Alma Mater Studiorum - Università di Bologna

DOTTORATO DI RICERCA IN

CHIMICA

Ciclo 34

Settore Concorsuale: 03/A2 - MODELLI E METODOLOGIE PER LE SCIENZE CHIMICHE

Settore Scientifico Disciplinare: CHIM/02 - CHIMICA FISICA

MODELLING UV RESPONSE OF BIOMOLECULES: FROM ELECTRONIC STRUCTURE TO AB-INITIO DYNAMICS

Presentata da: Vishal Kumar Jaiswal

Coordinatore Dottorato

Luca Prodi

Supervisore

Marco Garavelli

Co-supervisore

Elisabetta Venuti

Esame finale anno 2022

papa ke liye

Abstract

The interaction of organic chromophores with light initiates ultrafast processes in the timescale of femtoseconds. An atomistic understanding of the mechanism driving such photoinduced reactions opens up the door to exploit them for our benefit. This thesis studies the interactions of ultraviolet light with the DNA/RNA molecules and the amino-acid tryptophan. Using some of the most accurate electronic structure methods and sophisticated environmental modelling, the works documented herein enable quantitative comparisons with cutting-edge experimental data. The relaxation pathways undertaken by the excited molecule are revealed through static and dynamical investigations of the excited-state potential energy surface. The profound role played by the dynamic response of the environment to guide the excitation in these timescales is addressed thoroughly.

Contents

List of Figures

List of Tables

Introduction

The interaction of ultraviolet light with biomolecules is responsible for triggering a diverse set of photochemical reactions from photosynthesis to photodamage of DNA. These interactions with UV light induce transitions to higher-lying electronic states of the system and the resulting dynamics involving transitions between these states govern the possible reaction processes. While these complex reactions may happen in nanosecond timescales, key mechanistic steps can already occur in the femtosecond regimes after photoexcitation. Internal conversions between the electronic states occurring in these early times after photoexcitation profoundly affect the fate of the behaviour of the excited system [1]. These femtosecond processes occurring at the timescales of the motions of the nuclei, are abundant in biological systems like the isomerization of rhodopsin triggering the mechanism of vision [2] or internal conversion back to the ground state in DNA/RNA nucleobases protecting them from UV induced damage [3].

A mechanistic understanding of these photoinduced processes holds the key to protection against them or exploiting them for our benefit. The experimental investigation of these processes occurring in femtosecond to picosecond regime is a non-trivial process needing ultrashort laser pulses to be able to resolve dynamics at these timescales. In the past 30 years development of such laser pulses and elaborate multi-pulse techniques have enabled the experimental resolution needed to track these processes [4, 5]. However, these spectroscopic techniques like femtosecond transient absorption employing sub 20-fs pulses [6] do not capture an image of the evolving molecule, but rather record its response to the perturbation with the sequence of laser pulses. This leads to time evolving spectral signals which are fingerprints of the underlying photoinduced processes occurring in the system. Thus, it requires solid theoretical modelling of the light-matter interactions to interpret the experimental signals.

In my PhD activity I have undertaken studies to model the interaction of ultraviolet (UV) light with biologically relevant organic chromophores, namely DNA/RNA nucleobases and tryptophan. All the studies included here focused on the computation of quantitatively accurate electronic properties within realistic environments to enable direct comparisons with modern ultrafast transient spectroscopic experiments. These works were performed under the aegis of Lightdynamics project supported by European Union's Horizon 2020 Research and Innovation Programme[7]. In these studies reported here I have tackled problems of increasing complexity utilizing the skills and techniques from foundational problems to tackle more complex cases. The works included in this thesis are arranged in order of complexity of computational effort.

Chapter 1

Modelling accurate electronic structure

1.1 Theory - *how to compute the electronic structure*

To model electronic interactions and the triggered photoinduced events, one needs to compute the excited-states of the system. These electronic states are the solutions to the time-independent Schrodinger equation.

$$H|\Psi_i\rangle = E_i|\Psi_i\rangle \tag{1.1}$$

where the Hamiltonian H contains all the interactions involving the electrons and nuclei(U terms in equation 1.3) and their kinetic energies(T terms in equation 1.3).

$$H = U_{nuc-nuc} + U_{nuc-el} + U_{el-el} + T_{el} + T_{nuc} \tag{1.2}$$

The solutions $|\Psi_i\rangle$ to this equation are the various eigenfunctions of the Hamiltonian operator. These wavefunctions encode all the information about the state and be used to extract physical observables. Any acceptable wavefunction for a system should be orthogonal to each other.

$$\langle\Psi_i|\Psi_j\rangle = 0 \tag{1.3}$$

Most quantum chemistry methods work in the Born-Oppenheimer approximation to solve the electronic structure of the system. This approximation results in clamping the nuclei at a fixed position and solving the schrodinger equation by removing the nuclear kinetic energy term from the Hamiltonian. This is known as the electronic Hamiltonian of the system

$$H_e = U_{nuc-nuc} + U_{nuc-el} + U_{el-el} + T_{el} \tag{1.4}$$

$$= \sum_{i,j} \frac{1}{R_{i,j}} - \sum_{i,j} \frac{Z_{\alpha,i}}{R_{\alpha,i}} + \sum_{\alpha,\beta} \frac{Z_{\alpha,\beta}}{R_{\alpha,\beta}} - \sum_{i} \frac{1}{2} \nabla_i^2 \tag{1.5}$$

where the index α, β denote nuclei and i, j denote the electrons.

Even after using the Born-Oppenheimer approximation to clamp down the nuclei, the Schrodinger equation can be solved analytically only for hydrogen-like systems with only one-electron. The solutions of these are the one-electron orbitals (χ_i) with associated energy eigenvalues. The problem with analytical solutions for multi-electron systems is inter-electron interaction term (U_{el-el}) in the Hamiltonian also called electron correlation. The road to a numerical solution of Schrodinger equation for these systems is by simplifying the electron correlation term.

1.1.1 Wavefunction-based methods

This section is a primer to wavefunction based methods of computing electronic structure focusing on RASSCF/RASPT2 protocols which are employed throughout the works documented in this thesis. Wavefunction based methods aim to obtain the wavefunction of the system which are eigenfunctions of the electronic Hamiltonian. The one electron orbitals (χ_i), which are solutions of the one-electron Hamiltonian can be used as a primitive component to build the orbitals of the multi-electron system(ϕ_l).

$$\phi_l(\boldsymbol{q_i}) = \left(\sum_{k=1}^{N} u_{kl} \chi_k(\boldsymbol{r_i}) \right) \theta(i) \tag{1.6}$$

The N-electron wavefunction of the system is built from these molecular orbitals. Since electrons are fermions, the resultant wavefunction needs to be invariant against electron permutations. This leads to the electronic wavefunction as a linear combination of anti-symmetrized products of molecular orbitals. These anti-symmetrized products can be expressed compactly in a determinant form called as Slater determinant.

$$\Psi^{el}(1,2,..,n) = \sum_{q}^{\infty} C_q \Psi_q^{SD} \;\; where \;\; \Psi_q^{SD} = \frac{1}{\sqrt{n!}} \begin{vmatrix} \phi_a(1) & \phi_a(2) & \phi_a(3) & \dots & \phi_a(n) \\ \phi_b(1) & \phi_b(2) & \phi_b(3) & \dots & \phi_b(n) \\ \vdots & \vdots & \ddots & & \vdots \\ \vdots & \vdots & & \ddots & \vdots \\ \phi_z(1) & \phi_z(2) & \phi_z(3) & \dots & \phi_z(n) \end{vmatrix} \tag{1.7}$$

Full Configuration Interaction As seen from the equation 1.7 the full electronic wavefunction of a system is a linear combination of many Slater determinants. This means more

than one arrangement of electrons in molecular orbitals is needed to properly express the multi-electron wavefunction in terms of one-electron orbitals. So, the full electronic wavefunction of any state(*even ground-state*) is multi-configurational in nature. The full wavefunction of the system in a particular basis of one-electron orbitals is expressed by linearly combining all possible electronic configurations. This is called as the **Full Configuration Interaction (Full CI)**. In this method, the wavefunction of the electronic state is a variational minima obtained by optimizing the coefficients C_q in equation 1.7 and u_{kl} in equation 1.6.

In computational implementation of these multi configurational methods, one works with Configuration State functions (CSF) rather than with Slater determinants. The CSF are spin and symmetry adapted linear combination of Slater determinants and using these avoids spin contamination issues in computations. Additionally most electronic structure programs express the primitive orbitals used to construct the basis-set with Gaussian functions called Gaussian-type orbitals(GTO). The true solutions of the hydrogenic Hamiltonian leads to Slater-type orbitals(STO), but integral operations using these are quite computationally inefficient. These integral operations computing electron correlation is the computational bottleneck of any electronic structure method. Using GTO leads to magnitudes of speedup in these kinds of integral computations compared to using STO. However, the functional form of GTO and STO differ significantly close to the nucleus. Thus to get accurate functional form of one-electron orbital several GTO (θ_t) have to linearly combined to generate atomic orbitals(χk).

$$\chi_k(\boldsymbol{r_i}) = \left(\sum_{t=1}^{N} d_{tk} \theta_t(\boldsymbol{s_i}) \right) \tag{1.8}$$

The computational efficiency of GTO means a sufficiently large number of them can be easily used to accurately model spatial electronic distribution.

CASSCF/RASSCF The Full CI method which fully describes the wavefunction in a given basis of orbitals is computationally quite expensive, because of the huge number of configurations generated when all possible arrangements of electrons in orbitals are considered. The variational minimimization of this needs optimization of huge number of coefficients which makes the problem numerically intractable. However, all the possible configurations do not participate equally in all the electronic states of the system. For electronic states lying in a given spectrum of energy, certain configurations will be dominant in the wavefunction. For example, for the ground-state and lowest excited-states, configuration involving zero occupations of core orbitals will contribute marginally to the wavefunctions of these states. This is also expressed as semantically partitioning of the electron correlation term into a static and dynamical part. The dominant configurations are said to describe static correlation and the rest describing dynamical correlation. Since the configurations describing dynamical correlation have small contributions to the wavefunction their effect can be described with perturbation correction to a wavefunction expressed only in terms of dominant configurations. These reduced number of configurations can be created by only allowing selected orbitals to generate various electronic arrangements. This

is done by partitioning the molecular orbitals into three spaces **Inactive**,**Active**, and **Virtual**. The inactive orbitals are always fully occupied and virtual orbitals always fully empty. Only the selection of orbitals in the active set called as **active-space** of the system are allowed to generate various configurations and the resulting procedure is called **CASSCF** (Complete-active space Self consistent field) [8, 9]. A CASSCF procedure is formally written as ($|m,n|$) which means the configurations are generated by permuting m electrons in n orbitals.

The flexibility of CASSCF procedure can be extended by truncating the number of excitations in the active-space. This extension of CASSCF is called RASSCF(Restricted Active Space SCF) [10]. In this procedure the active-space involving $q+n+s$ orbitals is furthur partitioned into three subspaces called RAS1, RAS2 and RAS3. The orbitals in RAS2 space can have all possible occupation numbers like the active-space orbitals of CASSCF. The RAS1 orbitals are doubly occupied except for a maximum number of holes and the RAS3 orbitals have zero occupation except for a maximum number of electrons. This is formally written as $p,q|m,n|r,s$, where there are q orbitals in RAS1, n orbitals in RAS2 and s orbitals in RAS3. The number of electrons involved in the RASSCF active-space are m and a maximum p holes are allowed in RAS1 and maximum r excitations are allowed in RAS3.

In photophysical studies one usually deals with multiple electronic states lying within a spectrum of energy rather than just a single electronic state. In these State-Averaged CASSCF is used to solve for the wavefunction of multiple electronic states. The variational problem in this case focuses on the minimisation of the state-averaged energy

$$E = \sum_P w_P E_P \tag{1.9}$$

where w_P is the weighted coefficient for individual state P. Usually equal weights are chosen for all the states in a state-averaged CASSCF computation.

CASPT2/RASPT2 RASSCF wavefunctions neglect the contributions of configurations involving orbitals residing in Inactive and Virtual spaces. The contribution of these configurations can be added as a perturbational correction. In perturbation theory the full Hamiltonian H can be decomposed into a zeroth-order part H_0 whose eigenstates Ψ_α^0 are known and the rest $V = H - H_0$ describing the perturbation. In CASPT2 theory the zeroth-order Hamiltonian is defined using the one-electron Fock operator ($\hat{f}$). The space spanned by the possible configurations is also partitioned into model-space P spanned by the CASSCF wavefunctions and orthogonal space Q spanned by the rest of configurations. The zeroth-order Hamiltonian is then constructed by using the corresponding projection operators $\hat{P}$ and $\hat{Q}$ to project the fock-operator onto these spaces.

$$\hat{H}^0 = \hat{P}\hat{f}\hat{P} + \hat{Q}\hat{f}\hat{Q} \tag{1.10}$$

In general, the CASSCF states are **not** eigenstates of this zeroth-order Hamiltonian. However, based on the magnitude of energetic separations of the CASSCF states and the couplings due to the zeroth order Hamiltonian, certain approximations can be employed which lead to various

flavours of CASPT2.

SS-CASPT2 These perturbational corrections can be implemented through CASPT2 [11, 12](or RASPT2 [13, 14]) methods which are based on second order perturbation theory. The CASPT2 corrections can be applied in two major type of flavours. SS-CASPT2(Single-State CASPT2) provides a second order correction to the energies of the CASSCF wavefunction though their interaction with the CASPT2 Hamiltonian. Since the CASSCF states do not include excitations involving Inactive/Virtual orbitals, the original CASSCF functions are not eigenfunctions of the CASPT2 zeroth-order Hamiltonian. But when the reference active-space captures major part of the static electron-correlation and the CASSCF states are well separated energetically, these states are good approximations to eigenfunctions of CASPT2 Hamiltonian and only corrections to their energies are needed to account for the effect of CASPT2 Hamiltonian. So the original CASSCF states are assumed to be eigenstates of the CASPT2 zeroth-order Hamiltonian, and only a second-order correction to their energy is computed.

(X)MS-CASPT2 When CASSCF states are energetically close or degenerate they can have strong interactions through the terms of CASPT2 Hamiltonian [15]. In this case, re-orthogonalization of the wavefunctions are needed to generate the orthogonal eigenfunctions of the CASPT2 Hamiltonian. These are implemented through the various multi-state versions of CASPT2. The original version of this was named as MS-CASPT2 [16] in which a *diagonal approximation* [17] was employed to simplify the CASPT2 computations. In this approximation a prior assumption is made that the original CASSCF states are eigenstates of the P space of the zeroth-order CASPT2 hamiltionian H^0. This means there are no couplings introduced between the CASSCF states due to the first term of equation 1.10. This allows for a state-specific partitioning of the Hamiltonian for each CASSCF state, allowing for which has proven to give more-accurate zeroth-order energies in case of states having differing nature or well separated in energy [18]. In XMS-CASPT2 this approximation isn't employed [18, 19]. This results in the same state-averaged zeroth order Hamiltonian for all the CASSCF states and initially the CASSCF states are rotated to diagonalize this part of the zeroth-order Hamiltonian. The advantages gained with the XMS-CASPT2 is smooth description of potential energy surfaces especially near regions of degeneracies [20]. However with smaller active-spaces, the use of a state-averaged zeroth-order Hamiltonian can lead to worse energetics compared to MS-CASPT2 [18].

1.1.2 Density Functional Theory(DFT) methods

Density functional theory is a road to electronic structure and properties separate from the wavefunction methods. In DFT the total electron density ρ is the central quantity which completely determines all the ground-state properties of the N-electron system. In their seminal paper [21] Hohenberg and Kohn showed that there exists a unique one to one mapping between an external potential (v_{ext}) and the ground-state electronic density ρ of N electrons in this potential. Since the external potential also uniquely defines the wavefunction of the ground-state Ψ_0, it means

there is unique mapping of the ground-state wavefunction Ψ_0 and ρ

$$v_{ext} \rightarrow \Psi_0 \rightarrow \rho \tag{1.11}$$

Since this is a unique and reversible mapping, it means the ground-state density ρ also completely determines the external potential v_{ext} which leads to the wavefunction Ψ_0 and thus all properties of the ground-state.

$$\rho \rightarrow v_{ext} \rightarrow \Psi_0 \rightarrow everything \tag{1.12}$$

The energy functional which gives the energy of the system is composed of the sum all the potential and kinetic energies of the Born-Oppenheimer electronic Hamiltonian

$$E[\rho] = T[\rho] + V_{ne}[\rho] + V_{ee}[\rho] + V_{nn}[\rho] \tag{1.13}$$

where terms on right are functionals for electronic kinetic-energy(T), nuclei-electron interaction(V_{ne}), electron-electron interaction(V_{ee}) and nuclei-electron interaction (V_{nn}). Hohenberg and Kohn defined an energy density functional comprised of just kinetic(T) and electron-electron interaction(V) and proved that there exists a variational theorem

$$F[\rho'] + \int v_{ext}\rho' \geq F[\rho] + \int v_{ext}\rho = E_0 \tag{1.14}$$

$$F[\rho] = T[\rho] + V_{ee}[\rho] \tag{1.15}$$

where ρ' is the density corresponding to some other external potential and E_0 is the true ground-state energy. This variational principle of DFT allows approaching the true energy through variational minimisation.

This is done through the Kohn-Sham approach [22], where a reference system of a single Slater determinant is considered. The total density of this system in terms of orthogonal orbitals ψ_i is

$$\rho = 2\sum_i |\psi_i|^2 \tag{1.16}$$

and its kinetic energy is

$$T_0 = -\frac{1}{2}\sum_i 2\int \psi_i^* \nabla \psi_i \tag{1.17}$$

and the Coulomb self energy is

$$J(\rho) = \frac{1}{2}\int\int \frac{\rho(1)\rho(2)}{r_{12}} d1d2. \tag{1.18}$$

.

The true Hohenberg-Kohn energy density functional F_ρ is expressed in terms of these func-

tionals and collecting the difference in the term E_{xc} called as exchange-correlation term.

$$F[\rho] = T_0[\rho] + J[\rho] + E_{xc}[\rho] \tag{1.19}$$

$$E_{xc}[\rho] = T[\rho] + V_{ee}[\rho] - T_0[\rho] - J[\rho] \tag{1.20}$$

The total Kohn-Sham energy can be expressed as

$$E[\rho] = T_0[\rho] + \int v_{ext}\rho + J[\rho] + E_{xc}[\rho] \tag{1.21}$$

whose variational minimization in terms of orbitals leads to the Kohn-Sham orbital equations

$$-\frac{1}{2}\nabla^2\psi_i + v_{KS}\psi_i = \epsilon_i\psi_i \tag{1.22}$$

with the Kohn-Sham effective potential v_{KS} given by

$$v_{KS} = v_{ext} + v_{el} + \frac{\delta E_{xc}}{\delta\rho} \tag{1.23}$$

Since for the Slater determinant J and T_0 are known, E_{xc} is the only unknown functional in the equations. Unfortunately the general form E_{xc} is not known and over the years a huge bazaar of functionals have arisen which sit on different rungs of ladder to heaven of chemical accuracy [23].

1.1.3 Time-Dependent Density functional theory

The time-dependent route to obtain excited-states requires the time-evolution of the wavefunction under the action of external electromagnetic radiation. In their seminal paper [24] Runge and Gross outlined the basic principles enabling the computation of excited-states in the formalism of density functional theory of using electron density as a central quantity. The Runge-Gross theorem, states that the time-dependent density of the system determines the time-dependent external potential and the wave-function with a phase-factor. Therefore the time-dependent wavefunction is a functional of the density of the system.

$$\rho(r,t) \longleftrightarrow v[\rho](r,t) + C(t) \longleftrightarrow \Psi[\rho](r,t)\exp^{-i\alpha(t)} \tag{1.24}$$

Runge and Gross also formulated a variational principle based on the quantum-mechanical action integral $A[\rho(r,t)]$ to enable the computation of the exact time-dependent density of the system. The action integral is given by

$$A[\rho(r,t)] = \int_{t_0}^{t_1} dt\langle\Psi[\rho(r,t)]i\frac{\partial}{\partial t} - H(r,t)|\Psi|\rho(r,t)]\langle \tag{1.25}$$

The stationary point of this action-integral with respect to variance of density gives the exact time-dependent density of the system.

The expression of the time dependent density in terms of electron orbitals is done through time-dependent formulation of Kohn-Sham equations . A fictitious system of non-interacting particles have the same electron density as the real system can be expressed as a linear combination of one-electron orbitals.

$$\rho(r,t) + \sum_{i=1}^{n} |\phi_i(r,t)|^2 \tag{1.26}$$

This leads to the expression of the time-dependent Kohn-Sham equations

$$i\hbar \frac{\partial \phi_i(\vec{r},t)}{\partial t} = \hat{h}^{KS}[\rho(t)]\phi_i(\vec{r},t) \tag{1.27}$$

where the Kohn-Sham $\hat{h}^{KS}$ operator now has dependence on the external potential $v_{ext}(t)$.

The computation of excited-states can be done by propagating the density through time-dependent Kohn-Sham equations. But most implementations of TD-DFT employs the assumptions of Linear Response of a system. This leads to a non-Hermitian eigenvalue equation for an excitation energy ω_i

$$\begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{B*} & \mathbf{A*} \end{bmatrix} \begin{pmatrix} \mathbf{X}_I \\ \mathbf{Y}_I \end{pmatrix} = \omega_I \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \tag{1.28}$$

where the elements of matrices A and B are given by

$$\begin{aligned} A_{ijab} &= \delta_{ij}\delta_{ab}(\epsilon_a - \epsilon_i) + \langle ij|ab\rangle + \langle ij|f_{xc}|ab\rangle \\ B_{ijab} &= \langle ib|aj\rangle + \langle ib|f_{xc}|aj\rangle \end{aligned} \tag{1.29}$$

where matrix elements $\mathbf{X}$ and $\mathbf{Y}$ describe excitations and de-excitations and f_{xc} is the exchange-correlation kernel. The exchange-correlation kernels developed for ground-state DFT can be used if one employs an adiabatic approximation which neglects the *memory effect* of the time-evolution of density.

An additional approximation employed in TD-DFT computations neglects the de-excitations in the eignenvalue problem posed in equation 1.28 called as the Tamm-Dancoff approximation. This simplifies the the eigenvalue problem to

$$\mathbf{AX} = \omega\mathbf{X} \tag{1.30}$$

1.2 Project - *Benchmarking TD-DFT methods for Transient Spectra of Nucleobases*

In this work an extensive characterisation of singlet excited manifold of the five canonical nucleobases (1.1) was performed with TDDFT and RASPT2 based methods. The work focused on the

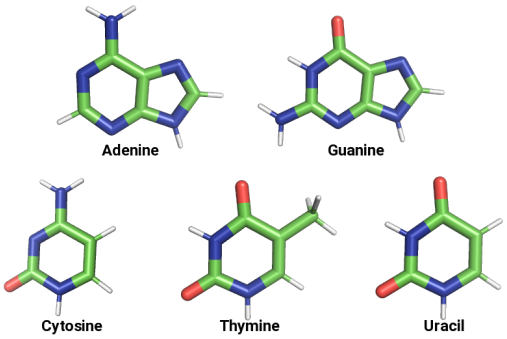

Figure 1.1: Nucleobases used in this study

theoretical modelling of excited-state absorption signals which are obtained in ultrafast transient spectroscopic experiments [2, 4, 25–27]. RASPT2 computations provide a reference benchmark for these computations against which TDDFT flavours can be assessed for their suitability. Two commonly used TDDFT functionals B3LYP [28, 29] and CAM-B3LYP [30] were employed in this study to model transient spectroscopic signals arising from the lowest singlet excited-states of the five nucleobases. Two basis-sets 6-31G**[31–35] and 6-311++G**[31–33] were employed in the TD-DFT computations. For pyrimidines(cytosine, uracil,thymine) only the absorption arising from the lowest singlet excited-state was modelled, while for purines(adenine and guanine) ESA signals were modelled from the two lowest singlet excited-states. The computations were performed in gas-phase on ground-state minimum structures optimised at CASSCF level.

The RASPT2 computations used as benchmark here were performed by Nenov et al.[36–38] in a previously published study using a systematically expanded active-space to capture maximum possible static correlation in the MCSCF computations. This enables quantitative accuracy comparable with experimental data after SS-CASPT2 corrections to the energy. Only $\pi\pi^*$ states were considered in this work as symmetry considerations render $n\pi*$ states as dark in both linear absorption and ESA from the lowest $\pi\pi^*$ state. The active-space in these computations included all the valence π orbitals of the system in the RAS2 space allowing to capture the full electronic correlation this space. This active-space was then augmented by adding 12 additional π orbitals in RAS3 space and allowing upto two excitation in these orbitals. This results in variational estimate in the energetic contribution arising from these orbitals and avoids the overestimation of their energetic contribution in the subsequent CASPT2 computation. This recipe of using an expanded active-space CASSCF+SS-CASPT2 results in accurate vertical excitation energies without using the empirical IPEA correction [39] which is commonly used in CASPT2 computations to obtain agreement with experimental values, but whose use for organic molecules has been discouraged in recent works [40].

1.2.1 Adenine

The RASPT2 linear absorption signals (**Figure 1.2**) show excellent agreement with the available experimental data validating their usage as the reference to benchmark TDDFT computations.

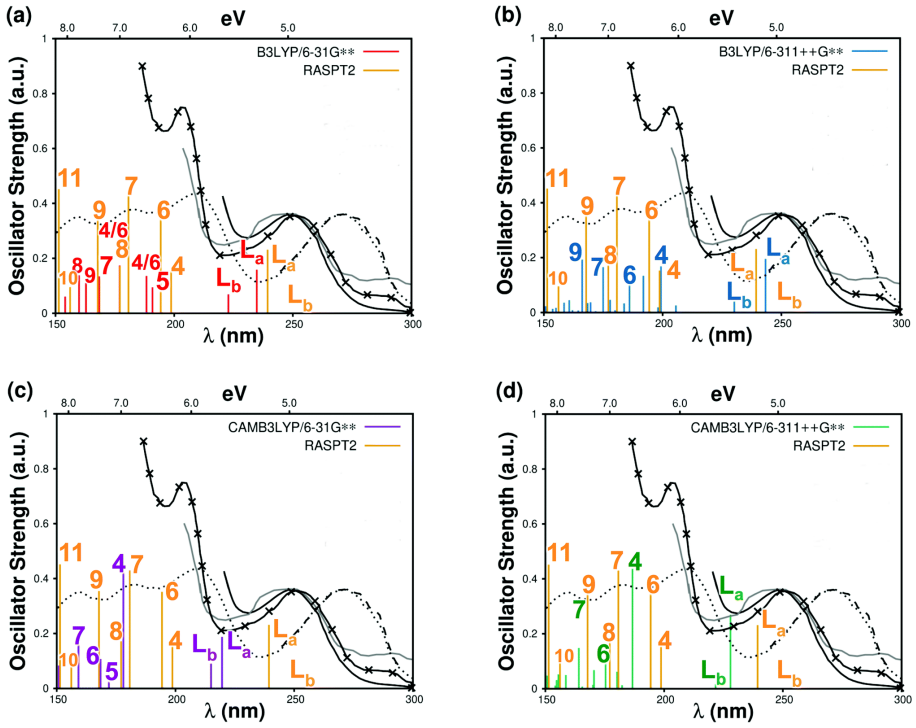

Figure 1.2: Experimental absorption spectrum in vapour phase (solid black line from [41]), solid gray line from [42]), matrix-isolated spectra (crossed black line from [43]) and from sublimed films (dotted black line from [44] and [45])., compared with vertical $S_0 \rightarrow S_n$ excitations (colored sticks) of adenine in gas-phase. Reference theoretical values at RASPT2 (yellow sticks) are compared with (a) B3LYP/6-31G**, (b) B3LYP/6-311++G**, (c) CAM-B3LYP/6-31G** and (d) CAM-B3LYP/6-311++G**. Only $\pi\pi^*$ states are labeled, according to their root number (*i.e.* relative energy) in the reference RASPT2 computations.

The multiple experimental plots from different sources shown in **Figure 1.2** have been scaled to match with intensity of the first absorption peak, and correspond to relative intensity in arbitrary units. The figure compares all the TDDFT flavours used in this study (two DFT functionals and basis-sets) against the RASPT2 reference data. The relative blue-shifted signals obtained when using smaller 6-31G**[31–35] basis-set and CAM-B3LYP functional reflect a universal trend seen also in subsequent computations. As seen in **Figure 1.2** , TDDFT computations in all cases result in reversed order of the two lowest excited-states of adenine (labeled La and Lb) in disagreement with RASPT2 and other correlated methods like coupled-cluster. At higher energy windows above 5.5 eV where higher absorption bands of adenine are present, several excitations are located by RASPT2 computations. In this spectral window too the best agreement with RASPT2 is obtained by B3LYP functional using the larger 6-311++G**[31–33] basis-set.

The ESA signals computed from the two lowest singlet states of adenine (La and Lb) show striking differences(**Figure 1.3**). In the near-IR probing window (1000-1200nm), the La-ESA shows a prominent band due to overlapping signals from transitions to states labeled as 4,5 and 6 in RASPT2 contrasting with a much weaker transition in the case of Lb-ESA. The trends

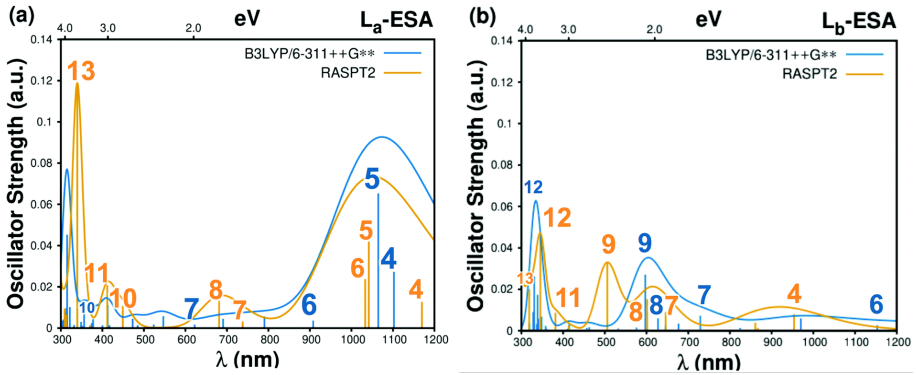

Figure 1.3: Computed ESAs associated to vertical $S_n \rightarrow S_m$ excitations (colored sticks) and corresponding convoluted spectra (colored lines) for adenine in gas-phase. Reference theoretical values at RASPT2 (yellow sticks) are compared with B3LYP/6-311++G** (blue sticks) computations for (a) S_n = La and (b) S_n = Lb. Only $\pi\pi^*$ states are labeled, according to root numbers in the reference RASPT2 computations.

and positions of ESA signals in this probing window are reasonably accurate when employing the B3LYP/6-311++G** basis-set. In the UV probing window RASPT2 computations show more intense signals arising from La-ESA. In this energy window transitions to states having double excitation are involved (state 13), which can not be accounted by TDDFT computations. Hence, TDDFT computations miss some excited-states lying in this energy window leading to comparatively weaker ESA signals in case of La-ESA. Overall B3LYP/6-311++G** provides convoluted spectra which are in qualitative agreement with RASPT2 calculations for excited states lying below 7.5 eV (as for linear absorption), and therefore for spectral windows probing the ESAs of La and Lb states in both near-IR and Vis regions. At higher-energies, in the UV probing window, TD-DFT computations are generally facing the limitation of missing double excitations, particularly important for the La-ESAs.

1.2.2 Guanine

In contrast to adenine, the two lowest singlet states of guanine have the same energetic order in TDDFT and RASPT2 computations (**Figure1.4**). The energy of Lb state is in good agreement with RASPT2 but the energy of La is slightly blue-shifted. It is to be remarked here that the La/Lb notation is used to be consistent with previous literature, but these states involve different electronic excitations in adenine vs guanine. The higher experimental linear absorption bands feature a shoulder at 200 nm and an intense band at 176 nm all of which can be located with RASPT2 transitions. At TDDFT level, the all the transitions corresponding to these bands are located. However, the transitions corresponding to the most intense band at 176 nm is slightly blue-shifted with lower oscillator strength at TDDFT level.

In the near-IR probing window at RASPT2 level, Lb-ESA features two transitions to states 5 and 7 while La-ESA has only one significant transition to state 4. At B3LYP/6-311++G** level, these corresponding transitions are blue-shifted in case of Lb-ESA and red-shifted in case of La-

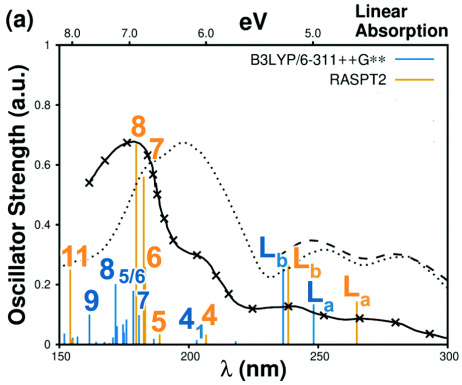

Figure 1.4: Experimental absorption spectra of guanine in nitrogen-matrix (crossed black line from [46]), from sublimed films (dotted black line from [44] dashed black line from [45]) and computed vertical $S_0 \rightarrow S_n$ excitations (colored sticks) of guanine in gas-phase. Reference theoretical values at RASPT2 (yellow sticks) are compared with B3LYP/6-311++G** (blue sticks).

ESA with weaker oscillator strengths. In the near-UV probing region, while reference RASPT2 computations of La-ESAs feature multiple $\pi\pi^*$ transitions with high TDMs (states 11, 12, 13), all TD-DFT methods predict only the transition involving state 11. In the case of Lb-ESAs, the manifold of bright states predicted by RASPT2 are not found at the TD-DFT levels in this probing window, as they are most likely shifted to higher energies or they have some contribution from double excitations. As remarked earlier, employing CAM-B3LYP functional blue-shifts all ESAs of both La and Lb states with respect to B3LYP.

1.2.3 Pyrimidines

The single-ring structure of pyrimidines leads to a simpler singlet state manifold due to fewer transitions in the same energetic window than purines. In the energetic window considered till 150 nm for linear absorption, there are five transitions located at RASPT2 level for all the pyrimidines(*incase of uracil the last transition at 147 nm falls outside the plotting window*). All these corresponding transitions are located at B3LYP/6-311++G** level. When comparing against RASPT2 energies there is a red-shift of B3LYP energies for all the transitions in case of cytosine. For the other two nucleobases, there are no such systematic discrepancies. In case of uracil the all B3LYP energies match quite well with RASPT2 while for thymidine state 5 displays significantly blue-shift compared to RASPT2. The use of smaller basis-set or CAM-B3LYP functional leads to blue-shifts in energies consistent with that of purines.

Only the ESA from the lowest singlet is considered for pyrimidines as the second excited-state is well separated energetically from the lowest singlet excited state. The agreement between RASPT2 and B3LYP/6-311++G** level is reasonably good for the near-IR to Visible probing window. In these probing windows RASPT2 locates weak transitions from the lowest singlet excited-state. All the corresponding transitions in this probing window are located at B3LYP with red-shifts in energetic positions. In the UV probing window more intense ESA signals are

located at RASPT2 level. In case of uracil while the visual agreement is good between RASPT2 and B3LYP the underlying transitions are different. Thymine has multiple significant transitions in the UV window at RASPT2 level in contrast to B3LYP which locates only a single intense ESA signal. For cytosine the agreement between RASPT2 and B3LYP is better in the UV window with both methods locating an intense signal involving the same transition.

1.2.4 Conclusions

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

This work has been published as Vishal K. Jaiswal, Javier Segarra-Martí, Marco Marazzi, Elena Zvereva, Xavier Assfeld, Antonio Monari, Marco Garavelli, and Ivan Rivalta. First-principles characterization of the singlet excited state manifold in dna/rna nucleobases. *Phys. Chem. Chem. Phys.*, 22:15496–15508, 2020. doi: 10.1039/D0CP01823F. URL `http://dx.doi.org/10.1039/D0CP01823F`

Table S1: Adenine $S_0 \rightarrow S_n$ vertical transition energies (in eV) and corresponding oscillator strengths in parentheses at various levels of theory.

State (n+1)	RASPT2	B3LYP 6-31G**	B3LYP 6-311++G**	CAMB3LYP 6-31G**	CAMB3LYP 6-311++G**
2 (Lb)	4.94(*0.00*)	5.56(*0.07*)	5.38(*0.04*)	5.77(*0.08*)	5.59(*0.02*)
3 (La)	5.18(*0.23*)	5.28(*0.16*)	5.09(*0.20*)	5.64(*0.18*)	5.44(*0.27*)
4	6.24(*0.15*)	-	6.22(*0.17*)	6.96(*0.42*)	6.64(*0.44*)
5	6.37(*0.00*)	6.49(*0.09*)	6.26(*0.02*)	7.21(*0.03*)	7.29(*0.01*)
6	6.38(*0.35*)	6.58(*0.14*),7.00(*0.17*)[#]	6.66(*0.10*)	7.36(*0.11*)	7.07(*0.10*)
7	6.86(*0.43*)	7.36(*0.13*)	7.08(*0.17*)	7.79(*0.16*)	7.56(*0.15*)
8	7.00(*0.17*)	7.75(*0.13*)	7.36(*0.03*)	8.51(*0.10*)	8.51(*0.00*)
9	7.39(*0.35*)	7.61(*0.11*)	7.46(*0.19*)	8.22(*0.08*)	7.97(*0.05*)
10	7.94(*0.09*)	8.87(*0.04*)	8.57(*0.00*)	9.59(*0.01*)	9.36(*0.03*)
11	8.19(*0.45*)	9.10(*0.18*)	8.37(*0.12*)	9.29(*0.13*)	9.24(*0.08*)
12	8.53(*0.13*)	9.27(*0.03*)	9.13(*0.03*)		
13	8.83(*0.02*)				

[#] The two values relate to the presence of two states featuring the same primary electronic excitation that associates to the corresponding state at RASPT2 reference

Table S2: Adenine $L_a \rightarrow S_n$ vertical transition energies (in eV) and corresponding oscillator strengths in parentheses at various levels of theory.

State (n+1)	RASPT2	B3LYP 6-31G**	B3LYP 6-311++G**	CAMB3LYP 6-31G**	CAMB3LYP 6-311++G**
4	1.06(*0.01*)	-	1.12*(0.03)*	1.32(*0.01*)	1.20(*0.00*)
5	1.19(*0.04*)	1.22(*0.06*)	1.16(*0.07*)	1.56(*0.07*)	1.85(*0.01*)
6	1.20(*0.02*)	1.30(*0.02*),1.72(*0.01*)[#]	1.57(*0.01*)	1.71(*0.01*)	1.63(*0.02*)
7	1.68(*0.00*)	2.08(*0.00*)	1.99(*0.00*)	2.14(*0.00*)	2.12(*0.00*)
8	1.82(*0.01*)	2.47(*0.01*)	2.27(*0.01*)	2.87(*0.02*)	3.07(*0.00*)
9	2.21(*0.00*)	2.33(*0.00*)	2.37(*0.00*)	2.57(*0.01*)	2.54(*0.01*)
10	2.76(*0.01*)	3.60(*0.01*)	3.48(*0.01*)	3.94(*0.03*)	3.92(*0.02*)
11	3.01(*0.02*)	3.83(*0.00*)	3.28(*0.00*)	3.65(*0.01*)	3.80(*0.00*)
12	3.35(*0.00*)	3.99(*0.01*)	4.04(*0.00*)		
13	3.65(*0.12*)				

[#] The two values relate to the presence of two states featuring the same primary electronic excitation that associates to the corresponding state at RASPT2 reference

Table S3: Adenine $L_b \rightarrow S_n$ vertical transition energies (in eV) and corresponding oscillator strengths in parentheses at various levels of theory.

State (n+1)	RASPT2	B3LYP 6-31G**	B3LYP 6-311++G**	CAMB3LYP 6-31G**	CAMB3LYP 6-311++G**
4	1.30(*0.01*)	-	*0.83(0.00)*	1.20(*0.01*)	1.05(*0.00*)
5	1.43(*0.00*)	0.93(*0.00*)	0.87(*0.00*)	1.44(*0.01*)	1.70(*0.01*)
6	1.44(*0.00*)	1.02(*0.00*),1.44(*0.00*)[#]	1.28(*0.01*)	1.59(*0.01*)	1.48(*0.01*)
7	1.92(*0.01*)	1.80(*0.01*)	1.70(*0.01*)	2.02(*0.01*)	1.96(*0.01*)
8	2.06(*0.02*)	2.19(*0.02*)	1.98(*0.01*)	2.75(*0.01*)	2.92(*0.00*)
9	2.45(*0.03*)	2.05(*0.00*)	2.08(*0.03*)	2.45(*0.02*)	2.38(*0.00*)
10	3.00(*0.00*)	3.31(*0.01*)	3.19(*0.00*)	3.82(*0.01*)	3.77(*0.00*)
11	3.25(*0.01*)	3.54(*0.00*)	2.99(*0.00*)	3.53(*0.00*)	3.64(*0.00*)
12	3.59(*0.05*)	3.71(*0.03*)	3.74(*0.03*)		
13	3.89(*0.02*)				

[#] The two values relate to the presence of two states featuring the same primary electronic excitation that associates to the corresponding state at RASPT2 reference

Table S4: Guanine $S_0 \rightarrow S_n$ vertical transition energies (in eV) and corresponding oscillator strengths in parentheses at various levels of theory.

State (n+1)	RASPT2	B3LYP 6-31G**	B3LYP 6-311++G**	CAMB3LYP 6-31G**	CAMB3LYP 6-311++G**
2 (La)	4.68(*0.14*)	5.20(*0.16*)	4.99(*0.13*)	5.45(*0.18*)	5.24(*0.16*)
3 (Lb)	5.20(*0.27*)	5.52(*0.19*)	5.24(*0.26*)	5.94(*0.27*)	5.63(*0.33*)
4	6.00(*0.03*)	6.47(*0.00*)	6.11(*0.02*),6.17(0.00)[#]	7.01(*0.01*)	6.59(*0.02*)
5	6.57(*0.26*)	7.22(*0.29*)	-	7.81(*0.15*)	7.55(*0.10*)
6	6.77(*0.56*)	7.35(*0.17*)	-	7.87(*0.25*)	7.53(*0.13*)
7	6.79(*0.67*)	7.11(*0.05*)	6.86(*0.10*)	7.65(*0.30*)	7.33(*0.24*)
8	6.90(*0.01*)	7.56(*0.14*)	7.22(*0.20*)	8.06(*0.06*)	7.84(*0.10*)
9	7.32(*0.02*)	7.91(*0.10*)	7.67(*0.10*)	8.47(*0.07*)	8.13(*0.03*)
10	8.00(*0.25*)	8.20(*0.00*)	7.82(*0.00*)	9.06(*0.03*)	8.71(*0.05*)
11	8.04(*0.04*)	8.70(*0.07*)	8.27(*0.15*)	9.17(*0.03*)	8.82(*0.02*)
12	8.28(*0.01*)	8.52(*0.02*)	8.16(*0.04*)	9.79(*0.06*)	
13	8.51(*0.06*)	9.78(*0.03*)		10.38(*0.03*)	
14	8.92(*0.08*)	10.01*(0.14)*		10.87(*0.00*)	

[#] The two values relate to the presence of two states featuring the same primary electronic excitation that associates to the corresponding state at RASPT2 reference

Table S5: Guanine $L_a \rightarrow S_n$ vertical transition energies (in eV) and corresponding oscillator strengths in parentheses at various levels of theory.

State (n+1)	RASPT2	B3LYP 6-31G**	B3LYP 6-311++G**	CAMB3LYP 6-31G**	CAMB3LYP 6-311++G**
4	1.32(0.01)	1.27(0.02)	1.11(*0.00*),1.17(*0.00*)[#]	1.57(0.02)	1.34(0.01)
5	1.89(0.01)	2.02(0.01)	-	2.36(0.01)	2.31(0.01)
6	2.09(0.03)	2.15(0.01)	-	2.42(0.02)	2.29(0.02)
7	2.11(0.01)	1.91(0.00)	1.86(0.00)	2.20(0.00)	2.09(0.00)
8	2.22(0.00)	2.36(0.00)	2.23(0.01)	2.62(0.00)	2.59(0.00)
9	2.64(0.03)	2.71(0.04)	2.68(0.04)	3.03(0.06)	2.88(0.02)
10	3.32(0.00)	3.00(0.01)	2.83(0.00)	3.61(0.01)	3.47(0.02)
11	3.36(0.02)	3.50(0.01)	3.27(0.02)	3.73(0.04)	3.58(0.02)
12	3.60(0.06)	3.32(0.01)	3.17(0.00)	4.35(0.00)	-
13	3.83(0.04)	4.58(0.00)	-	4.93(0.00)	-
14	4.24(0.01)	4.81(0.00)	-	5.42(0.05)	-

[#] The two values relate to the presence of two states featuring the same primary electronic excitation that associates to the corresponding state at RASPT2 reference

Table S6: Guanine $L_b \rightarrow S_n$ vertical transition energies (in eV) and corresponding oscillator strengths in parentheses at various levels of theory.

State (n+1)	RASPT2	B3LYP 6-31G**	B3LYP 6-311++G**	CAMB3LYP 6-31G**	CAMB3LYP 6-311++G**
4	0.80(0.03)	0.94(0.04)	0.86(*0.03*),0.92(*0.03*)[#]	1.07(0.05)	0.96(0.07)
5	1.37(0.04)	1.69(0.01)	-	1.87(0.01)	1.92(0.03)
6	1.57(0.00)	1.83(0.01)	-	1.93(0.02)	1.91(0.00)
7	1.59(0.03)	1.58(0.01)	1.61(0.01)	1.71(0.01)	1.70(0.01)
8	1.70(0.01)	2.03(0.01)	1.98(0.02)	2.12(0.02)	2.21(0.01)
9	2.12(0.00)	2.39(0.03)	2.43(0.01)	2.53(0.00)	2.50(0.00)
10	2.80(0.01)	2.67(0.03)	2.58(0.03)	3.12(0.11)	3.08(0.05)
11	2.84(0.07)	3.18(0.00)	3.02(0.00)	3.23(0.00)	3.20(0.03)
12	3.08(0.02)	2.99(0.00)	2.92(0.00)	3.85(0.01)	-
13	3.31(0.00)	4.25(0.00)	-	4.44(0.01)	-
14	3.72(0.03)	4.49(0.01)	-	4.93(0.02)	-

[#] The two values relate to the presence of two states featuring the same primary electronic excitation that associates to the corresponding state at RASPT2 reference

Table S7: Uracil $S_0 \rightarrow S_n$ vertical transition energies (in eV) and corresponding oscillator strengths in parentheses at various levels of theory.

State (n+1)	RASPT2	B3LYP 6-31G**	B3LYP 6-311++G**	CAMB3LYP 6-31G**	CAMB3LYP 6-311++G**
2	5.20(*0.17*)	5.41(*0.12*)	5.22(*0.12*)	5.70(*0.17*)	5.47(*0.18*)
3	6.18(*0.04*)	6.09(*0.04*)	5.97(*0.03*)	6.78(*0.04*)	6.63(*0.04*)
4	6.55(*0.18*)	6.85(*0.11*)	6.53(*0.12*)	7.29(*0.14*)	6.89(*0.16*)
5	7.39(*0.71*)	7.89(*0.17*)	7.52(*0.37*)	8.16(*0.41*)	7.88(*0.33*)
6	8.42(*0.05*)	9.27(*0.08*)	9.13(*0.10*)	10.01(*0.08*)	10.03(*0.05*)

Table S8: Uracil $S_1 \rightarrow S_n$ vertical transition energies (in eV) and corresponding oscillator strengths in parentheses at various levels of theory.

State (n+1)	RASPT2	B3LYP 6-31G**	B3LYP 6-311++G**	CAMB3LYP 6-31G**	CAMB3LYP 6-311++G**
4	1.35(*0.00*)	1.44(*0.01*)	1.31(*0.00*)	1.59(*0.01*)	1.42(*0.01*)
5	2.19(*0.01*)	2.39(*0.00*)	2.30(*0.00*)	2.46(*0.00*)	2.40(*0.00*)
6	3.22(*0.08*)	2.53(*0.02*)	3.91(*0.02*)	4.32(*0.05*)	4.56(*0.00*)
7	3.29(*0.00*)	3.14(*0.05*)	3.13(*0.05*)	3.47(*0.04*)	3.44(*0.07*)

Table S9: Thymine $S_0 \rightarrow S_n$ vertical transition energies (in eV) and corresponding oscillator strengths in parentheses at various levels of theory.

State (n+1)	RASPT2	B3LYP 6-31G**	B3LYP 6-311++G**	CAMB3LYP 6-31G**	CAMB3LYP 6-311++G**
2	5.00(*0.17*)	5.30(*0.12*)	5.06(*0.13*)	5.60(*0.18*)	5.33(*0.18*)
3	6.20(*0.11*)	6.36(*0.07*)	6.08(*0.05*)	6.90(*0.07*)	6.70(*0.05*)
4	6.51(*0.17*)	6.74(*0.14*)	6.46(*0.12*)	7.18(*0.18*)	6.83(*0.20*)
5	7.32(*0.89*)	7.91(*0.34*)	7.81(*0.31*)	8.27(*0.40*)	7.94(*0.23*)
6	8.06(*0.38*)	8.27(*0.06*)	8.06(*0.05*)	8.92(*0.07*)	8.70(*0.08*)

Table S10: Thymine $S_1 \rightarrow S_n$ vertical transition energies (in eV) and corresponding oscillator strengths in parentheses at various levels of theory.

State (n+1)	RASPT2	B3LYP 6-31G**	B3LYP 6-311++G**	CAMB3LYP 6-31G**	CAMB3LYP 6-311++G**
3	1.20(*0.01*)	0.96(*0.01*)	1.02(*0.01*)	1.30(*0.01*)	1.37(*0.01*)
4	1.51(*0.00*)	1.45(*0.00*)	1.40(*0.01*)	1.58(*0.01*)	1.51(*0.01*)
5	2.32(*0.00*)	2.61(*0.00*)	2.75(*0.00*)	2.67(*0.00*)	2.61(*0.00*)
6	3.06(*0.02*)	2.98(*0.09*)	3.00(*0.12*)	3.32(*0.10*)	3.37(*0.12*)

Table S11: Cytosine $S_0 \rightarrow S_n$ vertical transition energies (in eV) and corresponding oscillator strengths in parentheses at various levels of theory.

State (n+1)	RASPT2	B3LYP 6-31G**	B3LYP 6-311++G**	CAMB3LYP 6-31G**	CAMB3LYP 6-311++G**
2	4.66(*0.05*)	4.74(*0.13*)	5.65(*0.03*)	5.12(*0.05*)	5.00(*0.106*)
3	5.59(*0.10*)	5.66(*0.08*)	5.48(*0.06*)	6.15(*0.12*)	65.94(*0.11*)
4	6.46(*0.83*)	6.88(*0.23*)	6.34(*0.21*)	7.16(*0.46*)	6.70(*0.15*)
5	6.90(*0.30*)	7.10(*0.38*)	6.69(*0.33*)	7.46(*0.21*)	6.96(*0.11*)
6	8.01(*0.24*)	8.02(*0.14*)	7.87(*0.15*)	8.61(*0.18*)	8.45(*0.16*)

Table S12: Cytosine $S_1 \rightarrow S_n$ vertical transition energies (in eV) and corresponding oscillator strengths in parentheses at various levels of theory.

State (n+1)	RASPT2	B3LYP 6-31G**	B3LYP 6-311++G**	CAMB3LYP 6-31G**	CAMB3LYP 6-311++G**
3	0.93(*0.01*)	0.91(*0.01*)	0.82(*0.01*)	1.02(*0.01*)	0.94(*0.01*)
4	1.91(*0.00*)	2.14(*0.00*)	1.69(*0.00*)	2.04(*0.00*)	1.70(*0.00*)
5	2.24(*0.03*)	2.36(*0.01*)	2.03(*0.00*)	2.34(*0.01*)	1.96(*0.00*)
6	3.35(*0.09*)	3.28(*0.05*)	3.22(*0.05*)	3.49(*0.06*)	3.45(*0.08*)

ADENINE

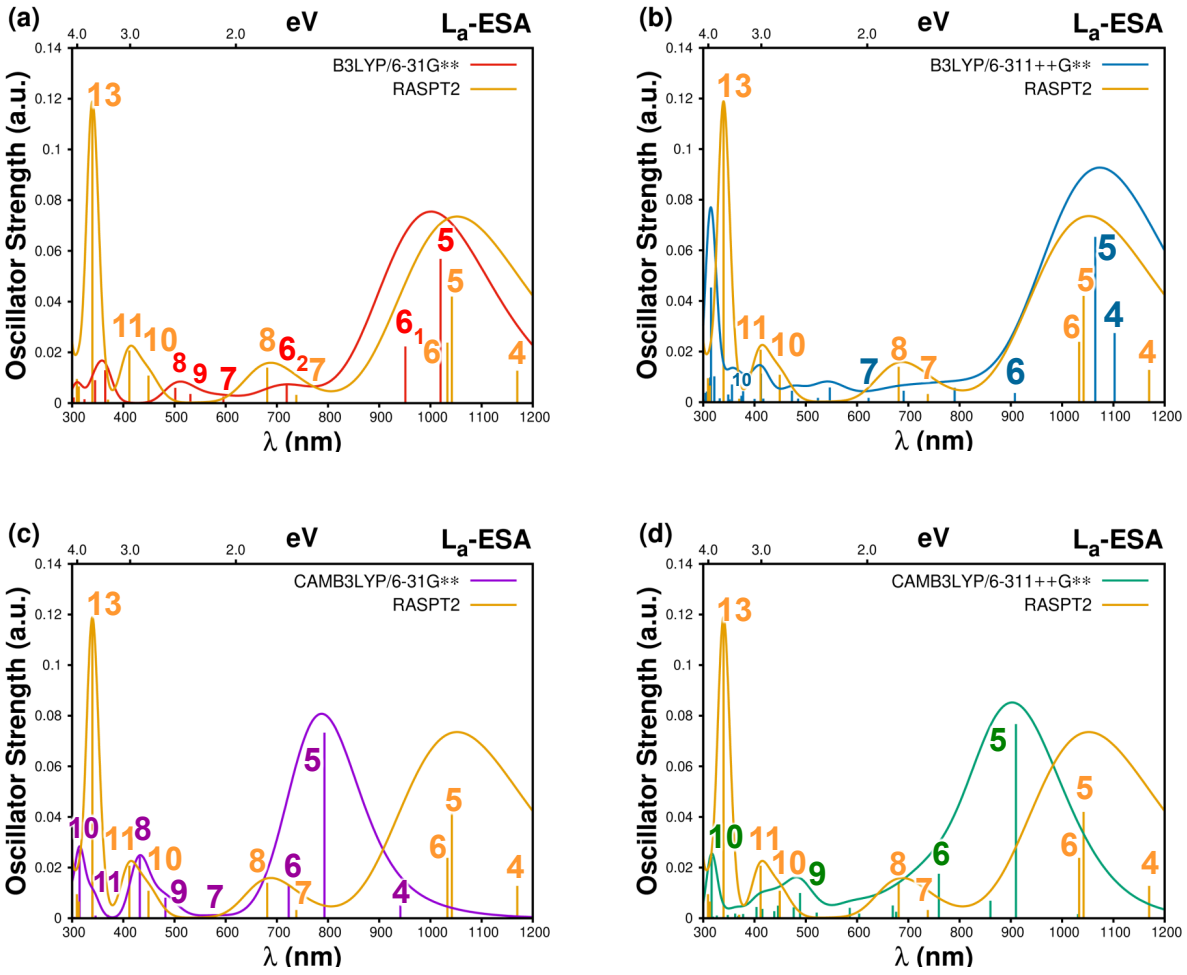

Figure S1. Computed ESAs associated to vertical $L_a \rightarrow S_m$ excitations (colored sticks) of adenine in gas-phase. Reference theoretical values at RASPT2 (yellow sticks) are compared with B3LYP/6-31G**(red sticks), B3LYP/6-311++G** (blue sticks), CAMB3LYP/6-31G**(violet sticks) and CAMB3LYP/6-311++G** (green sticks) and corresponding convoluted spectra (colored lines). Only ππ* states are labeled, according to root numbers in the reference RASPT2 computations (Table S2)

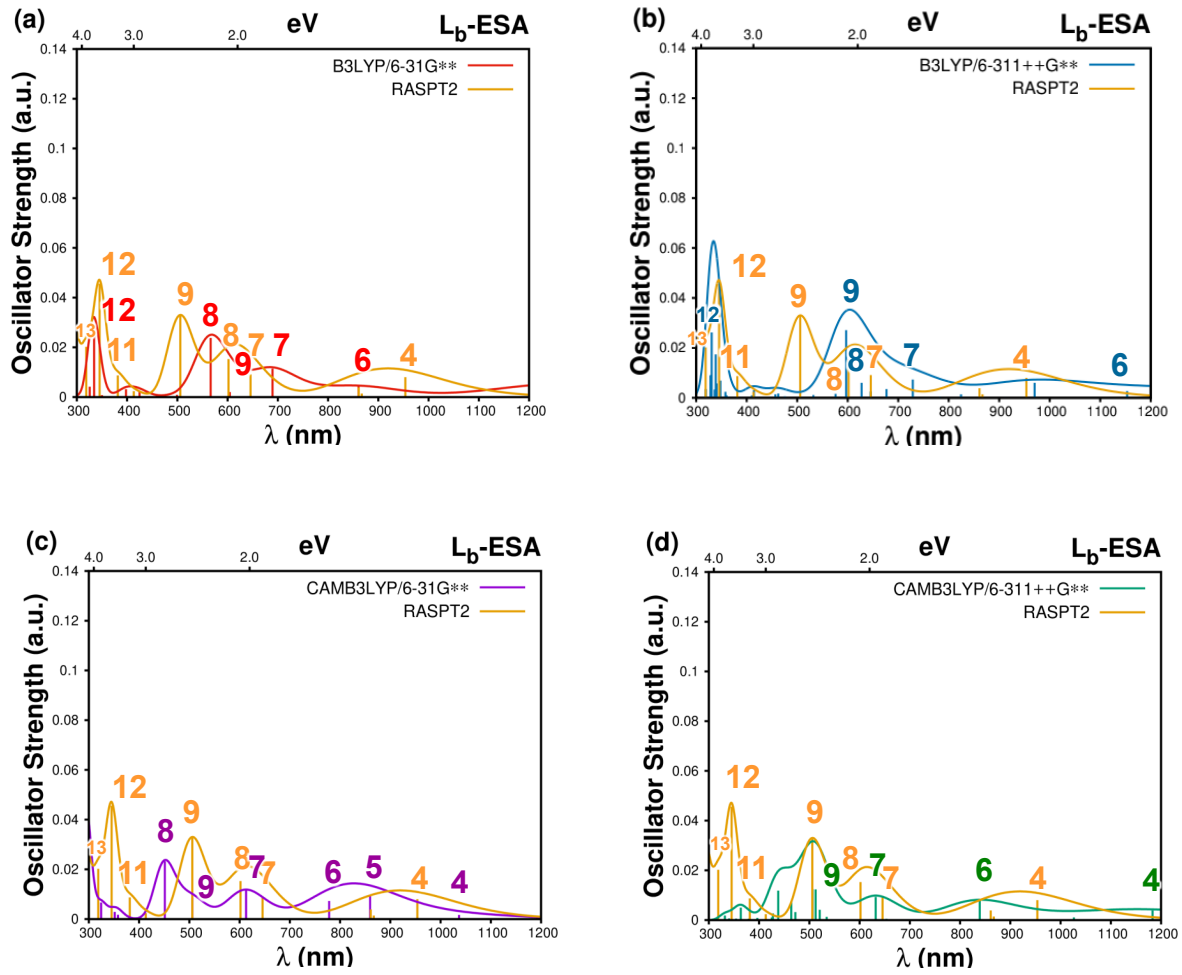

Figure S2. Computed ESAs associated to vertical $L_b \rightarrow S_m$ excitations (colored sticks) of adenine in gas-phase. Reference theoretical values at RASPT2 (yellow sticks) are compared with B3LYP/6-31G**(red sticks), B3LYP/6-311++G** (blue sticks), CAMB3LYP/6-31G**(violet sticks) and CAMB3LYP/6-311++G** (green sticks) and corresponding convoluted spectra (colored lines). Only ππ* states are labeled, according to root numbers in the reference RASPT2 computations (Table S3)

GUANINE

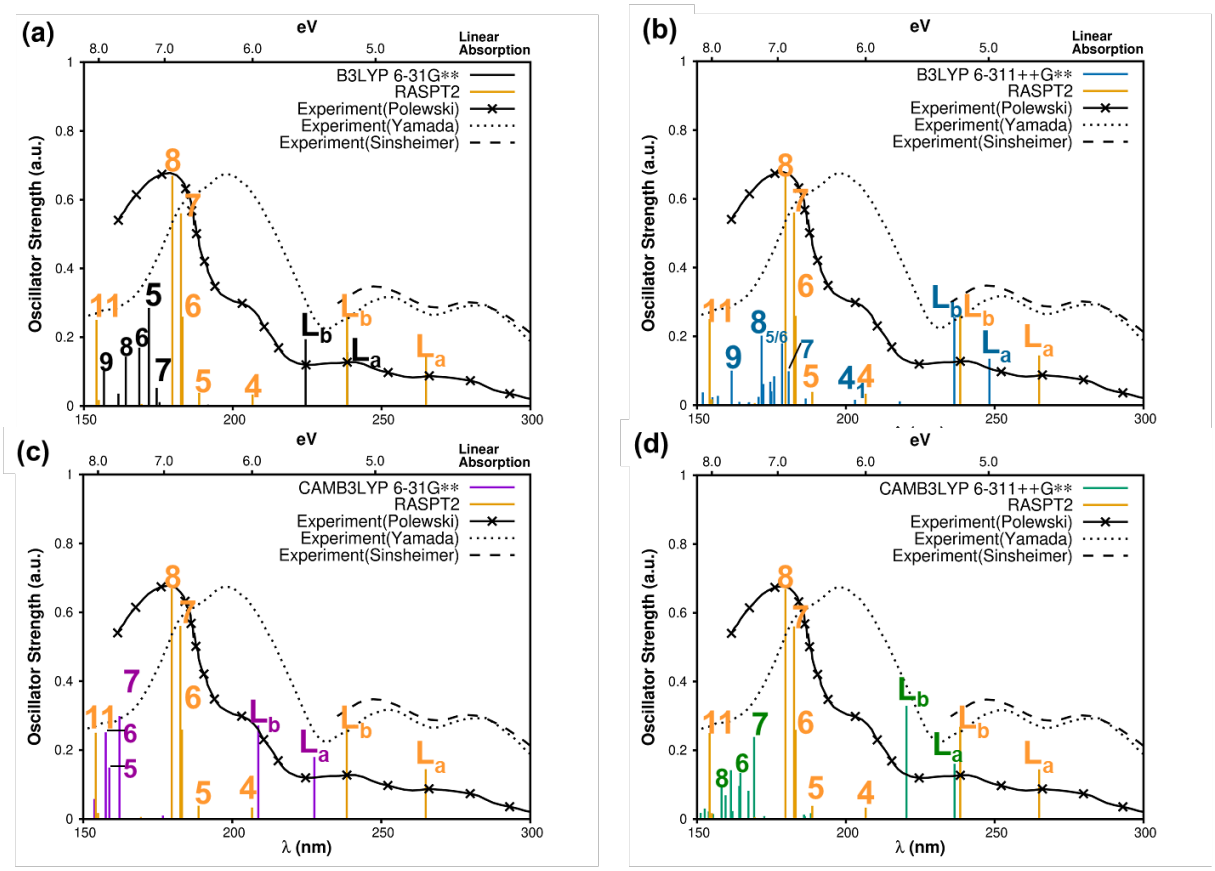

Figure S3. Computed spectra associated to vertical $S_0 \rightarrow S_n$ excitations (colored sticks) of guanine in gas-phase. Reference theoretical values at RASPT2 (yellow sticks) are compared with B3LYP/6-31G**(black sticks), B3LYP/6-311++G** (blue sticks), CAMB3LYP/6-31G**(violet sticks) and CAMB3LYP/6-311++G** (green sticks). Only $\pi\pi^*$ states are labeled, according to root numbers in the reference RASPT2 computations (Table S4)

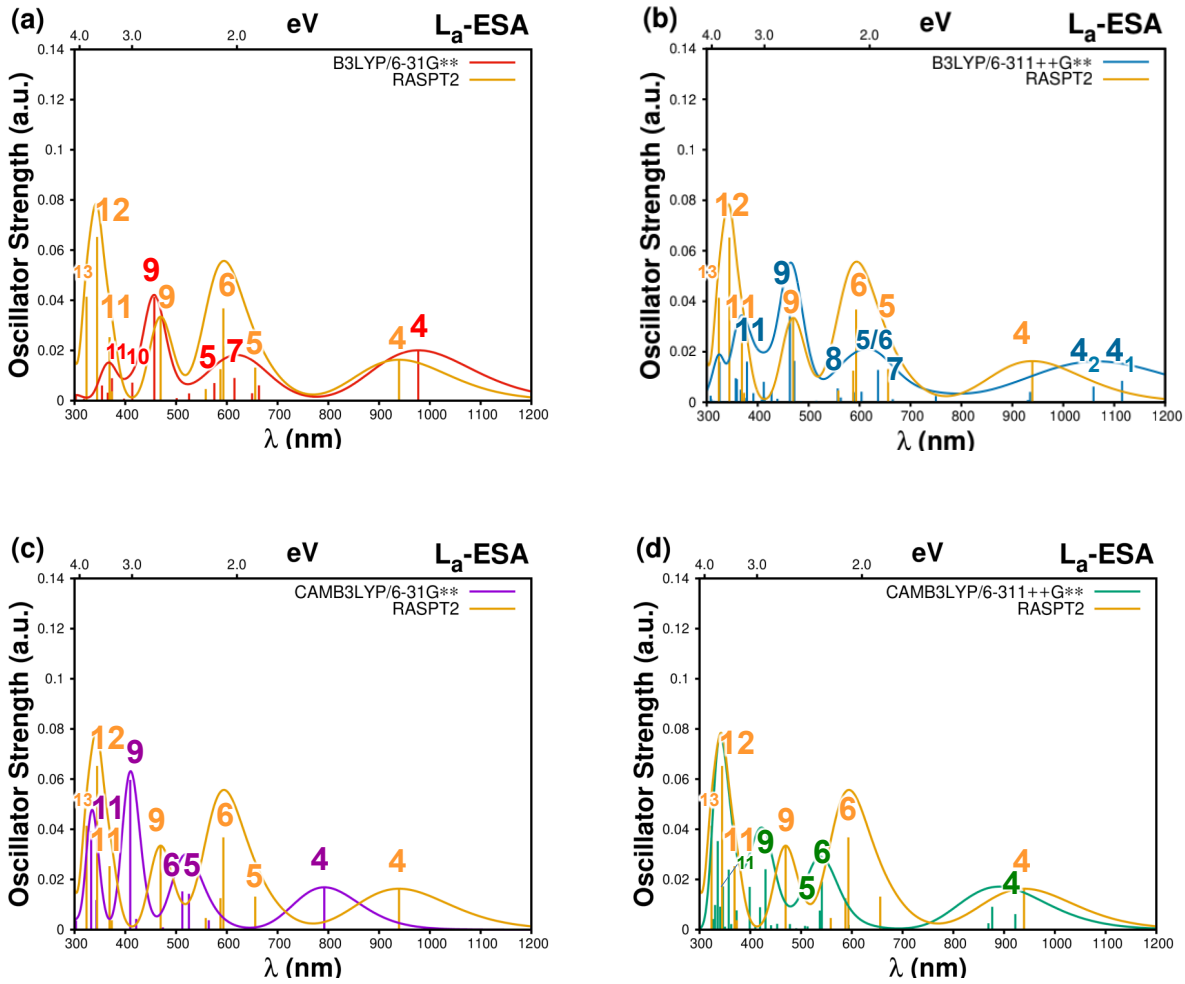

Figure S4. Computed ESAs associated to vertical $L_a \rightarrow S_m$ excitations (colored sticks) of guanine in gas-phase. Reference theoretical values at RASPT2 (yellow sticks) are compared with B3LYP/6-31G**(red sticks), B3LYP/6-311++G** (blue sticks), CAMB3LYP/6-31G**(violet sticks) and CAMB3LYP/6-311++G** (green sticks) and corresponding convoluted spectra (colored lines). Only ππ* states are labeled, according to root numbers in the reference RASPT2 computations (Table S5)

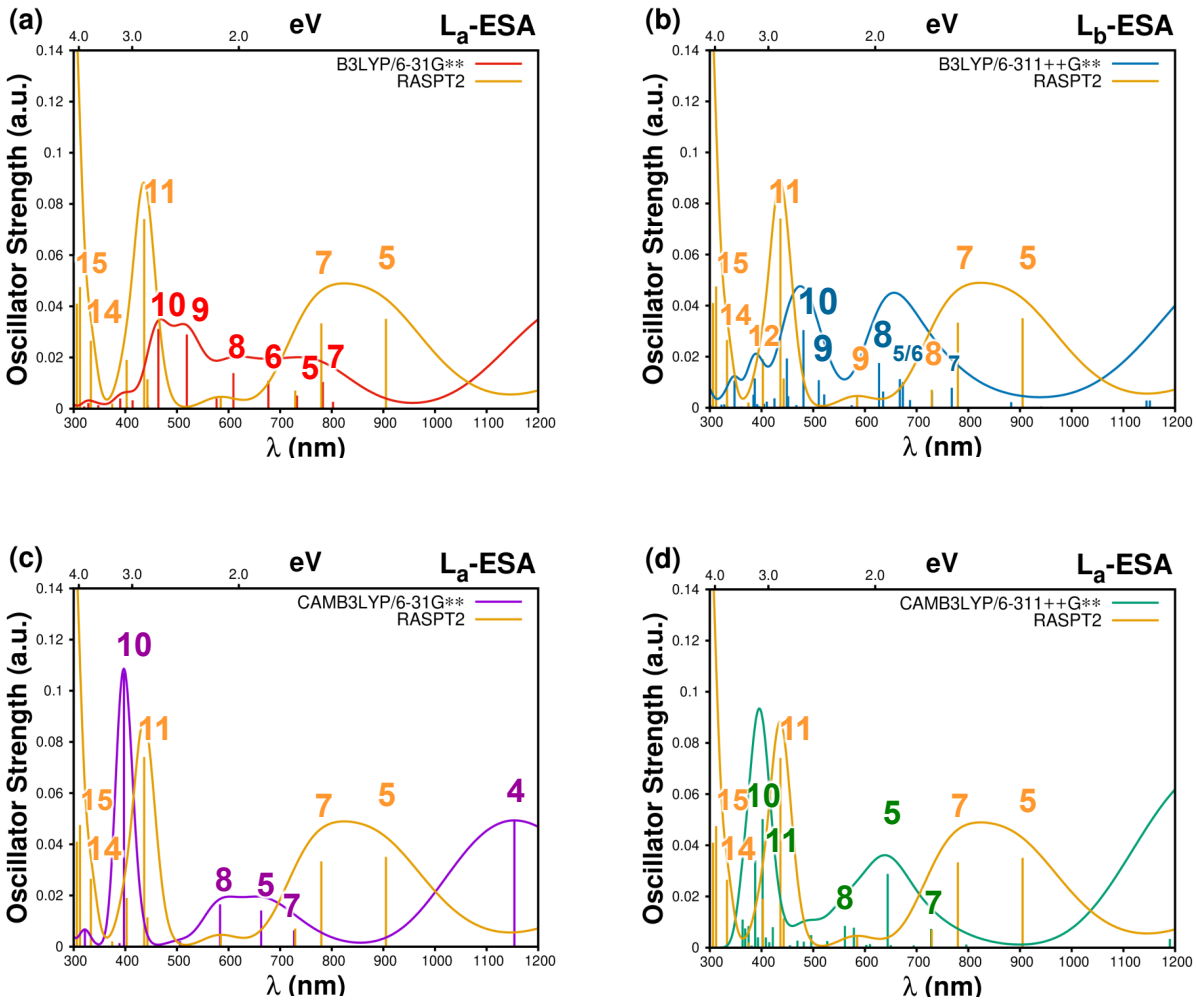

Figure S5. Computed ESAs associated to vertical $L_b \rightarrow S_m$ excitations (colored sticks) of guanine in gas-phase. Reference theoretical values at RASPT2 (yellow sticks) are compared with B3LYP/6-31G**(red sticks), B3LYP/6-311++G** (blue sticks), CAMB3LYP/6-31G**(violet sticks) and CAMB3LYP/6-311++G** (green sticks) and corresponding convoluted spectra (colored lines). Only ππ* states are labeled, according to root numbers in the reference RASPT2 computations (Table S6)

URACIL

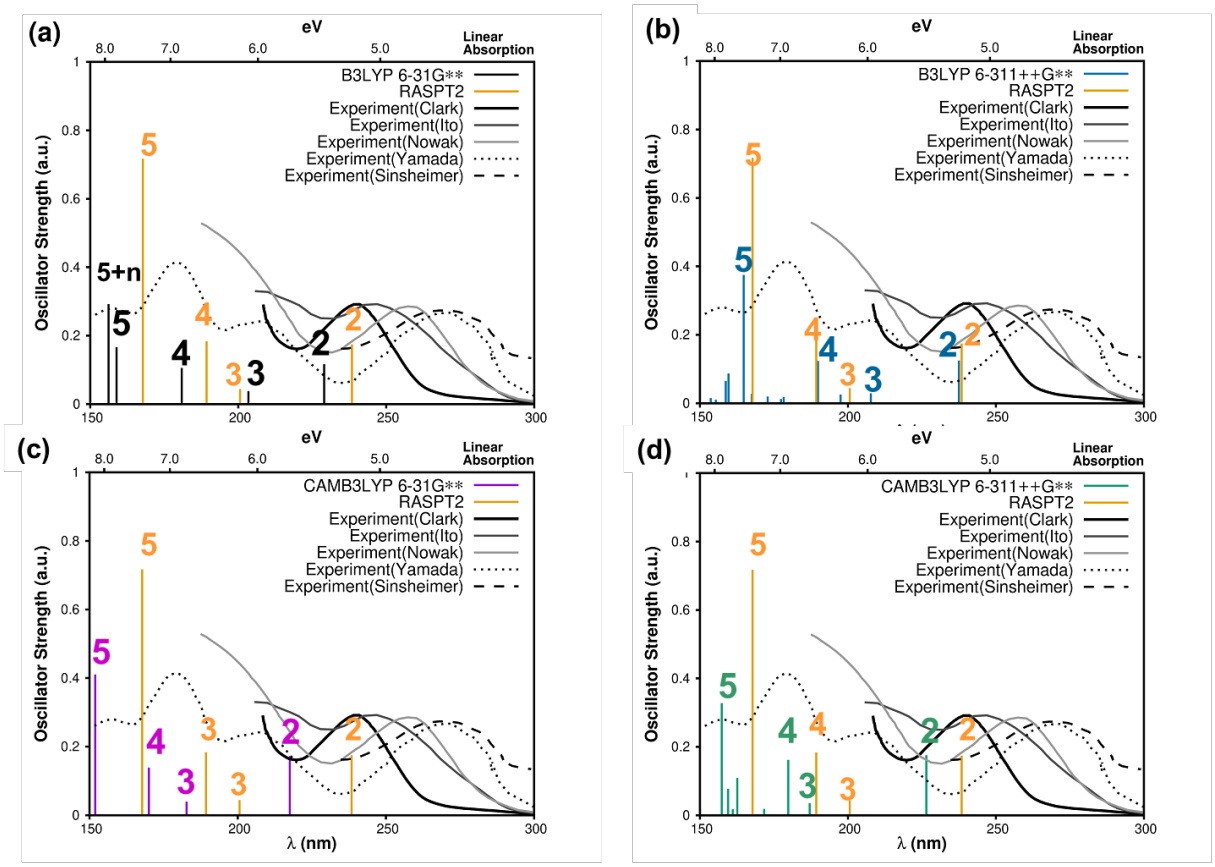

Figure S6. Computed spectra associated to vertical $S_0 \rightarrow S_n$ excitations (colored sticks) of uracil in gas-phase. Reference theoretical values at RASPT2 (yellow sticks) are compared with B3LYP/6-31G**(black sticks), B3LYP/6-311++G** (blue sticks), CAMB3LYP/6-31G**(violet sticks) and CAMB3LYP/6-311++G** (green sticks). Only $\pi\pi^*$ states are labeled, according to root numbers in the reference RASPT2 computations (Table S7)

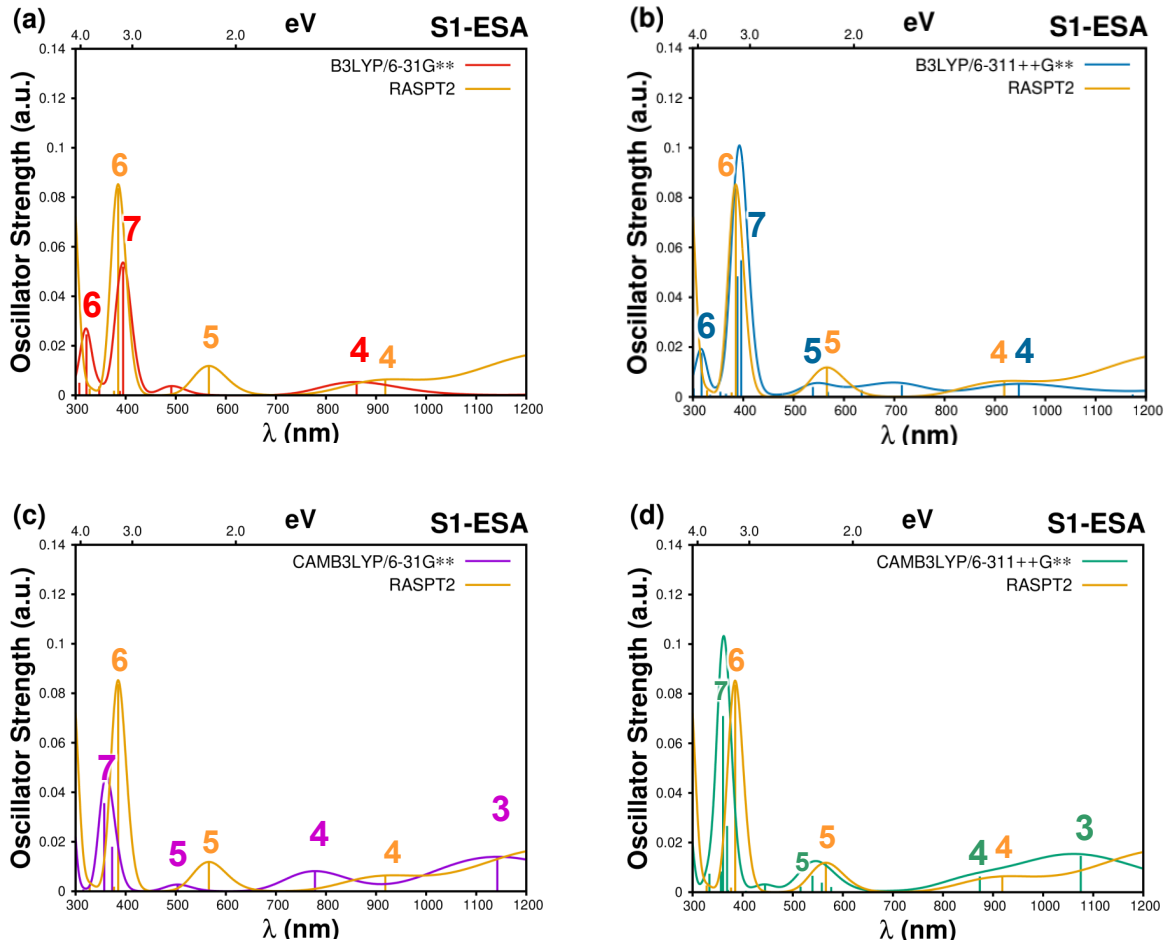

Figure S7. Computed ESAs associated to vertical $S_1 \rightarrow S_m$ excitations (colored sticks) of uracil in gas-phase. Reference theoretical values at RASPT2 (yellow sticks) are compared with B3LYP/6-31G**(red sticks), B3LYP/6-311++G** (blue sticks), CAMB3LYP/6-31G**(violet sticks) and CAMB3LYP/6-311++G** (green sticks) and corresponding convoluted spectra (colored lines). Only ππ* states are labeled, according to root numbers in the reference RASPT2 computations (Table S8)

THYMIME

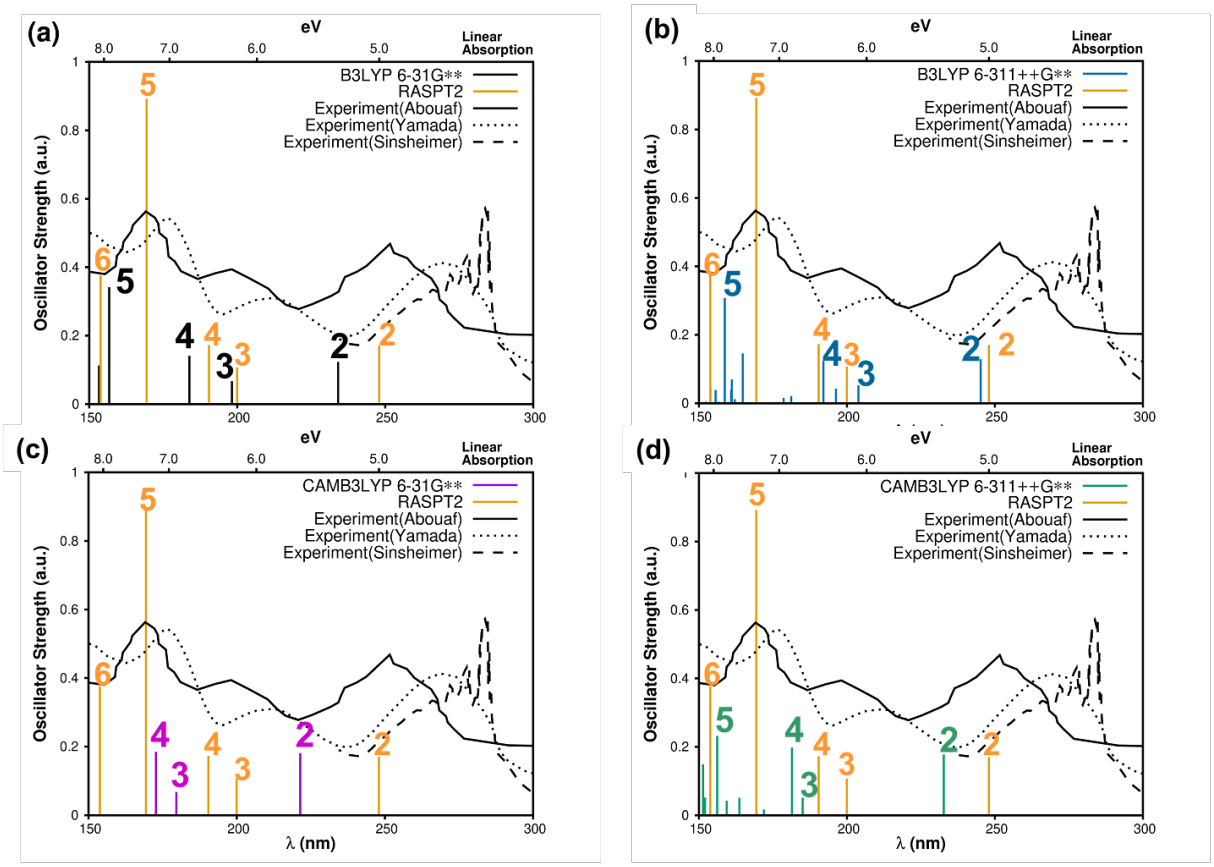

Figure S8. Computed spectra associated to vertical $S_0 \rightarrow S_n$ excitations (colored sticks) of thymine in gas-phase. Reference theoretical values at RASPT2 (yellow sticks) are compared with B3LYP/6-31G**(black sticks), B3LYP/6-311++G** (blue sticks), CAMB3LYP/6-31G**(violet sticks) and CAMB3LYP/6-311++G** (green sticks). Only ππ* states are labeled, according to root numbers in the reference RASPT2 computations (Table S9)

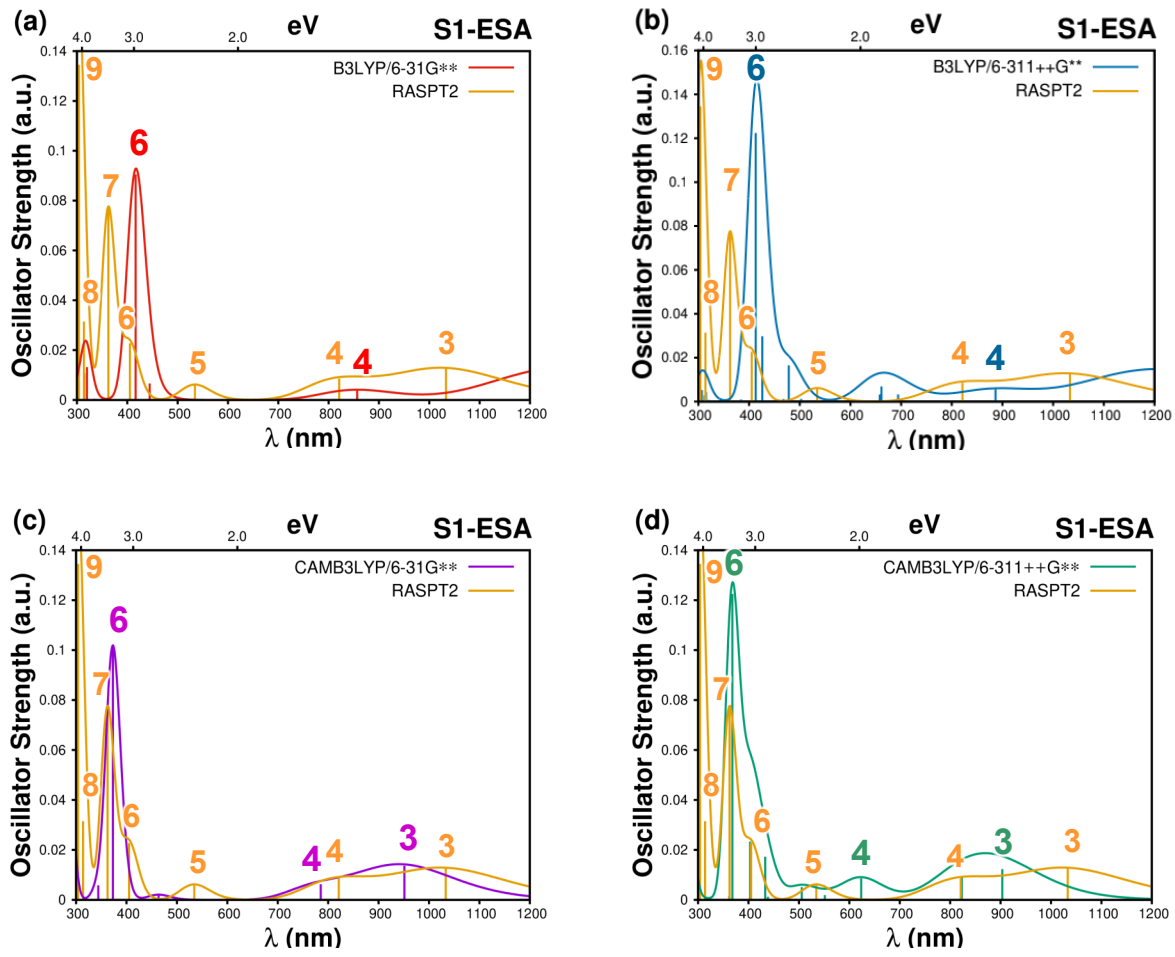

Figure S9. Computed ESAs associated to vertical $S_1 \rightarrow S_m$ excitations (colored sticks) of thymine in gas-phase. Reference theoretical values at RASPT2 (yellow sticks) are compared with B3LYP/6-31G**(red sticks), B3LYP/6-311++G** (blue sticks), CAMB3LYP/6-31G**(violet sticks) and CAMB3LYP/6-311++G** (green sticks) and corresponding convoluted spectra (colored lines). Only ππ* states are labeled, according to root numbers in the reference RASPT2 computations (Table S10)

CYTOSINE

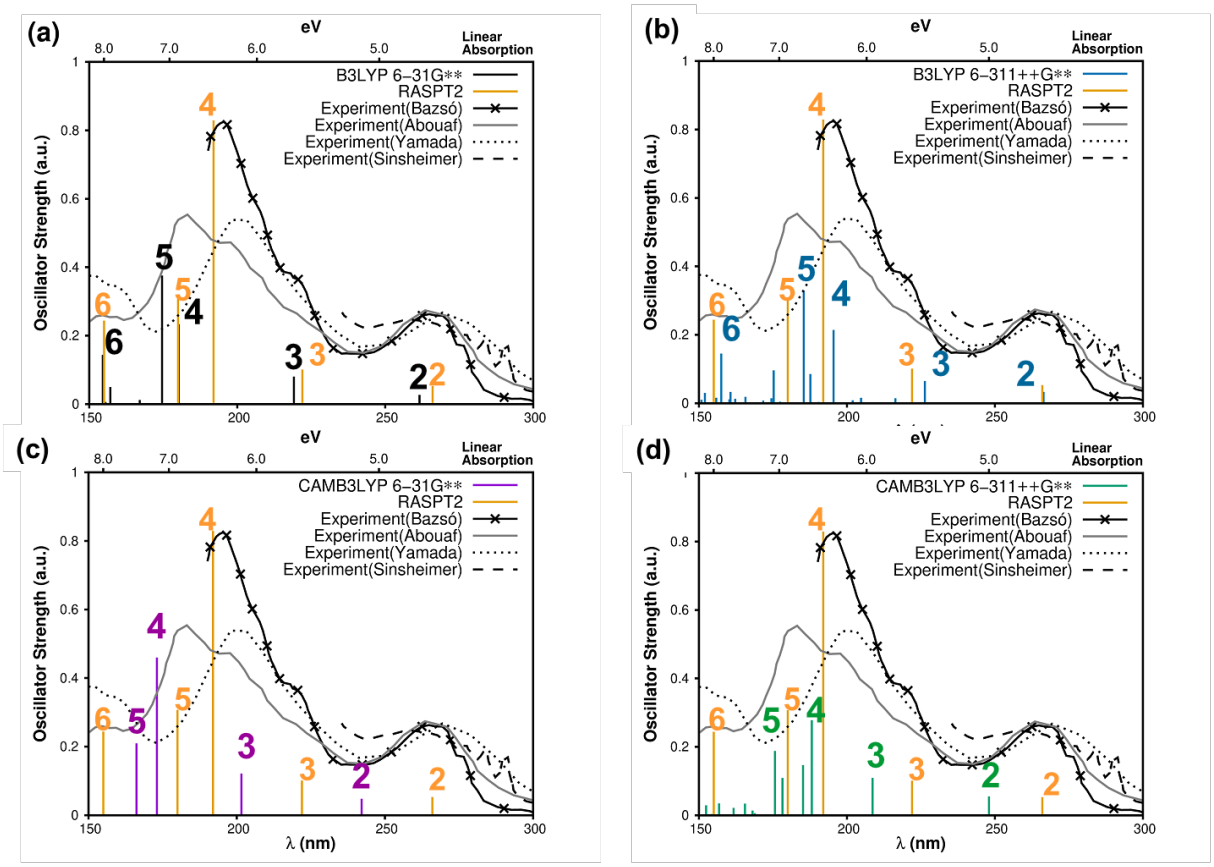

Figure S10. Computed spectra associated to vertical $S_0 \rightarrow S_n$ excitations (colored sticks) of cytosine in gas-phase. Reference theoretical values at RASPT2 (yellow sticks) are compared with B3LYP/6-31G**(black sticks), B3LYP/6-311++G** (blue sticks), CAMB3LYP/6-31G**(violet sticks) and CAMB3LYP/6-311++G** (green sticks). Only ππ* states are labeled, according to root numbers in the reference RASPT2 computations (Table S11)

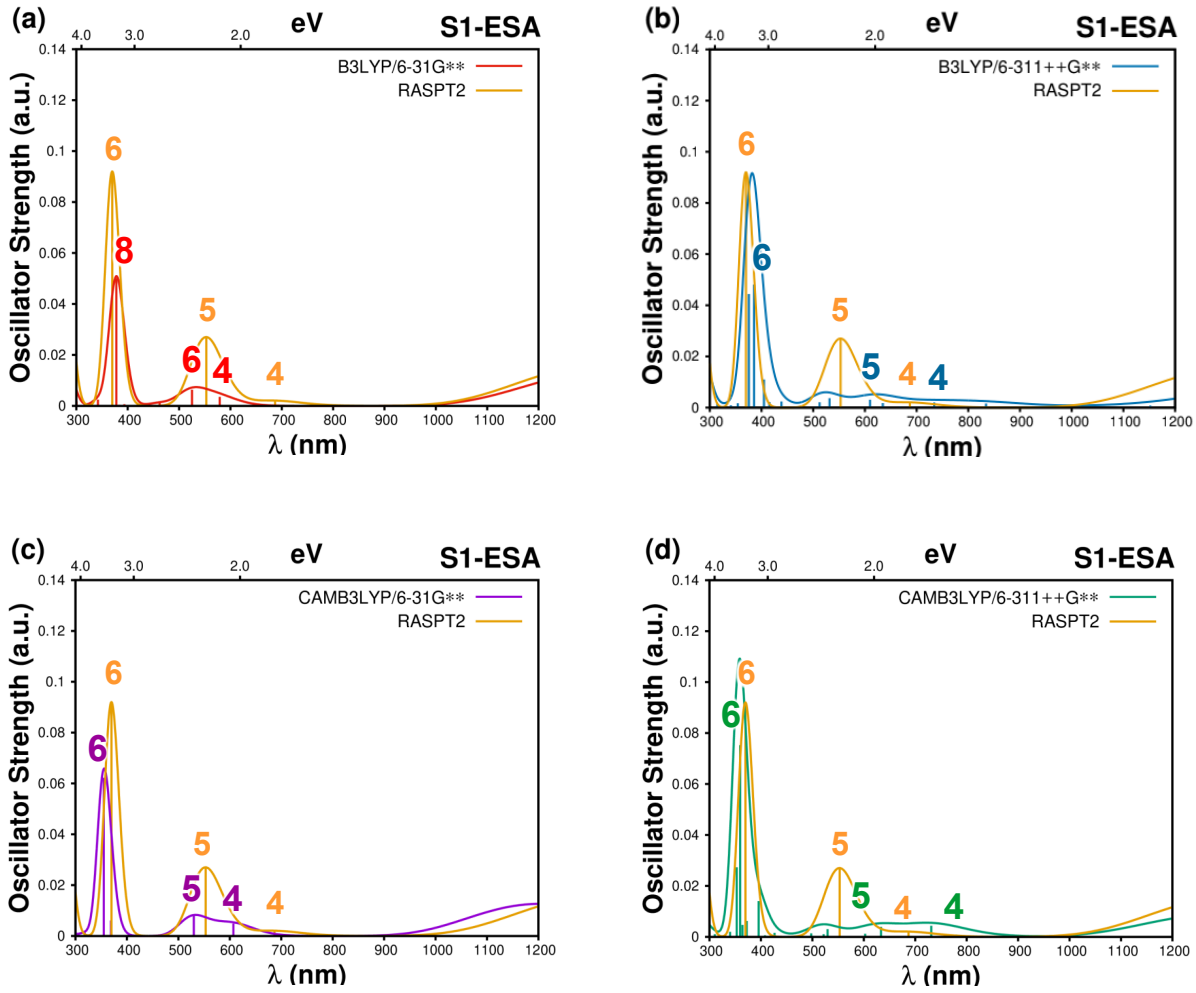

Figure S11. Computed ESAs associated to vertical $S_1 \rightarrow S_m$ excitations (colored sticks) of cytosine in gas-phase. Reference theoretical values at RASPT2 (yellow sticks) are compared with B3LYP/6-31G**(red sticks), B3LYP/6-311++G** (blue sticks), CAMB3LYP/6-31G**(violet sticks) and CAMB3LYP/6-311++G** (green sticks) and corresponding convoluted spectra (colored lines). Only ππ* states are labeled, according to root numbers in the reference RASPT2 computations (Table S12)

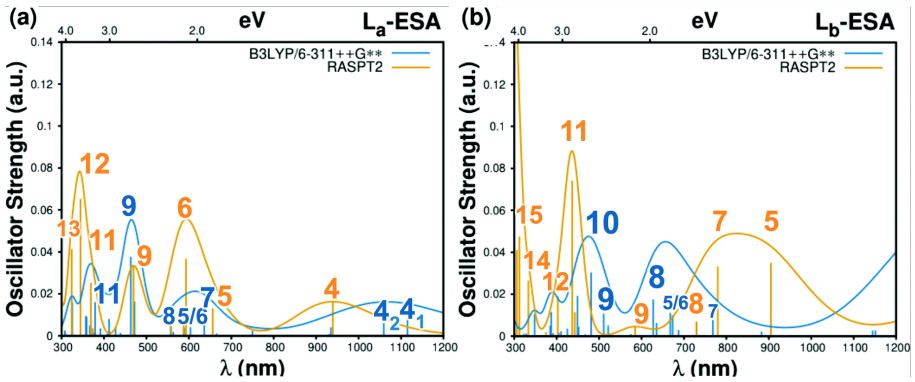

Figure 1.5: Computed ESAs associated to vertical $S_n \rightarrow S_m$ excitations (colored sticks) and corresponding convoluted spectra (colored lines) for guanine in gas-phase. Reference theoretical values at RASPT2 (yellow sticks) are compared with B3LYP/6-311++G** (blue sticks) computations for (a) S_n = La and (b) S_n = Lb. Only $\pi\pi^*$ states are labeled, according to root numbers in the reference RASPT2 computations.

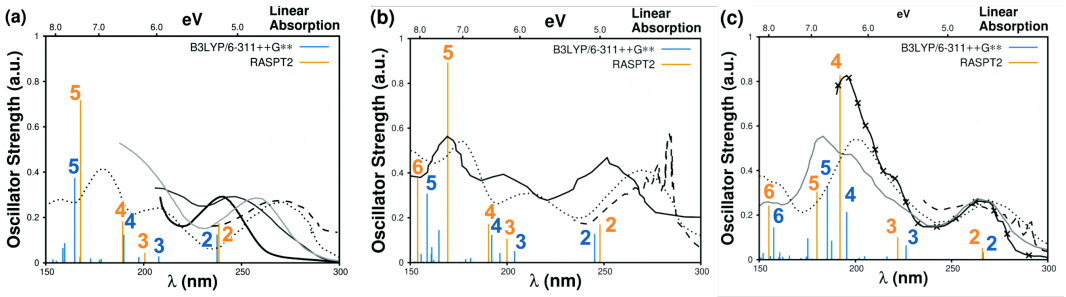

Figure 1.6: (a) Experimental absorption spectrum of uracil in vapor phase (solid black line from [41] solid dark-gray line from [47] and solid light-gray line from [48]), from sublimed films (dotted black line from [44] dashed black line from [45]), and computed vertical $S_0 \rightarrow S_n$ excitations (colored sticks) of uracil in gas-phase. (b) Experimental absorption spectrum of thymine using electron impact energy loss EEEL (solid black line from [49]), from sublimed films (dotted black line from [44] dashed black line from [45]) and computed vertical $S_0 \rightarrow S_n$ excitations (colored sticks) of thymine in gas-phase. (c) Experimental absorption spectrum of cytosine in argon-matrix (crossed black line from [50]), from EEEL spectra (gray black line from [51]), from sublimed films (dotted black line from [44] dashed black line from [45]) and computed vertical $S_0 \rightarrow S_n$ excitations (colored sticks) of cytosine in gas-phase.

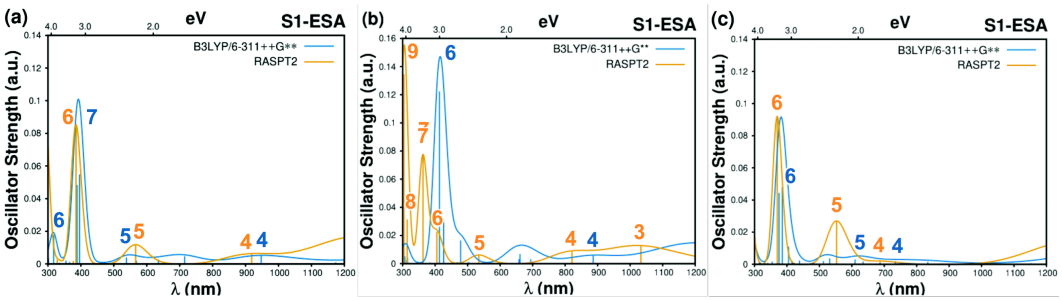

Figure 1.7: Computed ESAs associated to vertical $S_1 \rightarrow S_m$ excitations (colored sticks) and corresponding convoluted spectra (colored lines) for (a) uracil (b) thymine and (c) cytosine in gas-phase. Reference theoretical values at RASPT2 (yellow sticks) are compared with B3LYP/6-311++G** (blue sticks) computations. Only $\pi\pi^*$ states are labeled, according to root numbers in the reference RASPT2 computations

Chapter 2

Modelling the effect of environment on electronic structure

2.1 Theory - *how electronic structure is computed in realistic systems*

It is important to include the effect of environment when modelling any kind of photophysical phenomena. The definition of environment encompasses solvents or areas of the molecule not participating in the excitation directly, but nevertheless influencing the excited-states of the system. The static effect of the environment is twofold through energetic and entropic effects. The interaction with the environment changes the electronic energy landscape compared to gas-phase, stabilizing or destabilizing excited-states according their electronic density[53]. The additional degrees of freedom imparted by a solvent or a flexible part of an extended molecule leads to an entropic stabilization of the free energy landscape of the system. This additional flexibility also broadens the variance of the possible excitation energies([54, 55]). Therefore both of these effects need to be included to accurately model the photophysical processes in the chromophore. Among the many protocols employed to model the environment around a photoactive molecule or molecular fragment, we will refer to explicit approaches thereafter, i.e. the atoms of the environment are treated explicitly.

2.1.1 Sampling the environment

Molecular Dynamics

A system at room temperature is not fixed in geometry but adopts a multitude of configurations whose relative populations are dictated by the temperature. In a solvated system there are different arrangements of solvent molecules around the solute molecule,each leading to different values of excitation energy. To create the possible configurations adopted by the system in a

realistic ensemble, molecular dynamics or random sampling can be employed. For realistic systems molecular dynamics is an efficient way to create an ensemble representative of experimental samples. According to the ergodic hypothesis, an ensemble created by visiting structures along a molecular dynamics is equivalent to the ensemble in real space present in an experimental sample. Therefore any molecular dynamics aimed at recreating the experimental ensemble should be long enough to have realistic populations of various conformers.

The most important ingredient of classical molecular dynamics are the system dependent force-fields. These force-fields are mathematical functions decomposing all the interactions of the system involving nuclei and electrons into various bonding(covalent) and non-bonding(van der waals) interactions. Thus the energy of the system is re-expressed in terms of variations of bond-lengths, bond-angles and dihedral angles for covalent interactions and distance dependent electrostatic and van-der waals terms for non-covalent interactions.

$$E_{FF} = E_{stretch} + E_{bend} + E_{torsion} + E_{electrostatic} + E_{vdW} + E_{cross} \tag{2.1}$$

These functional forms of energy allow us to bypass the full quantum mechanical computation of interactions and enable molecular dynamics of molecular systems. The bonding interactions between atoms are described by the stretching and bending terms of the force-field while non-bonding interactions between the atoms are included in electrostatic and van-der Waals interactions. Certain force-fields like UFF or MMFF94 also contain additional cross-term parameters to describe the coupling between the various interactions.

The functional forms utilized in force field interactions have to parametrized to approximate the underlying potential energy surface.The parameters can be chosen either experimentally or through quantum mechanical computations. Experiments like X-ray crystallography of NMR can provide equilibrium distances, angles and other structural factors various bonds, while vibrations spectroscopy like infrared gives information about the harmonic frequency of various bonds. For paramatrizing dihedrals, experimental populations difference between various conformers or an adiabatic scan computed at quantum mechanical level can be used to optimize the parameters. An important consideration when parametrizing a force field is to avoid dangers of overfitting to the experimental or simulation data set. Similar atoms in similar environments like a carbon in linear aliphatic chain, or oxygen in an hydroxyl bond should have similar parameters across different systems. With such constraints one can then identify and classify various atom types and assign partial charges and parameters for Lennard-Jones interactions. Thus in additional to being accurate, force-fields should be general and transferable to have predictive power in new simulations.

Wigner Sampling

Most classical molecular dynamics schemes create an ensemble of nuclear conformations and velocities consistent with thermal equilibrium. This leads to an average energy of kbT being equally deposited into all modes of freedom. However for degrees of freedoms with zero point energy larger than kbT, this is in violation of quantum mechanics. Thus for stiffer bonds (high

frequency modes), molecular dynamics gives nuclear distributions which are narrower than the lowest quantum energy level[56]. To obtain the correct statistical distributions of nuclei Wigner's quasiprobability distribution can be utilized. The Wigner function maps the quantum mechanical wavefunction to a distribution in the phase space of position and momenta. For a quantum system with f degrees of freedom described by the wavefunction $\psi(q)$ the Wigner function is given by

$$W(q,p) = \frac{1}{(2\pi\hbar)^f} \int ds\, \psi^*(q - s/2)\, exp(ip \cdot s/\hbar)\, \psi(q + s/2) \tag{2.2}$$

where q and p denote coordinate and momenta and s is a dummy variable for the integration. Usually one is interested in sampling the ground state equilibrium distribution of structures, where harmonic approximations can be employed for the various normal modes of the system. These normal modes can be obtained at various levels of electronic structure theory after an optimization to locate the local minima and frequency computations to locate the normal modes. In these cases one can use the analytical expression for the canonical ensemble of harmonic oscillator as described by Wigner[57].

$$W(q,p,T) = \frac{1}{\pi\hbar} \tanh\left(\frac{\hbar\omega}{2k_BT}\right) \exp\left(-\frac{2}{\hbar\omega} \tanh\left(\frac{\hbar\omega}{2k_BT}\right)\left(\frac{p^2}{2\mu} + \frac{\mu\omega^2 q^2}{2}\right)\right) \tag{2.3}$$

2.1.2 QM/MM scheme

To model photoinduced process one needs to compute the electronic structure of the system. But accurate electronic structure computations are only possible for molecular systems with dozens of atoms. Therefore some approximations are needed to compute the electronic structure of realistic systems in physiological environments. Because the photoinduced processes are concentrated only in a subset of the system called chromophore, one can treat only this part with explicit quantum mechanic computations. The rest of the system defined as the environment can be treated in a more simplistic fashion. This partitioning of the total system called QM/MM scheme, was proposed by Warshel, Lwevitt and Karplus([58, 59]) to allow modelling complex systems using a multi-scale approach. In this approach the chromophore is described at quantum-mechanical level with nuclei and electrons while the rest of the system is modelled using classical force-fields. This QM/MM partitioning of the system imparts a great flexibility in modelling of complex systems. One can combine any electronic structure software/method with multitude of available models for the environment like fixed or polarizable charges.

The most important ingredient of a QM/MM scheme is the description of the interaction between QM(quantum described) and MM(classically described) part. All the QM/MM computations reported in this thesis used the COBRAMM software [60] which employs an electrostatic embedding scheme to include the effect of the environment on the QM part. In this scheme the environment in seen as point charges which are included in the one-electron Hamiltonian in the electronic structure computation. To compute the full energy of the system at QM/MM level care has to be taken to avoid any double counting of interaction terms. This implemented in COBRAMM through a subtractive scheme.

$$E_{QM/MM}(S) = E_{MM}(S) + E_{QM}(I + link) - E_{MM}(I + link) \tag{2.4}$$

where S is the whole system and I denotes the inner(QM) region. When QM/MM boundary crosses a covalent bond, a *link* atom is created to fulfil the valence requirements in the QM computation. As seen from above equation the QM/MM energy computation of the whole system can be done in 3 steps.

1. Calculation of the entire systems energy at the MM level of theory.
2. Computation of the inner system and link atom energies using a QM method.
3. Computation of the inner system and link atom energies with an MM method, and finally subtracting this term from the sum of the latter

2.1.3 Inclusion of solvent effects

The energies of the electronic states of a molecule are different in gas and condensed phases[61]. This is due to the interaction of the solvent with the chromophores leading to energetic shifts of the various electronic levels. Different types of electronic states interact differently with solvents. These differences can be due to the symmetry of frontier orbitals involved in excitation($\pi\pi^*$ vs $n\pi^*$) or due to the extent of localisation of excitation in multi-chromophoric system (local vs charge-transfer states). The differing nature of excitation leads to different relative shifts in energy compared to gas-phase and need to be modelled explicitly. The interaction with solvent can either be modelled through an atomistic description of the solvent arrangement(explicit solvent) or through a continuum model(implicit solvent). Explicit solvent models help to the optical response in a realistic solvent distribution can be used to compute the entropic effects of various solvent arrangements. However such realistic solvent arrangements have to generated beforehand by various sampling or molecular dynamics methods. On the other hand, implicit solvent models like PCM (polarizable continuum model [62–65]) don't need the generation of a particular solvent arrangement and can compute average properties in the solvent medium. However, it is impossible to model any time-dependent dynamical solvent response in these implicit models and one has to resort to an explicit solvent description for such tasks.

2.2 Project - *Linear Absorption of solvated cytidine nucleoside*

The excited states of nucleobases can decay through a variety of pathways[66–74]. These pathways can be unravelled at an atomistic level by ab-initio molecular dynamics simulations[67, 75–78]. However, for even a qualitative sampling of the decay pathways and corresponding rates, multiple trajectories are needed. The ensemble of points in phase-space on which these trajectories are started should correspond to the equilibrium nuclear distribution of ground state and the energy deposited in various modes. One can assess the quality of such a generated distribution

and associated computational protocols by comparison with steady-state averages of physical observables. Reproduction of experimental linear absorption is a good indicator for the quality of nuclear distribution and electronic structure calculation methods. In the present work linear absorption of solvated cytidine nucleoside(nucleobase + sugar) is modelled with explicit solvent. In this work, linear absorption of the ensemble by wavefunction based method(RASPT2) is computed. The effect of active of active-space at the CASSCF level on the electronic energy manifold at Franck-Condon geometry is analyzed.

2.2.1 Nuclear ensemble generation

The generation of nuclear ensemble is commenced by a molecular dynamics at classical level to generate a realistic water conformation around the solute. The molecular dynamics was performed by the ff99 RNA force-fields [79] augmented with bsc0[80] and OL3 corrections[81]. This is the default recommended RNA force-field in the Amber Molecular Dynamics package (loaded by parm10.dat file). The water were modelled in a rigid fashion with the TIP3P model which freezes the O-H bond-lengths and the H-O-H bond-angle allowing for integration steps of 2 fs during molecular dynamics. The simulation was done at 300K and constant pressure (NPT ensemble) with periodic boundary condition employing a Langevin thermostat. After the creation of the system and heating to 300K in steps of 50 K and equilibration for 100ps, a classical dynamics of 30ns was done to enable the Boltzmann population of the various microstates (solute conformers + solvent arrangements).

The dominant conformational features in the nucleoisdes are the relative orientations of the sugar with nucleobase and various puckering of the sugar moiety. The syn and anti conformers characterize the orientation of the sugar relative to the nucleobase. The dynamics results in a predominant population of the anti conformer which is equivalently distributed among two different sugar puckering(**Figure2.1(a)**). Since, these puckering are far away from the chromophoric nucleobase, they can be assumed to have similar effects on the electronic levels of the chromophore. A snapshot from the most populated syn-conformation is chosen for refinement at QM/MM level and the generation of wigner ensemble.

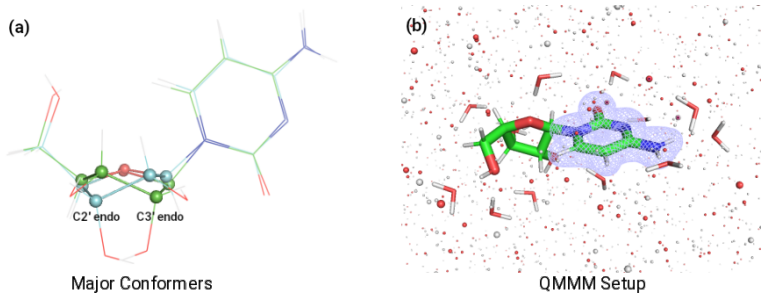

Figure 2.1: (a) Two equally populated anti conformers from molecular dynamics differing in the puckering of sugar rings. (b) QM/MM setup employed in the study. The mobile parts High and Medium are shown in sticks and the fixed part with dots. The High part used in electronic structure computations is emphasised with mesh surface.

The chosen snapshot is transformed into a spherical droplet of water and then optimised at MP2 level to locate the Franck-Condon structure. The QM/MM optimization is done by putting the sugar and nearest solvents in the medium part and the nulecobase in the high part (**Figure2.1(b)**). The rest of the solvent molecules are kept frozen during optimization. The frequencies computed upon the optimized structure are used to generate the Wigner ensemble at 300K temperature. After generating the distorted structures, a short solvent dynamics is performed to relax the solvent around the distorted solute geometries. This protocol depicted in Figure 2.2 also removes any bias towards the specific solvent arrangement in our chosen snapshot.

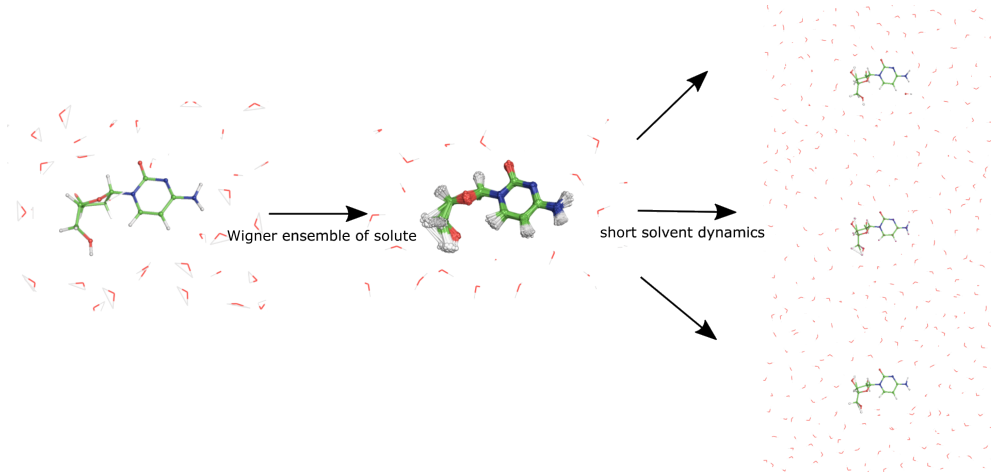

Figure 2.2: The protocol used to generate wigner-corrected statistical QM/MM ensemble. A wigner ensemble of solute is generated from the MP2 optimized Franck-Condon geometry. A short solvent dynamics is performed around the wigner structures, to generate decorrelated solvent ensembles.

2.2.2 Excited States at the Franck-Condon point

State	**|14,10|**			**|4,7|0|4,7|**		
	SS	**MS**	**XMS**	**SS**	**MS**	**XMS**
S_1 = H→L	4.23(*0.06*)	4.41(*0.21*)	4.39(*0.11*)	4.52(*0.07*)	4.54(*0.15*)	4.53(*0.13*)
S_2 = H-1→L	4.93(*0.10*)	4.80(*0.27*)	4.97(*0.33*)	5.19(*0.14*)	5.08(*0.30*)	5.13(*0.25*)
S_3 = $n_{Nitrogen}$→L	5.43(*0.01*)	5.58(*0.00*)	5.53(*0.01*)	5.63(*0.04*)	5.74(*0.01*)	5.76(*0.01*)
S_4 = H→L+1	5.60(*0.36*)	5.75(*0.29*)	5.68(*0.25*)	6.02(*0.61*)	5.98(*0.19*)	5.96(*0.11*)
S_5 = n_{Oxygen}→L	5.59(*0.02*)	5.97(*0.06*)	5.82(*0.01*)	6.05(*0.00*)	6.09(*0.02*)	6.02(*0.06*)
S_6 = n_{Oxygen}→L+1	6.08(*0.00*)	6.33(*0.00*)	6.05(*0.01*)	6.21(*0.00*)	6.28(*0.01*)	6.26(*0.07*)
S_7 = H-1→L+1	5.75(*0.63*)	6.23(*0.38*)	6.22(*0.47*)	6.13(*0.41*)	6.36(*0.62*)	6.37(*0.65*)
S_8 = $n_{Nitrogen}$→L+1	6.52(*0.01*)	6.77(*0.02*)	6.64(*0.03*)	6.78(*0.03*)	6.84(*0.03*)	6.89(*0.02*)

Table 2.1: Comparison of CASPT2 vertical excitation energies(*Oscillator Strengths*) at MP2 optimized Franck-Condon geometry of selected snapshot with full valence |14,10| and augmented 4,7|0|4,7 active space.

The vertical excited states at the Franck-Condon geometry of the selected snapshot are shown

in **Table 2.1** labelled as S_N(N = 1,2...). These manifold of states cover all the electronic excitations until 7 eV and consist of four bright $\pi\pi^*$ and four dark $n\pi^*$ states as seem in the natural transition orbitals of these states in **Figure 2.3**. These states are mostly comprised of excitations to the two lowest π^* virtual orbitals labelled as L and L+1. The $\pi\pi^*$ states remove electrons from occupied orbitals labelled as H and H-1 and the dark $n\pi^*$ states remove electrons from lone pairs on either nitrogen or oxygen atom.

In **Table 2.1** the excited states are computed with two different active spaces to assess the accuracy of the computation. The |14,10| active-space consists of the valence orbitals of the system involving 14 electrons in 10 orbitals. The bigger |4,7|0|4,7 active-space augments the previous by adding 4 additional virtual orbitals. To make this bigger active-space tractable, a RASSCF procedure is used with an empty RAS2 space by putting all occupied orbitals in RAS1 and all virtual in RAS3 and allowing 4 excitations between these spaces. To analyze the effect of different dynamical correlation methods on electronic energies, three different CASPT2 flavours SS(Single-State), MS(MultiState) and XMS(eXtended MultiState) are considered. SS-CASPT2 provides a correction of energy to the original CASSCF states while MS-CASPT2 and XMS-CASPT2 use the original CASSCF states as a reference to obtain new orthogonal states after diagaonalizing the CASPT2 Hamiltonian.

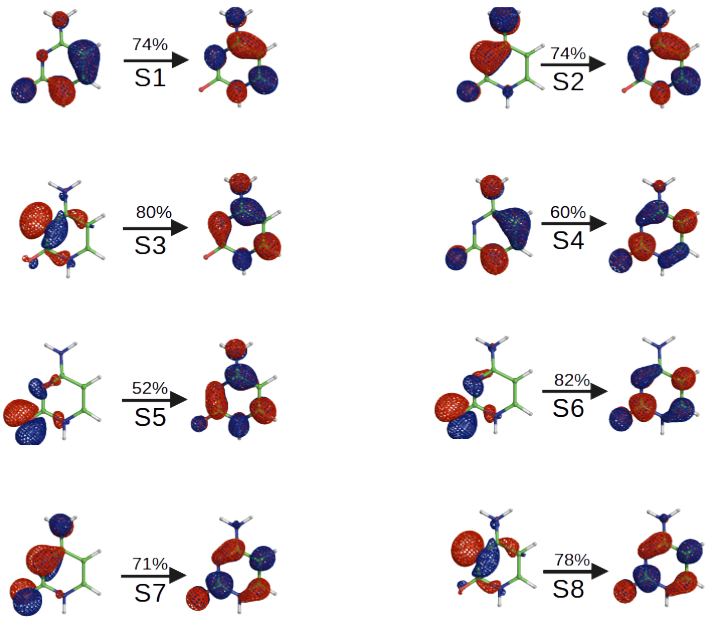

Figure 2.3: Natural Transition Orbitals of the electronic states of solvated cytidine.

The bigger active-spaces capture the contribution of more orbitals in a variational approach, the perturbation approximation at CASPT2 step is less severe for them than smaller active-spaces. Thus bigger active-spaces can be used as a benchmark to validate computations with smaller active-spaces. The XMS computation with the bigger RAS active-space in **Table 2.1** serves as the reference to compare other methods.

All the electronic structure methods outlined in **Table 2.1** agree on the nature of the three lowest states. These are two bright $\pi\pi^*$ states with a dark $n\pi^*$ state involving the lone pair on nitrogen. The energetic ordering of the higher lying states from S4 to S7 depend on the flavour of CASPT2 being employed. At SS-CASPT2 level S4 and S5 are almost degenerate at |14,10| active-space with an energetic gap of 0.01 eV and and the $\pi\pi^*$ labelled as S7 state is lower in energy than S6 which is a $n\pi^*$. These trends of SS-CASPT2 are preserved in the bigger active-space computation and differ with the XMS computation. At XMS level, S4 and S5 have an energetic gap of 0.06 eV with the augmented active space and the $n\pi^*$ state S6 is lower than the bright S7 state. These XMS trends are also reproduced with the smaller |14,10| active-space. The electronic state from $n_{Nitrogen}$→L+1 labelled as S8 is always the highest state in all the CASPT2 computations. These comparisons reveal that the smaller |14,10| active-space can be safely employed to perform non-adiabatic dynamics at XMS level as it preserves the energy order of the bigger active-space computation.

2.2.3 Excited-States from the Wigner Ensemble.

The wigner ensemble generated using the frequencies of the optimized Franck-Condon structure is used to construct the Linear Absorption spectra. Vertical excitations were computed on 500 structures of the ensemble and were broadened by an empirical value of 0.18 eV (standard deviation). It should be noted that such a computation does not aim to and cannot reproduce the experimental lineshape of the absorption spectra accurately. The experimental absorption lineshape encodes complex dynamical information abut environmental and early excited-state dynamics in addition to static distribution of excitation energies. The aim here is to have a reasonable approximation to the experimental spectra to validate the theoretical method of electronic structure.

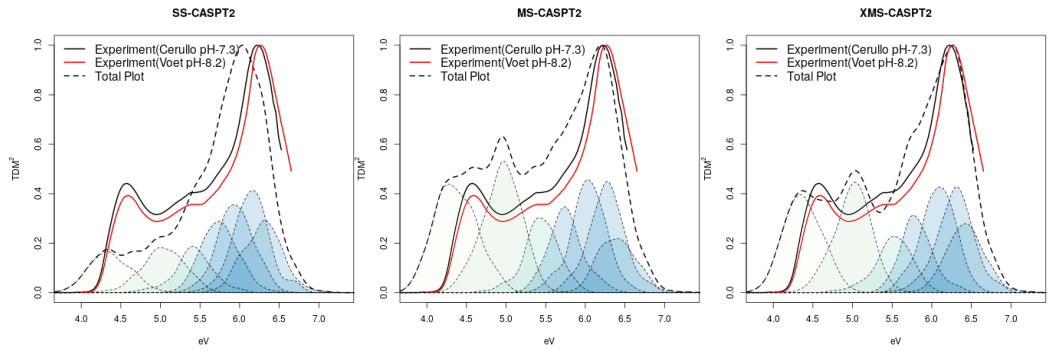

Figure 2.4: Linear absorption of the Wigner ensemble computed with SS-CASPT2,MS-CASPT2 and XMS-CASPT2 with |4,7|0|4,7 active-space in dotted line. The filled curves correspond to the adiabatic states of the wigner ensemble. Also shown are experimental plots recorded by Cerullo et al.(in solid black line) and Voet et al. [82] (in solid red line)

Figure 2.4 shows the linear absorption of the wigner ensemble computed with SS/MS/XMS-CASPT2 with the augmented |4,7|0|4,7 active-space in dotted lines with the experimental plots

recorded by Cerullo et. al and Voet et. al with solid lines. The experimental linear absorption recorded till 7.0 eV shows two prominent peaks at 4.54 eV and 6.20 eV, with peaked shoulder between them slightly peaked at 5.39 eV. The total wigner absorption is shown decomposed into individual adiabatic states of the wigner ensemble. Since the geometrical distortions in structures of wigner ensemble can lead to change in energetic positions of $\pi\pi^*$ and $n\pi^*$ states, all the adiabatic states except the highest one have significant oscillator strengths in **Figure 2.4** It can be appreciated from **Figure2.4** that XMS-CASPT2 plot looks closest to the experimental plots. In all three plots the position of the first band is slightly underestimated by 0.1 eV with the theoretical peaks lying at 4.4 eV. The SS-CASPT2 plot has significantly underestimated transition dipole moment for the first band and energetic position of the highest band at 7 eV, while MS-CASPT2 overestimates the strength of the first transition.

With wave-function based methods, we can also analyze the theoretical linear absorption in terms of the diabatic-states .This diabatic characterisation of the electronic states focuses on the underlying mature of electronic excitation($\pi\pi^*/n\pi^*$) rather than the energetic ordering (S1/S2). This is done by computing the overlap of adiabatic states of the ensemble to the Franck-Condon states. In this way, every adiabatic state of the ensemble is re-expressed in terms of linear combination of the Franck-Condon states. These overlap matrices inform if the adiabatic states in the ensemble mix the Franck-Condon states or preserve their purity. In **Figure 2.5(a)** these states are then assigned to the diabatic state which has the largest wavefunction overlap with them. It can be seen by (**Figure 2.5(a)**) that even after the diabatic decomposition the $n\pi^*$ states have significant non-zero oscillator strength, which is due to there being truly mixed states $\pi\pi^*$-$n\pi^*$ states being present in the ensemble.

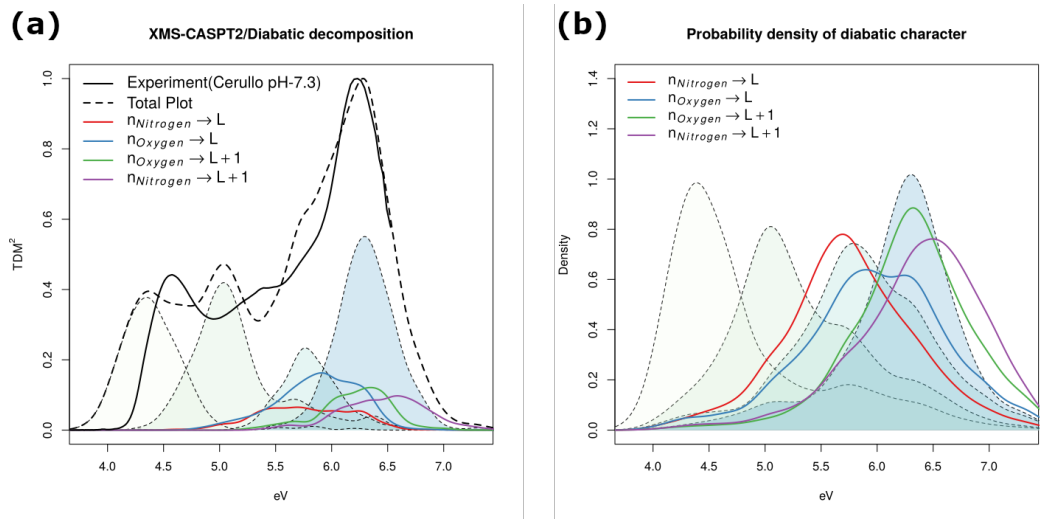

Figure 2.5: (a) Linear absorption of the Wigner ensemble computed with XMS-CASPT2/|4,7|0|4,7 active-space in dotted line decomposed into Franck-Condon diabatic states. The filled curve are the $\pi\pi^*$ states and the solid lines are the $n\pi^*$ states. (b) The probability density of energetic position of diabatic states in the wigner ensemble with filled curves being the $\pi\pi^*$ states and solid lines the $n\pi^*$ states.

To better analyze the energetic positions of the diabatic states in the ensemble, their energetic distribution is shown in **Figure 2.5(b)**. The first absorption band till 5.0 eV is mostly comprised of the lowest two $\pi\pi^*$ states. Above 5 eV the distributions of the $n\pi^*$ states are highly overlapped with the bright $\pi\pi^*$ states. The distribution of $n\pi^*$ state (S3 in Table 2.1) shows high overlap with the S4 $\pi\pi^*$ state. Similar high overlap of energetic distribution is seen in last bright state S7 and $n\pi^*$ state S6. The dark S5($n_{Oxygen}\rightarrow$L) state displays a comparatively broader energetic distribution having significant overlaps will all states except the lowest bright S1 state. This mixed nature of S5 is also reflected in the natural transition orbitals at the Franck-Condon geometry where the most predominant contribution is 52% as shown in **Figure 2.3**.

2.2.4 Conclusions

In this work a nuclear ensemble at 300K for the solvated cytidine molecule was created, and its linear absorption till 7 eV was analyzed using RASPT2 based methods. An expanded active-space augmenting the valence CASSCF active-sapce with four extra orbitals is necessary to obtain accurate energetics. The SS and MS versions of CASPT2 showed discrepancies in energetic positions when compared against the reference XMS-CASPT2 computation. This is also reflected in the linear absorption computed from the ensemble with XMS-CASPT2 showing the closest agreement with the experimental spectra. The effect of using a smaller active-space is not drastically adverse in the relative positions of the excited-states at the Franck-Condon geometry, which suggests that it could be safely employed with XMS-CASPT2 for non-adiabatic dynamics simulations.

Chapter 3

Photoinduced Reaction mechanisms from adiabatic surfaces

3.1 Theory - *how potential energy surfaces reveal photophysical pathways*

The use of Born-Oppenheimer approximation leads to the concept of a potential energy surface (PES) of the system [83]. By computing the solutions of the electronic Hamiltonian (Equation 1.4) at various nuclear geometries, one can obtain these surfaces. In contrast to standard thermochemistry, one has multiple PES corresponding to those of each electronic eigenstate. One should recall that these surfaces are solutions of the Hamiltonian which ignores the nuclear kinetic energy. Therefore these surfaces dictate *adiabatic motions* of the nuclei[84]. In other words, these surfaces dictate the motions of infinitely slow moving nuclei in the field of electrons. This adiabatic approximation is valid when these electronic PES are well separated in energy as is usually true around ground-state geometry.

However due to deformations of nuclear geometry, these surfaces can also become close to another and can cross each other. These crossings of PES are called as conical intersections when involving the states sharing the spin and multiplicity and they are ubiquitous in nature. In these regions the adiabatic approximation breaks down and the system can traverse from one PES to another through these conical intersections, thus working as doorways from higher to lower states. The interaction of system with light can force the necessary nuclear motions to induce these *nonadiabatic* processes. Indeed these nonadiabatic processes are the cornerstone of photophysics and photochemistry. The passage through these conical intersections can provide efficient pathways of ultrafast nonradiative relaxation [25] or can lead to formation of new photoproducts[1, 85].

3.1.1 Breakdown of Born-Oppenheimer approximation

The electronic wavefunctions ψ_i obtained by using Born-Oppenheimer approximation are the solutions to the electronic Hamiltonian Equation 1.4, obtained by ignoring the nuclear kinetic energy operator $T_{nuc} = \sum_\alpha \frac{-1}{2M_\alpha}\nabla_\alpha^2$ from the full Hamiltonian (Equation 1.3). Therefore these eigenstates ψ_i are not the solutions of full molecular Hamiltonian and interact with each other through the nuclear kinetic energy operator. This interaction can be written as matrix elements (notated as K_{ij}) of this operator

$$K_{ij} = \langle\psi_i|\hat{T}_{nuc}\psi_j|\rangle = \langle\psi_i|\sum_\alpha \frac{-1}{2M_\alpha}\nabla_\alpha^2\psi_j|\rangle \tag{3.1}$$

with the summation being carried over α nuclei. Using chain rule this expression becomes

$$K_{ij} = -\sum_\alpha \frac{1}{2m_\alpha}\left(\underbrace{\langle\psi_i|\nabla_\alpha^2\psi_j|\rangle}_{scalar\ coupl.} + 2\cdot\underbrace{\langle\psi_i|\nabla_\alpha\psi_j|\rangle}_{derivative\ coupl.}\cdot\nabla_\alpha\right) \tag{3.2}$$

Thus the full nonadiabatic coupling is composed of two terms derivative coupling f_{ij} and scalar coupling k_{ij}. The scalar coupling term $\mathbf{k}_{ij}$ can be expressed in terms of the derivative coupling term $\mathbf{f}_{ij}$

$$\langle\psi_i|\nabla^2\psi_j|\rangle = \nabla\cdot f_{ij} + f_{ij}\cdot f_{ij} \tag{3.3}$$

Therefore the non-adiabatic coupling is defined by the properties of the derivative coupling term f_{ij}. The value of the matrix element of f_{ij} term is given by

$$f_{ij} = \langle\psi_i|\nabla\psi_j|\rangle = \frac{\langle\psi_i|\nabla H\psi_j}{E_j - Ei} \tag{3.4}$$

Thus the derivative coupling term between the adiabatic states is inversely proportional to the energy-gap between them. When the electronic energy gap δE = 0 , the value of derivative coupling is infinity. These are regions of complete breakdown of Born-Oppenheimer approximation and appearance of conical intersections. In regions efficient non-radiative relaxation can take place due to pathways connecting the different Born-Oppenheimer surfaces.

3.1.2 Adiabatic vs diabatic states

The matrix **K** which couples the electronic eigenstates can be diagonalized by a unitary transformation of the adiabatic states ψ_i. These new states ϕ_i obtained as a result have zero nonadiabatic couplings, and are called *diabatic* states. The PES obtained by these states govern the motion of the nuclei with finite velocities. For a basis consisting of two adiabatic states such a transformation is given by

$$\begin{pmatrix}\psi_1\\ \psi_2\end{pmatrix} = \begin{pmatrix}cos(\theta) & -sin(\theta)\\ \sin(\theta) & cos(\theta)\end{pmatrix}\begin{pmatrix}\phi_1\\ \phi_2\end{pmatrix} \tag{3.5}$$

The angle θ is called as mixing angle for which the coupling $K_{ij} = 0$. These diabatic states are related to the character of the electronic state for example a $\pi\pi^*$ or $n\pi^*$ state.

3.1.3 Conical Intersections

Conical intersections are regions where two adiabatic surfaces cross each other. The large value of non-adiabatic coupling in these regions leads to efficient non-radiative transitions in the system. In a basis of two diabatic states, each adiabatic state can be expanded as

$$\begin{aligned} \psi_1 &= c_{11}\phi_1 + c_{21}\phi_2 \\ \psi_2 &= c_{12}\phi_1 + c_{22}\phi_2 \end{aligned} \tag{3.6}$$

Using these diabatic basis, the electronic Hamiltonian can be built in the basis of these diabatic states

$$\begin{aligned} H^e &= \begin{pmatrix} H_{11} & H_{12} \\ H_{21} & H_{22} \end{pmatrix} \\ H_{ij} &= \langle \phi_i | H^e \phi_j | \rangle \end{aligned} \tag{3.7}$$

The diagonalization the this Hamiltonian matrix gives us the resulting eigenvalues of the adiabatic state as

$$\begin{aligned} E_{1,2} &= \frac{H_{11} + H_{22} \pm \sqrt{(\Delta H)^2 + H_{12}^2}}{2} \\ \Delta H &= H_{11} - H_{22} \end{aligned} \tag{3.8}$$

From this there are two independent conditions for degeneracy i.e. $E_1 = E_2$, which are

$$\begin{aligned} \Delta H = H_{11} - H_{22} &= 0 \\ H_{12} &= 0 \end{aligned} \tag{3.9}$$

The existence of these two independent conditions implies that that two independently varying parameters are needed to obtain these points of degeneracy. In a diatomic system there is only one parameter which the bond-length between the atoms. These results in these conical intersections being present only for states of different symmetry where $H_{12} = 0$ and $\Delta H = 0$ being fulfilled for some particular bond-length. In systems with N atoms (N $>$ 2), there are N^{int} = 3N-6 independent parameters. Thus these conical intersections are ubiquitous in polyatomic systems. In these systems the degeneracy is preserved in N^{int}-2 directions to the first order forming an intersection seam along which the conical intersections are connected in a continuous path [86].

The degeneracy is lifted in two directions forming the branching-plane. To obtain these directions we can express the Hamiltonian matrix elements by Taylor series expansion around

the conical intersection point $\mathbf{R_0}$

$$\begin{aligned} H(\mathbf{R}) &= H(\mathbf{R_0}) + \nabla H(\mathbf{R_0}) \cdot \delta(\mathbf{R}) \\ \Delta H(\mathbf{R}) &= 0 + \nabla \Delta H(\mathbf{R_0}) \cdot \delta(\mathbf{R}) \\ H_{12}(\mathbf{R}) &= 0 + \nabla H_{12}(\mathbf{R_0}) \cdot \delta(\mathbf{R}) \end{aligned} \tag{3.10}$$

which leads to the degeneracy conditions as

$$\begin{aligned} \nabla \Delta H(\mathbf{R_0}) \cdot \delta(\mathbf{R}) = 0 \\ \nabla H_{12}(\mathbf{R_0}) \cdot \delta(\mathbf{R}) = 0 \end{aligned} \tag{3.11}$$

These two equations define the two directions $\vec{g}$ and $\vec{h}$ which lift the degeneracy around the conical intersection

$$\begin{aligned} \vec{g} &= \nabla \Delta H \\ \vec{h} &= \nabla H_{12} \end{aligned} \tag{3.12}$$

These two vectors span the branching space. The vector $\vec{g}$ which brings the diabatic states together is called as the *tuning* mode and $\vec{h}$ which couples the diabatic states is called as *coupling* mode.

Since the diabatic states are obtained by a unitary transform of adiabatic states, the same branching plane can also be defined in terms of adiabatic states and eigenvalues V_1 and V_2

$$\begin{aligned} \vec{X_1} &= \nabla_R (V_1 - V2) \\ \vec{X_2} &= (V_2 - V_1) \langle \psi_1 | \nabla_R \psi_2 \rangle \end{aligned} \tag{3.13}$$

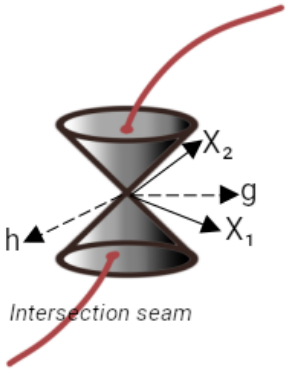

Figure 3.1: Conical intersection with the two branching-space vectors $\vec{g}$ and $\vec{h}$. Also shown in red is the intersection seam, which is perpendicular to the branching plane at the CI

The vectors X_1 is the *gradient difference* and X_2 is the derivative coupling vectors. These terms can be obtained from outputs of standard quantum chemistry programs.

3.2 Ultrafast photophysics of solvated tryptophan

3.2.1 Introduction

Tryptophan(Trp) is an amino acid widely used a probe to follow structural dynamics due to its strong fluorescence[87–92]. So, it has been a subject of intense photophysical investigation since decades. In particular, the ultrafast dynamics triggered by UV light have been a matter of interest in numerous experimental and theoretical studies [93–106]. In spite of a large amount of work, the primary processes occurring in the sub-50 fs time scale, which triggers the system response, are largely unexplored due to both technical challenges and the intrinsic electronic structure of Trp. The lowest electronic energy manifold of Tryptophan consists of two close-lying electronic states. These states labelled as L_a and L_b have different nature with L_a being a polar state and L_b a non-polar state. This difference in polarity is a consequence of the redistributed electronic density after excitation from ground-state as seen in **Figure 3.2**, where L_a has internal charge-transfer from the five to six ring while L_b has more evenly distributed excitation. Due to this different nature of polarity, these states experience a different response of the environment especially in polar media like water which can stabilise the L_a state. In particular, La is above Lb in the Franck-Condon (FC) region, but it is believed to become the lowest excited state collecting the population upon solvent relaxation.

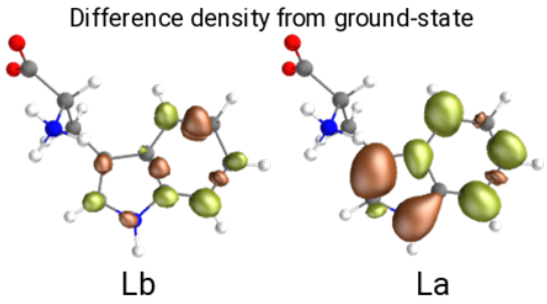

Figure 3.2: Difference density for Lb and La with respect to the ground state. Red and green signify regions of electron depletion and addition compared to ground state electron density, respectively.

3.2.2 Previous works on ultrafast dynamics

For this reason ultrafast spectral dynamics observed in polar solvents have been attributed to Lb→La internal conversion (IC). The time scale of this process has been a subject of debate in literature. Ruggiero et al[101] attributed a time constant of 1.6 ps, while Chergui et al. [102] assigned a sub-100 fs time constant followed by a bi-exponential solvent relaxation with timescale of 160 ± 40 fs and 1.02 ± 0.12 ps. Shen et al.[93] and Sharma et al.[106] also attributed the ps time constant to solvent relaxation dynamics, thus implying that the IC is a sub-ps event. Considering the ability of modern ultrafast lasers to coherently excite a superposition of vibronic states, it is illuminating to regard the photo-physics through a quantum wavepacket-based formalism. Approaching from this angle the analysis of their time-resolved fluorescence data, Yang et al.[103]

considered the state prepared after 290 nm (4.27 eV) excitation as a superposition of La and Lb states and postulated that the decay of the superposition as well as the Lb→La IC occur on a sub-100 fs time scale.

There is a distinct lack of theoretical studies on the electronic properties of solvated Trp. However, works done on indole, the chromophore of Trp, can shed light into the photo-physics triggered by UV light[105, 107–111]. Giussani et al.[105] mapped out the minimum energy paths in gas phase at CASSCF//CASPT2 level of theory, thereby demonstrating that in the absence of a polar solvent Lb is energetically more stable with respect to La. They reported two conical intersections (CI) between Lb and La hinting at a potential La→Lb population transfer immediately after photoexcitation that had remained elusive in time-resolved experiments. The existence of a strong vibronic coupling required for an efficient IC was demonstrated by Brand et al.[107, 108] Relying on non-adiabatic dynamics in explicit solvent using TD-DFT, Wohlgemuth et al.[104] obtained a time constant of 45 fs for the La→Lb IC, along with a minor repopulation of the ground state (GS) through a $\pi\sigma^*$ state accessed from La. [96, 112].

A careful examination of the literature summarized in the previous paragraphs suggests the intriguing idea that IC in Trp is characterized by a sub-ps, CI assisted, solvent sensitive population transfer back and forth between La and Lb. In the present work,an accurate characterization of the dynamics following photo-excitation of solvated Trp in the sub-5 ps regime is presented. The study combinies transient absorption (TA) spectroscopy with 30-fs temporal resolution (*courtesy Cerullo, Kabaciński et al.*) and theoretical computations at CASPT2 level incorporating solvent effects within a hybrid QM/MM setup. Definitive spectral fingerprints are obtained confirming the hypothesized sub-50 fs initial population transfer from La to Lb, which is followed by a reverse solvent-driven repopulation of the La state on the ps time scale. It is demonstrated that solvent reorganization dynamics dictate the electronic state order, and consequently, the direction of the population transfer. Additionally, it is shown that a vibronic coherence established upon photoexcitation and lasting several ps is a signature of the La/Lb coupling facilitating the non-adiabatic dynamics.

3.2.3 Optimization of ground-state

To locate the most populated ground-state structure, the major conformers of Tryptophan in its zwiterionic form were sampled by multiple runs of molecular dynamics in TIP3P water totalling upto 30 ns. The molecular dynamics were run at 300K with 1 atm of pressure (NPT ensemble). These major conformers of Trp can be differentiated based on the orientation of the indole chromophore w.r.t. to the alanyl side-chain containing the -NH_3^+ and -COO^- groups. The conformational analysis resulted in 3 major rotamers (labelled as **A**,**B** and **C**) which can each exist in two forms due to the rotation of the indole shown in **Figure 3.3**.

The geometry optimization of a selected snapshot of the most populated conformer was done at MP2/ANO-L-VDZP method, using a QMMM setup on a spherical droplet of radius 20 A°. The QMMM setup included the whole Trp molecule in High (full QM described), solvents within first two solvent shells in Medium (movable MM) , and the rest of the solvents in low(frozen

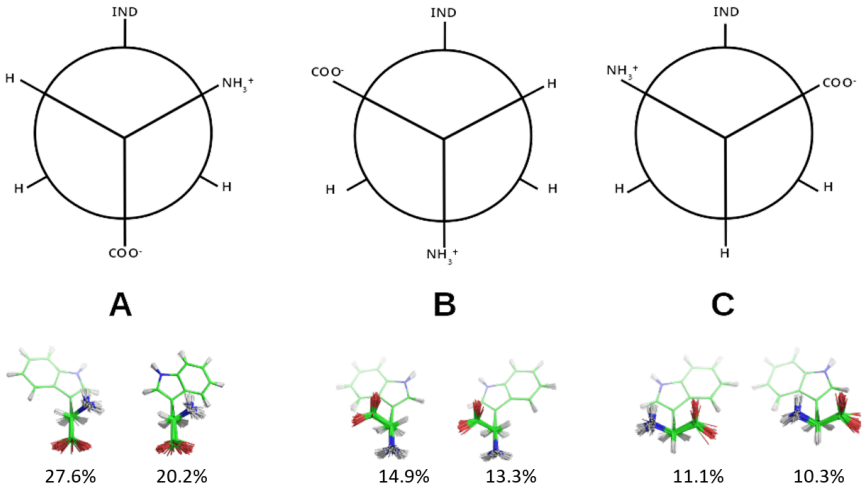

Figure 3.3: The major conformers of Tryptophan from molecular dynamics simulations with their populations. The 3 major rotamers are labelled as **A**,**B** and **C** and depicted through newman projections. Each rotamer can exist in two different forms due to rotation of the indole ring, which is also shown.

MM) layer (**Figure 3.4**).

3.2.4 Interpretation of experimental ultrafast pump-probe spectra

The computation of excited-states of Trp was performed using CASPT2 based methods. The minimal valence active-space employed 10 electrons in 9 π,π* orbitals |10,9|. This active-space was also employed for geometry optimization of the La/Lb state. To improve upon the accuracy, this active-space was expanded by adding 4 additional virtual orbitals in RAS3 subspace giving an active-space of 0|10,9|2,4. This expanded active-space was used to compute vertically excited-states for the simulation of transient spectra.

The photo-induced dynamics of Trp are investigated experimentally by pumping at 4.70 eV (264 nm) and 4.37 eV (284 nm). As seen in **Figure 3.5(a)** the spectral windows are chosen for the purpose of exciting predominantly either the La or the Lb electronic states which are the states involved in the IC following excitation of the lowest absorption band, in order to discriminate the decay paths associated with each state. **Figure 3.5(b)** shows the map for 4.37 eV pump (the map for 4.70 eV pump is shown in the SI, Figure 3.12). The two recorded transient maps show remarkable similarity across the 5 ps timescale, suggesting that the photo-induced dynamics are independent of the nature of the initially populated state. Positive signals are registered over the entire probe window indicating the presence of intense photoinduced absorption (PA) covering the stimulated emission (SE) from the transient species expected above 3 eV. In particular, the spectrum above 3 eV is characterized by the disappearance of a strong PA band at 3.35 eV (PA1) and the simultaneous appearance of a band at 3.75 eV (PA2) over the course of 5 ps (**Figure 3.5** c). **Figure 3.5 d and e** compares experiments (transient spectra at 50 fs and 5 ps delay times) with calculations (manifold of higher lying excited states bright from either Lb or La). This

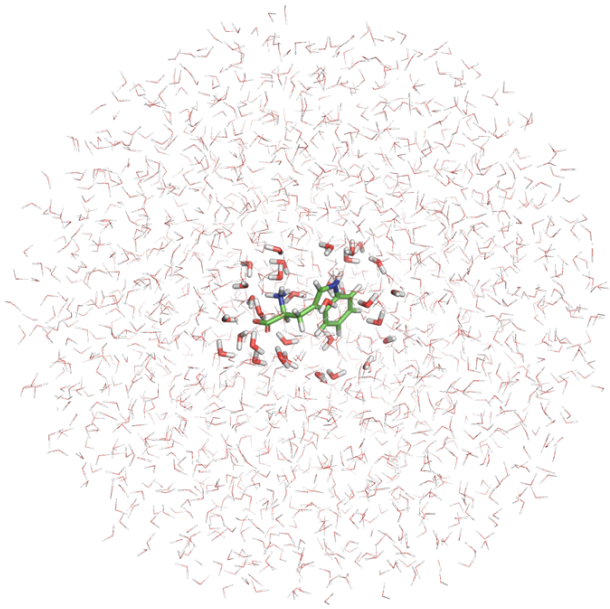

Figure 3.4: QMMM setup of solvated Trp

allows us to assign PA1 and PA2 bands as fingerprints of Lb and La, respectively and shows that Lb is populated at early times whereas at later times the population is in La. Further, less intense PA signals are observed below 3 eV where Lb and La exhibit a spectrum rich of weak absorption features. It deserves notice that this region is obscured by the PA coming from solvated electrons formed as a minor by-product due to the photoionization of the indole chromophore (dashed line in **Figure 3.5** (d)). The Trp+ cation is expected to contribute to the PA around 2 eV and 3.5 eV (Supplementary Figure. 3.15). Global analysis of the experimental TA data reveals two time constants: 220 fs and 1.1 ps. Based on the comparison with the calculations the faster time constant is assigned to the Lb→La transfer and that the longer time constant describes the solvent-assisted relaxation of the population in the La state.

The appearance of signals characteristic of the Lb state at very early times (50 fs) even after pumping predominantly the La state (4.70 eV pump) is a strong evidence of an ultrafast, CI mediated La→Lb transfer (addressable to an initial Lb greater stability) occurring at a sub-50 fs time scale, whose direct observation is obscured by a coherent artifact in the spectra at early times. This interpretation of back and forth La ↔ Lb population transfer implies a state ordering inversion on a sub-ps time scale. This is facilitated by the substantially different electronic nature of the Lb and La states.

As shown by the difference density with respect to the ground state (GS) in Figure 3.2, La is characterized by a significantly higher permanent dipole moment due to the intramolecular (5 → 6-membered ring) charge transfer nature of the transition. This contrasts with the Lb state, which is characterized by the delocalization of partial positive and negative charges resulting in a dipole moment with a magnitude similar to that of the GS. The larger dipole moment of

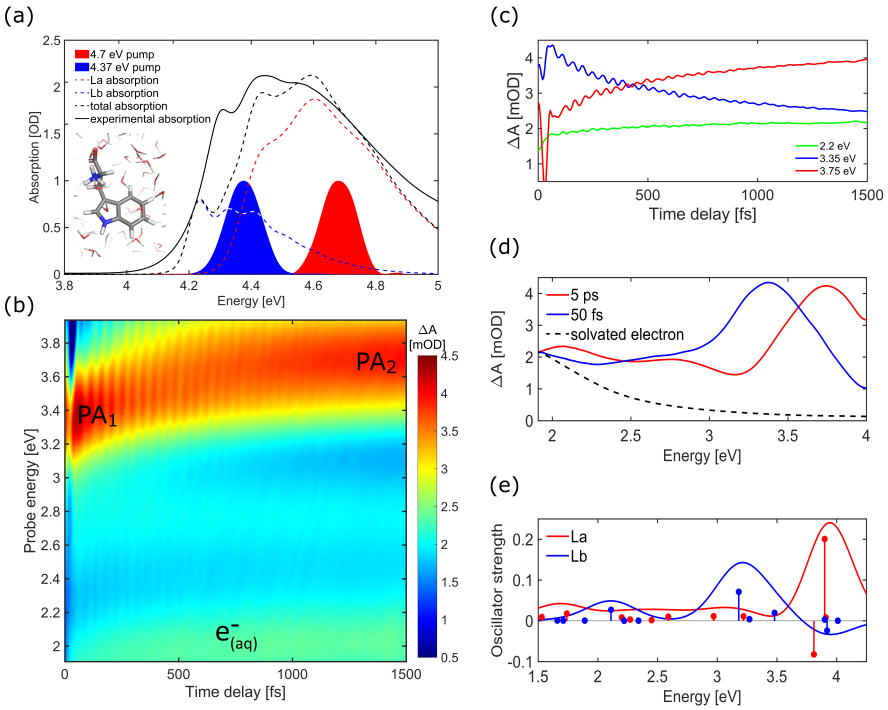

Figure 3.5: (a) Experimental and theoretical linear absorption computed at XMS-CASPT2 level of theory with an expanded active of (0|10,9|2,4) in a displaced harmonic oscillator formalism. The values of vertical excitations are in Table 3.3 The two pump windows are depicted with filled curves. (b) Experimental ΔA map following excitation at 4.37 eV. (c) Temporal dynamics at selected energies showing simultaneous decay of PA1 and rise of PA2 (d) Transient absorption spectra following excitation at 4.37 eV at selected time delays 50 fs and 5ps. (e) The theoretical spectra (positive for photoabsorption, negative for stimulated emission) from the Lb and La excited-state minima. *Experimental plots courtesy Piotr Kabaciński et al.*

La is expected to induce a significant dynamical response (relaxation) of a polar solvent such as water. The solvent relaxation would thus cause the stabilization of La with respect to Lb, which is initially lower in energy and is less sensitive to the solvent reorganization, eventually leading to inversion of the state ordering. In the following, the coupled solute-solvent dynamics are discussed in detail.

3.2.5 Branching plane at early-times

We begin the discussion by focusing on the electronic structure of Trp immediately after the interaction with the pump pulse. As the solvent is in equilibrium with the GS electronic density, La is above the Lb (ca. 0.3 eV vertical energy difference). The region of the potential energy surface (PES) relevant for the non-adiabatic dynamics at early times, i.e. before solvent relaxation kicks in, is conveniently displayed by means of the branching space around the Lb/La minimum energy CI (**Figure 3.6**). The branching plane is defined by a pair of vectors termed gradient

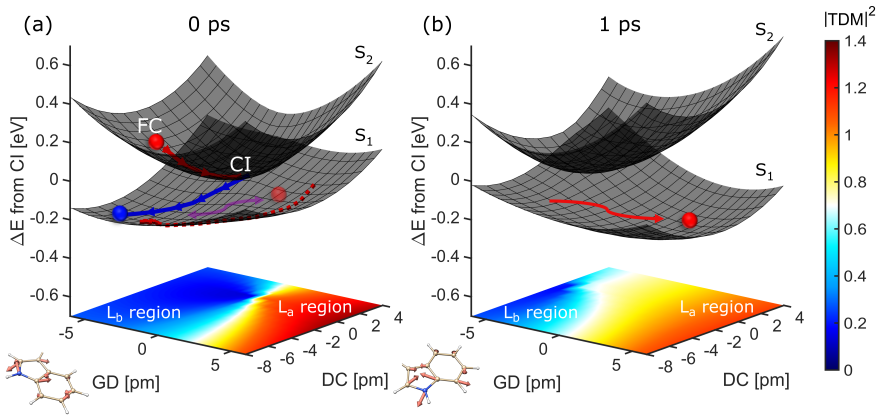

Figure 3.6: Branching space with **(a)** solvent relaxed to ground-state electron density representative of early times and **(b)** after 1ps of solvent dynamics after the interaction with the pump pulse. The geometrical deformations associated with the derivative coupling (DC) and gradient-difference(GD) vectors are also depicted with arrows. The xy-plane shows as a heat map of the $|\mathrm{TDM}|_2$ from the ground state to the lower adiabatic surface $\mathbf{S_1}$. Colours allow to characterize the nature of S_1 surface as La (red), Lb (blue) or mixed (cyan/white/yellow) on the basis of $|\mathrm{TDM}|^2$ for GS→S1 transition. **(a)** Red and blue lines denote the projection of minimum energy path from the FC point to the CI and from the CI to the Lb minimum, respectively. The dashed red line depicts the optimization from the La region to the Lb region on the lower surface. The tendency of a hot wavepacket on the S1 surface to explore La and Lb regions is shown schematically through a double-headed magenta arrow. **(b)** The stabilization of La region on S1 surface leads to transfer of population from the Lb region depicted schematically with a red arrow

difference(GD) and derivative coupling(DC), which lift the degeneracy between the electronic surfaces thereby giving rise to the characteristic double cone topology of the PES around the conical intersection (CI). To aid the discussion the electronic character of the lower adiabatic surface (termed S1) is depicted through the magnitude of the transition dipole moment from the GS clearly demarcating regions of La (red) and Lb (blue) character, as well as regions of strong wave function mixing (cyan/white/yellow).

The branching plane has been computed at multi-state level using the RMS-CASPT2 flavour implemented in OpenMolcas. Multi-state CASPT2 formalism attempt to diagonalize the hamiltonian at PT2 level by rotating the original CASSCF states. This is necessary to have a correct description of the topology of the adiabatic surfaces around conical intersection. The branching plane surfaces of the system at various levels of CASPT2 theory,single-state (SS), multi-state(MS), eXtended(XMS) and Rotated multi-state(RMS) are shown in the **Supplementary Figures 3.10 and 3.11**. It can be seen from the discussion in Supplementary that RMS-CASPT2 method shows the best agreement with reference XMS-CASPT2 with an expanded space. Therefore RMS-CASPT2 has been used to compute the adiabatic surfaces of the branching plane coordinates displayed here.

The minimum energy path initiated on the S2 surface at the FC geometry (i.e. in the La state) can reach the CI without encountering any energetic barrier (red line in **Figure 3.6(a)**). If we further proceed in the direction of the aforementioned MEP switching to the S1 surface,

essentially trying to emulate the behaviour of a momentum conserving wave packet across the CI, we encounter a region of predominant La character indicating that the CI crossing along the imaginary continuation of the MEP is to a large degree diabatic, i.e. character preserving. The La region does not display a local minimum that could lead to population trapping. Instead, the branching space topography indicates that the La/Lb vibronic coupling around the CI would allow for the relaxation to proceed on the S1 surface circumventing the CI and traversing the region of strong mixing in order to reach the energetically more stable Lb region. In support, a full-dimensional geometry optimization initiated in the La region on S1 ends up in the Lb region (projection on the branching plane represented through a dashed red line in **Figure 3.6(a)**).

The nature of the branching plane vectors indicates that the CI is reached along high frequency C-C and C-N stretching modes. In particular, the very first passage is possible within a $\frac{1}{4}$ of a period, i.e. as short as 5 fs after interaction with the pump, while the Lb region can be reached within just a single period. This elucidates why Lb fingerprints dominate the experimental spectrum already at early times even if pumped at 4.70 eV (**Figure 3.5(d,e)**), where absorption from Lb is negligible.

It deserves noting that the S1 surface is rather flat, with the Lb minimum region being only 0.1 eV more stable than the La allowing the "hot" wave packet arriving from S2 to spread and explore regions of both Lb and La character after excitation by the pump-pulse.

3.2.6 Branching plane with solvent relaxation through non-equilibrium molecular dynamics

The possibility to reside simultaneously in regions with Lb and La character is essential as it allows fractions of the wave packet in the La region to polarize the environment, thereby inducing large-scale solvent reorganization leading to electrostatic relaxation of the system. To study the coupled solute-solvent dynamics the response of the solvent to the electronic structure of either Lb or La was modelled by means of classical non-equilibrium dynamics. The Trp atomic charges were fitted to the electron density of the corresponding electronic state, whereas the effect of the solvent reorganization on the electronic structure of Trp was addressed by tracking the La-Lb energy gap during the dynamics. These non-equilibrium dynamics were carried in fully molecular mechanics scheme with AMBER by inserting Merz-Kollman charges[113] fitted to CASPT2 density using Multiwfn [114, 115] shown in Table 3.2. The dynamics were initialized on 100 decorrelated solvent snapshots obtained by solvent sampling around restrained solute. During the non-equilibrium dynamics, the solute was restrained by harmonic forces. Before computing CASPT2 energies on selected snapshots from dynamics, the respective CASPT2 optimized geometries of La and Lb minimum were re-inserted in place of MM solute. To avoid any kind of solvent bias 100 trajectories were run for 5000 fs and the reported values were computed taking the ensemble average over the 100 copies

Solvent relaxation around the Lb electron density preserves the state ordering (dashed line in Suppl Figure 3.16), that is the Lb region remains more stable than the La region on the S1 surface throughout. In contrast, solvent relaxation around the La electron density reveals that

inversion of the state ordering happens within 100 fs (solid line in **Figure 3.16**) and the adiabatic stabilization of La continues in the picosecond regime. This results in a drastic change in the topography of the potential energy surfaces as demonstrated by the branching plane after 1 ps of solvent reorganization dynamics (**Figure 3.6b**). Expectedly, the La region becomes more stable, which allows it to collect the S1 population through a back-transfer from Lb. This process is observed experimentally resulting into the decay and simultaneous rise of the Lb (PA1) and La (PA2) fingerprint signals, respectively, with a time constant of 220 fs.

3.2.7 Solvent relaxation effects on transient spectra

Considering the pivotal role PA1 and PA2 play in identifying the Lb and La states in the course of the photoinduced dynamics, it is of paramount concern to address the effect of the coupled solute-solvent dynamics on the spectral dynamics of the fingerprints. To this aim, the transient signals were computed at different times between 80 fs and 5 ps along the solvent reorganization dynamics. The spectra (**Figure 3.7**) reveal that the relaxation of the solvent around the La electron density does not lead to any notable spectral shift of the PA2 signal. Thus, we can discard the interpretation that the disappearance of the PA1 and simultaneous appearance of PA2 are consequence of a solvent induced blue-shift of a PA signature of the La state, thereby reinforcing the Lb→La population transfer interpretation.

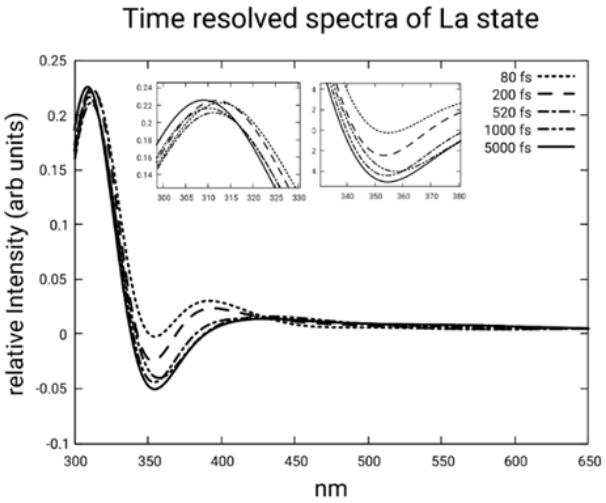

Figure 3.7: Transient spectra for the La-state computed with the ensemble of structures from non-equilibrium solvent dynamics at times.

3.2.8 La/Lb vibronic coherence

The high resolution of the experimental setup is able to record the coherent oscillations in the transient maps which encode additional information about the process. Such oscillations are signatures of vibrational dynamics either in the excited or in the ground state and are observed

as periodic modulation of the transient signal. **Figure 3.8(a)** shows the 2D oscillation map of the pump-probe spectra. The passage of the wave packet through a minimum on the PES results in a π-phase jump across the probing wavelengths leading to a switch in the sign of oscillation amplitude (*red to blue*) in **Figure 3.8(a)**. By locating the phase jump on top of the SE or GSB signals, we can associate it with either the excited state or the ground state vibrations being observed. The average location of the phase-jump can be clearly revealed as a node in the Fourier transform of the complete 2D oscillation map till 1000 fs(Figure 3.13). The position of the phase-jump is not static in Trp and its continuously red-shifting in time as seen in **Figure 3.8(c)** . This shift is reproduced nearly quantitatively by the evolution of La emission obtained from the non-equilibrium dynamics (**Figure 3.8(d)**), thus supporting assignment of these coherence to the excited state and evidencing that the transient signal above 3eV comes from the interplay of PA and SE.

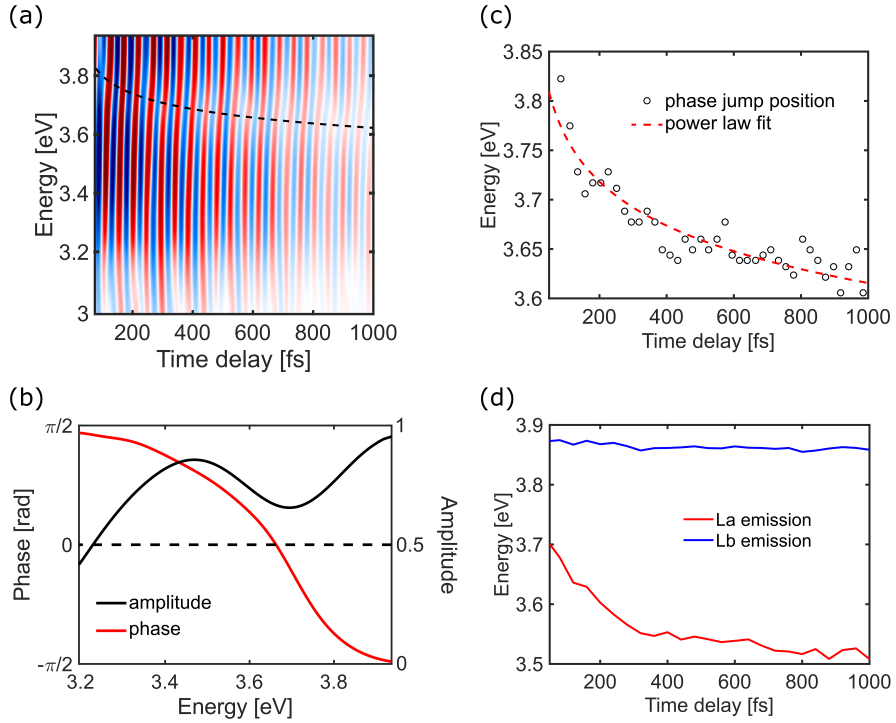

Figure 3.8: (a) Map of the oscillations recorded in the transient absorption spectroscopy map with the position of phase-jump in dashed black line (b) Amplitude and phase of fourier transfrom of oscillations Fourier transform of oscillations from panel (a) for 720 cm^{-1} frequency showing π phase-jump across the peaks at energies associated with stimulated emission(c) Experimental phase jump location fitted to a power-law decay over time revealing the vertical gap of excited-state minima. (d) Emission energies computed at SS-CASPT2 level along non-equilibrium relaxation of the solvent around the respective excited-states. *Experimental plots courtesy Piotr Kabaciński et al.*

A Fourier analysis of the oscillatory component reveals a 720 cm-1 mode dominating the entire probing wavelength. The nature of the underlying molecular vibrations are studied by

means of normal mode analysis. As it can be conjectured from the nature of the branching plane vectors (Figure 4) high frequency C-C and C-N stretching modes (>1000 cm-1), which are beyond the limits of the temporal resolution of our experimental setup, dominate the excited state vibrational dynamics. In particular, above 1000 cm-1 a mode with a frequency of 1588 cm-1 an be identified with a high Huang Rhys (HR) factor only in the La state (0.50 compared to HR value of 0.00 in the Lb Table 3.1).Consequently, this mode will be activated upon vertical excitation to the La state.This mode has the highest overlap of all normal modes with the GD vector of the branching plane which is reflected in the inversion of Lb/La state ordering in a scan along this mode (**Figure 3.9**).

Below 1000 cm^{-1} a mode with a frequency of 750 cm^{-1} can be identifies describing a distortion of the indole moiety from planarity which exhibits high HR factors of comparable magnitude in both Lb (0.34) and La (0.235). , This mode shows the biggest overlap with the DC vector of the branching plane among the excited modes (with significant Huang-Rhys factors) below 1000 cm-1, implying strong Lb/La wavefunction mixing in the direction associated with the activation of this mode.

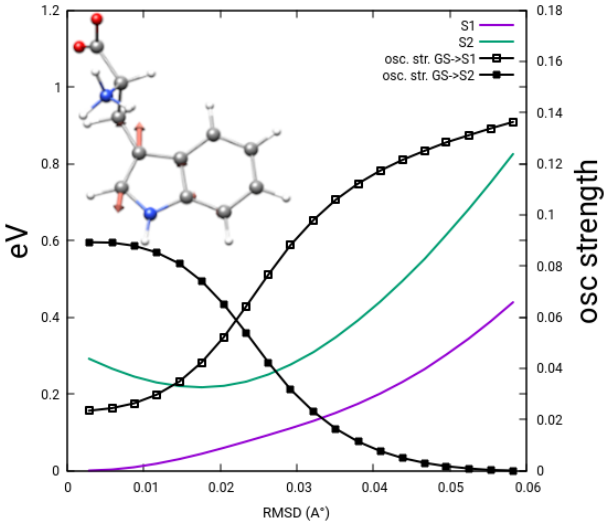

Figure 3.9: Scan along 1588 cm^{-1} mode at RMSPT2 |10,9| leads to an inversion of La/Lb through weakly avoided crossing. This is seen through reversal of the magnitudes of oscillator strengths for the GS->S1 transition

In summary, coherent dynamics along the 1588 cm^{-1} and 720 cm^{-1} modes are immediately initiated upon photoexcitation. The former facilitates ballistic access to the CI after pumping into the La state leading to quasi-minima in the La-region on S1 surface after the crossing (*represented through the red line in Figure 4 connecting the FC point on S2 and the meta-stable La region on S1 through the CI*) , whereas the latter in conjunction with aforementioned 1588cm^{-1} mode, leads to the effective La→Lb population transfer on the S1 surface circumventing the CI through the inter-state coupling region(*red dashed line connecting the La and Lb regions on S1*).). The observation that the 720 cm-1 vibrational coherence remains active on a ps time scale is evidence

that the branching plane vectors remain active and consistent with the interpretation that the wave packet does not remain trapped in the Lb valley. Instead, it coherently explores regions of Lb, La and of mixed character on the S1 surface through the activated modes including 1588cm-1 and 720cm-1 modes. This long living coherence, which leads to a simultaneous population of both states at early times, is responsible for triggering solvent relaxation to the La state that smoothly, but steadily, stabilizes the La state already from very early times, eventually leading to inversion of adiabatic state order and full population of the La state.

3.2.9 Conclusions.

By combining ultrafast transient spectroscopy in the UV with multiconfigurational CASPT2 level of theory within a hybrid QM/MM setup, this work demonstrated that the primary phtoinduced dynamics of water solvated tryptophan are driven by the two vibronically coupled lowest singlet states, La and Lb, whose different response to the polar solvent modulates the electronic population in the two state and dictate the evolution of the transient spectral signals.

Non-equilibrium dynamics show that the polar solvent responds differently to these two states, resulting in a dynamical inversion of their minima, responsible for the population dynamics observed in the picosecond regime. The evolution of the topology of the potential energy surfaces around the CI during the coupled solute-solvent dynamics reveals that the IC occurs over a two-step mechanism, a sub-50 fs La→Lb non-adiabatic population transfer followed by a 200 fs Lb→La adiabatic back-transfer mediated by the dynamical lowering of La below Lb due to solvent reorganization, which persists until ps time range, affecting exclusively the polar La state.

Back and forth La↔Lb population transfer has been proved based on theoretical models that allow to quantitatively reproduce, and interpret, the observed transient spectral features, namely: a) the La and Lb PA transient signals (**Figure 3.5(d,e)** and the La relaxation dynamics and the corresponding phase jump signal profile (**Figure 3.8c,d)**. Finally, an accurate analysis of the oscillatory pattern has allowed us to identify a photoactivated mode at 720 cm^{-1} which is responsible for ultrafast La→Lb population transfer and a system that is living as a superposition of La and Lb states, until full population of La due to its stabilization by solvent reorganization.

3.3 Supplementary

Discussion on computation of branching plane - *why use RMSPT2*

In **Figure 3.10** the topology with an expanded active-space 0|10,9|2,4 at XMS level(also referred as aug-XMSPT2) is considered as the reference when assessing the other surfaces computed at the smaller |10,9| active-space. With this consideration, the topology at RMS level with the smaller |10,9| active-space looks most similar to the reference showing a stabilization of Lb region compared to La-region. Curiously the topology of XMS at |10,9| active-space shows the opposite tendency to the reference with a slight stabilization in La-region. Single-state CASPT2 (SSPT2) does not describe the branching plane but locates a conical intersection seam instead, and the electronic character of the surfaces show little La/Lb mixing across the plane. It correctly describes the stabilization of Lb region versus La region in the Franck-Condon distribution of solvent. In contrast MS-CASPT2 has the most different topology compared to all the other methods showing a flat long valley connecting La and Lb regions along GD vector. An alternate view of the lower S1 surfaces are shown in **Figure 3.11**.

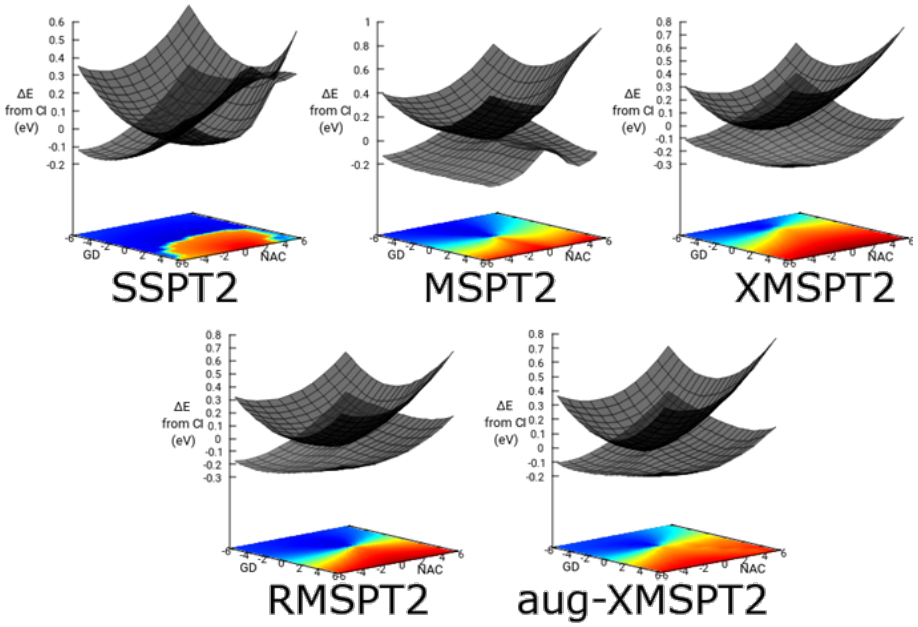

Figure 3.10: The topology of the adiabatic surfaces across the branching plane coordinates with initial solvent distribution at various levels of CASPT2. The color plot on xy plane is the TDM of the GS→S1 transition. The color coded TDM helps to characterize the regions of La(red) , Lb(blue) or mixed (cyan,white,yellow)on the lower surface. The computations have been done at single-state CASPT2(SSPT2) , and various flavors of multistate PT2 (MS,XMS and RMS). The reference computation is done at XMS-CASPT2 with an active-space of 10 electrons in 9 orbitals in RAS2 space augmented with 4 extra virtual orbitals in RAS3 space, referred as aug-XMSPT2. All the other (SS,MS,RMS and XMS) computations are done with active-space of 10 electrons in 9 orbitals |10,9|.

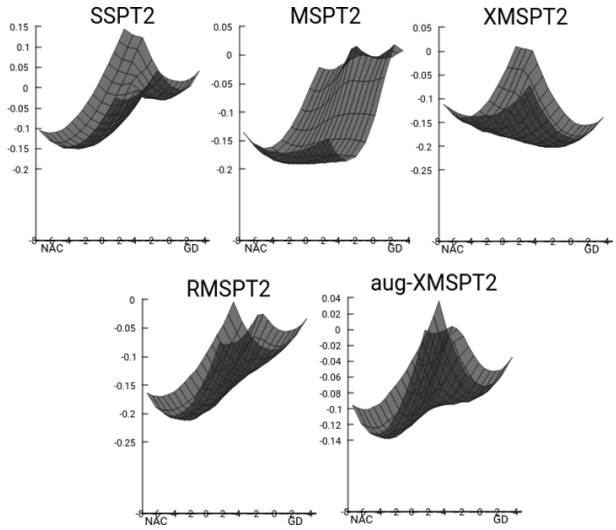

Figure 3.11: An alternate view of the the same lower S1 adiabatic surface depicted in Figure 3.10 to better appreciate the La vs Lb stabilization on the S1 surface across various levels of PT2 theory.

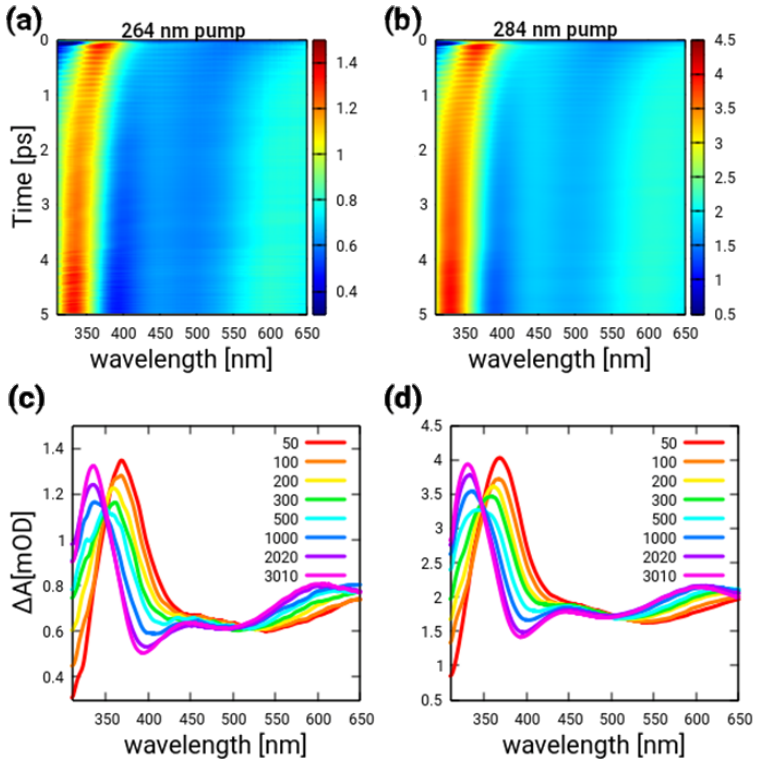

Figure 3.12: 2D ΔA map and ΔA spectra of Trp in pH 7.4 buffer solution following pump at 264nm and 284nm; (a) ΔA map following 264 nm pump; (b) ΔA map following 284 nm pump; (c) ΔA spectra following 264 nm pump and (d) ΔA spectra following 284 nm pump at different times, starting from 50 fs until 3ps. *Courtesy Prof. Giulio Cerullo et al.*

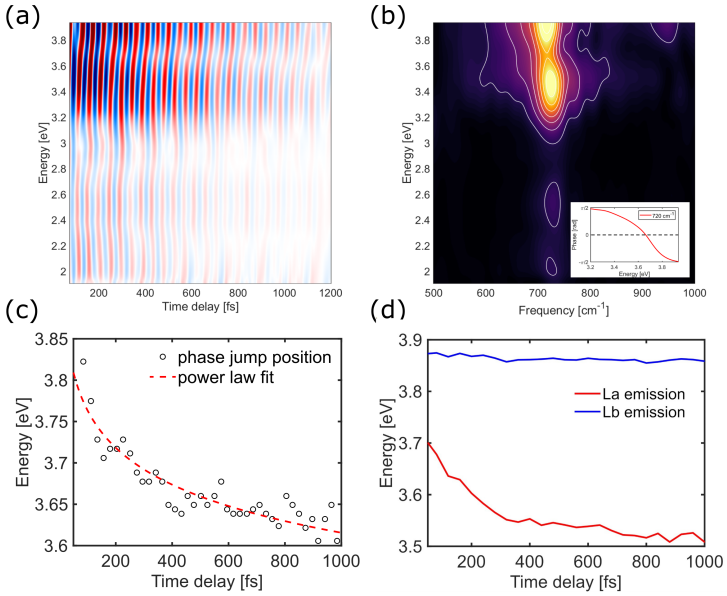

Figure 3.13: (a) Map of the oscillations recorded in the transient absorption spectroscopy map. (b)Fourier transform of the complete oscillation map till 1000 fs reveals the most active frequency at 720 cm^{-1} and average phase-jump position at 3.65 eV. (c) Experimental phase jump location fitted to a power-law decay over time revealing the vertical gap of excited-state minima. (d) Emission energies computed at SS-CASPT2 level along non-equilibrium relaxation of the solvent around the respective excited-states. *Experimental plots courtesy Piotr Kabaciński et al.*

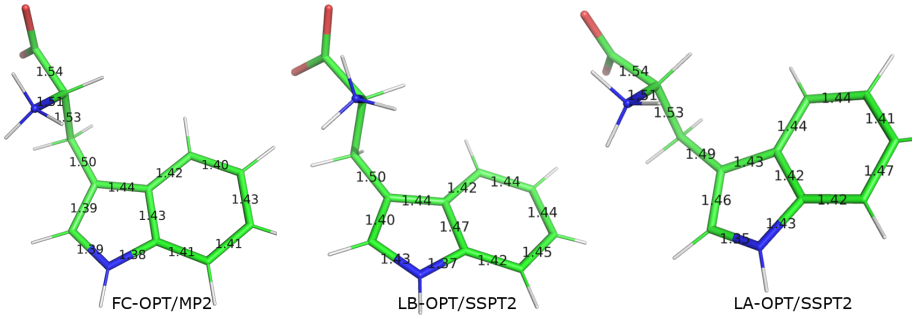

Figure 3.14: Optimised structures of the critical geometries of tryptophan.

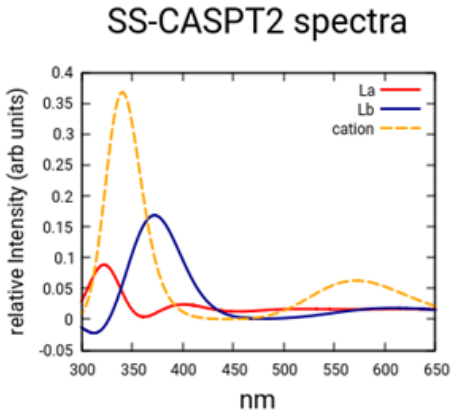

Figure 3.15: Transient spectra of La and Lb states along Linear Absorption from the cation (in dashed yellow) from its optimized minima

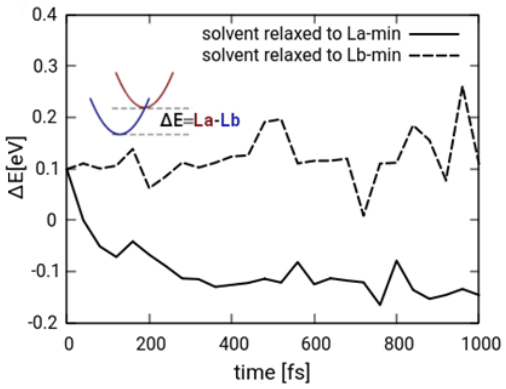

Figure 3.16: The time-evolution of the adiabatic energy gap between La and Lb at their respective minima in solvent relaxed around La state (solid line) versus Lb state (dashed line). Relaxation of solvent around La-min (solid line) leads to a level inversion of the excited state minima within 100 fs. In contrast, relaxation around Lb-state (dashed line) results in La minima always being above the Lb ones.

MP2 Frequency	Lb	La
30.2616	0.0034	0.0099
87.1706	0.003	0.1776
137.4788	0.233	0.5038
187.0386	0.2665	0.0156
197.9587	0.0062	0.057
218.1998	0.0543	0.5264
226.0516	0.0306	0.0037
233.3158	0.0059	0.0953
257.5746	0.0017	0.0015
339.0144	0.0058	0.0006
358.4333	0.0089	0.0308
409.7918	0.0147	0.0002
426.4277	0.0014	0.0014
471.9441	0.0108	0.0001
514.6048	0.0025	0.022
540.9589	0.015	0.0117
564.246	0.0198	0.0027
578.8857	0.1125	0.1221
591.0905	0.0129	0.0036
675.454	0.0017	0.004
708.6298	0.0246	0.0057
718.1794	0.02	0.0522
745.1002	0.0135	0.0413
752.5991	0.3409	0.2349
775.294	0.0448	0.0857
802.119	0.068	0.2762
817.6869	0.0001	0.033
838.0504	0.0134	0.007
871.3543	0.0009	0.1153
889.4647	0.0007	0.0016
924.2294	0.0028	0.0116
933.6901	0.0041	0.0101
943.974	0.0101	0.0973
989.6248	0.0032	0.0062
1024.383	0.1737	0.0283
1062.677	0.0037	0.0208
1097.252	0.008	0.0097
1125.912	0.0116	0.014

Table 3.1: Huang-Rhys factors of the La and Lb optimized minima at SS-CASPT2 level along the normal modes of ground-state MP2 frequencies

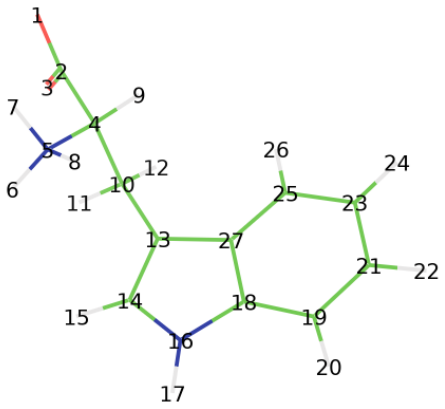

Figure 3.17: Atoms ids used in Table 3.2

Index	GS	Lb	La
1	-0.6748	-0.9412	-0.9465
2	0.7246	0.9499	0.9523
3	-0.7197	-0.9481	-0.9535
4	0.0755	0.3798	0.4193
5	-0.2905	-0.7443	-0.7265
6	0.2659	0.4462	0.4478
7	0.2659	0.4554	0.4457
8	0.2659	0.4177	0.4088
9	0.0663	0.0145	-0.0096
10	-0.0421	-0.2935	-0.3258
11	0.0486	0.1435	0.1406
12	0.0486	0.1249	0.1406
13	-0.1210	-0.1466	-0.4150
14	-0.1671	-0.1286	0.0215
15	0.2326	0.2228	0.2333
16	-0.3382	-0.6081	-0.4394
17	0.3466	0.4792	0.4464
18	0.1184	0.2593	0.1272
19	-0.2136	-0.4315	0.0891
20	0.1479	0.2283	0.1760
21	-0.1399	-0.2819	-0.5476
22	0.1395	0.1823	0.2179
23	-0.1970	-0.0718	-0.0580
24	0.1440	0.1446	0.1739
25	-0.2046	-0.2859	-0.3863
26	0.1422	0.1892	0.2191
27	0.0760	0.2434	0.1482

Table 3.2: Table showing the original AMBER charges as (**GS**) and CASPT2 density fitted **Lb** and **La** charges used for non-equilibrium dynamics. For atom ids refer to **Figure 3.17**

	|14,10|			|0,0|14,10|2,4|		
State	**SS**	**MS**	**XMS**	**RMS**	**SS**	**MS**	**XMS**	**RMS**
L_a	4.16(*0.02*)	4.13(*0.03*)	4.33(*0.03*)	4.15(*0.02*)	4.22(*0.02*)	4.20(*0.03*)	4.40(*0.03*)	4.22(*0.02*)
L_a	4.44(*0.98*)	4.49(*0.09*)	4.56(*0.08*)	4.47(*0.09*)	4.56(*0.09*)	4.58(*0.10*)	4.67(*0.10*)	4.57(*0.10*)

Table 3.3: Comparison of CASPT2 vertical excitation energies(*Oscillator Strengths*) at MP2 optimized Franck-Condon geometry of selected snapshot with full valence |14,10| and augmented 0|14,10|2,4 active space

Chapter 4

Time-dependent behaviour by semi-classical trajectories

4.1 Theory - *how to do molecular dynamics with multiple surfaces*

The time-evolution of a system can be simulated with the help of molecular dynamics. These simulations provide a way to realise the reaction paths that are undertaken by a system. These dynamical simulations give the actual time-constants that are associated with any process. In the formalism of quantum mechanics a reaction can be described as transfer of population from one eigenstate to another eigenstate. In UV-light induced phenomena this involves transfer of populations among various electronic states of the molecule. When these electronic populations transfer are slow compared to the intra-molecular vibrations, the reaction rate can be approximated through the Fermi Golden Rule. However, excitation by UV light prepares the system in a highly non-stationary state where the nuclei of the system are not relaxed to the new electronic density. Therefore if we want to understand the time-evolution of the system from this non-stationary state, we have to perform explicit dynamical computations to obtain the time-dependent behaviour of the system.

In classical physics molecular dynamics are performed by evaluating the forces and propagating the system forward in time through algorithms which integrate Newton's equation of motion eg. Velocity Verlet. These numerical methods propagate the system on the instantaneous potential energy surface. As we saw in previous Chapter, there are multiple potential surfaces in photophysics through which electronic transitions take place through conical intersections. Surface-hopping using fewest switches [116] , first proposed by Tully is a scheme which provides a method to perform non-adiabatic dynamics involving multiple potential energy surfaces.

4.1.1 Surface-Hopping with fewest switches

In surface-hopping method the nuclei always propagate on a single adiabatic surface at any given instant like classical dynamics. At every instant of time the nuclei carries with itself the wavefunctions of the various adiabatic states ψ_i and associated coefficients c_i whose squared modulus gives their population.

$$\Psi(\vec{R})\rangle = \sum_i c_i \psi_i(\vec{R})\rangle \tag{4.1}$$

A single trajectory is always started with 100% population in a single adiabatic state and zero in all others (at t=0 ,$c_i = 1$ for some $i = j$ and $c_i = 0$ for $i \neq k$). At every integration step propagating the nuclei through forces (given by matrix element $\vec{F}_{jj}(\vec{R})$), the populations are transferred between the various adiabatic states ψ_i through changes in the coefficients c_i, mediated by non-adiabatic coupling of these states (given by matrix element $\vec{d}_{jk}$).

$$\begin{aligned} \vec{F}_{jj}(\vec{R}) &= -\left\langle \psi_j | \frac{\partial H_{el}}{\partial \vec{R}} | \psi_j \right\rangle \\ \vec{d}_{jk} &= \frac{\vec{F}_{jk}(\vec{R})}{V_{kk}(\vec{R}) - V_{jj}(\vec{R})} \end{aligned} \tag{4.2}$$

where V_{jj} is the expectation value of the electronic Hamiltonian operator ($V_{jj} = \langle \psi_j | H_{el} | \psi_j \rangle$).

The population transfer between adiabatic wavefunctions is done by integrating the electronic Schrodinger equation for the c_i amplitudes.

$$ih\frac{\mathrm{d}c_j}{\mathrm{d}t} = V_{jj}(\vec{R})c_j - ih\sum_{k\alpha}\frac{P^\alpha}{M^\alpha} d^\alpha_{jk}(\vec{R})c_k \tag{4.3}$$

where M_α and P_α refer to the array of nuclear masses and momenta respectively.

The integration propagating the electronic populations in the various adiabatic states involves the derivative coupling between the wavefunctions which is usually highly localised in space. This means due to finite time-step of integration of nuclear positions, one can overstep such regions and erroneously fail to transfer adequate populations. Additionally computation of derivative coupling terms is not widely implemented for methods like CASPT2. This issue is resolved by using an alternative approach of Hammes-Schiffer and Tully [117] to instead evaluate overlap integrals of wavefunctions at adjacent time-step. Since the transfer of populations occurs due to along the velocity due to non-adiabatic coupling, one can insert the insert the velocity vector in the coupling and relate the transfer to the overlaps σ_{jk} at adjacent time-steps.

$$\begin{aligned} \frac{\mathrm{d}\vec{R}}{\mathrm{d}t} d_{jk} &= \langle \psi_i(r,t) | \frac{\mathrm{d}}{\mathrm{d}t} \psi_j(r,t) \rangle \\ &= \langle \psi_i(t) | \psi_j(t+dt) \rangle \\ &= \sigma_{jk} \end{aligned} \tag{4.4}$$

The overlaps integrals between the wavefunctions at adjacent steps contain the measure of how much the adiabatic wavefunctions change into one-another(*mix*) in the given time-step and thus can be used to transfer the populations.

The changes in populations ($|c_i|^2$) of the adiabatic states due to population transfer along dynamics leads to hopping of the trajectory to different adiabatic state changing the current active-state. Tully argued that the electronic populations of the adiabatic states should also reflect the ratio of trajectories with those active-states. So when the populations change in a time-step the trajectory should hop in accordance. The "fewest switches" needed to accomplish this can be fulfilled if hops are only allowed to states with increasing populations. The probability to hop from state j can be given by

$$p_{j\rightarrow} = -\frac{\Delta|c_j|^2}{|c_j(t_n)|^2} = \frac{|c_j(t_n)|^2 - |c_j(t_{n+1})|^2}{|c_j(t_n)|^2} \tag{4.5}$$

Using the equation for propagation of the coefficients 4.3 the probability to hop from state j to state k can be written as.

$$p_{j\rightarrow k} = \frac{1}{|c_j|^2}\sum_{\alpha}\frac{2P^{\alpha}}{M^{\alpha}}\mathrm{Re}(d^{\alpha}_{jk}(\vec{R})\bar{c_k}c_j)\Delta t \tag{4.6}$$

4.2 Ultrafast deactivation of uridine

4.2.1 Introduction

The presence of non-radiative decay paths in nucleosides prevent photodamage from ultraviolet radiation[118–123]. Thus in-spite of having high absorption in the ultraviolet region, the potential damages are suppressed due to efficient relaxation of these molecules back to ground-state. These non-radiative relaxation mechanisms occur on ultrafast sub-picosecond timescales in isolated nucleosides [124–127]. These relaxation paths are mediated by passages through conical intersections (CI) [3]. Due to the presence of multiple excited-states in the absorption spectra, their involvements needs to be disentangled to unveil the decay channels present in the system.

The excited-state manifold of pyrimidines is composed of bright $\pi\pi^*$ and dark $n\pi^*$ states. The $n\pi^*$ states are not majorly populated upon excitation due to negligible GS$\rightarrow n\pi^*$ transition dipole moment. However, their involvement cannot be ruled out during non-adiabatic relaxation processes due to presence of $\pi\pi^*$-$n\pi^*$ conical intersections in the energetic vicinity of bright $\pi\pi^*$ states [128].

In gas-phase the lowest excited-state is assigned to the $n\pi^*$ states from theoretical computations in uracil and its associated nucleoside uridine[3]. In solvents, the $n\pi^*$ state is destabilised due to electrostatic interactions with the solvent molecules resulting in $\pi\pi^*$ state being the lowest excited-state [3]. A direct experimental proof of $\pi\pi^*$ vs $n\pi^*$ energetic orderings is difficult due to the dark nature of the $n\pi^*$.

The time-constants associated with excited-state decay is influenced by the underlying photophysics involving the $\pi\pi^*$ and $n\pi^*$ states. Most experimental transient absorption(TAS) and

fluorescence upconversion(FU) spectra describe the excited-state decay with one more more sub-picosecond and a longer several picosecond time-constant. These experimental dynamics are fitted with several time-constants ranging from sub-100 fs to several picoseconds. The hypothetical mechanisms describing these time-constants are detailed below.

Direct internal conversion (IC) to the ground state ($\pi\pi$* $\rightarrow$ S0)

Modern ultrafast measurements report a fast 100 fs time-constant in pyrimidines leading to an internal conversion to the ground-state. Theoretical computations have assigned this ultrafast IC due to access to a ethylene-like conical intersection between the ground-state and the lowest $\pi\pi^*$ singlet state [27, 129–131]. This conical intersection seam is characterised by a puckering of the aromatic ring with out of plane bending of hydrogen atoms [129, 132]. An additional decay mechanism associated to ring-opening due to N_1-C_2 bond cleavage has also been proposed in case of uracil [133–135].

IC to dark $n\pi$* state ($\pi\pi$* $\rightarrow$ $n\pi$*)

This mechanism proposes the involvement of the $n\pi^*$ states as an intermediate state in the internal conversion mechanism, with the shortest time-constants associated to a decay to this dark state through a $\pi\pi^*$-$n\pi^*$ conical intersection. The population in $n\pi^*$ state can decay through various proposed channels. A direct IC to ground state might occur through an energetically higher lying $n\pi^*$-ground state conical intersection[136, 137]. The $n\pi^*$ state can also lead to recrossing with the $\pi\pi^*$ state leading to $n\pi^\star \rightarrow \pi\pi^\star \rightarrow S_0$ mechanism of relaxation to the ground-state[127]. The $n\pi^*$ can also act as a gateway to the triplet states through intersystem crossing routes. [137, 138].

To investigate the mechanism of the ultrafast decay of UV excited uridine, surface-hopping dynamics with CASPT2 based methods employing $\pi\pi^*$ active-space of |10,8| (10 electrons in 8 orbitals) were performed by Nenov et al. [25] to interpret sub-30- fs transient absorption spectroscopic measurements in a joint theoretical and experimental work. The study identified that the after excitation by the pump-pulse, the lowest $\pi\pi^*$ singlet state decays within 100 fs top the ground-state through the GS-$\pi\pi^*$ conical intersection. More importantly, the simulations revealed a minimal involvement of $n\pi^*$ state with an upper estimate of 20% population transfer to this channel. The involvement of $n\pi^*$ state was estimated by re-computing the energetic along all trajectories with an expanded active-space taking into consideration nπ^* states followed by ad-hoc Tully FSSH in the basis of the $\pi\pi^*$ and $n\pi^*$ states.

In support of this study, the present author performed surface-hopping simulations on a subset of structures by explicitly including the nπ^* states with the full valence active space of |14,10| (14 electrons in 10 orbitals). These surface-hopping simulations were performed with a state-average of nine states at XMSPT2 level. This manifold covered all the lowest excited-states involving the lowest four $\pi\pi^*$ and four $n\pi^*$ states. These hopping simulations supported the minimal involvement of nπ^* state, conforming that the major decay path of uridine involves the GS-$\pi\pi^*$ conical intersection.

4.2.2 Excited-States at the Franck-Condon point

The excited-states of uridine at XMS-CASPT2 level are shown in Table 4.1 with Natural Transition Orbitals of the relevant GS→S_n transition shown in **Figure 4.1**. These transitions have been computed with an active-space of |14,10| with a QMMM formalism. The QMMM setup consisted on nucleobase in High Layer (Fully Quantum), and the effect of sugar and waters described through electrostatic embedding. For geometry optimization, the sugar and water closest waters in 3 A° layers were kept mobile (Medium Layer) while the rest of the system was kept frozen. The setup is similar to that described for cytidine in Chapter 2. The QMMM setup was used to obtain the optimized Franck-Condon structure at MP2/ANO-L-VDZP level.

State	**|14,10|**		
	SS	**MS**	**XMS**
S_1 = H→L	4.58(*0.17*)	4.70(*0.29*)	4.69(*0.23*)
S_2 = $n_{Oxygen1}$→L	5.02(*0.00*)	5.10(*0.00*)	4.94(*0.00*)
S_3 = H→L+1	5.64(*0.19*)	5.71(*0.18*)	5.54(*0.17*)
S_4 = H-1→L	5.83(*0.06*)	5.97(*0.20*)	5.89(*0.18*)
S_5 = $n_{Oxygen2}$→L+1	6.23(*0.00*)	6.24(*0.01*)	6.09(*0.00*)
S_6 = H-1→L+1	6.40(*0.79*)	6.60(*0.57*)	6.53(*0.64*)
S_7 = $n_{Oxygen2}$→L	6.98(*0.00*)	7.15(*0.00*)	6.62(*0.02*)
S_8 = $n_{Oxygen1}$→L+1	6.78(*0.01*)	6.96(*0.01*)	6.84(*0.00*)

Table 4.1: Vertical excitation energies and (*Oscillator Strengths*) at MP2 optimized Franck-Condon geometry of selected snapshot at SS,MS and XMS-CASPT2 with full valence |14,10| active-space used in surface-hopping simulations.

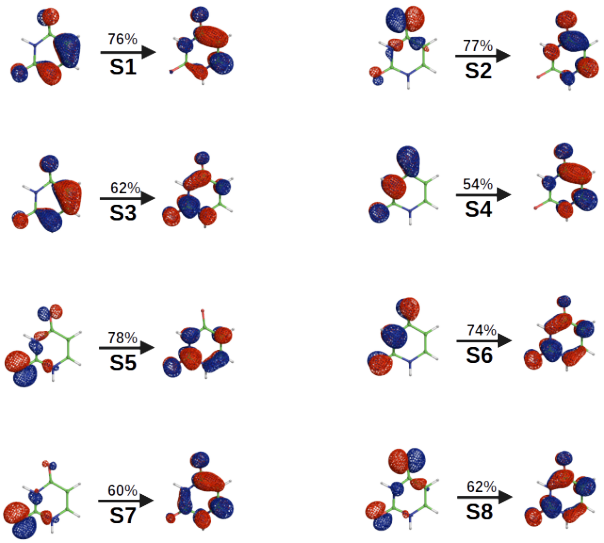

Figure 4.1: Natural Transition Orbitals of the electronic states of solvated uridine at XMS-CASPT2/|14,10|.

As seen from Table 4.1 and **Figure 4.1** the electronic manifold until 5 eV consists of a bright

$\pi\pi^*$ and a dark $n\pi^*$ state separated by 0.4 eV.

4.2.3 Technical details of surface-hopping simulations

The surface hopping simulations were performed for 25 trajectories until 500 fs following Newton's equations of motion for the nuclei and utilising hybrid QM/MM gradients at XMS-CASPT2 level with |14,10| active-space with a time-step of 0.5 fs. The surface-hopping algorithm employed Tully's fewest switches hopping algorithm with Tully-Hammes-Schiffer scheme using COBRAMM's parallel environment for computing numerical gradients through a two-point finite-difference formula. The state-averaging scheme covered the ground-state and 8 lowest $\pi\pi^*$ and $n\pi^*$ excited-states. The movable (MM) water layer comprised all water molecules within 5 Angstrom from the High Layer (nucleobase) to accommodate distortions of aromatic ring by puckering (**Figure 4.2**).

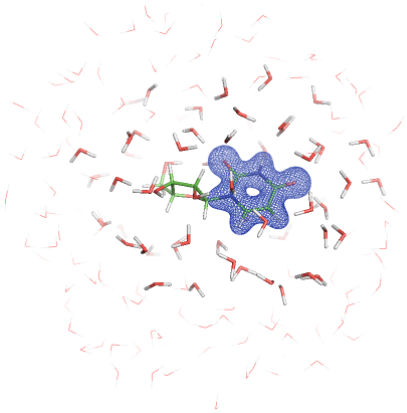

Figure 4.2: QMMM setup employed in surface-hopping simulations showing High (fully Quantum), Medium (movable MM) and Low (frozen MM) layers.

The expression for time-derivative coupling for states i and j employs the Hammes-Schiffer & Tully method and is approximated to finite differences. The change in electronic wavefunctions is resolved by computing overlap integrals between adiabatic wavefunctions at different time-steps using RASSI routine of OpenMolcas.

4.2.4 Results of Surface-hopping simulations

The surface-hopping simulations were run on a set of 25 trajectories at XMS-CASPT2 level, with the aim to compare with the simulations run at SS-CASPT2 level previously by Nenov et al. and assess the involvement of $n\pi^*$ state explicitly.24 trajectories hopped to the ground-state through the ethylene like GS-S1($\pi\pi^*$) conical intersection (**Figure 4.3**). This conical intersection has two quasi-symmetrical forms depicted in Figure 4.3 as **type1** and **type2**. The type1 conical intersection was reached with an average time of 128 fs in 15/25 trajectories and type2 with average time of 139 fs in 9/25 trajectories.

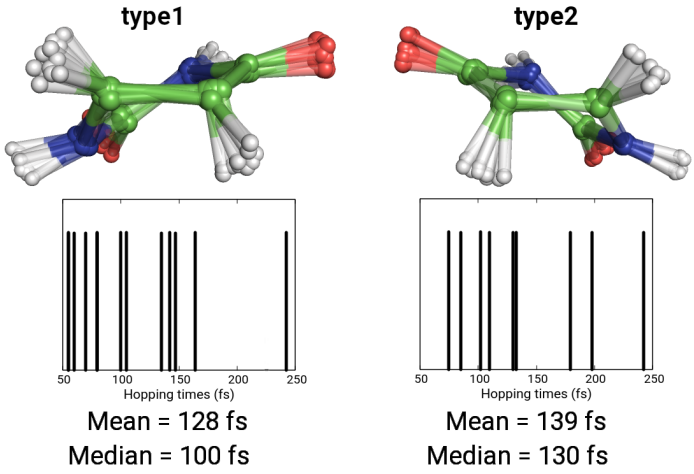

Figure 4.3: The two types of GS-S1($\pi\pi^*$) conical intersections from XMS-CASPT2 surface-hopping dynamics. Also reported are the mean and median of times of GS-S1($\pi\pi^*$) hop

To analyse the evolution of nuclear motions leading to the conical intersection seam, principal component analysis was performed on the trajectories till hopping time. The component with the largest eigenvalue is the dominant motion of the system in a linear approximation. The evolution of the trajectories along the largest principal component is depicted in **Figure 4.4**. The trajectories can be partitioned into three types of behaviour based on the time taken to reach the seam and the evolution of the principal reaction coordinate. The trajectories which are classified as showing ballistic decay show a monotonically evolution along the coordinate leading to the conical intersection seam in sub 100 fs. A lesser number of trajectories which take 100-300 fs classified as intermediate decay exhibit a few cycles of motion along the principal coordinate before monotonically reaching the conical intersection seam. A single trajectory showed exceptionally longer time to reach the CI seam in 426 fs. This trajectory displayed oscillatory motion along the reaction coordinate till 300 fs after which it reached the CI seam in a monotonic behaviour. In **Figure 4.5** examples of these three types of behaviour are shown with the evolution of potential energies along the non-adiabatic dynamics.

4.2.5 $n\pi^*$ involvement delays hop to ground-state.

The involvement of $n\pi^*$ state can trap the population and delay the time needed to access the $\pi\pi^*$-S_0 conical intersection. Since $n\pi^*$ state will have a different electronic density it can influence the nuclear motions differently and its presence is imprinted on the vibrational motions in the trajectory. Through the difference density of $\pi\pi^* \rightarrow n\pi^*$ transition, we see that population of the $n\pi$ state along dynamics will length to increased electronic density in the C5-C6 bond shown in **Figure 4.6**. This increased electronic density should lead to decrease in this bond-length when the active-state becomes $n\pi^*$ in nature. The distribution of C5-C6 bond-angle in the various trajectories, shows a peak at shorter bond-lengths for the longer time-decay trajectory, suggesting

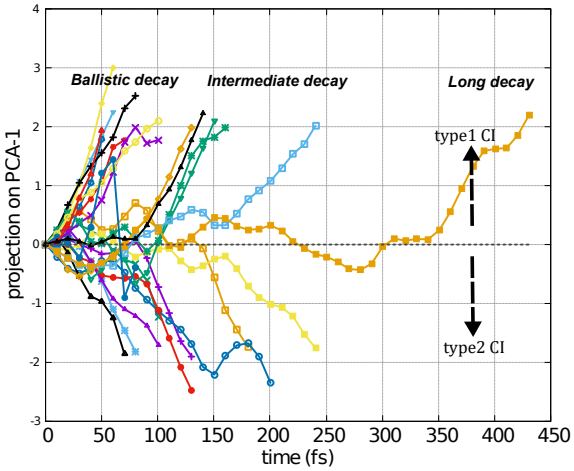

Figure 4.4: Projection on the first principal component along time for all the 24 trajectories which hopped to ground-state. The principal component analysis was performed from time=0 till ~10 fs after the hop to ground-state. The projections were shifted to match zero at t=0.

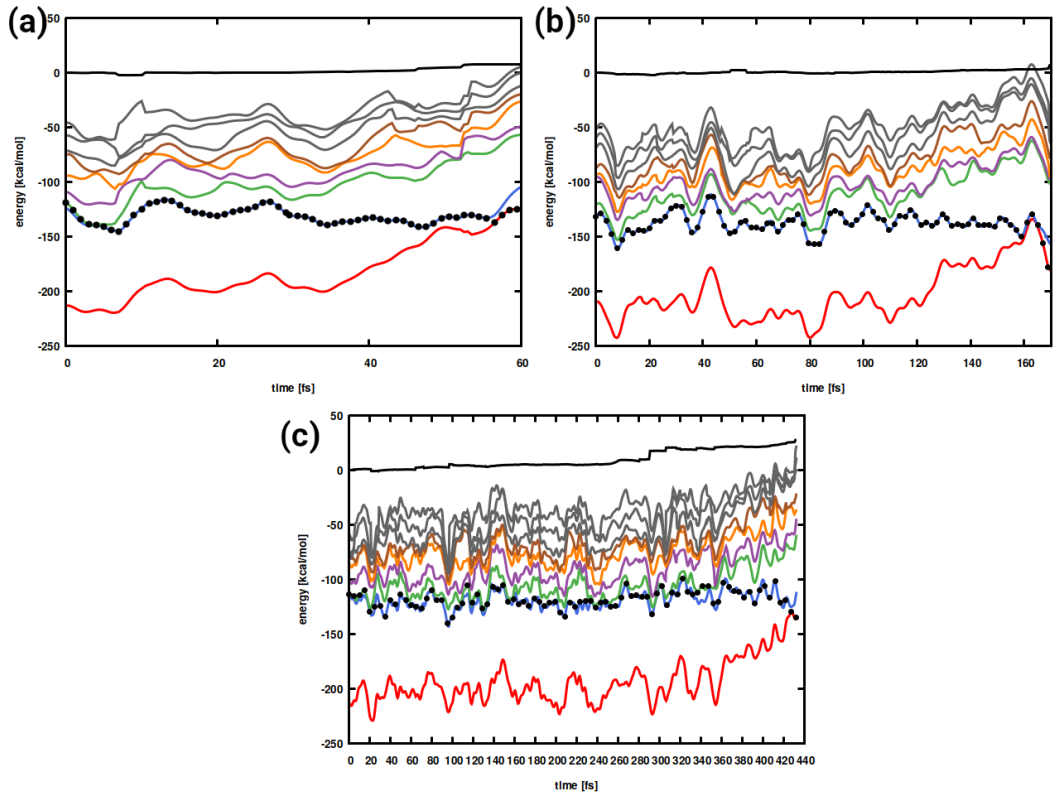

Figure 4.5: (a) Ballistic trajectory. (b) Intermediate trajectory (c) Long trajectory. The plots show the adiabatic potential energies of the 9 states included in the surface-hopping dynamics. States 1-6 are in colour and 7-9 are in grey. The active state is depicted wit black dots. The total energy (potential + kinetic) is also shown in solid black line.

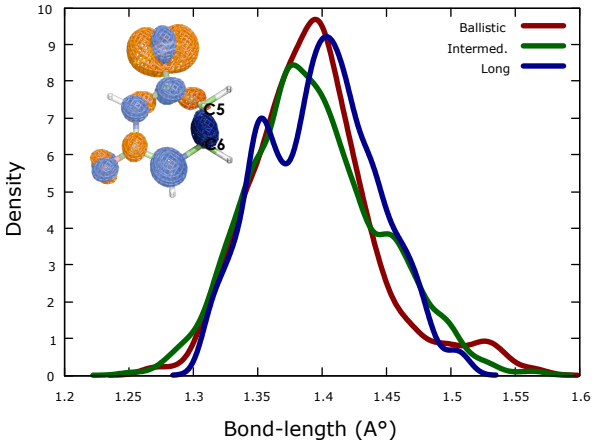

Figure 4.6: The distribution of C5-C6 bond-lengths for the ballistic , intermediate and long decay trajectories. The inset shows the difference density of $\pi\pi^*\rightarrow n\pi^*$ transition, with regions of electron addition in blue and electron depletion in red. The difference density implies a shortening of C5-C6 bond upon tranistion to the $n\pi^*$ state.

an increased $n\pi^*$ involvement in this trajectory. This suggests the presence of a $\pi\pi^*\rightarrow n\pi^*\rightarrow\pi\pi^*\rightarrow S_0$ channel might also be present in the system leading to longer times needed to decay to the ground-state.

4.2.6 Comparison with SS-CASPT2

The agreement of the SS-CASPT2 dynamics with the XMS-CASPT2 is quite satisfactory. The hopping times for the 24 trajectories which hopped to GS are shown in Table 4.2. For the trajectories which decay ballistically to the ground state in sub 100 fs the agreement in times between SS-CASPT2 and XMS-CASPT2 are excellent. This is to be expected as in these trajectories there is negligible involvement of $n\pi^*$ states. For XMS-CASPT2 trajectories which decay in $>$ 100 fs the discrepancy in time with SS-CASPT2 is increases to about 100 fs. This might be due to blocking of certain pathways in SS-CASPT2 trajectories due to removing the n-orbitals from the active-space which lead to them roaming around in $\pi\pi^*$ state for longer times before reaching the CI-seam. Finally as SS-CASPT2 trajectory cannot be trapped in $n\pi^*$ state the trajectory which decays in half a picosecond at XMS-CASPT2 reaches the CI in 200 fs at SS-CASPT2.

4.2.7 Comparison with Ultrafast experiment

Ultrafast experiments with sub-30 fs resolution performed by Cerullo et al. [25] reported two time constants of 97 fs and 1000 fs obtained by global analysis of the pump-probe transient map. The shorter time-constant corresponds to same timescale as taken by the XMS-CASPT2 trajectories to reach the conical intersection with the ground-state in a ballistic decay. The longer time-constant has been attributed to the involvement of the $n\pi^*$ state previously [139] providing

a gateway to the triplet state. Since the reported XMS-CASPT2 dynamics do not incorporate triplet state this channel cannot be simulated here. However, from the analysis in Figure 4.5 it can be seen that $n\pi^*$ state also acts as a trap for the system elongating the time in the singlet excited state and delaying the access to the conical intersection with the ground-state.

Index	**XMS-CASPT2**	**SS-CASPT2**
1	55	50
2	57	70
3	58	70
4	60	70
5	70	70
6	75	80
7	80	100
8	80	85
9	85	100
10	100	90
11	102	175
12	105	100
13	110	100
14	130	120
15	133	380
16	135	180
17	142	210
18	147	250
19	164	280
20	179	X
21	198	215
22	240	X
23	242	250
24	426	200

Table 4.2: Hopping to ground-state times (femtoseconds) at XMS-CASPT2 and SS-CASPT2 level.

Conclusions

In this PhD thesis I documented four major projects modelling UV-induced photophysical phenomena in diverse biomolecules. These works illustrate the need for accurate computation of electronic energy manifolds of systems in realistic environments to complement and interpret modern ultrafast spectroscopies.

While accurate electronic structure methods like CASSCF/CASPT2 can be computationally quite expensive, their use is necessary to have a sound description of all photochemically relevant states. As exemplified in the work benchmarking nucleobases, these methods are indispensable in energy windows typically accessed in transient spectroscopies. The ultrafast processes are also profoundly affected by the environment around them which can modulate the electronic energy manifold and dictate the possible relaxation pathways. Thus, it is paramount to include their effects in the theoretical models to root the simulations in physical reality. The works reported here have accomplished this by using the hybrid QM/MM setup allowing to model the static and dynamical effects of the environment on excited-state processes. The marriage of high-quality electronic structure computations in this hybrid QM/MM setup allows us to model realistic time-dependent behaviour of the system through semi-classical trajectory hopping simulations. These dynamical simulations allow us to track possible relaxation pathways and assess the effects of structural heterogeneity or solvent relaxation on the excited-state process. In tryptophan the dynamical relaxation of polar aqueous solvent changes the potential energy landscape along time , thus modulating the populations of the lowest excited states. The effect of environment is also seen in the ultrafast relaxation of uridine in water where it results in the destabilization of $n\pi^*$ state resulting in its minimal involvement in the internal conversion to the ground-state. Thus, these studies underlie the importance of accurate theoretical modelling including the environment as a mandatory tool to confidently disentangle complex photophysics and enable quantitative comparisons with experimental spectra and their interpretation.

Bibliography

[1] Marco Garavelli, Paolo Celani, Monica Fato, Michael J. Bearpark, Barry R. Smith, Massimo Olivucci, and Michael A. Robb. Relaxation paths from a conical intersection: The mechanism of product formation in the cyclohexadiene/hexatriene photochemical interconversion. *The Journal of Physical Chemistry A*, 101(11):2023–2032, 1997. doi: 10.1021/jp961554k. URL https://doi.org/10.1021/jp961554k.

[2] Dario Polli, Piero Altoè, Oliver Weingart, Katelyn M. Spillane, Cristian Manzoni, Daniele Brida, Gaia Tomasello, Giorgio Orlandi, Philipp Kukura, Richard A. Mathies, Marco Garavelli, and Giulio Cerullo. Conical intersection dynamics of the primary photoisomerization event in vision. *Nature*, 467(7314):440–443, Sep 2010. ISSN 1476-4687. doi: 10.1038/nature09346. URL https://doi.org/10.1038/nature09346.

[3] Roberto Improta, Fabrizio Santoro, and Lluís Blancafort. Quantum mechanical studies on the photophysics and the photochemistry of nucleic acids and nucleobases. *Chemical Reviews*, 116(6):3540–3593, 2016. doi: 10.1021/acs.chemrev.5b00444. URL https://doi.org/10.1021/acs.chemrev.5b00444. PMID: 26928320.

[4] Margherita Maiuri, Marco Garavelli, and Giulio Cerullo. Ultrafast spectroscopy: State of the art and open challenges. *Journal of the American Chemical Society*, 142(1):3–15, 2020. doi: 10.1021/jacs.9b10533. URL https://doi.org/10.1021/jacs.9b10533. PMID: 31800225.

[5] Irene Conti, Giulio Cerullo, Artur Nenov, and Marco Garavelli. Ultrafast spectroscopy of photoactive molecular systems from first principles: Where we stand today and where we are going. *Journal of the American Chemical Society*, 142(38):16117–16139, 2020.

[6] Rocío Borrego-Varillas, Lucia Ganzer, Giulio Cerullo, and Cristian Manzoni. Ultraviolet transient absorption spectrometer with sub-20-fs time resolution. *Applied Sciences*, 8 (6), 2018. ISSN 2076-3417. doi: 10.3390/app8060989. URL https://www.mdpi.com/2076-3417/8/6/989.

[7] Lightdynamics supported by h2020-msca-itn-2017 (etn), 2018-2022. URL https://www.lightdynamics.eu/.

[8] Björn O. Roos. *The Complete Active Space Self-Consistent Field Method and its Applications in Electronic Structure Calculations*, pages 399–445. John Wiley & Sons, Ltd, 1987. ISBN 9780470142943. doi: https://doi.org/10.1002/9780470142943.ch7. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/9780470142943.ch7.

[9] Björn O. Roos, Peter R. Taylor, and Per E.M. Sigbahn. A complete active space scf method (casscf) using a density matrix formulated super-ci approach. *Chemical Physics*, 48(2):157–173, 1980. ISSN 0301-0104. doi: https://doi.org/10.1016/0301-0104(80)80045-0. URL https://www.sciencedirect.com/science/article/pii/0301010480800450.

[10] Per Aake. Malmqvist, Alistair. Rendell, and Bjoern O. Roos. The restricted active space self-consistent-field method, implemented with a split graph unitary group approach. *The Journal of Physical Chemistry*, 94(14):5477–5482, 1990. doi: 10.1021/j100377a011. URL https://doi.org/10.1021/j100377a011.

[11] Kerstin Andersson, Per-Åke Malmqvist, and Björn O. Roos. Second-order perturbation theory with a complete active space self-consistent field reference function. *The Journal of Chemical Physics*, 96(2):1218–1226, 1992. doi: 10.1063/1.462209. URL https://doi.org/10.1063/1.462209.

[12] Kerstin. Andersson, Per Aake. Malmqvist, Bjoern O. Roos, Andrzej J. Sadlej, and Krzysztof. Wolinski. Second-order perturbation theory with a casscf reference function. *The Journal of Physical Chemistry*, 94(14):5483–5488, 1990. doi: 10.1021/j100377a012. URL https://doi.org/10.1021/j100377a012.

[13] Vicenta Sauri, Luis Serrano-Andrés, Abdul Rehaman Moughal Shahi, Laura Gagliardi, Steven Vancoillie, and Kristine Pierloot. Multiconfigurational second-order perturbation theory restricted active space (raspt2) method for electronic excited states: A benchmark study. *Journal of Chemical Theory and Computation*, 7(1):153–168, 2011. doi: 10.1021/ct100478d. URL https://doi.org/10.1021/ct100478d. PMID: 26606229.

[14] Per Åke Malmqvist, Kristine Pierloot, Abdul Rehaman Moughal Shahi, Christopher J. Cramer, and Laura Gagliardi. The restricted active space followed by second-order perturbation theory method: Theory and application to the study of cuo2 and cu2o2 systems. *The Journal of Chemical Physics*, 128(20):204109, 2008. doi: 10.1063/1.2920188. URL https://doi.org/10.1063/1.2920188.

[15] Jean-Paul Malrieu, Jean-Louis Heully, and Andréi Zaitsevskii. Multiconfigurational second-order perturbative methods: Overview and comparison of basic properties. *Theoretica chimica acta*, 90(2):167–187, Jan 1995. ISSN 1432-2234. doi: 10.1007/BF01113846. URL https://doi.org/10.1007/BF01113846.

[16] James Finley, Per Åke Malmqvist, Björn O. Roos, and Luis Serrano-Andrés. The multi-state caspt2 method. *Chemical Physics Letters*, 288(2):299–306, 1998. ISSN

0009-2614. doi: https://doi.org/10.1016/S0009-2614(98)00252-8. URL https://www.sciencedirect.com/science/article/pii/S0009261498002528.

[17] Alexander A. Granovsky. Extended multi-configuration quasi-degenerate perturbation theory: The new approach to multi-state multi-reference perturbation theory. *The Journal of Chemical Physics*, 134(21):214113, 2011. doi: 10.1063/1.3596699. URL https://doi.org/10.1063/1.3596699.

[18] Stefano Battaglia and Roland Lindh. Extended dynamically weighted caspt2: The best of two worlds. *Journal of Chemical Theory and Computation*, 16(3):1555–1567, 2020. doi: 10.1021/acs.jctc.9b01129. URL https://doi.org/10.1021/acs.jctc.9b01129. PMID: 32027802.

[19] Toru Shiozaki, Werner Győrffy, Paolo Celani, and Hans-Joachim Werner. Communication: Extended multi-state complete active space second-order perturbation theory: Energy and nuclear gradients. *The Journal of Chemical Physics*, 135(8):081106, 2011. doi: 10.1063/1.3633329. URL https://doi.org/10.1063/1.3633329.

[20] Saumik Sen and Igor Schapiro. A comprehensive benchmark of the xms-caspt2 method for the photochemistry of a retinal chromophore model. *Molecular Physics*, 116(19-20): 2571–2582, 2018. doi: 10.1080/00268976.2018.1501112. URL https://doi.org/10.1080/00268976.2018.1501112.

[21] P. Hohenberg and W. Kohn. Inhomogeneous electron gas. *Phys. Rev.*, 136:B864–B871, Nov 1964. doi: 10.1103/PhysRev.136.B864. URL https://link.aps.org/doi/10.1103/PhysRev.136.B864.

[22] W. Kohn and L. J. Sham. Self-consistent equations including exchange and correlation effects. *Phys. Rev.*, 140:A1133–A1138, Nov 1965. doi: 10.1103/PhysRev.140.A1133. URL https://link.aps.org/doi/10.1103/PhysRev.140.A1133.

[23] John P. Perdew and Karla Schmidt. Jacob's ladder of density functional approximations for the exchange-correlation energy. *AIP Conference Proceedings*, 577(1):1–20, 2001. doi: 10.1063/1.1390175. URL https://aip.scitation.org/doi/abs/10.1063/1.1390175.

[24] Erich Runge and E. K. U. Gross. Density-functional theory for time-dependent systems. *Phys. Rev. Lett.*, 52:997–1000, Mar 1984. doi: 10.1103/PhysRevLett.52.997. URL https://link.aps.org/doi/10.1103/PhysRevLett.52.997.

[25] Rocío Borrego-Varillas, Artur Nenov, Piotr Kabaciński, Irene Conti, Lucia Ganzer, Aurelio Oriana, Vishal Kumar Jaiswal, Ines Delfino, Oliver Weingart, Cristian Manzoni, Ivan Rivalta, Marco Garavelli, and Giulio Cerullo. Tracking excited state decay mechanisms of pyrimidine nucleosides in real time. *Nature Communications*, 12(1):7285, Dec 2021. ISSN 2041-1723. doi: 10.1038/s41467-021-27535-7. URL https://doi.org/10.1038/s41467-021-27535-7.

[26] Piotr Kabaciński, Marco Romanelli, Eveliina Ponkkonen, Vishal Kumar Jaiswal, Thomas Carell, Marco Garavelli, Giulio Cerullo, and Irene Conti. Unified description of ultrafast excited state decay processes in epigenetic deoxycytidine derivatives. *The Journal of Physical Chemistry Letters*, 12(45):11070–11077, 2021. doi: 10.1021/acs.jpclett.1c02909. URL https://doi.org/10.1021/acs.jpclett.1c02909. PMID: 34748341.

[27] Matthew M. Brister and Carlos E. Crespo-Hernández. Excited-state dynamics in the rna nucleotide uridine 5′-monophosphate investigated using femtosecond broadband transient absorption spectroscopy. *The Journal of Physical Chemistry Letters*, 10(9):2156–2161, 2019. doi: 10.1021/acs.jpclett.9b00492. URL https://doi.org/10.1021/acs.jpclett.9b00492.

[28] Axel D. Becke. Density-functional thermochemistry. iii. the role of exact exchange. *The Journal of Chemical Physics*, 98(7):5648–5652, 1993. doi: 10.1063/1.464913. URL https://doi.org/10.1063/1.464913.

[29] Chengteh Lee, Weitao Yang, and Robert G. Parr. Development of the colle-salvetti correlation-energy formula into a functional of the electron density. *Phys. Rev. B*, 37: 785–789, Jan 1988. doi: 10.1103/PhysRevB.37.785. URL https://link.aps.org/doi/10.1103/PhysRevB.37.785.

[30] Takeshi Yanai, David P Tew, and Nicholas C Handy. A new hybrid exchange–correlation functional using the coulomb-attenuating method (cam-b3lyp). *Chemical Physics Letters*, 393(1):51–57, 2004. ISSN 0009-2614. doi: https://doi.org/10.1016/j.cplett.2004.06.011. URL https://www.sciencedirect.com/science/article/pii/S0009261404008620.

[31] W. J. Hehre, R. Ditchfield, and J. A. Pople. Self—consistent molecular orbital methods. xii. further extensions of gaussian—type basis sets for use in molecular orbital studies of organic molecules. *The Journal of Chemical Physics*, 56(5):2257–2261, 1972. doi: 10.1063/1.1677527. URL https://doi.org/10.1063/1.1677527.

[32] P. C. Hariharan and J. A. Pople. The influence of polarization functions on molecular orbital hydrogenation energies. *Theoretica chimica acta*, 28(3):213–222, Sep 1973. ISSN 1432-2234. doi: 10.1007/BF00533485. URL https://doi.org/10.1007/BF00533485.

[33] Timothy Clark, Jayaraman Chandrasekhar, Günther W. Spitznagel, and Paul Von Ragué Schleyer. Efficient diffuse function-augmented basis sets for anion calculations. iii. the 3-21+g basis set for first-row elements, li–f. *Journal of Computational Chemistry*, 4(3):294–301, 1983. doi: https://doi.org/10.1002/jcc.540040303. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/jcc.540040303.

[34] James D. Dill and John A. Pople. Self-consistent molecular orbital methods. xv. extended gaussian-type basis sets for lithium, beryllium, and boron. *The Journal of Chemical Physics*, 62(7):2921–2923, 1975. doi: 10.1063/1.430801. URL https://doi.org/10.1063/1.430801.

[35] Michelle M. Francl, William J. Pietro, Warren J. Hehre, J. Stephen Binkley, Mark S. Gordon, Douglas J. DeFrees, and John A. Pople. Self-consistent molecular orbital methods. xxiii. a polarization-type basis set for second-row elements. *The Journal of Chemical Physics*, 77(7):3654–3665, 1982. doi: 10.1063/1.444267. URL `https://doi.org/10.1063/1.444267`.

[36] Angelo Giussani, Javier Segarra-Martí, Artur Nenov, Ivan Rivalta, Alessandra Tolomelli, Shaul Mukamel, and Marco Garavelli. Spectroscopic fingerprints of dna/rna pyrimidine nucleobases in third-order nonlinear electronic spectra. *Theoretical Chemistry Accounts*, 135(5):121, Apr 2016. ISSN 1432-2234. doi: 10.1007/s00214-016-1867-z. URL `https://doi.org/10.1007/s00214-016-1867-z`.

[37] Javier Segarra-Martí, Ana J. Pepino, Artur Nenov, Shaul Mukamel, Marco Garavelli, and Ivan Rivalta. The highly excited-state manifold of guanine: calibration for nonlinear electronic spectroscopy simulations. *Theoretical Chemistry Accounts*, 137(3):47, Mar 2018. ISSN 1432-2234. doi: 10.1007/s00214-018-2225-0. URL `https://doi.org/10.1007/s00214-018-2225-0`.

[38] Artur Nenov, Angelo Giussani, Javier Segarra-Martí, Vishal K. Jaiswal, Ivan Rivalta, Giulio Cerullo, Shaul Mukamel, and Marco Garavelli. Modeling the high-energy electronic state manifold of adenine: Calibration for nonlinear electronic spectroscopy. *The Journal of Chemical Physics*, 142(21):212443, 2015. doi: 10.1063/1.4921016. URL `https://doi.org/10.1063/1.4921016`.

[39] Giovanni Ghigo, Björn O. Roos, and Per Åke Malmqvist. A modified definition of the zeroth-order hamiltonian in multiconfigurational perturbation theory (caspt2). *Chemical Physics Letters*, 396(1):142–149, 2004. ISSN 0009-2614. doi: https://doi.org/10.1016/j.cplett.2004.08.032. URL `https://www.sciencedirect.com/science/article/pii/S0009261404012242`.

[40] J. Patrick Zobel, Juan J. Nogueira, and Leticia González. The ipea dilemma in caspt2. *Chem. Sci.*, 8:1482–1499, 2017. doi: 10.1039/C6SC03759C. URL `http://dx.doi.org/10.1039/C6SC03759C`.

[41] Leigh B. Clark, Gary G. Peschel, and Ignacio Tinoco. Vapor spectra and heats of vaporization of some purine and pyrimidine bases1. *The Journal of Physical Chemistry*, 69(10):3615–3618, Oct 1965. ISSN 0022-3654. doi: 10.1021/j100894a063. URL `https://doi.org/10.1021/j100894a063`.

[42] Liang. Li and David M. Lubman. Ultraviolet-visible absorption spectra of biological molecules in the gas phase using pulsed laser-induced volatilization enhancement in a diode array spectrophotometer. *Analytical Chemistry*, 59(20):2538–2541, Oct 1987. ISSN 0003-2700. doi: 10.1021/ac00147a022. URL `https://doi.org/10.1021/ac00147a022`.

[43] Krzysztof Polewski, David Zinger, John Trunk, and John C. Sutherland. Ultraviolet absorption and luminescence of matrix-isolated adenine. *Radiation Physics and Chemistry*, 80(10):1092–1098, 2011. ISSN 0969-806X. doi: https://doi.org/10.1016/j.radphyschem.2011.02.007. URL https://www.sciencedirect.com/science/article/pii/S0969806X11000478. Synchrotron radiation studies in Poland-The 10th International School and Symposium on Synchrotron Radiation in Natural Science (ISSRNS).

[44] Takenori Yamada and Hideo Fukutome. Vacuum ultraviolet absorption spectra of sublimed films of nucleic acid bases. *Biopolymers*, 6(1):43–54, 1968. doi: https://doi.org/10.1002/bip.1968.360060104. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/bip.1968.360060104.

[45] Robert L. Sinsheimer, J. F. M. Scott, John R. Loofbourow, B Best, Roger Hastings, E Vollmer, and M. Westergaard. Ultraviolet absorption spectra at reduced temperatures. ii. pyrimidines and purines. *The Journal of biological chemistry*, 187 1:313–24, 1950.

[46] Krzysztof Polewski, David Zinger, John Trunk, Denise C. Monteleone, and John C. Sutherland. Fluorescence of matrix isolated guanine and 7-methylguanine. *Journal of Photochemistry and Photobiology B: Biology*, 24(3):169–177, 1994. ISSN 1011-1344. doi: https://doi.org/10.1016/1011-1344(94)07018-0. URL https://www.sciencedirect.com/science/article/pii/1011134494070180.

[47] Maciej J. Nowak, Krystyna Szczepaniak, Andrzej Barski, and David Shugar. Spectroscopic studies on vapour phase tautomerism of natural bases found in nucleic acids. *Zeitschrift für Naturforschung C*, 33(11-12):876–883, 1978. doi: doi:10.1515/znc-1978-11-1213. URL https://doi.org/10.1515/znc-1978-11-1213.

[48] Yuko Tsuchiya, Teruhiko Tamura, Masaaki Fujii, and Mitsuo Ito. Keto-enol tautomer of uracil and thymine. *The Journal of Physical Chemistry*, 92(7):1760–1765, 1988. doi: 10.1021/j100318a013. URL https://doi.org/10.1021/j100318a013.

[49] R Abouaf, J Pommier, and H Dunet. Electronic and vibrational excitation in gas phase thymine and 5-bromouracil by electron impact. *Chemical Physics Letters*, 381(3):486–494, 2003. ISSN 0009-2614. doi: https://doi.org/10.1016/j.cplett.2003.09.121. URL https://www.sciencedirect.com/science/article/pii/S0009261403016944.

[50] Gábor Bazsó, György Tarczay, Géza Fogarasi, and Péter G. Szalay. Tautomers of cytosine and their excited electronic states: a matrix isolation spectroscopic and quantum chemical study. *Phys. Chem. Chem. Phys.*, 13:6799–6807, 2011. doi: 10.1039/C0CP02354J. URL http://dx.doi.org/10.1039/C0CP02354J.

[51] Robert Abouaf, Jacqueline Pommier, Henri Dunet, Phung Quan, Pham-Cam Nam, and Minh Tho Nguyen. The triplet state of cytosine and its derivatives: Electron impact and quantum chemical study. *The Journal of Chemical Physics*, 121(23):11668–11674, 2004. doi: 10.1063/1.1812533. URL https://doi.org/10.1063/1.1812533.

[52] Vishal K. Jaiswal, Javier Segarra-Martí, Marco Marazzi, Elena Zvereva, Xavier Assfeld, Antonio Monari, Marco Garavelli, and Ivan Rivalta. First-principles characterization of the singlet excited state manifold in dna/rna nucleobases. *Phys. Chem. Chem. Phys.*, 22:15496–15508, 2020. doi: 10.1039/D0CP01823F. URL `http://dx.doi.org/10.1039/D0CP01823F`.

[53] Martina De Vetta, Maximilian F. S. J. Menger, Juan J. Nogueira, and Leticia González. Solvent effects on electronically excited states: Qm/continuum versus qm/explicit models. *The Journal of Physical Chemistry B*, 122(11):2975–2984, 2018. doi: 10.1021/acs.jpcb.7b12560. URL `https://doi.org/10.1021/acs.jpcb.7b12560`. PMID: 29481750.

[54] Tim J. Zuehlsdorff and Christine M. Isborn. Modeling absorption spectra of molecules in solution. *International Journal of Quantum Chemistry*, 119(1):e25719, 2019. doi: https://doi.org/10.1002/qua.25719. URL `https://onlinelibrary.wiley.com/doi/abs/10.1002/qua.25719`.

[55] Javier Segarra-Martí, Francesco Segatta, Tristan A. Mackenzie, Artur Nenov, Ivan Rivalta, Michael J. Bearpark, and Marco Garavelli. Modeling multidimensional spectral lineshapes from first principles: application to water-solvated adenine. *Faraday Discuss.*, 221:219–244, 2020. doi: 10.1039/C9FD00072K. URL `http://dx.doi.org/10.1039/C9FD00072K`.

[56] Mario Barbatti and Kakali Sen. Effects of different initial condition samplings on photodynamics and spectrum of pyrrole. *International Journal of Quantum Chemistry*, 116(10):762–771, 2016. doi: https://doi.org/10.1002/qua.25049. URL `https://onlinelibrary.wiley.com/doi/abs/10.1002/qua.25049`.

[57] J. Patrick Zobel, Juan J. Nogueira, and Leticia González. Finite-temperature wigner phase-space sampling and temperature effects on the excited-state dynamics of 2-nitronaphthalene. *Phys. Chem. Chem. Phys.*, 21:13906–13915, 2019. doi: 10.1039/C8CP03273D. URL `http://dx.doi.org/10.1039/C8CP03273D`.

[58] A. Warshel and M. Levitt. Theoretical studies of enzymic reactions: Dielectric, electrostatic and steric stabilization of the carbonium ion in the reaction of lysozyme. *Journal of Molecular Biology*, 103(2):227–249, 1976. ISSN 0022-2836. doi: https://doi.org/10.1016/0022-2836(76)90311-9. URL `https://www.sciencedirect.com/science/article/pii/0022283676903119`.

[59] J. ANDREW MCCAMMON, BRUCE R. GELIN, MARTIN KARPLUS, and PETER G. WOLYNES. The hinge-bending mode in lysozyme. *Nature*, 262(5566):325–326, Jul 1976. ISSN 1476-4687. doi: 10.1038/262325a0. URL `https://doi.org/10.1038/262325a0`.

[60] Oliver Weingart, Artur Nenov, Piero Altoè, Ivan Rivalta, Javier Segarra-Martí, Irina Dokukina, and Marco Garavelli. COBRAMM 2.0 — A software interface for tailoring molecular electronic structure calculations and running nanoscale (QM/MM) simulations. *Journal of Molecular Modeling*, 24(9):271, 2018. ISSN 0948-5023. doi: 10.1007/s00894-018-3769-6. URL `https://doi.org/10.1007/s00894-018-3769-6`.

[61] Roberto Improta and Vincenzo Barone. *Excited States Behavior of Nucleobases in Solution: Insights from Computational Studies*, pages 329–357. Springer International Publishing, Cham, 2015. ISBN 978-3-319-13371-3. doi: 10.1007/128_2013_524. URL https://doi.org/10.1007/128_2013_524.

[62] Maurizio Cossi, Vincenzo Barone, Roberto Cammi, and Jacopo Tomasi. Ab initio study of solvated molecules: a new implementation of the polarizable continuum model. *Chemical Physics Letters*, 255(4):327–335, 1996. ISSN 0009-2614. doi: https://doi.org/10.1016/0009-2614(96)00349-1. URL https://www.sciencedirect.com/science/article/pii/0009261496003491.

[63] Maurizio Cossi, Giovanni Scalmani, Nadia Rega, and Vincenzo Barone. New developments in the polarizable continuum model for quantum mechanical and classical calculations on molecules in solution. *The Journal of Chemical Physics*, 117(1):43–54, 2002. doi: 10.1063/1.1480445. URL https://doi.org/10.1063/1.1480445.

[64] A. Klamt and G. Schüürmann. Cosmo: a new approach to dielectric screening in solvents with explicit expressions for the screening energy and its gradient. *J. Chem. Soc., Perkin Trans. 2*, pages 799–805, 1993. doi: 10.1039/P29930000799. URL http://dx.doi.org/10.1039/P29930000799.

[65] Maurizio Cossi, Nadia Rega, Giovanni Scalmani, and Vincenzo Barone. Energies, structures, and electronic properties of molecules in solution with the c-pcm solvation model. *Journal of Computational Chemistry*, 24(6):669–681, 2003. doi: https://doi.org/10.1002/jcc.10189. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/jcc.10189.

[66] Boiko Cohen, Patrick M. Hare, and Bern Kohler. Ultrafast excited-state dynamics of adenine and monomethylated adenines in solution: implications for the nonradiative decay mechanism. *Journal of the American Chemical Society*, 125(44):13594–13601, October 2003. doi: 10.1021/ja035628z. URL https://doi.org/10.1021/ja035628z.

[67] Hanneli R. Hudock and Todd J. Martínez. Excited-state dynamics of cytosine reveal multiple intrinsic subpicosecond pathways. *ChemPhysChem*, 9(17):2486–2490, December 2008. doi: 10.1002/cphc.200800649. URL https://doi.org/10.1002/cphc.200800649.

[68] Matthew M. Brister and Carlos E. Crespo-Hernández. Excited-state dynamics in the RNA nucleotide uridine 5'-monophosphate investigated using femtosecond broadband transient absorption spectroscopy. *The Journal of Physical Chemistry Letters*, 10(9):2156–2161, April 2019. doi: 10.1021/acs.jpclett.9b00492. URL https://doi.org/10.1021/acs.jpclett.9b00492.

[69] Jean-Marc L. Pecourt, Jorge Peon, and Bern Kohler. DNA excited-state dynamics: ultrafast internal conversion and vibrational cooling in a series of nucleosides. *Journal of the American Chemical Society*, 123(42):10370–10378, October 2001. doi: 10.1021/ja0161453. URL https://doi.org/10.1021/ja0161453.

[70] Piotr Kabaciński, Marco Romanelli, Eveliina Ponkkonen, Vishal Kumar Jaiswal, Thomas Carell, Marco Garavelli, Giulio Cerullo, and Irene Conti. Unified description of ultrafast excited state decay processes in epigenetic deoxycytidine derivatives. *The Journal of Physical Chemistry Letters*, 12(45):11070–11077, November 2021. doi: 10.1021/acs.jpclett.1c02909. URL `https://doi.org/10.1021/acs.jpclett.1c02909`.

[71] Irene Conti, Piero Altoè, Marco Stenta, Marco Garavelli, and Giorgio Orlandi. Adenine deactivation in DNA resolved at the CASPT2//CASSCF/AMBER level. *Physical Chemistry Chemical Physics*, 12(19):5016, 2010. doi: 10.1039/b926608a. URL `https://doi.org/10.1039/b926608a`.

[72] Thomas Gustavsson, Nilmoni Sarkar, Ignacio Vayá, M. Consuelo Jiménez, Dimitra Markovitsi, and Roberto Improta. A joint experimental/theoretical study of the ultrafast excited state deactivation of deoxyadenosine and 9-methyladenine in water and acetonitrile. *Photochem. Photobiol. Sci.*, 12:1375–1386, 2013. doi: 10.1039/C3PP50060H. URL `http://dx.doi.org/10.1039/C3PP50060H`.

[73] Carlos E. Crespo-Hernández, Lara Martínez-Fernández, Clemens Rauer, Christian Reichardt, Sebastian Mai, Marvin Pollum, Philipp Marquetand, Leticia González, and Inés Corral. Electronic and structural elements that regulate the excited-state dynamics in purine nucleobase derivatives. *Journal of the American Chemical Society*, 137(13):4368–4381, 2015. doi: 10.1021/ja512536c. URL `https://doi.org/10.1021/ja512536c`. PMID: 25763596.

[74] Karl Kleinermanns, Dana Nachtigallová, and Mattanjah S. de Vries. Excited state dynamics of DNA bases. *International Reviews in Physical Chemistry*, 32(2):308–342, June 2013. doi: 10.1080/0144235x.2012.760884. URL `https://doi.org/10.1080/0144235x.2012.760884`.

[75] M. Barbatti, A. J. A. Aquino, J. J. Szymczak, D. Nachtigallova, P. Hobza, and H. Lischka. Relaxation mechanisms of UV-photoexcited DNA and RNA nucleobases. *Proceedings of the National Academy of Sciences*, 107(50):21453–21458, November 2010. doi: 10.1073/pnas.1014982107. URL `https://doi.org/10.1073/pnas.1014982107`.

[76] Martin Richter, Philipp Marquetand, Jesús González-Vázquez, Ignacio Sola, and Leticia González. Femtosecond intersystem crossing in the DNA nucleobase cytosine. *The Journal of Physical Chemistry Letters*, 3(21):3090–3095, October 2012. doi: 10.1021/jz301312h. URL `https://doi.org/10.1021/jz301312h`.

[77] David Picconi, Vincenzo Barone, Alessandro Lami, Fabrizio Santoro, and Roberto Improta. The interplay between pipi*/npi* excited states in gas-phase thymine: A quantum dynamical study. *ChemPhysChem*, 12(10):1957–1968, May 2011. doi: 10.1002/cphc.201001080. URL `https://doi.org/10.1002/cphc.201001080`.

[78] Martha Yaghoubi Jouybari, Yanli Liu, Roberto Improta, and Fabrizio Santoro. Ultrafast dynamics of the two lowest bright excited states of cytosine and 1-methylcytosine: A quantum dynamical study. *Journal of Chemical Theory and Computation*, 16(9):5792–5808, July 2020. doi: 10.1021/acs.jctc.0c00455. URL https://doi.org/10.1021/acs.jctc.0c00455.

[79] Junmei Wang, Piotr Cieplak, and Peter A. Kollman. How well does a restrained electrostatic potential (RESP) model perform in calculating conformational energies of organic and biological molecules? *Journal of Computational Chemistry*, 21(12):1049–1074, 2000. doi: 10.1002/1096-987x(200009)21:12<1049::aid-jcc3>3.0.co;2-f. URL https://doi.org/10.1002/1096-987x(200009)21:12<1049::aid-jcc3>3.0.co;2-f.

[80] Alberto Pérez, Iván Marchán, Daniel Svozil, Jiri Sponer, Thomas E. Cheatham, Charles A. Laughton, and Modesto Orozco. Refinement of the AMBER force field for nucleic acids: Improving the description of α/γ conformers. *Biophysical Journal*, 92(11):3817–3829, June 2007. doi: 10.1529/biophysj.106.097782. URL https://doi.org/10.1529/biophysj.106.097782.

[81] Marie Zgarbová, Michal Otyepka, Jiří Šponer, Arnošt Mládek, Pavel Banáš, Thomas E. Cheatham, and Petr Jurečka. Refinement of the cornell et al. nucleic acids force field based on reference quantum chemical calculations of glycosidic torsion profiles. *Journal of Chemical Theory and Computation*, 7(9):2886–2902, August 2011. doi: 10.1021/ct200162x. URL https://doi.org/10.1021/ct200162x.

[82] D. Voet, W. B. Gratzer, R. A. Cox, and Paul Doty. Absorption spectra of nucleotides, polynucleotides, and nucleic acids in the far ultraviolet. *Biopolymers*, 1(3):193–208, 1963. doi: https://doi.org/10.1002/bip.360010302. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/bip.360010302.

[83] J. N. Murrell. The potential energy surfaces of polyatomic molecules. In *Novel Chemical Effects of Electronics Behaviour*, pages 93–146. Springer Berlin Heidelberg, 1977. doi: 10.1007/3-540-08014-7_3. URL https://doi.org/10.1007/3-540-08014-7_3.

[84] Nicholas C. Handy and Aaron M. Lee. The adiabatic approximation. *Chemical Physics Letters*, 252(5):425–430, 1996. ISSN 0009-2614. doi: https://doi.org/10.1016/0009-2614(96)00171-6. URL https://www.sciencedirect.com/science/article/pii/0009261496001716.

[85] M. Garavelli, C. S. Page, P. Celani, M. Olivucci, W. E. Schmid, S. A. Trushin, and W. Fuss. Reaction path of a sub-200 fs photochemical electrocyclic reaction. *The Journal of Physical Chemistry A*, 105(18):4458–4469, 2001. doi: 10.1021/jp010359p. URL https://doi.org/10.1021/jp010359p.

[86] Toshifumi Mori and Todd. J. Martínez. Exploring the conical intersection seam: The seam space nudged elastic band method. *Journal of Chemical Theory and Computation*, 9(2):

1155–1163, 2013. doi: 10.1021/ct300892t. URL https://doi.org/10.1021/ct300892t. PMID: 26588758.

[87] Jianhua Xu, Binbin Chen, Patrik Callis, Pedro L Muiño, Henriëtte Rozeboom, Jaap Broos, Dmitri Toptygin, Ludwig Brand, and Jay R Knutson. Picosecond Fluorescence Dynamics of Tryptophan and 5-Fluorotryptophan in Monellin: Slow Water–Protein Relaxation Unmasked. *The Journal of Physical Chemistry B*, 119(11):4230–4239, mar 2015. ISSN 1520-6106. doi: 10.1021/acs.jpcb.5b01651. URL https://doi.org/10.1021/acs.jpcb.5b01651.

[88] Yangzhong Qin, Chih-Wei Chang, Lijuan Wang, and Dongping Zhong. Validation of Response Function Construction and Probing Heterogeneous Protein Hydration by Intrinsic Tryptophan. *The Journal of Physical Chemistry B*, 116(45):13320–13330, nov 2012. ISSN 1520-6106. doi: 10.1021/jp305118n. URL https://doi.org/10.1021/jp305118n.

[89] R Bryn Fenwick, David Oyen, H Jane Dyson, and Peter E Wright. Slow Dynamics of Tryptophan–Water Networks in Proteins. *Journal of the American Chemical Society*, 140 (2):675–682, jan 2018. ISSN 0002-7863. doi: 10.1021/jacs.7b09974. URL https://doi.org/10.1021/jacs.7b09974.

[90] Patrick Houston, Nicolas Macro, Minhee Kang, Long Chen, Jin Yang, Lijuan Wang, Zhengrong Wu, and Dongping Zhong. Ultrafast Dynamics of Water–Protein Coupled Motions around the Surface of Eye Crystallin. *Journal of the American Chemical Society*, 142(8):3997–4007, feb 2020. ISSN 0002-7863. doi: 10.1021/jacs.9b13506. URL https://doi.org/10.1021/jacs.9b13506.

[91] Tanping Li, Ali A Hassanali, Ya-Ting Kao, Dongping Zhong, and Sherwin J Singer. Hydration Dynamics and Time Scales of Coupled Water Protein Fluctuations. *Journal of the American Chemical Society*, 129(11):3376–3382, mar 2007. ISSN 0002-7863. doi: 10.1021/ja0685957. URL https://doi.org/10.1021/ja0685957.

[92] Jin Yang, Yafang Wang, Lijuan Wang, and Dongping Zhong. Mapping Hydration Dynamics around a β-Barrel Protein. *Journal of the American Chemical Society*, 139(12):4399–4408, mar 2017. ISSN 0002-7863. doi: 10.1021/jacs.6b12463. URL https://doi.org/10.1021/jacs.6b12463.

[93] Xiaohua Shen and Jay R Knutson. Subpicosecond Fluorescence Spectra of Tryptophan in Water. *The Journal of Physical Chemistry B*, 105(26):6260–6265, jul 2001. ISSN 1520-6106. doi: 10.1021/jp010384v. URL https://doi.org/10.1021/jp010384v.

[94] J Léonard, D Sharma, B Szafarowicz, K Torgasin, and S Haacke. Formation dynamics and nature of tryptophan's primary photoproduct in aqueous solution. *Physical Chemistry Chemical Physics*, 12(48):15744–15750, 2010. ISSN 1463-9076. doi: 10.1039/C0CP00615G. URL http://dx.doi.org/10.1039/C0CP00615G.

[95] A L Sobolewski, W Domcke, C Dedonder-Lardeux, and C Jouvet. Excited-state hydrogen detachment and hydrogen transfer driven by repulsive $1\pi\sigma^*$ states: A new paradigm for nonradiative decay in aromatic biomolecules. *Physical Chemistry Chemical Physics*, 4(7): 1093–1100, 2002. ISSN 1463-9076. doi: 10.1039/B110941N. URL `http://dx.doi.org/10.1039/B110941N`.

[96] G Grégoire, C Jouvet, C Dedonder, and A L Sobolewski. On the role of dissociative $\pi\sigma^*$ states in the photochemistry of protonated tryptamine and tryptophan: An ab initio study. *Chemical Physics*, 324(2):398–404, 2006. ISSN 0301-0104. doi: https://doi.org/10.1016/j.chemphys.2005.11.005. URL `https://www.sciencedirect.com/science/article/pii/S0301010405005665`.

[97] H Kang, C Jouvet, C Dedonder-Lardeux, S Martrenchard, G Grégoire, C Desfrançois, J.-P. Schermann, M Barat, and J A Fayeton. Ultrafast deactivation mechanisms of protonated aromatic amino acids following UV excitation. *Physical Chemistry Chemical Physics*, 7(2): 394–398, 2005. ISSN 1463-9076. doi: 10.1039/B414986F. URL `http://dx.doi.org/10.1039/B414986F`.

[98] Virginia Ovejas, Marta Fernández-Fernández, Raúl Montero, Fernando Castaño, and Asier Longarte. Ultrafast Nonradiative Relaxation Channels of Tryptophan. *The Journal of Physical Chemistry Letters*, 4(11):1928–1932, jun 2013. doi: 10.1021/jz400810j. URL `https://doi.org/10.1021/jz400810j`.

[99] Patrik R Callis and Tiqing Liu. Quantitative Prediction of Fluorescence Quantum Yields for Tryptophan in Proteins. *The Journal of Physical Chemistry B*, 108(14):4248–4259, apr 2004. ISSN 1520-6106. doi: 10.1021/jp0310551. URL `https://doi.org/10.1021/jp0310551`.

[100] Pedro L Muiño and Patrik R Callis. Solvent Effects on the Fluorescence Quenching of Tryptophan by Amides via Electron Transfer. Experimental and Computational Studies. *The Journal of Physical Chemistry B*, 113(9):2572–2577, mar 2009. ISSN 1520-6106. doi: 10.1021/jp711513b. URL `https://doi.org/10.1021/jp711513b`.

[101] Anthony J Ruggiero, David C Todd, and Graham R Fleming. Subpicosecond fluorescence anisotropy studies of tryptophan in water. *Journal of the American Chemical Society*, 112(3):1003–1014, jan 1990. ISSN 0002-7863. doi: 10.1021/ja00159a017. URL `https://doi.org/10.1021/ja00159a017`.

[102] Ahmad Ajdarzadeh, Cristina Consani, Olivier Bräm, Andreas Tortschanoff, Andrea Cannizzo, and Majed Chergui. Ultraviolet transient absorption, transient grating and photon echo studies of aqueous tryptophan. *Chemical Physics*, 422:47–52, 2013. ISSN 0301-0104. doi: https://doi.org/10.1016/j.chemphys.2013.01.036. URL `https://www.sciencedirect.com/science/article/pii/S0301010413000657`.

[103] Jin Yang, Luyuan Zhang, Lijuan Wang, and Dongping Zhong. Femtosecond Conical Intersection Dynamics of Tryptophan in Proteins and Validation of Slowdown of Hydration Layer Dynamics. *Journal of the American Chemical Society*, 134(40):16460–16463, oct 2012. ISSN 0002-7863. doi: 10.1021/ja305283j. URL `https://doi.org/10.1021/ja305283j`.

[104] Matthias Wohlgemuth, Vlasta Bonačić-Koutecký, and Roland Mitrić. Time-dependent density functional theory excited state nonadiabatic dynamics combined with quantum mechanical/molecular mechanical approach: Photodynamics of indole in water. *The Journal of Chemical Physics*, 135(5):54105, aug 2011. ISSN 0021-9606. doi: 10.1063/1.3622563. URL `https://doi.org/10.1063/1.3622563`.

[105] Angelo Giussani, Manuela Merchán, Daniel Roca-Sanjuán, and Roland Lindh. Essential on the Photophysics and Photochemistry of the Indole Chromophore by Using a Totally Unconstrained Theoretical Approach. *Journal of Chemical Theory and Computation*, 7 (12):4088–4096, dec 2011. ISSN 1549-9618. doi: 10.1021/ct200646r. URL `https://doi.org/10.1021/ct200646r`.

[106] Divya Sharma, Jérémie Léonard, and Stefan Haacke. Ultrafast excited-state dynamics of tryptophan in water observed by transient absorption spectroscopy. *Chemical Physics Letters*, 489(1):99–102, 2010. ISSN 0009-2614. doi: https://doi.org/10.1016/j.cplett.2010.02.057. URL `https://www.sciencedirect.com/science/article/pii/S0009261410002502`.

[107] Jochen Küpper, David W Pratt, W Leo Meerts, Christian Brand, Jörg Tatchen, and Michael Schmitt. Vibronic coupling in indole: II. Investigation of the 1La–1Lb interaction using rotationally resolved electronic spectroscopy. *Physical Chemistry Chemical Physics*, 12(19):4980–4988, 2010. ISSN 1463-9076. doi: 10.1039/C001778G. URL `http://dx.doi.org/10.1039/C001778G`.

[108] Christian Brand, Jochen Küpper, David W Pratt, W Leo Meerts, Daniel Krügler, Jörg Tatchen, and Michael Schmitt. Vibronic coupling in indole: I. Theoretical description of the 1La–1Lb interaction and the electronic spectrum. *Physical Chemistry Chemical Physics*, 12(19):4968–4979, 2010. ISSN 1463-9076. doi: 10.1039/C001776K. URL `http://dx.doi.org/10.1039/C001776K`.

[109] Gaurav Kumar, Anirban Roy, Ryan S McMullen, Shanmukh Kutagulla, and Stephen E Bradforth. The influence of aqueous solvent on the electronic structure and non-adiabatic dynamics of indole explored by liquid-jet photoelectron spectroscopy. *Faraday Discussions*, 212(0):359–381, 2018. ISSN 1359-6640. doi: 10.1039/C8FD00123E. URL `http://dx.doi.org/10.1039/C8FD00123E`.

[110] T J Godfrey, Hui Yu, and Susanne Ullrich. Investigation of electronically excited indole relaxation dynamics via photoionization and fragmentation pump-probe spectroscopy. *The*

Journal of Chemical Physics, 141(4):44314, jul 2014. ISSN 0021-9606. doi: 10.1063/1.4890875. URL https://doi.org/10.1063/1.4890875.

[111] T J Godfrey, Hui Yu, Michael S Biddle, and Susanne Ullrich. A wavelength dependent investigation of the indole photophysics via ionization and fragmentation pump–probe spectroscopies. *Physical Chemistry Chemical Physics*, 17(38):25197–25209, 2015. ISSN 1463-9076. doi: 10.1039/C5CP02975A. URL http://dx.doi.org/10.1039/C5CP02975A.

[112] Andrzej L Sobolewski and Wolfgang Domcke. Ab initio investigations on the photophysics of indole. *Chemical Physics Letters*, 315(3):293–298, 1999. ISSN 0009-2614. doi: https://doi.org/10.1016/S0009-2614(99)01249-X. URL https://www.sciencedirect.com/science/article/pii/S000926149901249X.

[113] U Chandra Singh and Peter A Kollman. An approach to computing electrostatic charges for molecules. *Journal of Computational Chemistry*, 5(2):129–145, apr 1984. ISSN 0192-8651. doi: https://doi.org/10.1002/jcc.540050204. URL https://doi.org/10.1002/jcc.540050204.

[114] Tian Lu and Feiwu Chen. Multiwfn: A multifunctional wavefunction analyzer. *Journal of Computational Chemistry*, 33(5):580–592, feb 2012. ISSN 0192-8651. doi: https://doi.org/10.1002/jcc.22885. URL https://doi.org/10.1002/jcc.22885.

[115] Jun Zhang. libreta: Computerized Optimization and Code Synthesis for Electron Repulsion Integral Evaluation. *Journal of Chemical Theory and Computation*, 14(2):572–587, feb 2018. ISSN 1549-9618. doi: 10.1021/acs.jctc.7b00788. URL https://doi.org/10.1021/acs.jctc.7b00788.

[116] John C. Tully. Molecular dynamics with electronic transitions. *The Journal of Chemical Physics*, 93(2):1061–1071, 1990. doi: 10.1063/1.459170. URL https://doi.org/10.1063/1.459170.

[117] Sharon Hammes-Schiffer and John C Tully. Proton transfer in solution: Molecular dynamics with quantum transitions. *The Journal of chemical physics*, 101(6):4657–4667, 1994.

[118] Jean-Marc L Pecourt, Jorge Peon, and Bern Kohler. Ultrafast internal conversion of electronically excited rna and dna nucleosides in water. *Journal of the American Chemical Society*, 122(38):9348–9349, 2000.

[119] Jean-Marc L Pecourt, Jorge Peon, and Bern Kohler. Dna excited-state dynamics: Ultrafast internal conversion and vibrational cooling in a series of nucleosides. *Journal of the American Chemical Society*, 123(42):10370–10378, 2001.

[120] Jorge Peon and Ahmed H Zewail. Dna/rna nucleotides and nucleosides: direct measurement of excited-state lifetimes by femtosecond fluorescence up-conversion. *Chemical physics letters*, 348(3-4):255–262, 2001.

[121] Delphine Onidas, Dimitra Markovitsi, Sylvie Marguet, Alexei Sharonov, and Thomas Gustavsson. Fluorescence properties of dna nucleosides and nucleotides: a refined steady-state and femtosecond investigation. *The Journal of Physical Chemistry B*, 106(43):11367–11374, 2002.

[122] Susanne Ullrich, Thomas Schultz, Marek Z Zgierski, and Albert Stolow. Electronic relaxation dynamics in dna and rna bases studied by time-resolved photoelectron spectroscopy. *Physical Chemistry Chemical Physics*, 6(10):2796–2801, 2004.

[123] Hanneli R Hudock, Benjamin G Levine, Alexis L Thompson, Helmut Satzger, David Townsend, N Gador, Susanne Ullrich, Albert Stolow, and Todd J Martinez. Ab initio molecular dynamics and time-resolved photoelectron spectroscopy of electronically excited uracil and thymine. *The Journal of Physical Chemistry A*, 111(34):8500–8508, 2007.

[124] Chris T Middleton, Kimberly de La Harpe, Charlene Su, Yu Kay Law, Carlos E Crespo-Hernández, and Bern Kohler. Dna excited-state dynamics: from single bases to the double helix. *Annual review of physical chemistry*, 60:217–239, 2009.

[125] Bern Kohler. Nonradiative decay mechanisms in dna model systems. *The Journal of Physical Chemistry Letters*, 1(13):2047–2053, 2010.

[126] Carlos E Crespo-Hernández, Boiko Cohen, Patrick M Hare, and Bern Kohler. Ultrafast excited-state dynamics in nucleic acids. *Chemical reviews*, 104(4):1977–2020, 2004.

[127] Irene Conti and Marco Garavelli. Evolution of the excitonic state of dna stacked thymines: intrabase π $\pi^*\rightarrow$ s0 decay paths account for ultrafast (subpicosecond) and longer ($>$ 100 ps) deactivations. *The journal of physical chemistry letters*, 9(9):2373–2379, 2018.

[128] Spiridoula Matsika. Radiationless decay of excited states of uracil through conical intersections. *The Journal of Physical Chemistry A*, 108(37):7584–7590, 2004.

[129] Franziska Buchner, Akira Nakayama, Shohei Yamazaki, Hans-Hermann Ritze, and Andrea Lübcke. Excited-state relaxation of hydrated thymine and thymidine measured by liquid-jet photoelectron spectroscopy: Experiment and simulation. *Journal of the American Chemical Society*, 137(8):2931–2938, 2015. doi: 10.1021/ja511108u. URL `https://doi.org/10.1021/ja511108u`. PMID: 25671554.

[130] Clélia Canuel, Michel Mons, François Piuzzi, Benjamin Tardivel, Iliana Dimicoli, and Mohamed Elhanine. Excited states dynamics of dna and rna bases: Characterization of a stepwise deactivation pathway in the gas phase. *The Journal of Chemical Physics*, 122(7): 074316, 2005. doi: 10.1063/1.1850469. URL `https://doi.org/10.1063/1.1850469`.

[131] Ana Julieta Pepino, Javier Segarra-Martí, Artur Nenov, Roberto Improta, and Marco Garavelli. Resolving ultrafast photoinduced deactivations in water-solvated pyrimidine nucleosides. *The Journal of Physical Chemistry Letters*, 8(8):1777–1783, 2017. doi:

10.1021/acs.jpclett.7b00316. URL https://doi.org/10.1021/acs.jpclett.7b00316. PMID: 28346789.

[132] Akira Nakayama, Gaku Arai, Shohei Yamazaki, and Tetsuya Taketsugu. Solvent effects on the ultrafast nonradiative deactivation mechanisms of thymine in aqueous solution: Excited-state qm/mm molecular dynamics simulations. *The Journal of Chemical Physics*, 139(21):214304, 2013. doi: 10.1063/1.4833563. URL https://doi.org/10.1063/1.4833563.

[133] Omair Ghafur, Stuart W. Crane, Michal Ryszka, Jana Bockova, Andre Rebelo, Lisa Saalbach, Simone De Camillis, Jason B. Greenwood, Samuel Eden, and Dave Townsend. Ultraviolet relaxation dynamics in uracil: Time-resolved photoion yield studies using a laser-based thermal desorption source. *The Journal of Chemical Physics*, 149(3):034301, 2018. doi: 10.1063/1.5034419. URL https://doi.org/10.1063/1.5034419.

[134] B. Barc, M. Ryszka, J. Spurrell, M. Dampc, P. Limão-Vieira, R. Parajuli, N. J. Mason, and S. Eden. Multi-photon ionization and fragmentation of uracil: Neutral excited-state ring opening and hydration effects. *The Journal of Chemical Physics*, 139(24):244311, 2013. doi: 10.1063/1.4851476. URL https://doi.org/10.1063/1.4851476.

[135] Dana Nachtigallová, Adélia J. A. Aquino, Jaroslaw J. Szymczak, Mario Barbatti, Pavel Hobza, and Hans Lischka. Nonadiabatic dynamics of uracil: Population split among different decay mechanisms. *The Journal of Physical Chemistry A*, 115(21):5247–5255, 2011. doi: 10.1021/jp201327w. URL https://doi.org/10.1021/jp201327w. PMID: 21548626.

[136] Hans Lischka, Mario Barbatti, Farhan Siddique, Anita Das, and Adelia JA Aquino. The effect of hydrogen bonding on the nonadiabatic dynamics of a thymine-water cluster. *Chemical Physics*, 515:472–479, 2018.

[137] Matthew M Brister and Carlos E Crespo-Hernández. Excited-state dynamics in the rna nucleotide uridine 5'-monophosphate investigated using femtosecond broadband transient absorption spectroscopy. *The journal of physical chemistry letters*, 10(9):2156–2161, 2019.

[138] Juan José Serrano-Pérez, Remedios Gonzalez-Luque, Manuela Merchan, and Luis Serrano-Andres. On the intrinsic population of the lowest triplet state of thymine. *The Journal of Physical Chemistry B*, 111(41):11880–11883, 2007.

[139] Patrick M. Hare, Carlos E. Crespo-Hernández, and Bern Kohler. Internal conversion to the electronic ground state occurs via two distinct pathways for pyrimidine bases in aqueous solution. *Proceedings of the National Academy of Sciences*, 104(2):435–440, 2007. ISSN 0027-8424. doi: 10.1073/pnas.0608055104. URL https://www.pnas.org/content/104/2/435.

Acknowledgements

My journey in photochemistry started when I joined the group of Prof Marco Garavelli in 2012 for a short summer internship, as a green aimless chemistry student. In this sense the association with Prof Garavelli for almost a decade has shaped the entirety of my academic career. His contribution to shaping my professional life has been immense. With his joyful and helpful nature, his mentorship has been a guiding light and a positive stimulant throughout my doctoral work.

The works documented in this thesis would not have been possible without the constant guidance of Artur. He has always been available to clear doubts and brainstorm on ideas with infinite patience. A constant friend during my time in Bologna, his influences are imprinted on every work in this thesis. Irene was one of the first mentors to guide me in the group of Prof. Garavelli providing me with gentle introductions to field of computational chemistry and COBRAMM many years ago. The collaboration with her has been enriching and helped me grow into a competent? researcher. Ivan provided his invaluable mentorship during his supervision in Lyon and warming me up for a doctoral work in photochemistry. His gregarious nature, big heart and infinite generosity have helped to navigate the ebbs encountered in the research journey.

The group at POLIMI under leadership of Prof. Giulio Cerullo have provided access to cutting edge spectroscopic data which have been used throughout in this thesis. I am grateful to Prof. Cerullo and his research group for hosting me a month at his lab and introducing me to the world of experimental ultrafast spectroscopy. Special thanks goes to Piotr Kabaciński who has been my constant point of reference to understand the ultrafast experimental world. His quest for perfection in experiments, figures and analysis have always proven to be a motivating factor in my works. I am also grateful to Vasilis and Pranav for giving me company in Milan.

Flavia, Lorenzo, Francesco, Silvia, Davide, Nicola, Simone, Andrea, Moshsen, Marzio, Mario, Marco,Muhammad, Samira, Elham, Sonia, Giacomo, Luca and Silvia from the physical chemistry group at Toso have been a wonderful company during these years providing mental and material support. A unique camraderie exists here which makes work a place of pleasure and the people a source of joy. I would like to also thank the coordinators of my PhD program Prof. Domenica and Dr. Elisabetta for piloting our doctoral cycle through very unconventional times.

I am indebted to the LightDynamics umbrella financed by European Horizon2020 grant **H2020-MSCA-ITN-2017 (ETN)** and grateful for the connections it has fostered with set of unique and talented individuals. I would like to acknowledge all my fellow ESRs Davide, Marilu, Piotr ,Evelinna, Luis, Haritha, Martha, Evangelos, Zhang, Katerina, Filipp, Irene, Mark and Elena for providing a jovial company during the various seminars we attended together.

The invaluable knowledge shared by the various PIs in the LightDynamics network has illuminated the farthest corners of the mind suggesting new avenues of research. Foremost I would like to thank Prof. Roberto Improta who has guided this network and Dr. James Green for administrating and herding us to be a productive group despite of inertia. I am grateful to Prof. Dimitra, Prof. Fabrizio, Prof. Leticia, Dr. Susan and all other bosses for the invaluable distillation of cutting edge information in diverse fields of photobiology and photochemistry.

My existence as a researcher is owed to the sacrifices, help and and encouragement of my brother Vijit and parents Ravindra and Sushma. Their contribution is invaluable to measure and impossible to express. Their constant support in my journey has enabled me to pursue an academic career.

Lastly a special thanks to the vibrant community of FOSS for providing me with all the tools to undertake research and write this thesis. Special mentions to GNU/Linux system, R statistical environment and the LaTeX document system.

Vishal Kumar Jaiswal, Bologna, January 2022