we compute transient absorption (TA) spectra at 100 fs pump-probe delay time, at both the perturbative and the CASSCF level, and compare them with recently published experimental data.[79] Figure 7 shows a comparison between the experimental signals recorded in the 300−650 nm window, and four simulated ESA curves, which reproduce the $L_a \rightarrow HL^2$ band in: the perturbative model (red curve), the brute-force CASSCF model (blue curve), and two modified CASSCF models obtained by removing the cross-term $J_{L_aHL^2}(\omega)$ contributions (orange curve), and by employing a different $J^H_{ab}(\omega)$ spectral densities (dashed blue curve). We note that the computed lineshape obtained by including all spectral density components (i.e. coming from the autocorrelation functions of the L_a and HL^2 states, as well as the cross-correlation term between the two states) compares significantly well with experiment, at both the perturbative and CASSCF level (as the disorder parameters of the models compares well with each other, see Table 2). It is remarkable how, by removing the cross-term $J_{L_aHL^2}(\omega)$ (Fig.7, orange line), the linewidth becomes significantly broader. ~~We note that the perturbative approach results in a better description than at the higher level CASSCF level.~~ We note that the dashed blue curve, which was obtained employing $J^H_{ab}(\omega)$ spectral densities computed in the first snapshot of the MD and within the (16,12)-3nN active space, is considerably broader compared to the other curves, due to the higher coupling to intra-molecular vibrations found in the gradient of the HL^2 state found in this active space. ~~This is to be ascribed to the significantly higher coupling to intra-molecular vibrations found in the gradient of the HL^2 state in computations performed within the (16,12)-3nN active space, which enter in the description of the CASSCF model lineshape~~. This translates in almost double the intensity of the high frequency part of the spectral density $J^H_{HL^2HL^2}(\omega)$ in (16,12)-3nN with respect to (16,12)-2nN ($\lambda^H_{HL^2} = 14500\text{cm}^{-1}$, to be compared with $\lambda^H_{HL^2} = 8980\text{cm}^{-1}$ obtained in the (16,12)-2nN active space). The reason for this overestimation and the consequent worse agreement to experiments, is to be ascribed to the reduced quality in the description of the HL^2 state in the given solvent snapshot and within (16,12)-

3nN active space. ~~(used in the perturbative computations). The contributions to the low frequency part of the spectral density, $J^L_{L_aHL^2}(\omega)$ and $J^L_{HL^2HL^2}(\omega)$, were instead comparable in the two setups. This last observation~~

The results obtained for that TA band at 100 fs confirm the astonishing accuracy of the perturbative approach in reproducing the results of higher level computations~~, while the overall better agreement of the perturbative lineshape with respect to the CASSCF one,~~ while the overall better agreement of the models employing spectral densities $J^H(\omega)$ obtained within the (16,12)-2nN active space, underlines the importance of a proper description of the states for the vibronic coupling calculation. ~~To this respect, the use of (16,12) 2nN active space, seems to produce a better description of the adenine manifold of states than the (16,12) 3nN active space.~~ Finally, the intensity recorded in the experiments above 500 nm is to be ascribed to ESA signals to other excited states reached from L_a, that we have not included in our computations.

In Figure 8 we show a sequence of simulated 2DES contributions at time $t_2 = 0$, which highlight the non-trivial lineshape that can be observed in 2DES experiments in the absorption region of the L_a state. We separate GSB and SE emission from ESA contributions and further compare the perturvative and CASSCF models. Due to the absence of experimental data at short delay times, we present spectra for a pump-probe delay $t_2 = 0$ fs. GSB contributions (Figure 8(c-d)) are mainly localised on the diagonal of the map, at excitation and detection wavelength around 280 nm (no shift of the transition energy has been applied here). It gives a largely circular contribution also at time $t_2 = 0$, as the contribution of inhomogeneous broadening is not strong enough to produce an asymmetric shape. SE contributions (Figure 8(c-d)) appear as a positive tail, spreading to the red side of the window, from around 300 nm to 450 nm, and overlapping with the negative ESA signals (Figure 8(a-b)), which extends from 380 nm to the red edge of the simulated spectrum. The shape of GSB and SE peaks is similarly described by the two methods. A minor difference in the ESA contributions intensity comes from the slightly reduced norm of the transition dipole moment found when averaging CASSCF results over the (5,000) MD trajectory configurations considered for the brute-force QM/MM approach (as detailed in Table 1).

~~As previously observed in the analysis of the transient absorption of Figure 7, ESA contributions differ in the two maps, with the CASSCF ESA characterized by a broad peak with a reduced intensity. This is mostly due to the diverse coupling intensities between the HL^2 state and the adenine normal modes at the two levels of theory employed, reflecting the different description of the state with (16,12) 2nN and (16,12) 3nN. A minor component of the difference between the two ESA signals comes from the slightly reduced norm of the transition dipole moment found when averaging CASSCF results over the (5,000) MD trajectory configurations considered for the brute-force QM/MM approach (as detailed in Table 1). The difference between the two ESA signals is highlighted in their respective insets.~~

5 Conclusions

In this discussion we explore two different approaches to simulate spectral lineshapes in linear and third-order non-linear optical spectroscopies. Lineshapes are modulated by the coupling between the electronic transitions and the molecular normal modes, as well as by static and dynamic effects of solvation. Our extension of the CDC method with multireference electronic structure calculations allows computing disorder properties of a large number of system states, enabling its applicability to the simulation of 2DES.

In the first part of this paper we lay out a theoretical framework for the characterisation of these various contributions from first principles. Intra-molecular contributions were accounted for in the framework of the displaced harmonic oscillator. Concerning static and dynamic disorder we confront two methods, both relying on a sequence of snapshots extracted from an MD simulation and representing the solute through atom centered point charges: one that considers the effects of the different solvent arrangements around the the states energies, obtained as the Coulomb coupling term between the solute's (fixed) gas-phase states and the solvent charges, and another where full QM/MM computations (in which orbital relaxation is allowed) obtain the response of the solute electronic structure to the different solvent arrangements. The former, relying on a single gas phase reference computation, does not require further electronic structure theory evaluations and solely needs ~5 minutes per snapshot for evaluating the interaction with the reference state densities to obtain the perturbed energies along the MD. The latter, on the other hand, requires one QM/MM computation per extracted snapshot (5,000) and each of these costs ~6 hours for CASSCF and an additional 3 hours when performing CASPT2 on top, thus requiring a hundred times more computing time compared to the perturbative approach.

The two methods were tested and compared by simulating linear absorption, transient absorption and 2DES spectra. In all the three cases, the perturbative approach performed outstandingly well. Beyond its higher computational cost, we showed the difficulties that appear in applying a brute force approach along the trajectory, with the occurrence of state swapping and mixing, which make it necessary to take actions to properly identify the states of interest within the manifold of computed states along the MD trajectory. The perturbative method, which modifies only the state energies without influencing their order, does not suffer from this issue. Nonetheless, the fact that only state energies are perturbed, makes it unable to describe the variability of the transition dipole moment along the trajectory, which is instead captured by the brute-force approach.

Even if the here simulated spectra are mostly affected by intramolecular vibrations (high-frequency spectral density window), the computed static and dynamic disorder coming from the interaction between the adenine molecule and its environment give rise to non-negligible contributions to the lineshape. We expect that the importance of the proper evaluation of these terms should become even more apparent ~~in the visible IR spectral region, and~~ when a larger number of states, characterized by sig-

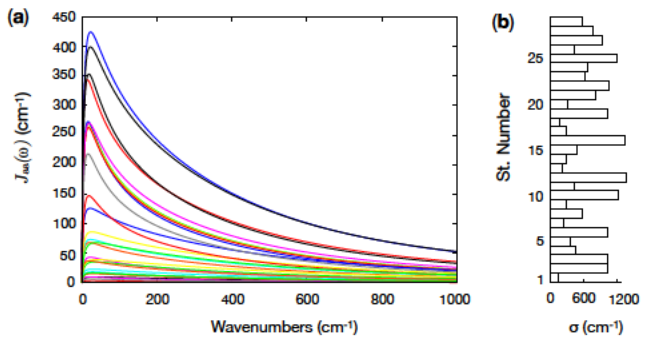

Fig. 9 Comparison of the spectral densities (a) and of the static disorder width σ (b) for all the 30 computed adenine states, in the framework of the perturbative model. The notable state-dependence of the system-bath coupling strength is apparent.

nificantly different coupling with the environment, are included. This is made apparent by looking at the large variability of the spectral densities $J^L_{aa}(\omega)$ (and of the static disorder σ_a) computed for the all the 30 states with the perturbative model, shown in Figure 9, whose peak heights vary by two orders of magnitude.

We therefore suggest that a promising strategy for the first principles description of the lineshape should not in general rely on the brute force computation of costly high-level computations on each snapshot of a generated MD trajectory. It should instead focus on producing very accurate reference states at a single point, on top of which both the perturbative approach and the evaluation of the coupling with the intra-molecular normal modes will be optimal. In this regard, a step towards increasing the reliability of the perturbative approach would be to rely on state densities obtained for a typical water distribution and/or an average description of the solvent via embedding models. In this way, the strength of the perturbative coupling will be reduced, thus enhancing the accuracy of the results.

6 Conflicts of interest

There are no conflicts to declare.

Acknowledgements

J.S.-M. acknowledges support from the European Commission through the Marie Curie actions (*AttoDNA*, FP8-MSCA-IF, grant n° 747662) and the Imperial College Research Computing Service (DOI: 10.14469/hpc/2232) where most of the computations were carried out. I. R. gratefully acknowledges support from the Agence National de la Recherche (FEMTO-2DNA, ANR-15-CE29-0010). M. G. acknowledges support of the European Union's Horizon 2020 research and innovation programme under grant agreement No. 765266 (LightDyNAmics).

Notes and references

1 R. G. Breene Jr, *Theories of spectral line shape*, Wiley, New York, 1981.

2 A. Nenov, A. Giussani, B. P. Fingerhut, I. Rivalta, E. Dumont, S. Mukamel and M. Garavelli, *Phys. Chem. Chem. Phys.*, 2015, **17**, 30925–30936.

3 A. Nenov, R. Borrego-Varillas, A. Oriana, L. Ganzer, F. Segatta, I. Conti, J. Segarra-Marti, J. Omachi, M. Dapor, S. Taioli, C. Manzoni, S. Mukamel, G. Cerullo and M. Garavelli, *The Journal of Physical Chemistry Letters*, 2018, **9**, 1534–1541.

4 D. Abramavicius, B. Palmieri, D. V. Voronine, F. Šanda and S. Mukamel, *Chemical Reviews*, 2009, **109**, 2350–2408.

5 J. Segarra-Martí, S. Mukamel, M. Garavelli, A. Nenov and I. Rivalta, *Top. Curr. Chem.*, 2018, **376**, 24.

6 M. Barbatti and K. Sen, *International Journal of Quantum Chemistry*, 2015, **116**, 762–771.

7 J. P. Zobel, M. Heindl, J. J. Nogueira and L. González, *Journal of Chemical Theory and Computation*, 2018, **14**, 3205–3217.

8 F. J. A. Ferrer, J. Cerezo, E. Stendardo, R. Improta and F. Santoro, *Journal of Chemical Theory and Computation*, 2013, **9**, 2072–2082.

9 F. J. A. Ferrer and F. Santoro, *Physical Chemistry Chemical Physics*, 2012, **14**, 13549.

10 A. Baiardi, J. Bloino and V. Barone, *Journal of Chemical Theory and Computation*, 2013, **9**, 4097–4115.

11 L. Zanetti-Polzi, M. Aschi, I. Daidone and A. Amadei, *Chemical Physics Letters*, 2017, **669**, 119–124.

12 L. Zanetti-Polzi, S. D. Galdo, I. Daidone, M. D'Abramo, V. Barone, M. Aschi and A. Amadei, *Physical Chemistry Chemical Physics*, 2018, **20**, 24369–24378.

13 M. D'Abramo, M. Aschi and A. Amadei, *The Journal of Chemical Physics*, 2014, **140**, 164104.

14 A. Nenov, I. Conti, R. Borrego-Varillas, G. Cerullo and M. Garavelli, *Chemical Physics*, 2018, **515**, 643–653.

15 M. Barbatti, A. J. A. Aquino and H. Lischka, *Physical Chemistry Chemical Physics*, 2010, **12**, 4959.

16 A. J. Pepino, J. Segarra-Martí, A. Nenov, R. Improta and M. Garavelli, *The Journal of Physical Chemistry Letters*, 2017, **8**, 1777–1783.

17 S. Shim, P. Rebentrost, S. Valleau and A. Aspuru-Guzik, *Biophysical Journal*, 2012, **102**, 649–660.

18 M. K. Lee, P. Huo and D. F. Coker, *Annual Review of Physical Chemistry*, 2016, **67**, 639–668.

19 M. K. Lee, K. B. Bravaya and D. F. Coker, *Journal of the American Chemical Society*, 2017, **139**, 7803–7814.

20 J. Segarra-Martí, E. Zvereva, M. Marazzi, J. Brazard, E. Dumont, X. Assfeld, S. Haacke, M. Garavelli, A. Monari, J. Léonard and I. Rivalta, *Journal of Chemical Theory and Computation*, 2018, **14**, 2570–2585.

21 A. Nenov, A. Giussani, J. Segarra-Marti, V. K. Jaiswal, I. Rivalta, G. Cerullo, S. Mukamel and M. Garavelli, *The Journal of Chemical Physics*, 2015, **142**, 212443.

22 I. Rivalta, A. Nenov, O. Weingart, G. Cerullo, M. Garavelli and S. Mukamel, *The Journal of Physical Chemistry B*, 2014, **118**, 8396–8405.

23 A. Giussani, J. Segarra-Martí, A. Nenov, I. Rivalta, A. Tolomelli, S. Mukamel and M. Garavelli, *Theoretical Chemistry Accounts*, 2016, **135**, 121.

24 J. Segarra-Martí, A. J. Pepino, A. Nenov, S. Mukamel, M. Gar-

avelli and I. Rivalta, *Theoretical Chemistry Accounts*, 2018, **137**, 47.
25 R. Borrego-Varillas, D. C. Teles-Ferreira, A. Nenov, I. Conti, L. Ganzer, C. Manzoni, M. Garavelli, A. M. de Paula and G. Cerullo, *Journal of the American Chemical Society*, 2018, **140**, 16087–16093.
26 A. Picchiotti, A. Nenov, A. Giussani, V. I. Prokhorenko, R. J. D. Miller, S. Mukamel and M. Garavelli, *The Journal of Physical Chemistry Letters*, 2019, **0**, DOI:10.1021/acs.jpclett.9b01325.
27 J. Segarra-Martí, V. K. Jaiswal, A. J. Pepino, A. Giussani, A. Nenov, S. Mukamel, M. Garavelli and I. Rivalta, *Faraday Discuss.*, 2018, **207**, 233–250.
28 T. A. A. Oliver, *Royal Society Open Science*, 2018, **5**, 171425.
29 M. Cho, *Chemical Reviews*, 2008, **108**, 1331–1418.
30 D. M. Jonas, *Annual Review of Physical Chemistry*, 2018, **69**, 327–352.
31 I. Conti, M. Garavelli and G. Orlandi, *Journal of the American Chemical Society*, 2009, **131**, 16108–16118.
32 I. Conti, P. Altoè, M. Stenta, M. Garavelli and G. Orlandi, *Phys. Chem. Chem. Phys.*, 2010, **12**, 5016–5023.
33 A. A. Beckstead, Y. Zhang, M. S. de Vries and B. Kohler, *Phys. Chem. Chem. Phys.*, 2016, **18**, 24228–24238.
34 S. Boldissar and M. S. de Vries, *Phys. Chem. Chem. Phys.*, 2018, **20**, 9701–9716.
35 B. K. McFarland, J. P. Farrell, S. Miyabe, F. Tarantelli, A. Aguilar, N. Berrah, C. Bostedt, J. D. Bozek, P. H. Bucksbaum, J. C. Castagna, R. N. Coffee, J. P. Cryan, L. Fang, R. Feifel, K. J. Gaffney, J. M. Glownia, T. J. Martinez, M. Mucke, B. Murphy, A. Natan, T. Osipov, V. S. Petrović, S. Schorb, T. Schultz, L. S. Spector, M. Swiggers, I. Tenney, S. Wang, J. L. White, W. White and M. Gühr, *Nature Communications*, 2014, **5**, 4235.
36 T. J. A. Wolf, F. Holzmeier, I. Wagner, N. Berrah, C. Bostedt, J. Bozek, P. Bucksbaum, R. Coffee, J. Cryan, J. Farrell, R. Feifel, T. J. Martinez, B. McFarland, M. Mucke, S. Nandi, F. Tarantelli, I. Fischer and M. Gühr, *Applied Sciences*, 2017, **7**, 681.
37 T. J. A. Wolf and M. Gühr, *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 2019, **377**, 20170473.
38 C. hung Tseng, S. Matsika and T. C. Weinacht, *Opt. Express*, 2009, **17**, 18788–18793.
39 J. Chen, Y. Zhang and B. Kohler, *Top. Curr. Chem.*, 2015, **356**, 39–87.
40 T. Takaya, C. Su, K. de La Harpe, C. E. Crespo-Hernández and B. Kohler, *Proceedings of the National Academy of Sciences*, 2008, **105**, 10285–10290.
41 T. Gustavsson, R. Improta and D. Markovitsi, *The Journal of Physical Chemistry Letters*, 2010, **1**, 2025–2030.
42 D. Markovitsi, T. Gustavsson and I. Vayá, *The Journal of Physical Chemistry Letters*, 2010, **1**, 3271–3276.
43 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.
44 R. Islampour and S. Mukamel, *Chemical Physics Letters*, 1984, **107**, 239 – 244.
45 L. Martínez-Fernández, A. J. Pepino, J. Segarra-Martí, J. Jovaišaitė, I. Vaya, A. Nenov, D. Markovitsi, T. Gustavsson, A. Banyasz, M. Garavelli and R. Improta, *Journal of the American Chemical Society*, 2017, **139**, 7780–7791.
46 A. J. Pepino, J. Segarra-Martí, A. Nenov, I. Rivalta, R. Improta and M. Garavelli, *Phys. Chem. Chem. Phys.*, 2018, **20**, 6877–6890.
47 J. Segarra-Martí, E. Zvereva, M. Marazzi, J. Brazard, E. Dumont, X. Assfeld, S. Haacke, M. Garavelli, A. Monari, J. Léonard and I. Rivalta, *Journal of Chemical Theory and Computation*, 2018, **14**, 2570–2585.
48 W.-M. Kwok, C. Ma and D. L. Phillips, *Journal of the American Chemical Society*, 2006, **128**, 11894–11905.
49 M. K. Lee and D. F. Coker, *The Journal of Physical Chemistry Letters*, 2016, **7**, 3171–3178.
50 S. Valleau, A. Eisfeld and A. Aspuru-Guzik, *The Journal of Chemical Physics*, 2012, **137**, 224103.
51 J. Adolphs and T. Renger, *Biophysical Journal*, 2006, **91**, 2778 – 2797.
52 X. Wang, G. Ritschel, S. Wüster and A. Eisfeld, *Physical Chemistry Chemical Physics*, 2015, **17**, 25629–25641.
53 D. A. Case, T. E. Cheatham III, T. Darden, H. Gohlke, R. Luo, K. M. Merz Jr., A. Onufriev, C. Simmerling, B. Wang and R. J. Woods, *Journal of Computational Chemistry*, 2005, **26**, 1668–1688.
54 R. Salomon-Ferrer, D. A. Case and R. C. Walker, *Wiley Interdisciplinary Reviews: Computational Molecular Science*, 2013, **3**, 198–210.
55 W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey and M. L. Klein, *The Journal of Chemical Physics*, 1983, **79**, 926–935.
56 J. Wang, R. M. Wolf, J. W. Caldwell, P. A. Kollman and D. A. Case, *Journal of Computational Chemistry*, 2004, **25**, 1157–1174.
57 C. I. Bayly, P. Cieplak, W. Cornell and P. A. Kollman, *The Journal of Physical Chemistry*, 1993, **97**, 10269–10280.
58 F. Aquilante, J. Autschbach, R. K. Carlson, L. F. Chibotaru, M. G. Delcey, L. De Vico, I. Fdez. Galván, N. Ferré, L. M. Frutos, L. Gagliardi, M. Garavelli, A. Giussani, C. E. Hoyer, G. Li Manni, H. Lischka, D. Ma, P. Å. 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. P. Vysotskiy, O. Weingart, F. Zapata and R. Lindh, *Journal of Computational Chemistry*, 2016, **37**, 506–541.
59 P.-O. Widmark, P.-Å. Malmqvist and B. O. Roos, *Theor. Chim. Acta*, 1990, **77**, 291.
60 P.-O. Widmark, B. J. Persson and B. O. Roos, *Theor. Chim. Acta*, 1991, **79**, 419.
61 F. Aquilante, R. Lindh and T. Bondo Pedersen, *J. Chem. Phys.*, 2007, **127**, 114107.

62 F. Aquilante, T. B. Pedersen and R. Lindh, *J. Chem. Phys.*, 2007, **126**, 194106.

63 F. Aquilante, T. B. Pedersen, R. Lindh, B. O. Roos, A. Sánchez de Merás and H. Koch, *J. Chem. Phys.*, 2008, **129**, 024113.

64 J. Böstrom, V. Veryazov, F. Aquilante, T. B. Pedersen and R. Lindh, *International Journal of Quantum Chemistry*, 2014, **114**, 321–327.

65 M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, G. A. Petersson, H. Nakatsuji, X. Li, M. Caricato, A. V. Marenich, J. Bloino, B. G. Janesko, R. Gomperts, B. Mennucci, H. P. Hratchian, J. V. Ortiz, A. F. Izmaylov, J. L. Sonnenberg, D. Williams-Young, F. Ding, F. Lipparini, F. Egidi, J. Goings, B. Peng, A. Petrone, T. Henderson, D. Ranasinghe, V. G. Zakrzewski, J. Gao, N. Rega, G. Zheng, W. Liang, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, K. Throssell, J. A. Montgomery, Jr., J. E. Peralta, F. Ogliaro, M. J. Bearpark, J. J. Heyd, E. N. Brothers, K. N. Kudin, V. N. Staroverov, T. A. Keith, R. Kobayashi, J. Normand, K. Raghavachari, A. P. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, J. M. Millam, M. Klene, C. Adamo, R. Cammi, J. W. Ochterski, R. L. Martin, K. Morokuma, O. Farkas, J. B. Foresman and D. J. Fox, *Gaussian 16 Revision B.01*, 2016, Gaussian Inc. Wallingford CT.

66 O. Weingart, A. Nenov, P. Altoè, I. Rivalta, J. Segarra-Martí, I. Dokukina and M. Garavelli, *Journal of Molecular Modeling*, 2018, **24**, 271.

67 K. Andersson, P. A. Malmqvist, B. O. Roos, A. J. Sadlej and K. Wolinski, *J. Phys. Chem.*, 1990, **94**, 5483–5488.

68 K. Andersson, P.-Å. Malmqvist and B. O. Roos, *J. Chem. Phys.*, 1992, **96**, 1218–1226.

69 D. Roca-Sanjuán, F. Aquilante and R. Lindh, *WIRES Comput. Mol. Sci.*, 2012, **2**, 585–603.

70 N. Forsberg and P.-Å. Malmqvist, *Chem. Phys. Lett.*, 1997, **274**, 196 – 204.

71 G. Ghigo, B. O. Roos and P.-Å. Malmqvist, *Chem. Phys. Lett.*, 2004, **396**, 142 – 149.

72 P.-Å. Malmqvist and B. O. Roos, *Chem. Phys. Lett.*, 1989, **155**, 189–194.

73 G. Hermann, V. Pohl, J. C. Tremblay, B. Paulus, H.-C. Hege and A. Schild, *Journal of Computational Chemistry*, 2016, **37**, 1511–1520.

74 C. T. Middleton, K. de La Harpe, C. Su, Y. K. Law, C. E. Crespo-Hernández and B. Kohler, *Annual Review of Physical Chemistry*, 2009, **60**, 217–239.

75 F. Segatta, L. Cupellini, S. Jurinovich, S. Mukamel, M. Dapor, S. Taioli, M. Garavelli and B. Mennucci, *Journal of the American Chemical Society*, 2017, **139**, 7558–7567.

76 F. Segatta, 2018.

77 D. Loco and L. Cupellini, *International Journal of Quantum Chemistry*, 2018, **119**, e25726.

78 J. Wuttke, *Algorithms*, 2012, **5**, 604–628.

79 R. Borrego-Varillas, G. Cerullo and D. Markovitsi, *The Journal of Physical Chemistry Letters*, 2019, **10**, 1639–1643.

80 A. A. Granovsky, *J. Chem. Phys.*, 2011, **134**, 214113.