



Scuola Superiore Meridionale
Università degli Studi di Napoli
Federico II

Doctoral Thesis in
Molecular Sciences for Earth and Space

Photoinduced Dynamics of DNA-Protein Model Systems in Condensed Phase

---

Author:
Gabriele Iuzzolino

Supervisors:
Prof. Nadia Rega
Dr. Fulvio Perrella

XXXVII CYCLE

Contents

Abstract **I**

Introduction **1**

I.1 Nucleic acid-protein photo-crosslinking 3

1 Microsolvation and Photophysics of 5-benzyluracil in methanol **5**

1.1 Methods 7

1.1.1 The Atom-centered Density Matrix Propagation formalism 7

1.1.2 The ONIOM partition scheme and Non-Periodic Boundary Conditions 10

1.1.3 UV-Vis spectrum simulation 14

1.2 Results 15

1.2.1 Optical absorption from minimum energy structures . . . 15

1.2.2 Conformational and solvation equilibrium 23

1.2.3 Optical absorption spectrum from AIMD sampling 29

Appendices 36

1.A Optical data from CAM-B3LYP/6-31+G(d,p) minima . . 36

1.B NTO analysis of transitions from minimum energy structures 37

1.C NTO analysis of transitions from AIMD sampling 41

2 Ultrafast Transient Fluorescence of 5-benzyluracil **46**

2.1 Introduction 46

2.1.1 Fluorescence Up-Conversion . . . 47
2.1.2 Transient emission spectroscopy of 5BU . . . 47
2.2 Methods . . . 49
2.2.1 Non adiabatic molecular dynamics . . . 49
2.2.2 Time-resolved emission simulation . . . 54
2.3 A novel protocol to compare simulation and experiment . . . 55
2.3.1 A challenging system . . . 56
2.3.2 Our strategy . . . 57
2.4 Results . . . 62
2.4.1 Transient fluorescence spectrum . . . 62
2.4.2 Population analysis . . . 64
2.4.3 Time-resolved conformational analysis . . . 68

3 High-energy Photoreactivity of 5- and 6-benzyluracil 72
3.1 Methods . . . 74
3.1.1 Photoionization simulation . . . 74
3.1.2 Photofragmentation simulation . . . 76
3.2 Results . . . 77
3.2.1 Simulated photo-electron spectra . . . 77
3.2.2 Fragmentation pathways . . . 82
Appendices . . . 93
3.A PES and PEPICO experimental spectra . . . 93
3.B Conformational analysis . . . 94
3.C Additional PES data . . . 99
3.D Additional fragmentation pathways . . . 101

4 Integrating nonadiabatic dynamics with multilayer extrapolative schemes: towards a new Newton-X/Gaussian16 interface 106
4.1 The Newton-X New Series / Gaussian16 ONIOM interface . . . 108
4.2 Generation of initial conditions in the explicit solvent . . . 111

Publications List CXV

List of Figures CXVI

List of Tables CXIX

Bibliography CXX

Abstract

Interaction of light with matter is a very useful tool to unveil the properties of biological system *in vivo*. Across the electromagnetic spectrum, biomolecules can undergo photoabsorption, stimulated and induced emission, synthesis of covalent bonds and, at higher energies, photoionization and photolysis and rearrangement. Several spectroscopic technique rely on them to study the living matter: steady state and transient optical spectroscopy, fluorescence upconversion, photoelectron and photoelectron-photoion coincidence spectroscopy. Model system have been developed to study the effect of such type of reaction on system in condensed phase and shed light on the analogous behaviour of complex biopolimers. The modeling of such system is vital for the improvement of spesctroscopic techniques; at the same time, it require advanced computational methods, able to consider the flexibility of such systems, the effect of the environment, but also investigate ultrafast time-scales and the coupling between the nuclear motion and electron dynamics. In this thesis work, we focused on two model systems of nucleic acid-protein photo-crosslinking, which are prototype systems of the experimental and computational challenges faced in the field. We accurately simulated their thermal equilibrium in liquid phase and found out that their photophysics heavily depends on the microsolvation; we simulated their ultrafast fluorescence by non-adiabatic dynamics and developed a new computational protocol compare it with low-time resolution experimental data; we efficiently simulated their photoionization and photofragmentation pathways. Finally, a new software interface was developed to perform non-adiabatic molecular dynamics with multilayered extrapolative partition schemes.

Introduction

The interaction of light with matter can induce multiple different reactions, which critically depend on the nature of the interacting system and the energy of the exciting radiation. In the years, this has pushed the development of the science of spectroscopy, which investigates the properties of chemical systems by analyzing their response to irradiation. Based on the energy, radiation can be classified between ionizing and non-ionizing.

Non-ionizing radiation, encompassing infrared, visible and near-ultraviolet (UV) light, can induce transitions among bound electronic, and roto-vibrational states: for their very nature, transitions induced by optical radiation are typically non-destructive and reverse on the sub-second timescale after the irradiation is stopped. Their non-invasivity and non-destructive nature makes them great tools to investigate the properties of biological systems: optical techniques applied in the biochemistry include absorption, spontaneous and stimulated emission, both in their steady-state and transient version.

On the other hand, light can induce irreversible reaction, including covalent bond synthesis and breaking: this can happen both with optical radiation - since bound excited states can have lower or even zero barrier towards specific reactions with respect to the ground stare - or with ionizing radiation - which promotes the system to an open-shell state, favoring bond breaking and rearrangement. In biological systems, such radiations can directly modify the structure of biopolimers like nucleic acids and proteins, or producing reactive oxygen species which then cause further damage. High-energy spectroscopic techniques, like photoelectron spectroscopy (PES), photoelectron-photoion con-

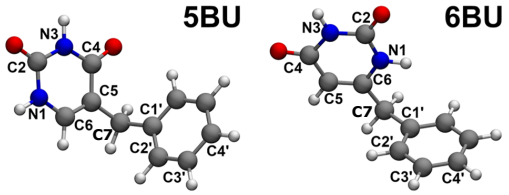


Fig. 1: Three-dimensional structures of 5- and 6-benzyluracil, with atom labels. [1]

cidence spectropy (PEPICO), X-ray absorption (XAS) and fluorescence (XFS) spectroscopy are used to simulate the photodamage and study the chemical mechanisms that allow nucleic acids to prevent it (photo-stability). On the other hand, photo-crosslinking reactions can be used to stabilize nucleoprotein complex and allow the isolation.

In this context, theoretical and computational chemistry can offer decisive insights about the molecular mechanisms influencing the absorption, fluorescence, photoionization and fragmentation of biomolecules in complex environments. Such understanding will assist the development of spectroscopic techniques and our understanding of the photodamage in living systems.

In this thesis work, we developed computational protocols to study the photoinduced dynamics of two systems developed to model Nucleic Acid-protein Photo-CrossLinking (NAPCL), named 5- and 6-benzyluracil. They are prototype of a series of challenges requiring advanced computational methods: they are flexible molecules formed by two interacting chromophore rings, able to form hydrogen bonds with protic solvents. When irradiated by UV light, they can undergo a photocyclization reaction, whose mechanism is still to be fully understood. Their photophysics is heavily influenced by the reciprocal orientation of the two rings and microsolvation of their heteroatoms: in chapter 1, this is studied via *ab initio* molecular dynamics in hybrid implicit/explicit solvent; they exhibit fluorescence decaying on the sub-ps time scale and characterized by the coupling of several excited states, which is described in chapter 2 by the means of non-adiabatic molecular dynamics. Chapter 3 explores their possible pathways of photo-ionization and photolysis, upon irradiation by ionizing radiation, via semiempirical molecular dynamics. Chapter 4, finally, reports the

development of a new interface to perform non-adiabatic molecular dynamics with extrapolative partition schemes. Here we detail their usefulness as model system of photoinduced reactions between nucleobasis and aminoacid residues.

I.1 Nucleic acid-protein photo-crosslinking

Interactions between nucleic acids and proteins on the molecular scale govern the most basic processes of life, from DNA replication and repair to protein expression and synthesis.[2–12] However, it is challenging to isolate nucleoprotein complexes with null or small perturbations, where these last ones in their turn might disrupt the very nature of these complexes. Moreover, very intricate, and indeed not fully understood, is the role of light (i.e. UV irradiation) with protein-nucleic acid complexes. Light can cause the crosslinking of the protein to the nucleic acid; a covalent modification that can drastically change (usually for the worst) the biological roles of such compounds.[13–16] Thus, it is very important to understand how light interacts with nucleoprotein complexes, starting from their optical properties and moving towards the study of the photo-induced crosslinking reaction mechanism.

The complete comprehension of protein - nucleic acid interactions and their interplay with light has been limited because, upon irradiation, an extensive protein and nucleic acid damage is often observed and the involved reactions are not very efficient. Advanced spectroscopic techniques and model compounds have been developed (and sometimes combined together) to better study this matter. Nucleic acid-protein laser-based photo-crosslinking is, for example, a promising technique to isolate nucleoprotein complexes in their biological conformation. It employs UV femtosecond-stimulated lasers to induce *in vivo* the formation of covalent bonds - named crosslinks - between a nucleobase and an amino acid residue. In this way, it is easier to stabilize the biological conformation of the complex and to allow it to survive the following isolation. In addition, the employment of ultra-short irradiation times minimizes side reactions and allows to characterize even transient interactions.[17–22]

The resulting complex reaction mixtures and conditions also make it dif-

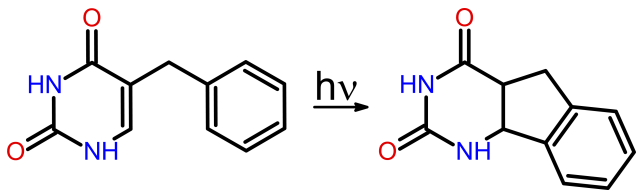


Fig. 2: Photo-cyclization of 5-benzyluracil to 1,2-indaneuracil.[28]

ficult to identify specific aminoacid/base cross-links.[23,24] A common proposed strategy is to attach a pyrimidine base to the tyrosine, phenylalanine, and tryptophan side chains, using a short linker to mimic the proximity and orientation in nucleoprotein complexes.[25]

More in details, this thesis work focuses on two isomers of benzyluracil: 5-benzyluracil (5BU) and 6-benzyluracil (6BU). They are formed by a uracil and a phenyl ring, linking by a methylene bridge; they can undergo a photoinduced crosslinking reactions leading to the synthesis of 1,2-indaneuracil and 2,1-indaneuracil, respectively, allowing to model the photo-addiction of phenylalanine to thymine or uracil, with the advantage to have better reaction yields than the nucleobase/amino acid mixture.[25] This reaction has been already shown to be possible in methanol, by photoexcitation between 254 and 266 nm,[25–27] corresponding to the maximum of the first band of the UV-Vis spectrum.

Chapter 1

Microsolvation and Photophysics of 5-benzyluracil in methanol

5BU has been previously investigated theoretically *via* both wave-function and DFT based approaches, further proving that 5BU is a good candidate as a model compound for studying the optical properties of nucleoprotein complexes and their photo-reactivity.[26,29,30] 5BU photo-cyclization pathway appears to be influenced by multiple factors: Sun *et al.*[25] proposed a mechanism relying on electron transfer between the chromophores and proton exchange with the solvent; according to Valadan *et al.*[27], the absorption involves exclusively the $S_1 \leftarrow S_0$ transition, but the following reactive pathway is determined by a strong coupling among low-lying states and the reciprocal orientation of the two aromatic rings.

However, the photophysics of such a model compound is still far to be fully understood. It is still debated which one of the low-lying excited states of 5BU is responsible for the absorption preceding the photoreaction, given the different results of the previous theoretical studies, relying on minimum energy structures.[26,27,29,30] Indeed, gas-phase simulations[29,30] considered the $S_2 \leftarrow S_0$ transition to be the only bright one; on the contrary, in methanol solution the absorption is either attributed entirely to the $S_2 \leftarrow S_0$ transition,[27] or can involve different states depending on the level of theory.[30] These different findings reflect

on the hypothesized system evolution in the excited state, where single and multiple crossings among electronic states can be predicted. Such results can be further influenced by the role of the solvent and how this one is modeled and by the thermal effects that are not fully taken into account when few minimum energy structures are analyzed.[31]

This chapter describes the optical absorption of 5BU in methanol solution at room temperature, by including both the explicit solute-solvent interactions and the finite temperature effects with a combined computational and experimental approach. For this aim, a complete study of both the conformational space accessible to the solute and the dynamics of the solute-solvent interactions was conducted. In such a way, an exhaustive knowledge of the ground-state equilibrium of the system in solution at room temperature was gathered, to provide the molecular insights of the absorption spectrum. A room temperature *ab initio* molecular dynamics (AIMD) trajectory was performed, with a hybrid explicit/implicit solvation scheme, combining quantum/molecular mechanics (QM/MM) and non-periodic boundary conditions (NPBCs).[32–35] Then, we simulated the optical absorption from configurations extracted from the AIMD,[31] by linear response time-dependent DFT (LR TD-DFT).[36–38] Such protocol is detailed in sections 1.1.1 to 1.1.3.

Multilayered quantum/classical/continuum MD are necessary to accurately describe both the ensemble averages, microsolvation dynamics and absorption spectra of chromophores in solution.[31,35,39–51] This is particularly true for 5BU in methanol, since 5BU is a flexible solute with two interacting aromatic rings and four heteroatoms, while methanol is a flexible, polar and protic solvent: they thus require, at the same time, the accurate treatment of the solute degrees of freedom and the efficient description of the solvent viscosity and mobility; moreover, we have to take into account both specific, polarization and electrostatic solute-solvent interactions.[35,52–58]

This methodology, taking into account several conformations and solute-solvent interactions, allowed us to unveil the role of different excited states in the overall 5BU observed photophysical properties, as described in section 1.2.

1.1 Methods

1.1.1 The Atom-centered Density Matrix Propagation formalism

As aforementioned, we collected a ground-state AIMD trajectory of 5BU in methanol at room temperature, with the hybrid ONIOM/NPBCs partition scheme described in section 1.1.2, by using a development version of the Gaussian electronic structure software package.[59] We exploited the Atom-centered Density Matrix Propagation (ADMP)[60–64] formalism, which avoids the density matrix convergence at each step, by instead propagates it along with the nuclear degrees of freedom, and is detailed in the following. A (fictitious) mass is required in the density matrix equation of motion. A tensorial mass weighting scheme was employed:[60] valence orbitals mass was set to 0.1 amu bohr^2, whereas core orbitals were weighted according to the scheme described in ref. 60. The ONIOM/ADMP/NPBCs AIMD was carried out for 20 ps, with a time step of 0.2 fs, in a NVT ensemble. A 298 K temperature was kept constant by velocity rescaling every 1 ps.

A preliminary equilibration was carried out before production AIMD simulation. 5BU was placed into a 9582 methanol molecules cubic box. A MM Langevin dynamics was performed (with NAMD software[65]) in the NpT ensemble ($T = 298$ K, $p = 1$ atm) with periodic boundary conditions, applying a Langevin piston to keep also the pressure constant.[66,67] GAFF was employed: GAFF atom types were assigned to solute atoms, as well as HF/6-31G(d) RESP charges. Solvent equilibration (fixing the solute coordinates) was first performed for 10 ns, then also solute degrees of freedom were allowed to propagate for 1 ns. Then, solvent molecules within 20 Å from the solute center of mass were selected. An AIMD equilibration of the selected system was performed in the NVT ensemble for 1 ps, applying a velocity rescaling every 250 fs, with the same partition scheme (ONIOM, non-periodic boundary conditions) and level of theory (B3LYP/6-31G(d,p) of the following production simulation.

ADMP is an *ab initio* method to describe the evolution of electron density in time without having to pay the computational cost of a convergence procedure.

It assigns electron density with a fictitious mass and propagates it with a classical equation of motion; in fact, ADMP does not respect the Born-Oppenheimer (BO) approximation and instead treats electron density and nuclei as two subsystem sharing the same classical phase space. Its core concept is borrowed by Car-Parrinello method [68–71] which however assigns the fictitious mass to Kohn-Sham orbitals, instead of density matrix, and thus has to use plane wave base functions, pure functionals and periodic boundary conditions. ADMP, on the contrary, has access to localized gaussian base functions, non-periodic boundary conditions (NPBCs) - both better fitting non-periodic molecular systems - hybrid functionals - increasing accuracy with respect to pure ones - and QM/MM partition - increasing computational efficiency.[64] ADMP defines the shared nuclear and electronic classical phase space by the following extended Lagrangian function:[60]

$$\mathcal{L} = \frac{1}{2}Tr[\boldsymbol{V^T M V}] + \frac{1}{2}Tr([\underline{\mu}^{1/4}\boldsymbol{W}\underline{\mu}^{1/4}]^2) - E(\boldsymbol{R},\boldsymbol{P}) - Tr[\boldsymbol{\Lambda P P} - \boldsymbol{P})] \quad (1.1)$$

where $\mathbf{M}$, $\mathbf{R}$, and $\mathbf{V}$ are the nuclear masses, positions, and velocities, respectively, while $\mathbf{P}$, $\mathbf{W}$ and $\underline{\mu}$ are the density matrix, density matrix velocity, and the fictitious mass matrix for the electronic degrees of freedom, respectively; constraints on the total number of electrons, N_e, and on the idempotency of the density matrix are imposed using a Lagrangian multiplier matrix $\mathbf{\Lambda}$. Defining fictitious mass as a matrix, rather than a scalar, allows reducing the mass of density matrix elements for the valence orbitals with respect to core ones: this further reduces the computational cost, with no significant loss in accuracy, since core electrons are hardly, if any, involved in chemical reactivity.[60]

The Euler–Lagrange equations for the propagation of the density matrix and the nuclei are, respectively:[60]

$$\underline{\mu}^{1/2}\frac{d^2\boldsymbol{P}}{dt^2}\underline{\mu}^{1/2} = -\left[\left.\frac{\partial E(\boldsymbol{R},\boldsymbol{P})}{\partial \boldsymbol{P}}\right|_{\boldsymbol{R}} + \boldsymbol{\Lambda P} + \boldsymbol{P\Lambda} - \boldsymbol{\Lambda}\right] \quad (1.2)$$

$$\boldsymbol{M}\frac{d^2\boldsymbol{R}}{dt^2} = -\left.\frac{\partial E(\boldsymbol{R},\boldsymbol{P})}{\partial \boldsymbol{R}}\right|_{\boldsymbol{P}} \quad (1.3)$$

It can be proven[60] that eq. 1.1 represents a conservative system, i.e. :

$$\frac{\partial \mathcal{H}}{\partial t} = 0 \tag{1.4}$$

where $\mathcal{H}$ is the conjugated hamiltonian. Equation 1.4 represents an important result, since conservation of energy is a condition for the good behavior of a dynamics. Propagating the electron density is, however, an approximation that has to be controlled: this control has to be done during the dynamics itself, using the BO method as a reference. Here are the variables one must check during the ADMP dynamics:

- The adiabaticity index
- The Frobenius norm

but does not distance itself too much from it: it oscillates around a BO PES. To control the distance from a BO trajectory - namely the 'adiabaticity' - we have to control the fictitious mass : a fictitious mass too big would be unphysical, while a fictitious mass to small would produce an exceedingly high acceleration. In a sense, it's possible to compare it to the time-step: they are two variables the operator can control and must pay attention to set neither too big nor too small; indeed, one of the advantages of ADMP over Car-Parrinello is the possibility to use a smaller mass and thus a larger time step, saving computations.[60]

After setting appropriate fictitious mass and time step, the good behavior of the trajectory is monitored by some parameters. The first one is the adiabaticity index, defined by:

$$\frac{\partial \mathcal{H}_{fictitious}}{\partial t} = Tr[\boldsymbol{W}\underline{\mu}^{1/2}\frac{d^2\boldsymbol{P}}{dt^2}\underline{\mu}^{1/2}] \tag{1.5}$$

If we partition the total energy into real and fictitious contributions:[60]

$$\frac{\partial \mathcal{H}}{\partial t} = \frac{\partial \mathcal{H}_{real}}{\partial t} + \frac{\partial \mathcal{H}_{fictitious}}{\partial t} = 0 \Rightarrow \frac{\partial \mathcal{H}_{real}}{\partial t} = -\frac{\partial \mathcal{H}_{fictitious}}{\partial t} \tag{1.6}$$

and since in a BO trajectory $\partial\mathcal{H}/\partial t = \partial\mathcal{H}_{real}/\partial t = 0$, it follows that $\partial\mathcal{H}_{ficititious}/\partial t$

measures adiabaticity of the ADMP trajectory and specifically must be close to zero;[60] from eq. 1.5, we also see the adiabaticity index depends on the fictitious mass.

Other parameters are the total energy E_{tot} - which must not diverge, since our system must be conservative - the nuclear kinetic energy $E_{kin,C}$ and the fictitious kinetic energy $E_{kin,P}$ - which must be much smaller than $E_{kin,C}$, since a too high 'temperature' of the density matrix would cause unphysical energy exchange with the nuclear part.[60] The last parameter is the Frobenius norm of the commutator between the Fock matrix and the density matrix, $|||[\boldsymbol{F}, \boldsymbol{P}]|||$: in a BO trajectory it is identically zero, since the SCF method calculates the latter from the first; thus a physical ADMP dynamics requires it to always be close to zero.[60]

Finally, the choice of initial velocity of the density matrix elements must in some sense be consistent with the choice of initial velocities of the nuclei, again to avoid unphysical thermal exchange between the two subsystems: this can be done by converging the initial density matrix by a SCF calculation, to then propagate the nuclear positions for a single time step and recaculate the density matrix.[60]

1.1.2 The ONIOM partition scheme and Non-Periodic Boundary Conditions

The 5BU-methanol system was simulated by a multilayered QM/MM/implicit partition, combining the N-layered integrated molecular orbital and molecular mechanics (ONIOM)[72–74] extrapolative scheme with non-periodic boundary conditions (NPBCs).[32–35,41,43,75–83] The solute, treated at the B3LYP[84–87]/6-31G(d,p) level of theory, was put at the center of a spherical cavity with radius of 20 Å, including 356 methanol molecules, described with the General AMBER Force Field (GAFF) model.[88]

DFT potentials appear in this context the method of choice, because of their good compromise between cost and accuracy for organic/biological system.[35,89–101]

The cavity radius was determined by the equilibration procedure detailed

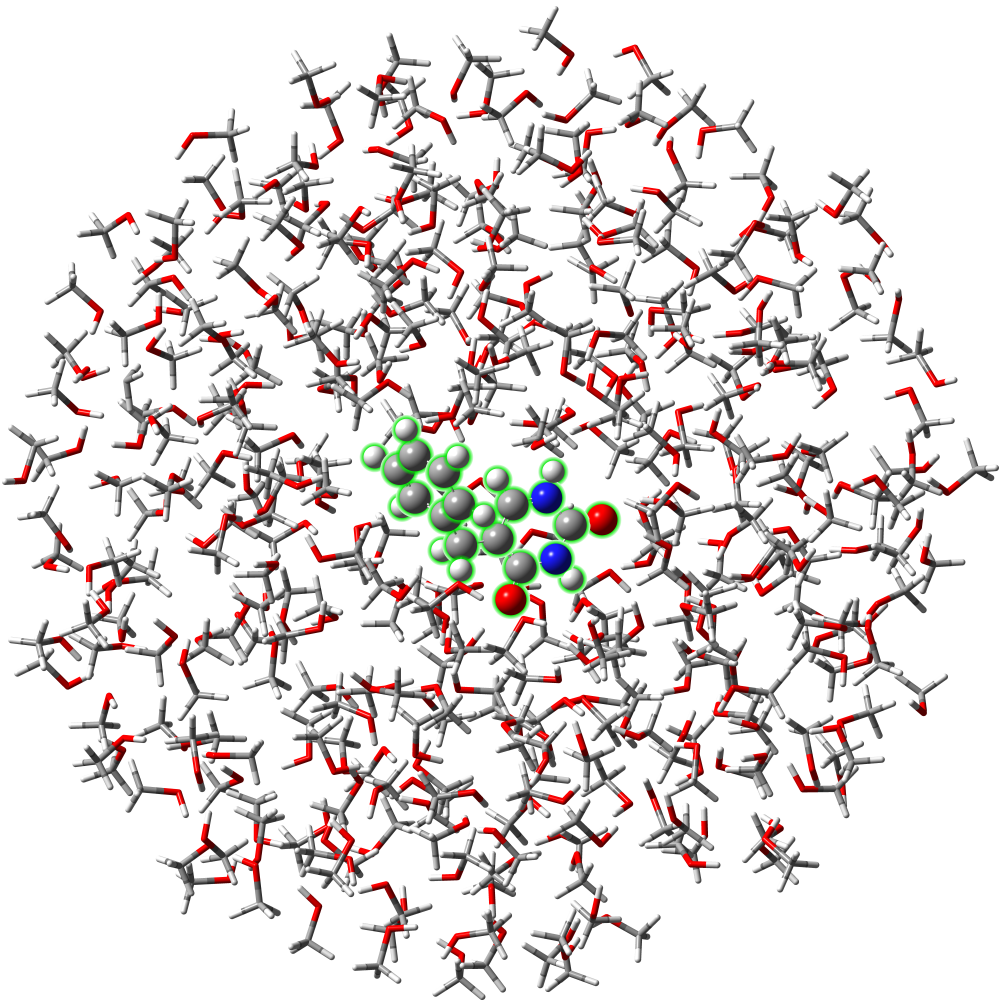


Fig. 1.1: The multi-layer scheme employed in ground state AIMD: the DFT layer (green-highlighted balls and sticks) includes the 5BU solute and the MM layer (wireframe) the explicit solvent; NPBCs are applied to the sphere. [28]

in section 1.1.1. NPBCs were enforced on the cavity by the conductor-like polarizable continuum model (C-PCM) self-consistent reaction field[102–105] (*via* the ONIOM/PCM-X[106–108] formalism) and the dispersion-repulsion potential named GLOB,[32–35,41,43,79,81–83] specifically designed by some of us for flexible solvents, which is detailed in section 1.1.2.

The Generalized Liquid Optimized Boundary

For a NVT ensemble, the constant-volume free energy A of the simulated system can be expressed as:

$$A(\boldsymbol{R}) = E(\boldsymbol{P}, \boldsymbol{R}) + W(\boldsymbol{P}, \boldsymbol{R}) \tag{1.7}$$

where $E(\boldsymbol{P}, \boldsymbol{R})$ is the potential energy of the explicit system in the nuclear configuration $\boldsymbol{R}$, and $W(\boldsymbol{P}, \boldsymbol{R})$ is the "mean field" contribution accounting for the interactions with the implicit environment; the mutual polarization of the explicit and implicit systems is explicitly considered through the one-electron density matrix $\boldsymbol{P}$. The mean field contribution W can be decomposed according to Ben–Naim's scheme of the solvation process:[109]

$$W(\boldsymbol{R}) = W_{\text{disp-rep}}(\boldsymbol{R}) + W_{\text{elec}}(\boldsymbol{R}) + W_{\text{cav}}(\boldsymbol{R}) \tag{1.8}$$

where W_{cav} is the cavitation free energy (constant in our NVT model), W_{elec} accounts for long-range electrostatic interactions with the bulk solvent, included through the C-PCM self-consistent reaction field, $W_{\text{disp-rep}}$ accounts for the short-range dispersion–repulsion interactions between explicit and bulk solvent, allowing to initialize and mantain a physical solvent distribution, as well as acting as a confining potential.

The GLOB (Generalized Liquid Optimized Boundary) is a radial potential $W_{\text{disp-rep}}(r)$ acting on the center of mass of each explicit solvent molecule and given by the sum of a set of N_g Gaussian functions, g_i, with r_i centers equally

spaced along the radius of the spherical cavity:

$$W_{\text{disp-rep}}(r) = \sum_{i}^{N_g} g_i(r - r_i) = \sum_{i}^{N_g} \lambda_i e^{-\frac{(r-r_i)^2}{2\sigma^2}} \tag{1.9}$$

Each g_i(r) function has the same spread, 2σ, but a variable height, λ_i, which is determined empirically from a test MD simulation using the overall numerical density of the solvent as a reference: at each step of the simulation, the local density is computed on each spherical layer. Each λ_i value is then updated, in order to to maintain the local density as close as possible to the reference value. This process converges in about few nanoseconds, when the $W_{disp-rep}(r)$ function is optimized with no further significant changes, corresponding to an average of the instantaneous interactions accounted for during the simulations.[32–35,79,81] This methodology has been developed and widely employed by some of us to model systems in both water[32–34,41,43,79,81–83] and methanol solution.[35]

The ONIOM extrapolative method

The energy of the atomistic system - $E(\boldsymbol{P}, \boldsymbol{R})$ in eq. 1.7 - can be expressed, according to the ONIOM scheme, as:[72–74]

$$E(\text{ONIOM}n) = \sum_{i=1}^{n} E(\text{Layer}_i, \text{Method}_{n+1-i}) - \sum_{i=2}^{n} E(\text{Layer}_i, \text{Method}_{n+2-i}) \tag{1.10}$$

which, for a two-layer ONIOM(QM/MM) calculation, becomes:

$$E(\boldsymbol{P}, \boldsymbol{R}) = E^{\text{model,QM}}(\boldsymbol{P}, \boldsymbol{R}) + E^{\text{real,MM}}(\boldsymbol{R}) - E^{\text{model,MM}}(\boldsymbol{R}) \tag{1.11}$$

where $E^{\text{real,MM}}$ is the energy of the whole explicit system calculated at MM level, while E^{model} is the energy of the so-called *model system*, treated at quantum mechanical level. The electrostatic interaction between the two layers was treated by including the atomic charges of the MM layer into the hamiltonian of the model sub-system (electronic embedding).

The combination of ONIOM and NPBCs provides a well defined, single

valued and differentiable potential to perform multilayered AIMD simulations in explicit solvents. Such hybrid solute-solvent QM/MM approach gives a good compromise between accuracy and computational cost for the equilibrium and non-equilibrium simulation of medium-large size systems in solution, including specific solute-solvent interactions.[35,90,110–119]

1.1.3 UV-Vis spectrum simulation

The intensity of the electronic absorption spectrum $L(\omega)$ can be calculated (assuming a classical nuclear motion) as:[41,120]

$$L(\omega) = \hbar \int d\boldsymbol{Q} \rho_i(\boldsymbol{Q}, T) |\mu_{ij}(\boldsymbol{Q})|^2 \delta(V_i(\boldsymbol{Q}) - V_j(\boldsymbol{Q}) + \hbar\omega) \tag{1.12}$$

where $\boldsymbol{Q}$ is a set of nuclear coordinates and $\rho_i(\boldsymbol{Q}, T)$ is the probability density of configuration $\boldsymbol{Q}$ in the initial electronic state, which could be obtained by a MD trajectory on the $V_i(\boldsymbol{Q})$ potential energy surface. Further approximating such density of states with a discrete set of N_p configurations $\boldsymbol{Q}_l$ extracted from the MD, the spectrum $L(\omega)$ features the highest intensity at frequencies ω such that $\hbar\omega = |V_i(\boldsymbol{Q}_l) - V_j(\boldsymbol{Q}_l)|$:

$$\begin{aligned} L(\omega) &= \sum_j L_j(\omega) \\ L_j(\omega) &= \frac{1}{N_p} \sum_{l=1}^{N_p} f_{ij}(\boldsymbol{Q}_l)\, w\big(\hbar\omega - \Delta V_{ij}(\boldsymbol{Q}_l)\big) \end{aligned} \tag{1.13}$$

where $L_j(\omega)$ is the spectrum contribution by the the $j \leftarrow i$ transition (i being the ground state), N_p is the number of sampled configurations, $\boldsymbol{Q}_l$ is the l-th configuration, $f_{ij}(\boldsymbol{Q}_l)$ is the oscillator strength of the $j \leftarrow i$ transition at the l-th configuration, $\Delta V_{ij}(\boldsymbol{Q}_l)$ is the vertical transition energy, and $w\big(\hbar\omega - \Delta V_{ij}(\boldsymbol{Q}_l)\big)$ is a broadening function, peaked at $\hbar\omega = \Delta V_{ij}$, which is here a Gaussian broadening with a $s^2 = 0.001\,\mathrm{eV}^2$ variance:

$$w\big(\hbar\omega - \Delta V_{ij}(\boldsymbol{Q}_l)\big) = \exp\left[-\frac{1}{2}\left(\frac{\hbar\omega - \Delta V_{ij}(\boldsymbol{Q}_l)}{s}\right)^2\right] \tag{1.14}$$

This approach is able to simulate absorption spectra of systems of significant size including thermal broadening, but neglecting the vibronic fine structure.[41,120]

In order to simulate 5BU spectrum in methanol solution, 2001 uncorrelated configurations were extracted from the AIMD trajectory at regular time intervals (10 fs). The latter was chosen as a compromise between the computational cost of excited states calculation and the accurate sampling of 5BU torsional degrees of freedom.[31] The first 8 excited states were calculated for each configuration through linear response time-dependent DFT.[36–38] An ONIOM hybrid scheme was used: 5BU and the four methanol molecules closest to the solvation sites (possibly forming hydrogen bonds) were modeled at the QM TD-CAM-B3LYP/6-31+G(d,p) level,[84–87,121] and the remaining ones at MM level through the GAFF[88] force field (see fig. 1.2).

In these regards, the CAM-B3LYP functional appears suitable to describe the photophysics of (substituted) pyrimidines,[26,30,122–125] with particular regard to their charge-transfer states. Electrostatic interactions by the bulk solvent were treated with the C-PCM implicit solvent model, placed at the boundary of the spherical solvent box.

1.2 Results

1.2.1 Optical absorption from minimum energy structures

We first describe the optical properties of 5BU in methanol by adopting a static picture, i.e. by performing TD-DFT calculations on ground-state minimum energy structures, differing in 5BU geometries, solvation schemes and levels of theory. First, we chose two conformations that, according to literature,[27] are adopted by the 5BU molecule: the former has the rings arranged in an approximately orthogonal manner (*flag-like* structure, shortened to "F"), while the latter has the rings approximately facing each other (*butterfly-like* structure, shortened to "B") - see fig. 1.4. It is expected that the relative orientation of the two rings could influence the amount of orbital interaction and so the excited

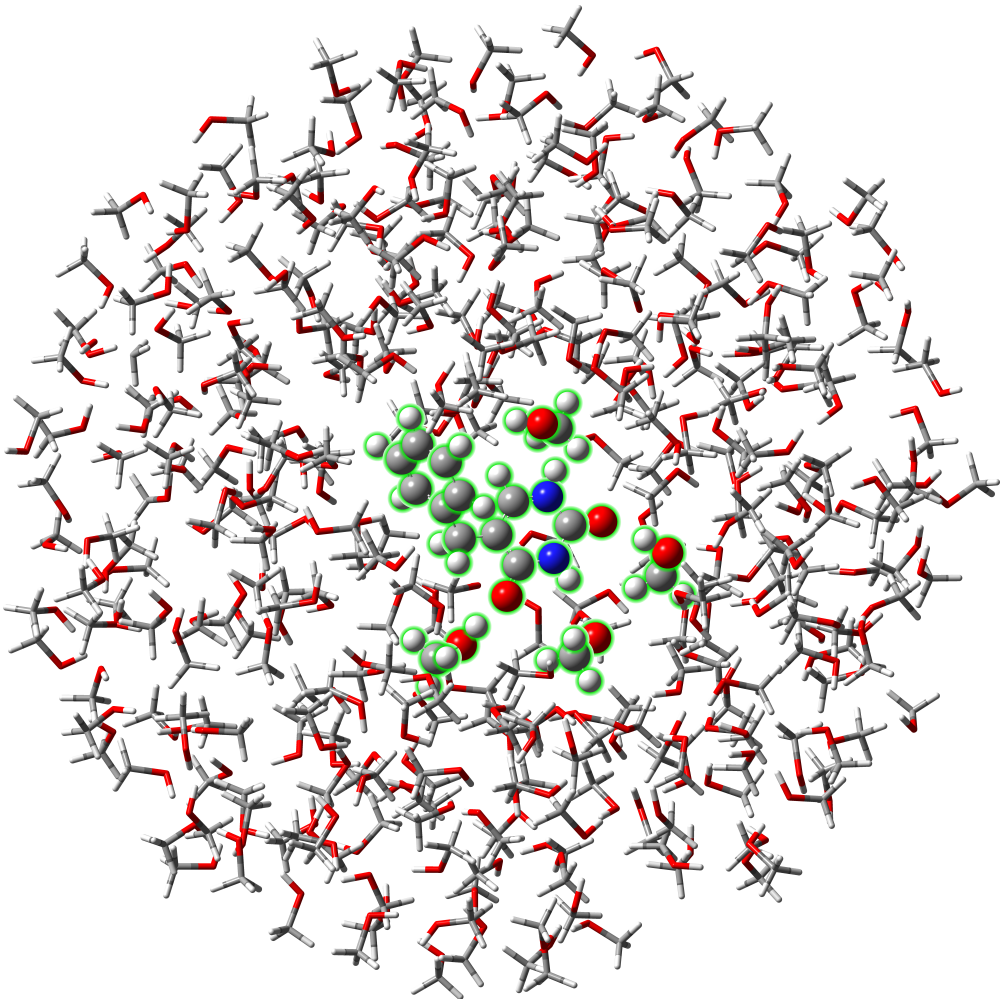


Fig. 1.2: The multi-layer scheme employed in excited states calculations: the DFT layer (green-highlighted balls and sticks) includes the 5BU solute and four proximal methanol molecules, while the MM layer (wireframe) the remaining explicit solvent.[28]

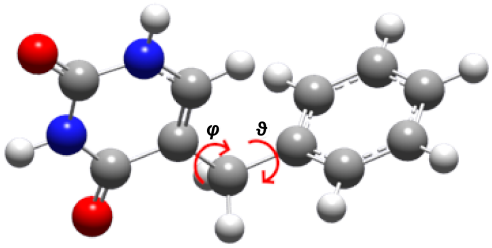


Fig. 1.3: The two rotation described by the dihedrals ϑ and φ.[28]

states character. Moreover, to evaluate the role of solvent effects, three different solvation schemes were employed: first, we simulated the isolated molecule; then, we added the effect of the methanol bulk solvent by an implicit description (C-PCM potential); finally, we included both the implicit solvent and four explicit methanol molecules, each one interacting with a solute hydrogen-bond site. All the optimizations were performed at the B3LYP/6-31G(d,p) level of theory, also employed for AIMD simulations (see also following discussion). Structures were also optimized at CAM-B3LYP/6-31+G(d,p) level for comparison, since this density functional was previously adopted in literature for this system.[27,26] For each ground-state minimum energy structure, time-dependent calculations were performed at the TD-CAM-B3LYP/6-31+G(d,p) level, with the same solvation scheme used in the relative optimization. All calculations were performed with the Gaussian 16 software package.[126]

The minimum energy structures optimized at the B3LYP/6-31G(d,p) level are shown in fig. 1.4: the left and the right columns show the butterfly-like and the flag-like energy minima, respectively; the three rows show instead the structures optimized respectively in the gas phase (figs. 1.4a and 1.4b), in the methanol treated implicitly (figs. 1.4c and 1.4d), and in methanol described in a hybrid manner, i.e. by four methanol molecules and bulk implicit solvent (figs. 1.4e and 1.4f). The analogous structures optimized at CAM-B3LYP/6-31+G(d,p) level are shown in fig. 1.5, for each solvation scheme, in the same order.

The two dihedral angles ϑ and φ mainly represent 5BU conformation: the

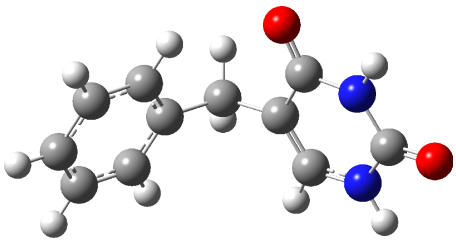
former is defined by the carbon atoms C5, C7, C1' and C2' (see fig. 1 for atom numbering) and describes the rotation of the phenyl ring, while the latter is defined by the carbon atoms C6, C5, C7 and C1' and describes the rotation of the uracil ring (see fig. 1.3). Their values for each optimized structure are reported in table 1.1.

In the following, we compare results of the structures optimization at the B3LYP/6-31G(d,p) level in different solvation schemes; we also compare them with those obtained at the CAM-B3LYP/6-31+G(d,p) level, and with those reported in literature.

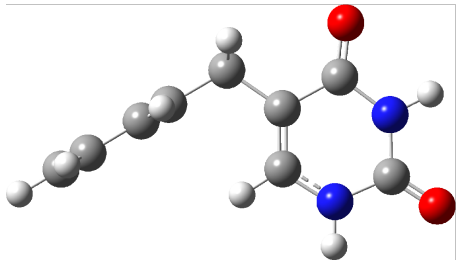
Two different ground-state energy minima were actually found for each solvation scheme, in accordance with previous works.[27] The orthogonal ($\varphi \approx 0°$ and $\vartheta \approx 90°$, table 1.1) arrangement of the two rings is retained by flag-like structures in almost all solvation models (figs. 1.4b and 1.4d) and for both levels of theory (figs. 1.5b, 1.5d and 1.5f). A slight rotation of the uracil ring ($\varphi = -21.57°$) is observed for B3LYP/6-31G(d,p) optimized F conformer in methanol cluster (fig. 1.4f).

Gas-phase butterfly-like minimum energy structures (figs. 1.4a and 1.5a have instead a symmetric arrangement of the two rings, with both dihedral angles at around 90° (table 1.1). This is consistent with gas-phase structures optimized at different levels of theory (such as M06/cc-pVTZ and SA-CASSCF(20,14)/DZP)[29,30] proposed in recent literature. In implicit methanol, this symmetric arrangement is indeed preserved at the B3LYP/6-31G(d,p) level (fig. 1.4c). However, the addition of explicit solvent molecules induces a slight phenyl ring rotation ($\vartheta = -60.69°$, see fig. 1.4e), confirmed also by the CAM-B3LYP/6-31+G(d,p) optimized B conformer with the same solvation scheme ($\vartheta = -56.96°$, see fig. 1.5e). The CAM-B3LYP/6-31+G(d,p) minimum in implicit methanol also presents a similar rotation ($\vartheta = -46.8°$, see fig. 1.5c), likely due to solvent-induced anisotropic polarization effects; this is consistent with geometry optimizations from literature.[26]

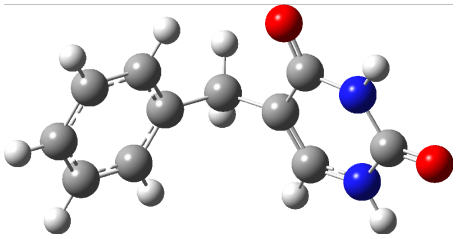
We now discuss the main features of 5BU lowest energy excited states in solution, computed at TD-CAM-B3LYP/6-31+G(d,p) level from B3LYP/6-31G(d,p) minimum energy structures. Their vertical excitation energies (VEE)



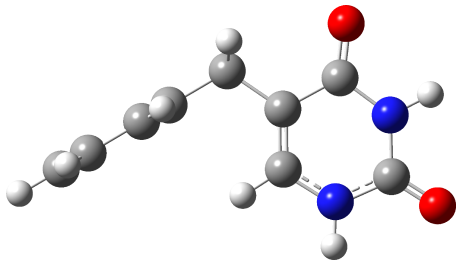
(a) B energy minimum in gas.



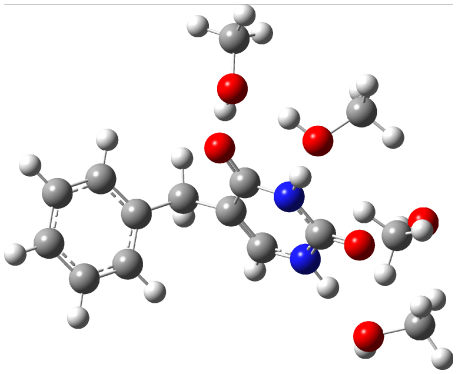
(b) F energy minimum in gas.



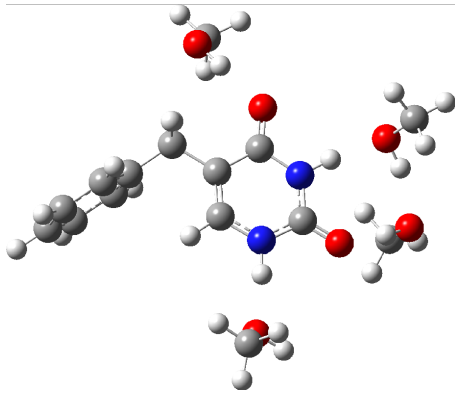
(c) B energy minimum in implicit methanol.



(d) F energy minimum in implicit methanol.

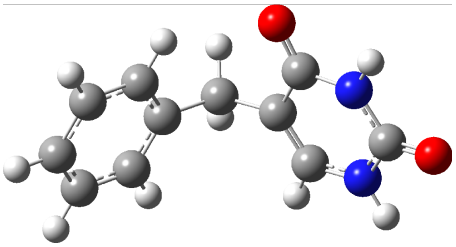


(e) B energy minimum in methanol cluster.

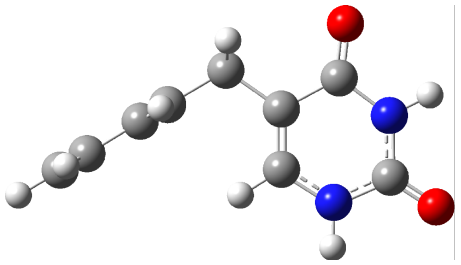


(f) F energy minimum in methanol cluster.

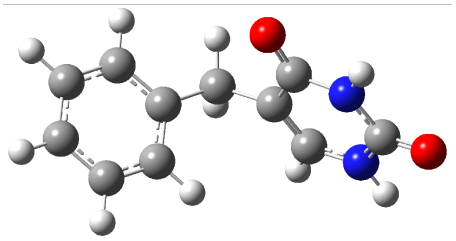
Fig. 1.4: 5BU conformers optimized at the B3LYP/6-31G(d,p) level.[28]



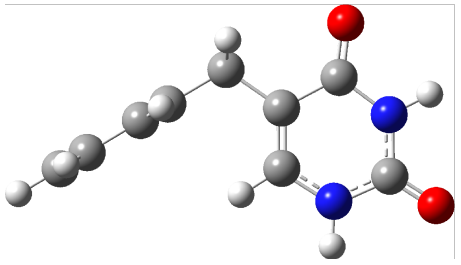
(a) B energy minimum in gas.



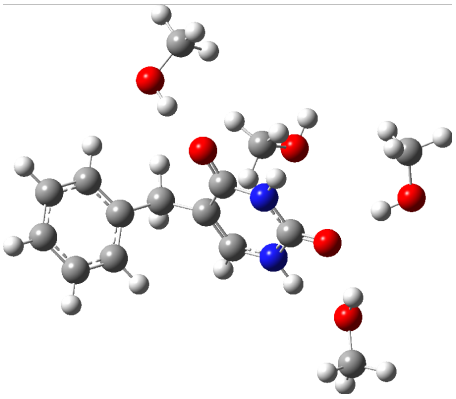
(b) F energy minimum in gas.



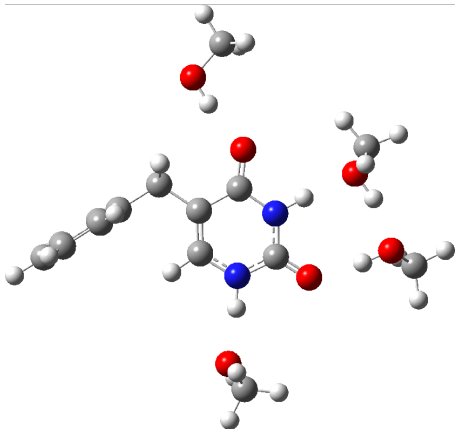
(c) B energy minimum in implicit methanol.



(d) F energy minimum in implicit methanol.



(e) B energy minimum in methanol cluster.



(f) F energy minimum in methanol cluster.

Fig. 1.5: 5BU conformers optimized at the CAM-B3LYP/6-31+G(d,p) level. [28]

Table 1.1: Values of the dihedral angles ϑ and φ (degrees) of 5BU minimum energy structures, optimized at the B3LYP/6-31G(d,p) (upper table) or CAM-B3LYP/6-31+G(d,p) (lower table) level, for B (first two columns) and F (last two columns) conformers, in gas phase (first row), implicit methanol (second row) and 5BU-methanol cluster (third row)

B3LYP/6-31G(d,p)	B minimum		F minimum	
	φ	ϑ	φ	ϑ
5BU	93.75	−102.98	0.02	−89.58
5BU/C-PCM	96.63	−95.78	0.01	−89.80
5BU+4MeOH/C-PCM	107.75	−60.69	−21.57	−79.94
CAM-B3LYP/6-31+G(d,p)	B minimum		F minimum	
	φ	ϑ	φ	ϑ
5BU	95.05	−100.68	0.03	−89.66
5BU/C-PCM	108.96	−46.8	0.01	−89.89
5BU+4MeOH/C-PCM	106.46	−56.96	0.18	−90.03

and oscillator strengths are reported in table 1.2, while the natural transition orbital (NTO) pairs [127] describing the first two excitations are displayed in appendix 1.B. Excited states from CAM-B3LYP/6-31+G(d,p) optimized structures were also computed for further comparison and confirmation, and reported in the appendix (table 1.7).

For each conformer optimized in gas phase, S_2 is the only low-lying bright state ($f = 0.155$ and $f = 0.189$ in F and B conformers, respectively, see table 1.2) and is dominated by a uracil $U\pi\pi^*$ transition. In the B conformer, such transition is mixed with a charge-transfer transition from a phenyl ring π orbital to a uracil π^* orbital ($P\pi U\pi^*$) (see fig. 1.14). In contrast, for both conformers S_1 has a dark uracil $n\pi^*$ character, where the n orbital is localized on the oxygen #4 (fig. 1.13, please refer to fig. 1 for atoms labels). Interestingly, the inclusion of solvent effects (either implicitly or through cluster models) increases the energy of the dark $Un\pi^*$ transition and stabilizes the bright $U\pi\pi^*$ state. Therefore, in methanol solution the S_1 state has a main $U\pi\pi^*$ character (figs. 1.15 and 1.17) and becomes the bright state (table 1.2). Conversely, the S_2 state acquires a dark $Un\pi^*$ character (including an implicit solvent model, fig. 1.16), or a dark $P\pi\pi^*$ one with a minimal $Un\pi^*$ contribution in a cluster/implicit solvent model (fig. 1.18), while the $Un\pi^*$ character (S_1 in gas phase) is shown

Table 1.2: VEE (eV) and oscillator strengths of the first four electronic transitions of 5BU, calculated at TD-CAM-B3LYP/6-31+G(d,p) level, for the two structures F (top table) and B (bottom table) optimized in the ground state at the B3LYP/6-31G(d,p) level, in gas phase (left column), implicit methanol (central column) and cluster (right column)

F structure			
Transition		V.E.E. (f)	
	5BU	5BU/C-PCM	5BU+4MeOH/C-PCM
$S_1 \leftarrow S_0$	5.12 (0.000)	5.16 (0.218)	5.04 (0.226)
$S_2 \leftarrow S_0$	5.26 (0.155)	5.31 (0.000)	5.39 (0.005)
$S_3 \leftarrow S_0$	5.41 (0.000)	5.42 (0.000)	5.45 (0.003)
$S_4 \leftarrow S_0$	5.99 (0.000)	6.04 (0.024)	5.98 (0.010)

B structure			
Transition		V.E.E. (f)	
	5BU	5BU/C-PCM	5BU+4MeOH/C-PCM
$S_1 \leftarrow S_0$	5.10 (0.004)	5.09 (0.265)	5.04 (0.294)
$S_2 \leftarrow S_0$	5.14 (0.189)	5.28 (0.002)	5.38 (0.001)
$S_3 \leftarrow S_0$	5.38 (0.012)	5.39 (0.003)	5.43 (0.003)
$S_4 \leftarrow S_0$	5.74 (0.020)	5.84 (0.056)	5.87 (0.066)

instead by the S_3 state (NTOs not shown).

In particular, specific solvation effects (included in the cluster/implicit solvent model) induce a relevant increase (0.24 eV for the B conformer) of the adiabatic S_2 state energy, while S_1 energy is slightly reduced. This results in an increase of the S_2-S_1 energy gap, as well as a decrease of the S_3-S_2 gap, both for the B3LYP/6-31G(d,p) and the CAM-B3LYP/6-31+G(d,p) optimized structures.

For each solvation scheme, we also observe that the transitions from the B minimum are generally brighter and red-shifted with respect the corresponding ones from the F minimum; the same conformational effect is observed for the CAM-B3LYP/6-31+G(d,p) structures.

Comparing transitions calculated on structures optimized at B3LYP/6-31G(d,p) and CAM-B3LYP/6-31+G(d,p) levels, we observe overall very similar features of the excited states, both in terms of energy and brightness. The former level of theory was also employed in ADMP/ONIOM/NPBCs ab initio molecular dynamics to sample the conformation space for following excited states and spectrum simulations (section 1.2.3).

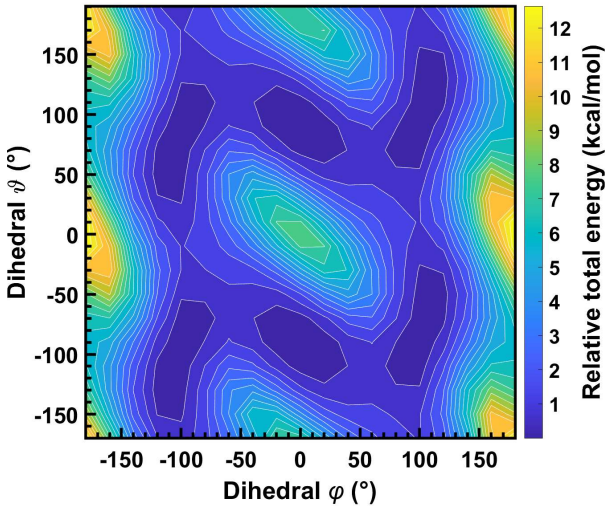


Fig. 1.6: Color map of 5BU total energy ($\mathrm{kcal\,mol^{-1}}$) at B3LYP/6-31G(d,p)/C-PCM level, as a function of the dihedrals ϑ (vertical axis) and φ (horizontal axis) (degrees), relative to the scan minimum energy. Energy is reported through a color scale.[28]

1.2.2 Conformational and solvation equilibrium

In order to better characterize and systematically explore 5BU conformational space and, in particular, other possible minimum energy structures, an energy scan in methanol with respect to ϑ and φ dihedral angles (defined in section 1.2.1) was performed. In particular, the energy was evaluated in the whole domain $\vartheta, \varphi \in [-180° : 180°]$ at B3LYP/6-31G(d,p)/C-PCM level of theory, starting the scan from the $(-89.8, 0.01)$ optimized F conformer.

The total energy of the system, relative to the scan minimum energy (in the following referred to as "R.T.E", for "relative total energy"), is reported in fig. 1.6 as a function of the dihedrals ϑ and φ. Because of the C_2 symmetry of the phenyl ring, the scan can be divided into two congruent halves: the upper one, where $\vartheta > 0$, and the lower one, where $\vartheta < 0$; we can thus restrict our analysis of the scan to the region $-180° < \vartheta < 0°$. Three wells (R.T.E. $< 1\,\mathrm{kcal\,mol^{-1}}$) can be recognized: (i) the well on the left, hereafter called "L", lies approximately between $-160° < \vartheta < -50°$ and $-120° < \varphi < -70°$ and its lowest point has $\vartheta = -89.8°$, $\varphi = -100.0°$ and R.T.E. $= 0.118\,\mathrm{kcal\,mol^{-1}}$;

(ii) the well at the center, hereafter named "C", ranges approximately between $-130° < \vartheta < -60°$ and $-30° < \varphi < 50°$ and its lowest point has $\vartheta = -89.8°$, $\varphi = 0.0°$ and R.T.E. $= 0\,\mathrm{kcal\,mol^{-1}}$ (the absolute minimum of the scan); (iii) the well on the right, named "R", lies approximately between $-130° < \vartheta < -10°$ and $70° < \varphi < 120°$ and its lowest point has $\vartheta = -89.8°$, $\varphi = 100.0°$ and R.T.E. $= 0.118\,\mathrm{kcal\,mol^{-1}}$.

Conformers belonging to L and R wells are actually mirror images (enantiomers). Thus, the energy scan shows that 5BU in methanol has just two distinct conformers: the flag-like minimum energy structure (fig. 1.4d), corresponding to the C well, and the butterfly-like minimum energy structure (fig. 1.4c), located in the R (or L) well, both previously characterized in section 1.2.1.

At the same time, the saddle point between the C and R wells ($\vartheta = -109.8°$, $\varphi = 60°$) has a R.T.E. $= 0.961\,\mathrm{kcal\,mol^{-1}}$, suggesting a low activation barrier for the interconversion of the two conformers: the ground-state equilibrium of 5BU appears to not be dominated by rigid structures, rather by a plethora of energetically close configurations: a dynamical analysis is thus necessary to infer a complete description of the excited states, overcoming the bias associated to the study of few minimum energy structures.

A conformational analysis of the solute from ADMP/ONIOM/ NPBCs AIMD was performed to understand the 5BU conformations actually visited at the equilibrium. The (ϑ, φ) two-dimensional distribution from the AIMD is plotted in fig. 1.7. We clearly distinguish two regions: one on the left, approximately located at $-130° < \vartheta < -30°$, $-60° < \varphi < 50°$, the other one on the right, approximately at $-170° < \vartheta < -50°$, $50° < \varphi < 130°$. These two regions correspond, respectively, to the aforementioned C and R wells of the energy scan (fig. 1.6) and so to the population of the flag-like and the butterfly-like conformers. Therefore, a sufficient sampling of both 5BU distinct conformers in methanol at room temperature is provided by the collected trajectory, as well as structural fluctuations within each conformer.

Solvent structuring and 5BU microsolvation was then studied, through radial distribution functions (RDFs) between solvent oxygen atoms and 5BU solvation sites, i.e. uracil O2, O4 (as hydrogen bond acceptors) and N1, N3 (as hydrogen

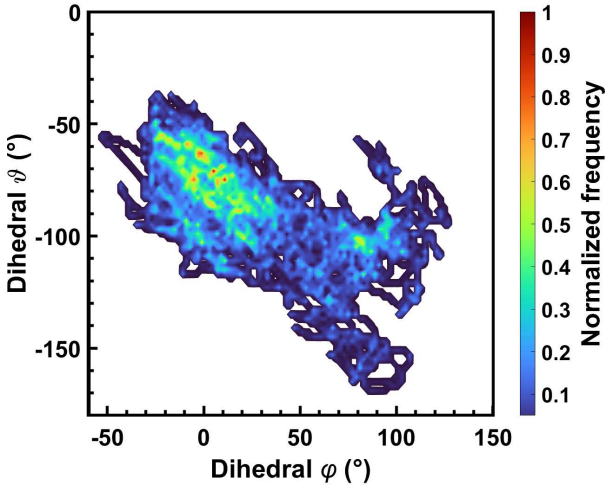


Fig. 1.7: (ϑ, φ) two-dimensional dihedrals distribution from 5BU AIMD. The color map indicates the number of AIMD frames within $1° \times 1°$ resolution, normalized with respect to the maximum. Hotter colours are used for more populated (ϑ, φ) values and white colour for less than 0.01 frequencies.[28]

bond donors, please see fig. 1.3 for atoms numbering). Clear 5BU - methanol specific interactions can be recognized. Looking at O2 RDF (fig. 1.8a), the first peak at 2.75 Å suggests a strong interaction by 1 - 2 methanol molecules (according to $g(r)$ radial integral) in the first solvation shell. The second peak at 3.55 Å is instead actually due to the methanol molecules solvating the other three heteroatoms: $g(r)$ integral (between 2 and 3) is compatible with the coordination numbers of N1, N3 and O4. O2 second solvation shell is represented by the peak at 7.25 Å, including, on average, about 14 solvent molecules. A third shell at 11.45 Å is moreover recognized. While O4 first solvation shell is located at the same distance than that of O2 (2.75 Å, fig. 1.8b), it is instead less populated (between 0 and 1 methanol molecules). This effect can be explained by the steric hindrance of the benzyl group, closer to O4. Moreover, a reduced exchange between O4 first and outer shells is observed, as suggested by $g(r)$ minimum values between the peaks (0.338 for O4, 0.797 for O2), likely due to the lower exposure to the solvent. The two low peaks at 3.75 and 4.75 Å actually reflect O2, N1 and N3 solvation. O4 second solvation shell (at 6.75 Å)

includes, on average, 9 - 10 solvent molecules. The third shell is located instead at 11.55 Å. N1 and N3 first solvation shells are at the same average distance (2.85 Å, figs. 1.8c and 1.8d, respectively) and show a comparable population (1 methanol molecule). Nevertheless, the first minimum in N3 RDF is much deeper and wider (0.295 and 0.580 $g(r)$ minimum values for N3 and N1, respectively). This suggests that the number of solvent molecules moving between the first two shells of N3 is, on average, less than that of N1. This is likely due to the position of N3 between the two oxygen atoms, where the interplay of three solvation sites causes a stronger local structuring of the solvent. The intense second peaks at 4.95 Å include the methanol molecules solvating O2 (N1 RDF) or O2 and O4 (N3 RDF). N3 second solvation shell is represented by the peak at 6.65 Å (fig. 1.8d) and includes, on average, about 18 methanol molecules. N1 RDF in the same region (fig. 1.8c) appears instead already converged to 1, suggesting a lower structuration.

In addition to the average information provided by the RDFs, we further investigated 5BU heteroatoms microsolvation to observe the occurrence of strong hydrogen bonds. A structural criterion was employed (based upon a donor-acceptor distance ≤ 4 Å and hydrogen-donor-acceptor angle $\leq 40°$). For each heteroatom and for each AIMD frame, the distance from the oxygen of the closest methanol molecule having an orientation suitable for hydrogen bonding (i.e., verifying the angle criterion) was evaluated. Distributions of such distance from the AIMD for each heteroatom are shown in fig. 1.9. These distributions reveal a mobile solvation equilibrium, exploring both fully solvated configurations and partially solvated ones. Indeed, we can consider a heteroatom unsolvated when the closest protic molecule with an adequate angle is too distant (> 4 Å). Therefore, O4 and N1 appear unsolvated in a significant fraction of the configurations explored by the AIMD, since their distributions (figs. 1.9b and 1.9c, respectively) show a significant population of > 4 Å distances.

According to the previous discussion, the collected 5BU ADMP/ONIOM/NPBCs AIMD simulation provided therefore a complete picture of both conformational and solvation equilibria in methanol at room temperature. Such ensemble allowed a detailed prediction of 5BU excited states and absorption spectra in such

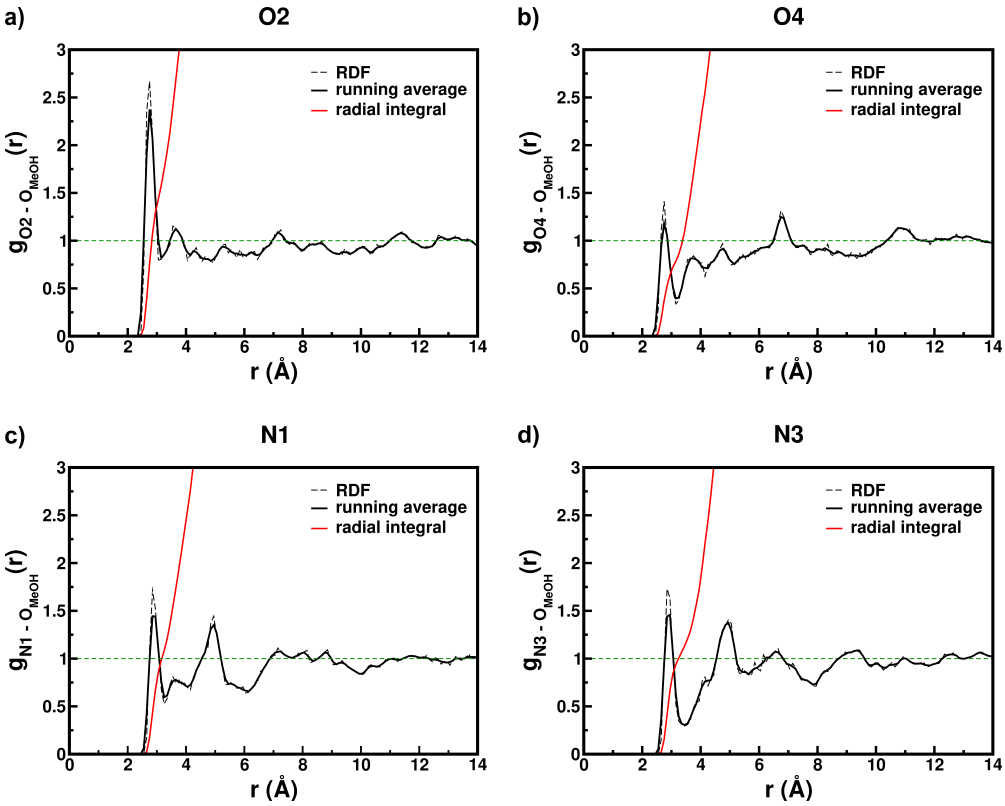


Fig. 1.8: RDFs (dashed black) of methanol oxygen atoms - 5BU heteroatoms distances (please refer to fig. 1.3 for atoms labels). The solid black line represents the 3-point $g(r)$ running average and the red line its radial integral; the green dashed straight line is set to 1, the value of $g(r)$ for a homogeneous distribution.[28]

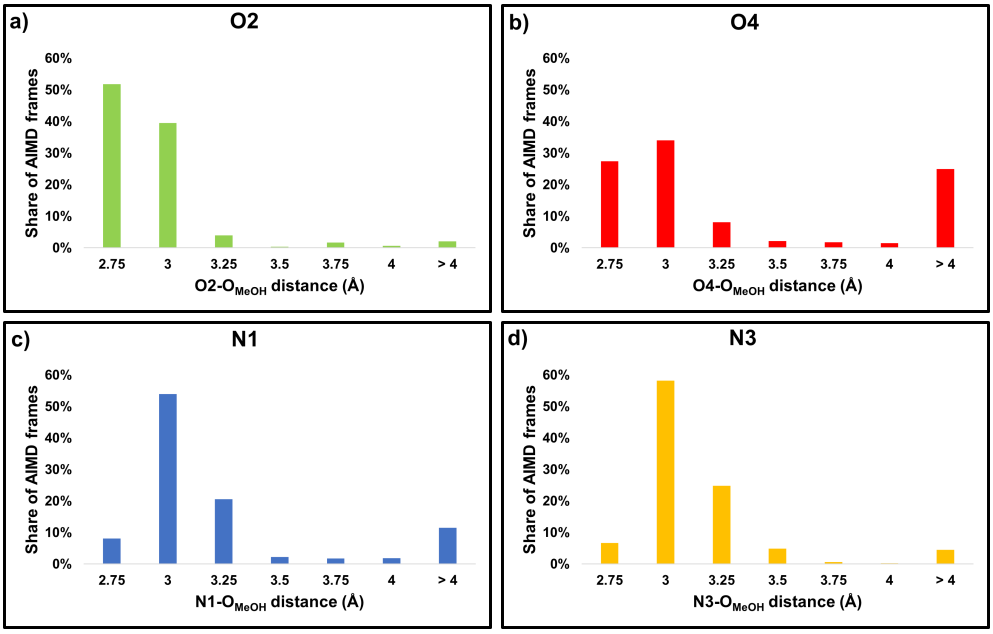


Fig. 1.9: Distributions of the distance of each 5BU heteroatom from the closest solvent oxygen verifying H_{MeOH}-O_{MeOH}-O or H-N-O_{MeOH} $< 40°$. Such criterion allows to select solvent molecules potentially able to participate in hydrogen bonds. Please refer to fig. 1.3 for atoms labels.[28]

conditions, as shown in the following section (section 1.2.3).

1.2.3 Optical absorption spectrum from AIMD sampling

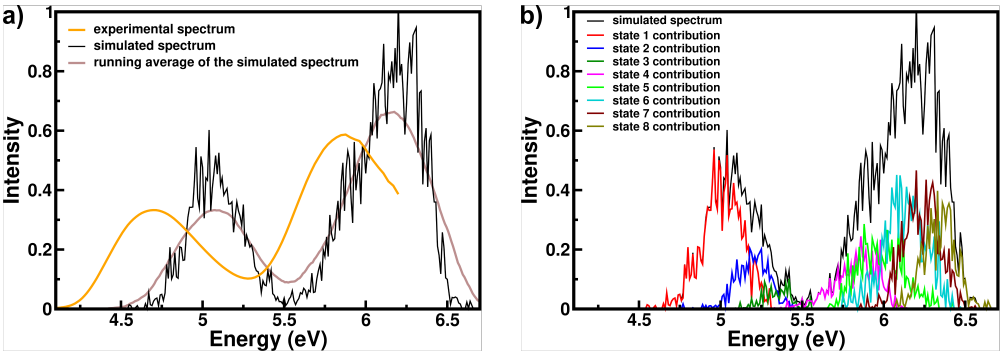


Fig. 1.10: TD-CAM-B3LYP/6-31+G(d,p) UV-Vis absorption spectrum of 5BU in methanol (black line). In plot a), its 50-points running average (brown line) and the experimental spectrum (orange) are also reported; in plot b), the contributions of the $S_1 \leftarrow S_0$ (red), $S_2 \leftarrow S_0$ (blue), $S_3 \leftarrow S_0$ (dark green), $S_4 \leftarrow S_0$ (magenta), $S_5 \leftarrow S_0$ (light green), $S_6 \leftarrow S_0$ (turquoise), $S_7 \leftarrow S_0$ (maroon) and $S_8 \leftarrow S_0$ (olive) transitions to the total spectrum are instead shown. The experimental spectrum is expressed as absorbance, while the simulated spectrum, its running average and states contributions are rescaled so that the (smoothed) calculated first band and the experimental one are comparable.[28]

The experimental UV-Vis absorption spectrum of 5BU in methanol was recorded by our collaborators Carlo Altucci and Mohammadhassan Valadan.[28] It exhibits two distinctive absorption bands, located at approximately 210 and 265 nm, corresponding to 5.90 and 4.68 eV, respectively. Our interest was to disentangle the main electronic transitions and the structural features determining the observed absorption. To this aim, we simulated the optical absorption spectrum of 5BU in methanol, by performing TD-DFT calculations on an ensemble of configurations extracted from the AIMD, according to the protocol detailed in section 1.1.3. As shown in fig. 1.10a, the band shape of the experimental spectrum is reproduced with a systematic blue shift, suggesting that the energy spacing among the excited states is correctly estimated. By recognizing the individual contributions of the different states, we attribute the lowest-energy band, which is the one excited in photo-cyclization,[26,27] mainly to the transi-

Table 1.3: S_1 (top table) and S_2 (down) excited states analysis at TD-CAM-B3LYP/6-31+G(d,p) level for five representative configurations sampled from 5BU QM/MM MD in methanol solution (see appendix 1.C). Column 1: 5BU conformer and solvation state of O4 and N1 heteroatoms (please refer to section 1.2.2 for the definition of the B and F conformers and to fig. 1 for atoms numbering). Column 2: S_1 and S_2 VEE (eV) and oscillator strength. Column 3: S_1 and S_2 characters according to hole-electron NTO pairs.

Configuration	$S_1 \leftarrow S_0$	
	V.E.E. (f)	Character
α (B, H-bonded O4 and N1)	4.92 (0.217)	$U\pi\pi^*$
β (F, H-bonded O4 and N1)	5.05 (0.179)	$U\pi\pi^*$
γ (B, H-bonded N1 but not O4)	4.83 (0.158)	$U\pi\pi^* + P\pi U\pi^*$
δ (F, H-bonded N1 but not O4)	5.12 (0.170)	$U\pi\pi^*$
ϵ (F, neither O4 or N1 H-bonded)	5.03 (0.024)	$Un\pi^*$

Configuration	$S_2 \leftarrow S_0$	
	V.E.E. (f)	Character
α (B, H-bonded O4 and N1)	5.25 (0.019)	$Un\pi^* + P\pi U\pi^*$
β (F, H-bonded O4 and N1)	5.27 (0.017)	$Un\pi^*$
γ (B, H-bonded N1 but not O4)	5.13 (0.054)	$Un\pi^*$
δ (F, H-bonded N1 but not O4)	5.30 (0.006)	$Un\pi^*$
ϵ (F, neither O4 or N1 H-bonded)	5.14 (0.102)	$U\pi\pi^*$

tion towards S_1, with also contributions from S_2 and S_3, the latter only found beyond the band maximum (see fig. 1.10b).

The different conformations and solvation states sampled by the QM/MM/NPBCs trajectory in methanol solution at room temperature (section 1.2.2) suggest that the adiabatic S_1 and S_2 lowest-energy states may actually exchange their character according to such parameters. We thus investigated the character of the first two excitations (through a NTO analysis) on a small subset of representative configurations extracted from the QM/MM/NPBCs MD, differing in the conformation and solvation state (please refer to fig. 1.19 in appendix for their structures). The features of S_1 and S_2 states calculated for such configurations are correlated to the structural parameters in table 1.3. NTO pairs isosurfaces can be instead found in figs. 1.20 to 1.24 in appendix.

Consistently with the calculations on optimized structures with either implicit or explicit solvation (section 1.2.1), S_1 is a bright $U\pi\pi^*$ state, while S_2 has a dark $Un\pi^*$ character (being the n orbital located on O4), due to specific methanol - 5BU interactions (especially on N1). Interestingly, in the ϵ config-

Table 1.4: First column: oscillator-strength-weighted AIMD-averages of the VEE of the first four transitions of 5BU in methanol. Second column: AIMD-averaged oscillator strengths of the same transitions. Second-last row: average energy of the points composing the first band of the computed spectrum in fig. 1.10, each one weighted for the respective intensity. Last row: energy of the maximum of the first band in the experimental spectrum.

Transition	V.E.E. (eV)	f
$S_1 \leftarrow S_0$	5.01	0.123
$S_2 \leftarrow S_0$	5.20	0.042
$S_3 \leftarrow S_0$	5.34	0.013
$S_4 \leftarrow S_0$	5.82	0.050
Average of the first band	5.06	
Experimental peak	4.68	

uration, where neither of the two heteroatoms O4 and N1 are solvated, S_1 and S_2 characters are swapped: the former is therefore a dark $Un\pi^*$ state, while the latter has instead a bright $U\pi\pi^*$ character (see fig. 1.24). We also observe that S_1 and S_2 of the butterfly-like representative configurations (α and γ) show also a small $P\pi U\pi^*$ charge-transfer character, which is absent in the corresponding transitions of flag-like configurations (see figs. 1.20 and 1.21).

We then compared the average values of energy and oscillator strength of the first four states from the AIMD (see table 1.4) with the values calculated from the minimum energy structures optimized in methanol cluster at the B3LYP/6-31G(d,p) level (i.e., the same level of the AIMD, see the third column of table 1.2). This comparison revealed consistent energies and spacing, but fairly different oscillator strength values: S_1 is the brightest state from both static and dynamics calculations, but its oscillator strength calculated from AIMD frames is on average only 0.123, compared with the values of 0.294 and 0.226 from minimum energy structures. S_2 and S_3, which from static calculations appeared to be completely dark states, from AIMD frames have instead an average oscillator strength of 0.042 and 0.013, respectively. S_2 state results therefore non-completely dark since it acquires in a reduced fraction of the sampled configurations a $U\pi\pi^*$ character. Besides S_1, S_2 is of particular interest since it contributes, together with S_1, to the first band, which is excited to induce 5BU photo-cyclization.[26,27]

The swap of the character the first two states depending upon 5BU confor-

mation and solvation can explain several trends in their brightness. Regarding the first excited state, the formation of a hydrogen bond between O4 and a solvent molecule appears to be an important feature. As discussed in section 1.2.2, some of the sampled configurations have at least one solvent molecule with distance and orientation suitable for a hydrogen bond with O4, while in the other ones O4 does not form hydrogen bonds (see fig. 1.9b). In the former group, the average S_1 oscillator strength is 0.146 and its distribution peaks between 0.15 and 0.18 (see the blue bars in fig. 1.11a); in the latter group, instead, the average S_1 oscillator strength is 0.069 and its distribution peaks between 0.00 and 0.03 (see the orange bars in fig. 1.11a). Analogously to O4, also N1 microsolvation (see fig. 1.9c) influences the brightness of S_1: S_1 is mostly a bright state when there is at least one solvent molecule with distance and orientation suitable for a hydrogen bond with N1, while it is darker when N1 does not form hydrogen bonds. In the former group, the average S_1 oscillator strength is 0.132 and its distribution peaks between 0.15 and 0.18 (see the green bars in fig. 1.11b); in the latter group, instead, the average S_1 oscillator strength is 0.059 and its distribution peaks between 0.00 and 0.03 (see the red bars in fig. 1.11b). Finally, the relative orientation of the uracil and phenyl rings also plays a role: when they are approximately orthogonal to each other (flag-like conformer, occupying the right half of the distribution in fig. 1.7, see as instance fig. 1.4b), S_1 oscillator strength has a bimodal distribution, with a dark peak between 0.00 and 0.03 and a bright peak between 0.15 and 0.18 (see the pink bars in fig. 1.11c). When the rings face each other (butterfly-like conformer, occupying the left half of the distribution in fig. 1.7, see as instance fig. 1.4a), the distribution dark peak disappears and S_1 is mostly a bright state, with a distribution peaked between 0.15 and 0.18 (see the brown bars in fig. 1.11c).

S_2 exhibits an opposite behaviour with respect to S_1, even if its variations are more subtle. Like S_1, indeed, its brightness mainly depends on the microsolvation of O4, but, contrarily to S_1, it is higher in the configurations where O4 does not form hydrogen bonds and decreases in those where a methanol molecule solvates O4. In the first group, S_2 oscillator strength has a bimodal distribution, with the highest peak between 0.00 and 0.03 and lower peak (with

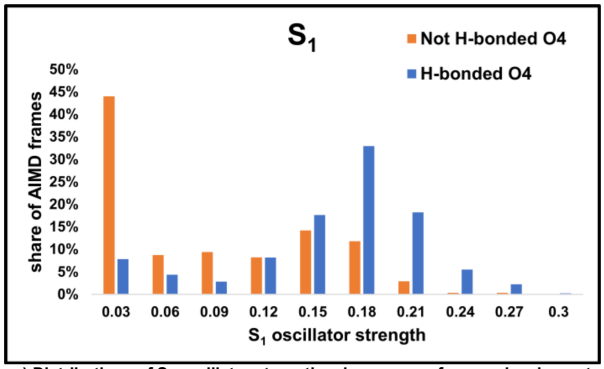
Table 1.5: Average S_1 oscillator strength for every considered cluster of AIMD frames

Cluster	Average f(S_1)	Cluster	Average f(S_1)
Not H-bonded O4	0.069	H-bonded O4	0.146
Not H-bonded N1	0.059	H-bonded N1	0.132
F conformer	0.123	B conformer	0.127

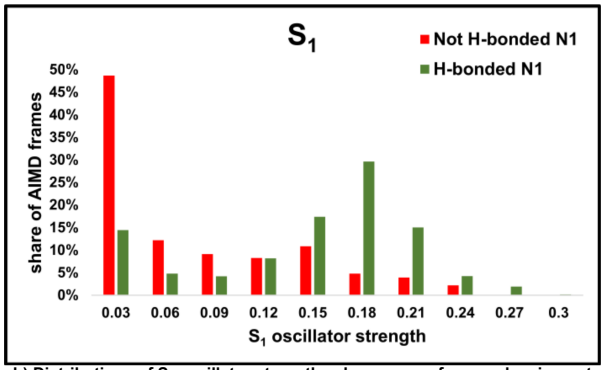
Table 1.6: Average S_2 oscillator strength for every considered cluster of AIMD frames

Cluster	Average f(S_2)	Cluster	Average f(S_2)
Not H-bonded O4	0.071	H-bonded O4	0.031
Not H-bonded N1	0.066	H-bonded N1	0.039
F conformer	0.042	B conformer	0.021

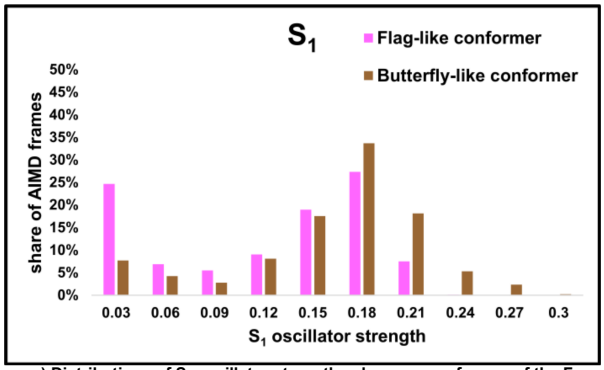
a population half of the first one) between 0.15 and 0.18. In the second group instead the distribution shows no bright peaks (see the blue bars in fig. 1.12a) and the second excited state is mostly dark, with an average oscillator strength of 0.0308. Again in contrast to the first state, the fraction of frames with a bright second excited state significantly increase when there are no nearby solvent molecules able to form a hydrogen bond with N1, while it is lower when N1 is solvated: in the first group, S_2 oscillator strength has a bimodal distribution, with the highest peak between 0.00 and 0.03 and a second one between 0.12 and 0.15, with a relative abundance half of the first one (see the red bars in fig. 1.12b). The second group does not show this bright peak in the distribution, but a mostly dark second excited state, with an average oscillator strength of 0.039 (see the green bars in fig. 1.12b). Finally, also the effect of the two rings arrangement is reversed with respect to S_1: the second excited state is brighter when the two rings are approximately orthogonal (flag-like conformer), while it is darker when they face each other (butterfly-like conformer) (see fig. 1.12c).



a) Distributions of S_1 oscillator strength values among frames showing not H-bonded O4 (orange), or among frames showing H-bonded O4 (blue).

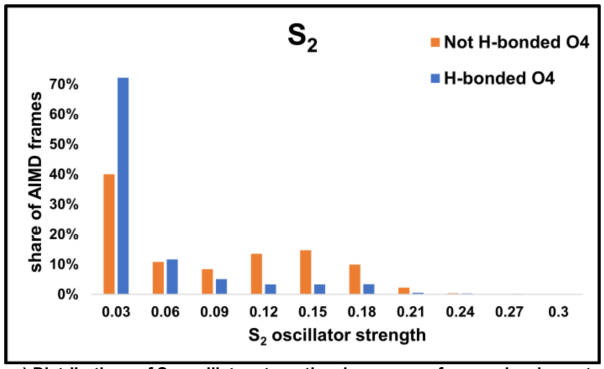


b) Distributions of S_1 oscillator strength values among frames showing not H-bonded N1 (red), or among frames showing H-bonded N1 (green).

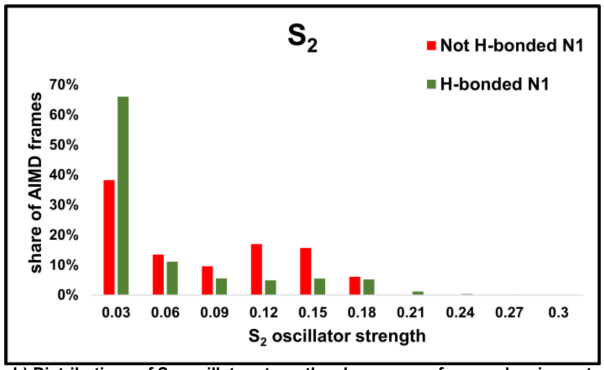


c) Distributions of S_1 oscillator strength values among frames of the F conformer (pink), or among frames of the B conformer (brown).

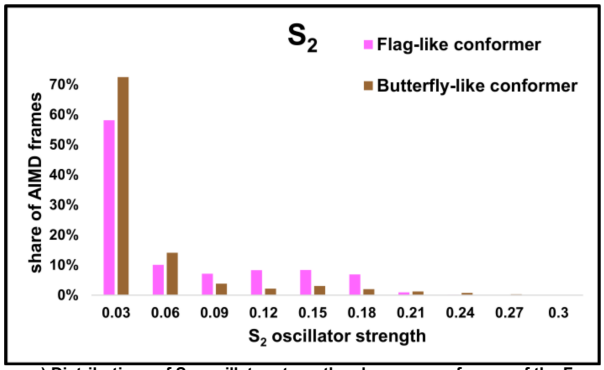
Fig. 1.11: Distributions of S_1 oscillator strength values among groups of frames taken from the AIMD.[28]



a) Distributions of S_2 oscillator strength values among frames showing not H-bonded O4 (orange), or among frames showing H-bonded O4 (blue).



b) Distributions of S_2 oscillator strength values among frames showing not H-bonded N1 (red), or among frames showing H-bonded N1 (green).



c) Distributions of S_2 oscillator strength values among frames of the F conformer (pink), or among frames of the B conformer (brown).

Fig. 1.12: Distributions of S_2 oscillator strength values among groups of frames taken from the AIMD.[28]

Appendices

1.A Optical data from CAM-B3LYP/6-31+G(d,p) minima

VEE and oscillator strengths of the four lowest energy 5BU states calculated on CAM-B3LYP/6-31+G(d,p) optimized structures (section 1.2.1) are reported in table 1.7 and discussed in section 1.2.1.

	F structure		
Transition		V.E.E. (f)	
	5BU	5BU/C-PCM	5BU+4MeOH/C-PCM
$S_1 \leftarrow S_0$	5.13 (0.000)	5.20 (0.216)	5.15 (0.241)
$S_2 \leftarrow S_0$	5.31 (0.153)	5.31 (0.000)	5.43 (0.000)
$S_3 \leftarrow S_0$	5.45 (0.000)	5.45 (0.000)	5.49 (0.000)
$S_4 \leftarrow S_0$	6.00 (0.000)	6.07 (0.020)	6.07 (0.024)
	B structure		
Transition		V.E.E. (f)	
	5BU	5BU/C-PCM	5BU+4MeOH/C-PCM
$S_1 \leftarrow S_0$	5.12 (0.002)	5.18 (0.294)	5.13 (0.306)
$S_2 \leftarrow S_0$	5.19 (0.193)	5.26 (0.002)	5.35 (0.001)
$S_3 \leftarrow S_0$	5.41 (0.010)	5.41 (0.003)	5.42 (0.004)
$S_4 \leftarrow S_0$	5.77 (0.022)	5.93 (0.045)	5.85 (0.048)

Table 1.7: VEE (eV) and oscillator strengths of the first four transitions of 5BU in gas phase (left), implicit methanol (center) and cluster (right), calculated at TD-CAM-B3LYP/6-31+G(d,p) level for the two structures F (top) and B (bottom), optimized at CAM-B3LYP/6-31+G(d,p) level.

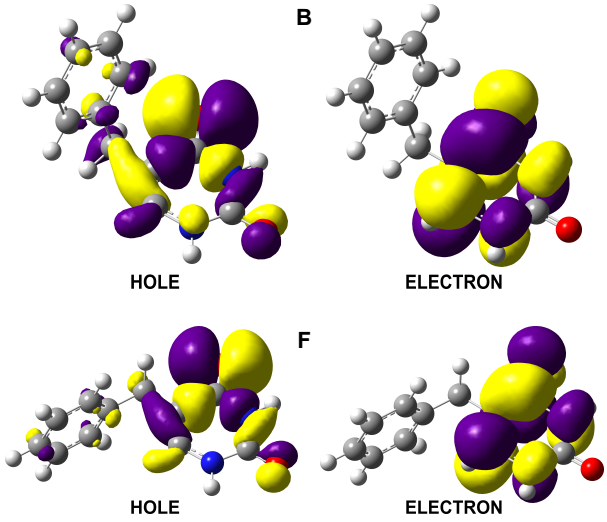


Fig. 1.13: Hole and electron isodensity plots (isovalue = 0.02) of S_1, calculated in gas phase (TD-CAM-B3LYP/6-31+G(d,p) level) for B (up) and F (down panel) structures. [28]

1.B NTO analysis of transitions from minimum energy structures

We performed a Natural Transition Orbitals (NTO)[127] analysis of the first two 5BU excitations, computed (at TD-CAM-B3LYP/6-31+G(d,p) level on B3LYP/6-31G(d,p) structures) for both the B and F conformers and for each solvation scheme. Such excitations can be described in terms of the most occupied hole-electron pairs of orbitals, discussed in section 1.2.1 and here shown for B and F structures, in gas phase (figs. 1.13 and 1.14), implicitly treated methanol (figs. 1.15 and 1.16) and methanol cluster (figs. 1.17 and 1.18).

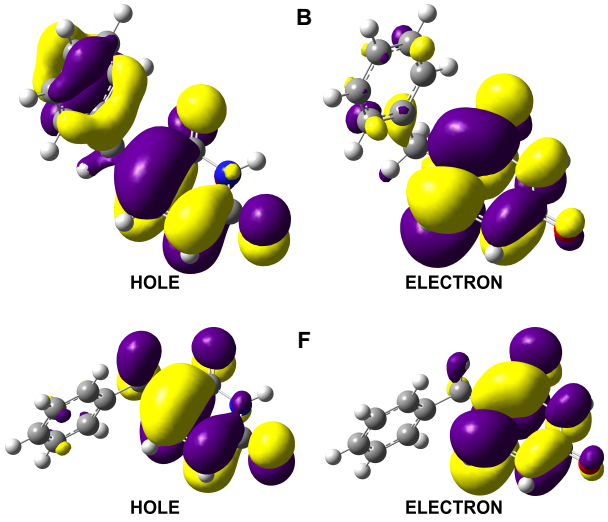


Fig. 1.14: Hole and electron isodensity plots (isovalue = 0.02) of S_2, calculated in gas phase (TD-CAM-B3LYP/6-31+G(d,p) level) for B (up) and F (down panel) structures.[28]

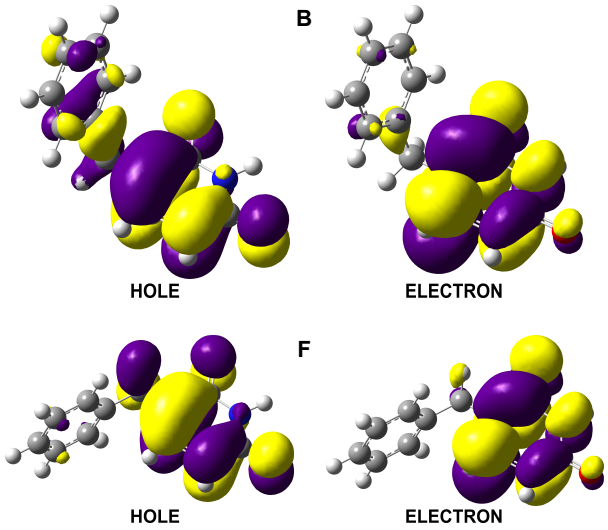


Fig. 1.15: Hole and electron isodensity plots (isovalue = 0.02) of S_1, calculated in implicit methanol (TD-CAM-B3LYP/6-31+G(d,p)/C-PCM level), for B (up) and F (down panel) structures.[28]

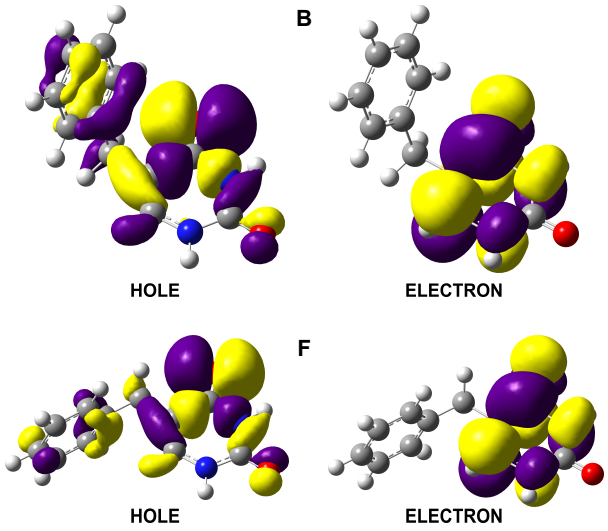


Fig. 1.16: Hole and electron isodensity plots (isovalue = 0.02) of S_2, calculated in implicit methanol (TD-CAM-B3LYP/6-31+G(d,p)/C-PCM level), for B (up) and F (down panel) structures. [28]

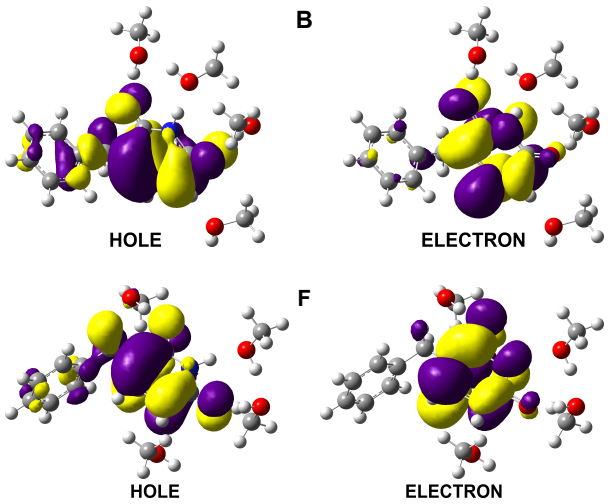


Fig. 1.17: Hole and electron isodensity plots (isovalue = 0.02) of S_1, calculated in methanol cluster (TD-CAM-B3LYP/6-31+G(d,p)/C-PCM level), for B (up) and F (down panel) structures. [28]

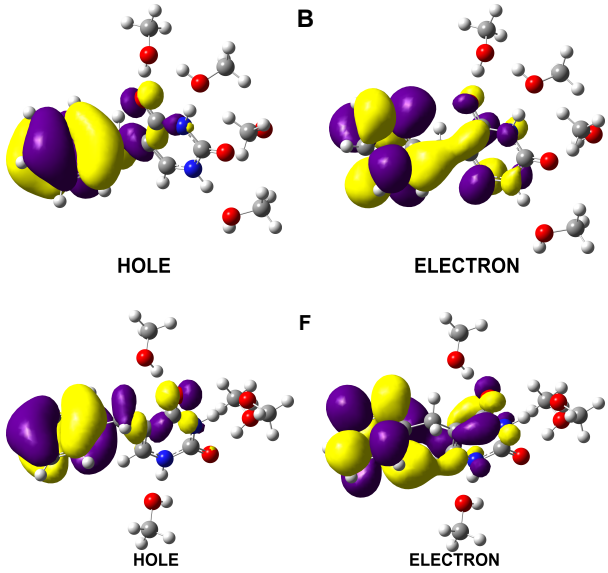
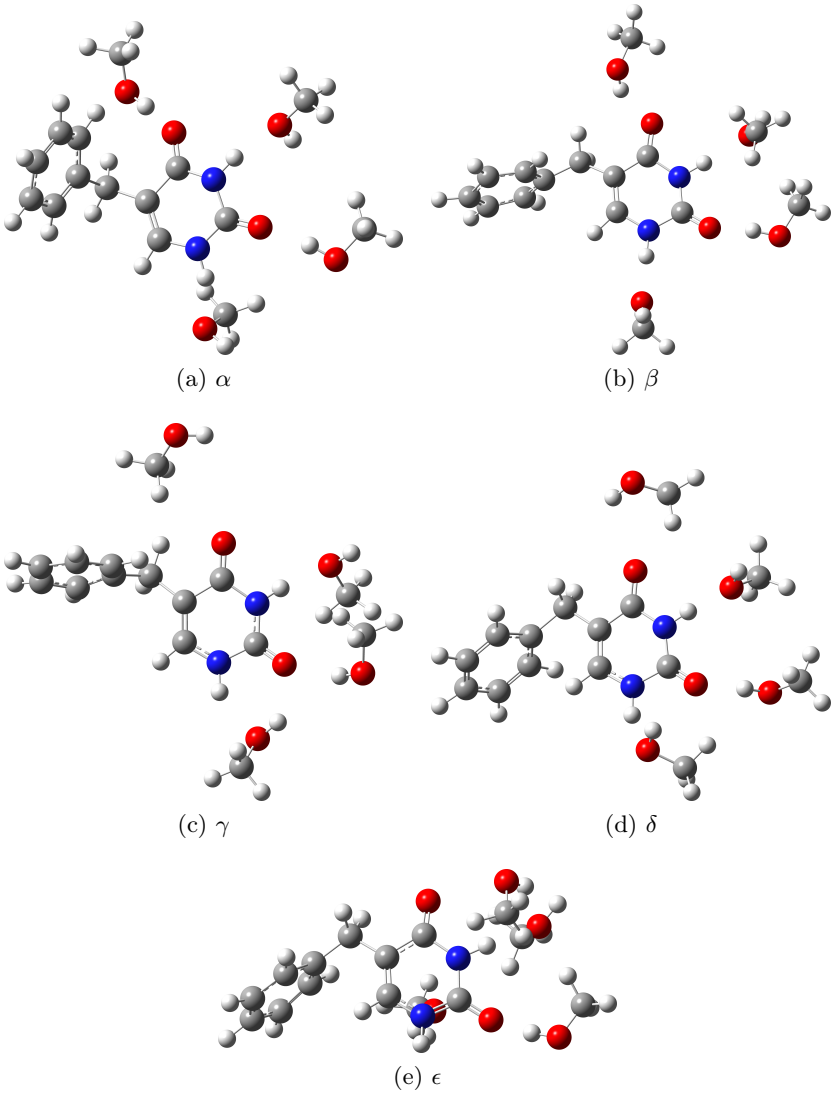


Fig. 1.18: Hole and electron isodensity plots (isovalue = 0.02) of S_2, calculated in methanol cluster (TD-CAM-B3LYP/6-31+G(d,p)/C-PCM level), for B (up) and F (down panel) structures.[28]

1.C NTO analysis of transitions from AIMD sampling

We performed a NTO analysis of the first two 5BU excitations, computed (at TD-CAM-B3LYP/6-31+G(d,p) level) on five configurations extracted from the MD trajectory (shown in fig. 1.19), representative of distinct solvation modes and conformations. The solute and four proximal methanol molecules were modeled at the aforementioned QM level, while the atomic charges of all the other solvent molecules were embedded into the hamiltonian (electronic embedding).

The character, energy and oscillator strength of the first two states are resumed in table 1.3 and discussed in section 1.2.3. Here we show the dominant NTO pairs for each computed transition of the five configurations: α (fig. 1.20), β (fig. 1.21), γ (fig. 1.22), δ (fig. 1.23) and ϵ (fig. 1.24).



(a) α (b) β (c) γ (d) δ (e) ϵ

Fig. 1.19: The five representative configurations extracted from AIMD for excited states analysis.[28]

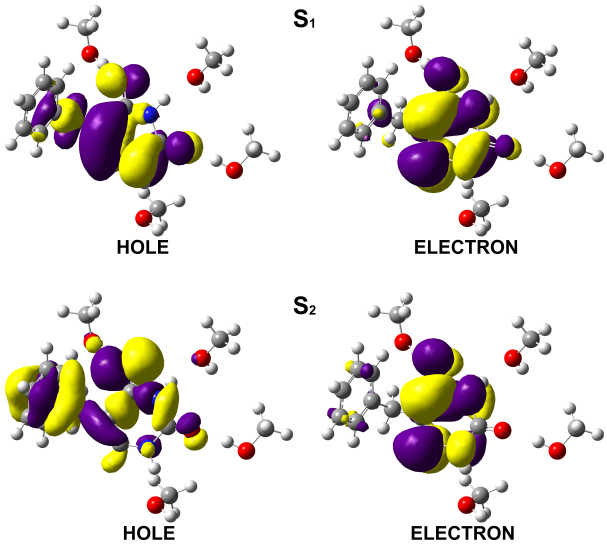


Fig. 1.20: Hole and electron isodensity plots (isovalue = 0.02) of S_1 (up) and S_2 (down), calculated for the α configuration (fig. 1.19a) extracted from the MD. [28]

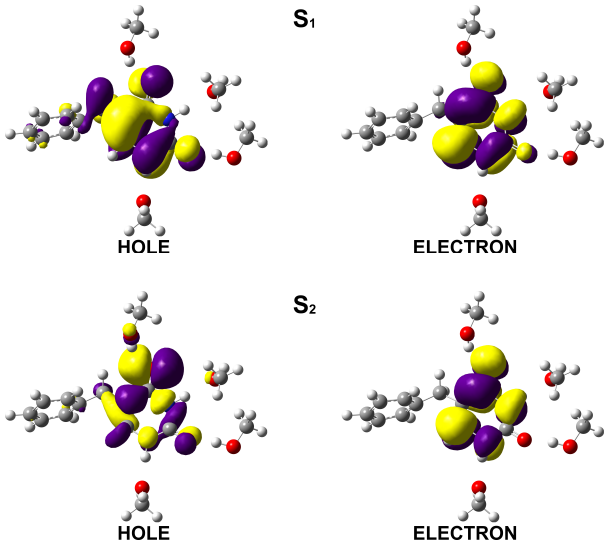


Fig. 1.21: Hole and electron isodensity plots (isovalue = 0.02) of S_1 (up) and S_2 (down), calculated for the β configuration (fig. 1.19b) extracted from the MD. [28]

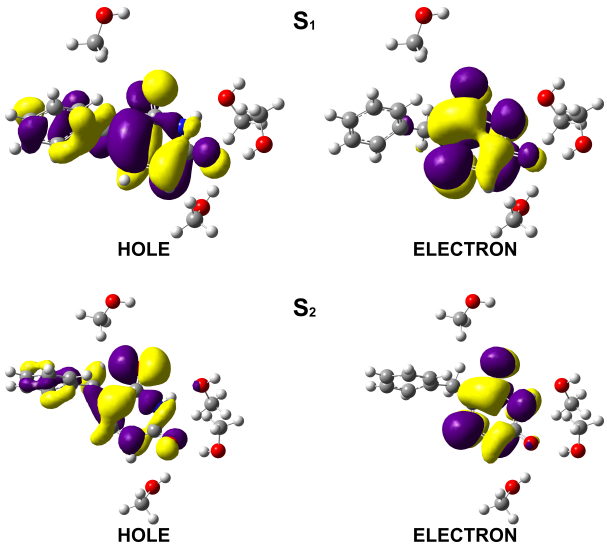


Fig. 1.22: Hole and electron isodensity plots (isovalue = 0.02) of S_1 (up) and S_2 (down), calculated for the γ configuration (fig. 1.19c) extracted from the MD. [28]

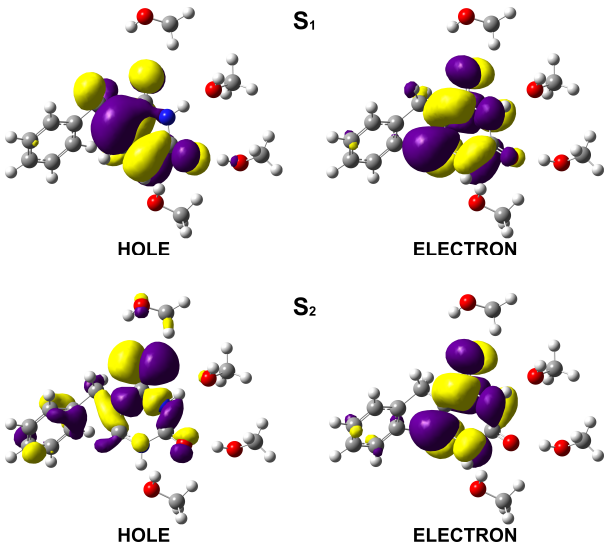


Fig. 1.23: Hole and electron isodensity plots (isovalue = 0.02) of S_1 (up) and S_2 (down), calculated for the δ configuration (fig. 1.19d) extracted from the MD. [28]

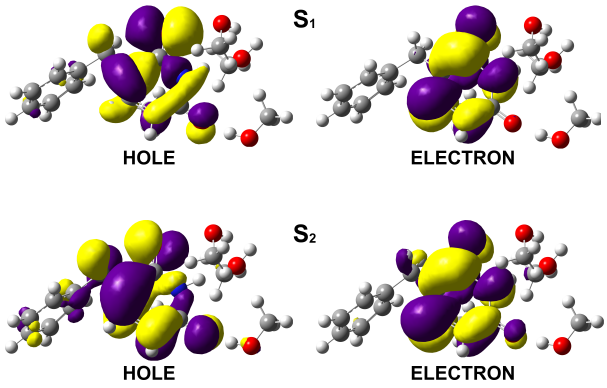


Fig. 1.24: Hole and electron isodensity plots (isovalue = 0.02) of S_1 (up) and S_2 (down), calculated for the ϵ configuration (fig. 1.19e) extracted from the MD.[28]

Chapter 2

Ultrafast Transient Fluorescence of 5-benzyluracil

2.1 Introduction

Recent progresses in computational chemistry have made feasible the reliable simulation of large, floppy molecules with strong nonadiabatic coupling, i.e. requiring to consider the coupling between nuclear and electronic motion.[128,129] 5BU is a very interesting example of such systems: as detailed in chapter 1, it has a complex electronic landscape, with three different transition which comprise the low-energy absorption band and span less than 1 eV; their brightness is heavily influenced by the orientation of the benzyl substituent and solvation effects. It is clear that a comprehensive analysis of the photoinduced dynamics of 5BU in methanol requires the use of Non-Adiabatic molecular Dynamics (NAD), in order to disentangle the effect of the different involved states.

In addition, the fluorescence of 5BU is a remarkable case study for ultrafast processes, which have fundamental importance in photochemistry and photobiology.[31,130] Indeed, Valadan *et al.*[27] recorded the transient and steady-state emission of 5BU in methanol, by the means of Fluorescence Up-Conversion (FlUC) and their results clearly show signs of a quick change in the electronic

landscape of 5BU in less than 300 fs after the excitation.

2.1.1 Fluorescence Up-Conversion

FlUC is a pump-probe technique made to record fluorescence on sub-ps intervals: the fluorescence of the sample is combined with a sub-ps impulse - named *gate pulse* - whose delay is changed at regular steps. The resulting signal is a series of sub-ps pulses called *up-converted fluorescence*: the frequency of each pulse is given by the sum of the frequency of the gate pulse and the fluorescence at the time of the delay; the intensity is given by the convolution of the sample fluorescence and gate pulse. The procedure is performed simultaneously and in parallel for each wavelength in the range of interest.[131]

The up-converted signal is typically a concave function, increasing in intensity during the excitation of the sample and then decreasing together with the decay of the sample fluorescence. The time constant(s) of the decay is computed by fitting the signal with the so-called *Instrument Response Function* (IRF), convoluted with an appropriate negative exponential function (in ref. 27, a biexponential):

$$f(t) = A_1\,\mathrm{IRF}(t) \otimes e^{-\frac{t}{\tau_1}} + A_2\,\mathrm{IRF}(t) \otimes e^{-\frac{t}{\tau_2}}$$

As a consequence, the IRF is also the limit of the time resolution in a FlUC experiment.

The time scale of a FlUC spectrum starts at the maximum of the up-converted signal - a forced choice, since the goal is to record the fluorescence decay; since the signal is recorded in parallel for each wavelength, kinetic traces computed at different wavelength can have different "time zero"; such difference is usually within the IRF.

2.1.2 Transient emission spectroscopy of 5BU

As aforementioned, Valadan *et al.*[27] recorded the transient emission spectrum of 5BU in methanol, with an IRF of 300 fs. The excitation wavelength was

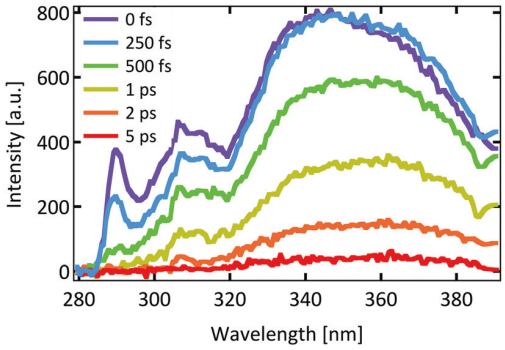


Fig. 2.1: Temporal evolution of transient emission spectra of 5BU in FlUC measurements for different time-delays between pump and gate pulse (From ref. 27).

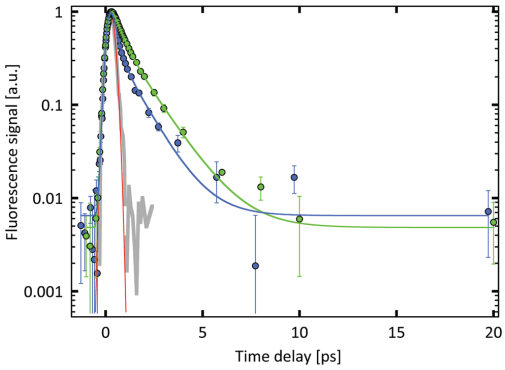


Fig. 2.2: Normalized kinetic traces recorded at 310 (blue) and 350 nm (green) of 5BU emission, compared with the IRF of the FlUC setup (grey) (From ref. 27).

266 nm, corresponding to the maximum of the low-energy band, which we have simulated and discussed in section 1.2.3.

The spectrum recorded within the IRF = 300 fs (labeled "0 fs" in fig. 2.1) is clearly indicative of a very different landscape with respect to the absorption, having two bands (at 310 and 350 nm) originating from the single absorption band at the exciting wavelength. Both of the two bands undergo then a ultrafast decay, losing more than half of their peak intensity at already 1 ps.

Valadan *et al.*[27] fitted the kinetic trace of the fluorescence decay, recorded at the maximum of each band and starting at the maximum of the FlUC signal (see fig. 2.2): using a biexponential function, they found, for each band, two

Table 2.1: Decay constants and the respective amplitudes obtained by the fit of the recorded kinetic trace for the maxima of the two bands.[27]

	310 nm	350 nm
τ_1	230 ± 40 fs (0.76)	440 ± 50 fs (0.53)
τ_2	1.2 ± 0.2 ps (0.24)	1.4 ± 0.1 ps (0.47)

significant time constants, one in the sub-ps range and the other one in the ps range (see table 2.1).

The aim of this project is thus to understand which molecular motions are responsible for the changes in the electronic landscape producing the high-energy fluorescence band in less than 300 fs and then the rapid decay of the fluorescence in less than 1 ps.

2.2 Methods

2.2.1 Non adiabatic molecular dynamics

Coupling between electronic and nuclear degrees of freedom is often an unavoidable aspect of the photoinduced evolution of chemical and biological systems; however, a full, simultaneous quantum treatment of nuclear and electronic motion is computationally unfeasible for systems of significant size.[132,133] Hence, Nonadiabatic Mixed Quantum-Classical (NA-MQC) dynamics is a family of related methodologies which treat the electronic problem at quantum level, while propagating the nuclear motion via classical trajectories[134–136]: in other words, the molecular wave function Φ is factorized in two functions - Ψ, describing the electrons, and χ, describing the nuclei.

In particular, the Trajectory Surface Hopping (TSH or SH) dynamics approximate the nuclear wave packet with a swarm of independent trajectories, each one following the classical Born-Oppenheimer PES of a specific state L:

$$\frac{d^2\boldsymbol{R}_\alpha}{dt^2} = -\frac{\nabla_\alpha H_{LL}}{M_\alpha} \tag{2.1}$$

where $\boldsymbol{R}$ and M are the coordinates and masses of the nuclei α; in the adi-

abatic basis, H_{LL} is the adiabatic energy E_L.[137,138] The electronic degrees of freedom are propagated via a separate, but coupled, equation of motion alongside the nuclear trajectory; at the end of each step, the observables and other needed quantities are computed on-the-fly; then, a stochastic algorithm is used to decide if the system will remain on the current electronic state L or hop to another one, which in principle can be higher or lower in energy. Indeed, a SH trajectory is like a single-surface trajectory for the whole propagation step, and this makes possible to use the same integrators and electronic structure methods developed for single-surface excited-state molecular dynamics - provided there is a method to compute the hopping probability. The random factor in the hopping also means that a single SH trajectory holds no physical value: average properties and common trends across most of the trajectories can instead reliably simulate the behavior of the system in the excited state, provided the number of trajectories is large enough to minimize the sampling error. Avoiding the need for precomputed PES is a great advantage of SH in terms of computational cost, even if it hinders its ability to describe quantum effects like tunneling.

By far, the most widespread formalism of SH is the Fewest-Switches Surface Hopping: the time-dependent electronic wavefunction is expanded as a linear combination of the electronic states K and the coefficient are propagated according to:[139]

$$\begin{aligned}
\boldsymbol{\Psi}(\boldsymbol{r}, t; \boldsymbol{R}) &= \sum_K c_K(t)\, \psi_K(\boldsymbol{r}; \boldsymbol{R}) \\
\frac{\mathrm{d}c_J}{\mathrm{d}t} &= \sum_K -c_K \left(i\hbar^{-1}\, H_{JK} + \sigma^{\mathrm{NAC}}_{JK} \right) \qquad (2.2) \\
H_{JK}(\boldsymbol{R}) &\equiv \left\langle \psi_J \,\middle|\, \hat{H}_e \,\middle|\, \psi_K \right\rangle \\
\sigma^{\mathrm{NAC}}_{JK}(\boldsymbol{R}) &\equiv \left\langle \psi_J \middle| \frac{\partial \psi_K}{\partial t} \right\rangle = \boldsymbol{d}_{JK} \cdot \boldsymbol{v} \qquad (2.3)
\end{aligned}$$

where $\boldsymbol{d}_{JK} \equiv \langle \psi_J \mid \nabla \psi_K \rangle$ is the non-adiabatic coupling vector and $\boldsymbol{v}$ is the classical nuclear velocity. FSSH makes actually use of two time steps, one

to integrate the nuclear equation of motion (EOM, eq. 2.1) and one for the electronic one (eq. 2.2): for each time step of the nuclear propagation (typically 100-500 as), the electronic EOM is valuated several times (typical time step of 5-10 as).

The FSSH algorithm computes the hopping probability from the coefficients of the expanded wavefunction in eq. 2.2; in the adiabatic base:[139]

$$P_{L\to J}^{\mathrm{FSSH}} = max\left[0, \frac{-2\Delta t}{|c_L|^2} Re\left(\sigma_{LJ}^{\mathrm{NAC}}\, c_J c_L^*\right)\right] \tag{2.4}$$

The first criterion to accept a hopping event from state L to J is that a sampled random number $0 \leq r_t \leq 1$ is less than the probability to hop from state L to J (times the number of K states lower than J), but greater than the sum of the probabilities of all the other possible hoppings from L:[134]

$$\sum_{K=1}^{J-1} P_{L\to K}^{\mathrm{FSSH}} < r_t \leq \sum_{K=1}^{J} P_{L\to J}^{\mathrm{FSSH}} \tag{2.5}$$

The second criterion is basically the conservation of energy: the energy lost or gained by the electronic subsystem must be recovered in terms of nuclear kinetic energy. The algorithm can thus prevent a hopping that would not preserve the total energy, or allow it by changing the nuclear velocity after the hop:[138] such rescaling should have the same direction of the NAC vector and intensity corresponding to the potential energy gap at the hopping time.[140] Softwares which cannot compute the NAC vectors usually use the direction of the momentum instead; in this case the modulus of the kinetic energy has to be reduced *ad hoc*, to avoid spending all the available kinetic energy for the hop.

In this project, we performed a FSSH molecular dynamics and adopted the partial coupling approximation:[141] at each time step, only the coupling with the currently occupied state are considered, while the NAC in eq. 2.3 not involving it are set to zero.

5BU was simulated at TD CAM-B3LYP/6-31G(d,p) level, in implicit methanol (C-PCM): TD-DFT is a widespread electronic structure method to be interfaced with mixed quantum-classical NAD and, when used correctly, has been proved

to be a good compromise between cost and accuracy to describe large and/or flexible systems, provided they have a negligible multi-reference character and are not near the crossing region with the ground state.[142–144] The choice of the density functional for this system has already been discussed in section 1.1.3; regarding the solvent, while an explicit solvent cluster could have been useful to tailor directional interactions, a QM treatment of the solvent would not have been feasible and the C-PCM has the advantage, with respect to QM/MM, of the mutual polarization between the solute electron density and the solvent response function.

We performed 50 trajectories, each one lasting for 1 ps: given the experimental fluorescence in fig. 2.1, this was considered a good simulation time to observe the fast component of the rise and decay of the fluorescence. The classical time step was set at 0.5 fs, the quantum time step at 0.025 fs; the hopping probabilities were obtained by computing the NACs; the Velocity Verlet[145,146] was used to integrate the EOM. The first 5 excited states were included, in order to fully cover the low-energy absorption band (see fig. 1.10).

Initial conditions were generated for both of the only two distinct conformers of 5BU in methanol, in equilibrium at room temperature, which have been described in section 1.2.2. For each conformer, a set of configurations was generated according to the quantum Wigner distribution, centered on the B3LYP/6-31G(d,p) minimum energy structure; the low-energy band of the absorption spectrum was computed from the Wigner distribution at the TD CAM-B3LYP/6-31G(d,p) level, in good agreement with the spectrum computed from the ONIOM/NPBC AIMD (see fig. 1.10). The initial conditions were filtered so that the VEE of the initial state fit inside an energy window of 1 eV, centered at the maximum of the absorption spectrum computed by the Wigner distribution. Energy conservation was ensured by rejecting trajectories with an energy drift larger than 0.5 eV with respect to either the previous or the initial step.

From each one of said configuration, the absorption was calculated at the CAM-B3LYP/6-31G(d,p) level for each one of the first 4 excited states . The average probability of each transition was calculated as the integral of the oscillator strength along the set of the initial conditions. From each set of initial

	S_1	S_2	S_3	S_4
B	34	7	1	1
F	2	3	1	1

Table 2.2: Number of trajectories initialized in each conformer (B and F, see section 1.2.1 for their definition) and excited state.

conditions, a number of trajectories was started; the number of trajectories from each set was weighted accordingly to the Boltzmann population of the conformers (computed from the relative total energies of their respective ground-state minima) and the average probability of each transition (computed as the integral of the oscillator strength along the set of the initial conditions), considering the first 4 transitions from the ground state. This method to generate the initial condition ensures that our trajectories are representative of the ground-state conformational equilibrium of 5BU, which is relevant to its photophysics, as detailed in section 1.2.2. The resulting set of initial conditions is summarized in table 2.2.

The FSSH dynamics was performed by the dedicated interface between the Newton-X Classical Series (NX CS) package, v. 2.4,[147,148] and the Gaussian 16 software:[126] for each step, the former integrates the equations of motion and computes the NACs, while the latter performs the TDDFT calculations and computes oscillator strengths, energies and forces.[149] The initial conditions were generated via the same interface: Newton-X computed the quantum harmonic oscillator Wigner distribution, Gaussian the absorption spectrum. Regarding the analysis of the trajectories, the energy and population of each involved adiabatic state were computed by Newton-X, while the emission spectrum, its fit and the evolution of structural parameters were computed via a suite of home-made Python scripts based on the NumPy,[150] SciPy[151] and scikit-learn[152] libraries. The Cremer-Pople[153] analysis was performed with a modified version of the Ring_Analysis.py module from the RING repository[154,155].

2.2.2 Time-resolved emission simulation

For every frame t of the NAD, the instantaneous emission spectrum was computed by taking the VEE and the oscillator strength of each excited state occupied by each trajectory. The emission intensity can be thus immediately computed via a trivial modification of eq. 1.13:[41,120]

$$L(\omega;t) = \frac{1}{N_t} \sum_{\substack{l=1 \\ j\neq 0}}^{N_t} f_{j0}(\boldsymbol{Q}_{l,t})\, w\big(\hbar\omega - \Delta V_{j0}(\boldsymbol{Q}_{l,t})\big) \tag{2.6}$$

where N_t is the total number of trajectories and, at time t:

- j is the state occupied by the trajectory l;
- $\boldsymbol{Q}_{l,t}$ is the configuration of the l-th trajectory.
- $f_{j0}(\boldsymbol{Q}_{l,t})$ and $\Delta V_{j0}(\boldsymbol{Q}_l)$ are the oscillator strength and the VEE of the $j \to S_0$ transition at the configuration $(\boldsymbol{Q}_{l,t})$, respectively: they generally differ from the ΔV_{0j} and f_{0j} of the inverse transition $j \leftarrow S_0$, since the solvent response is equilibrated with the excited-state electron density.
- $w\big(\hbar\omega - \Delta V_{j0}(\boldsymbol{Q}_{l,t})\big)$ is a broadening function, peaked at $\hbar\omega = \Delta V_{j0}(\boldsymbol{Q}_{l,t})$, which is here a Gaussian broadening with the same functional form as eq. 1.14.

2.3 A novel protocol to compare simulation and experiment

The very core of the scientific method is the ability to formulate a mathematical model to explain physical phenomena and compare its predictions with experimental observations. This is even more important for DFT calculations, given its dependency on a somewhat black-box exchange-correlation functional and its possible inaccuracies in the energetic order of the excited states. In addition, the computational cost of performing an *ab initio* molecular dynamics on medium-large systems, like 5BU, prevents the possibility of using a high-level computation as a reference.

The computational simulation of transient fluorescence decay is not a novel task: the main observable used to compare them with the experiment is the fit of the kinetic trace of the fluorescence - i.e. the series of emission intensities recorded with different time delays. The intensity used to compute the experimental fit is usually the absolute maximum of the spectrum, where there is the best signal/noise ratio; when a spectrum contains multiple bands, however, each maximum can be fitted independently, in order to follow the possibly different decay pathways happening at different energy ranges. Typically, decay phenomena are fitted with a negative exponential function, where the number of components is informative of the number of chemical-physical processes involved in the decay.

The fit of the kinetic trace is thus a quantitative measure highly informative of the dynamics of the system. It should not come as a surprise, therefore, its popularity as a benchmark for computational methods, often more than the band shape, which can be quite volatile for transient spectroscopy. On the computational side, for NAD studies it is common to fit the population of the involved excited states and compare it to the experimental fit of the kinetic trace.[156] Lan *et al.*[157] estimated the fluorescence lifetime by comparing the initial and final intensity computed during the NAD; other observables have also been used - depending on the available experimental data - like the instantaneous Stokes shift.[82]

2.3.1 A challenging system

In this context, 5BU represent a peculiar case: as mentioned in section 2.1.2, the two fluorescence bands have a different kinetics and thus have been fitted independently: the simulation should ideally be able to reproduce both fits, at the respective wavelength. Experimentally, the wavelenghts used to compute the fit are the maxima of each band at the first recorded spectrum (named "0 fs" in fig. 2.1). But here two problems arise: the definition of the bands and the comparison of the time scales.

The two emission bands are indeed rather thin (≈ 0.35 eV the high-energy band and ≈ 0.5 eV the low-energy one), very close (0.45 eV between the maxima) and partially overlapped (the intensity at the minimum between the bands is still more than two thirds the maximum of the lesser band):[27] it is not granted that TD-DFT manages to clearly distinguish them, or that the error on the energy is the same for both bands; they are actually so close that even the Gaussian broadening function (eq. 1.14) has to be carefully tailored to define the band shape.

If there was a 1:1 correspondence between the time of the simulation and the experiment, it might be possible to compare the band shape of the experimental and simulated spectra and hence judge the accuracy of the simulation and distinguish the two bands. But even understanding which time - or time interval - of the simulation is to be compared with which of the experiment is far from trivial:

1. As aforementioned in section 2.1.2, the "time zero" of the experimental spectrum (captured by the "0 fs" spectrum in fig. 2.1, which is also the starting time of the fit of the fluorescence) does not correspond with the beginning of the fluorescence kinetic trace, rather at its maximum (which is also the point captured by the "0 fs" spectrum in fig. 2.1): this means that not all the molecules are excited at the same time; for each time step, the recorded emission spectrum will be a superimposition of the fluorescence profile of molecules excited at time 0 (the majority) and the fluorescence profile of the ones excited before; our simulation, based on

the single molecule approximation, can only account for the former group;

2. As again written in section 2.1.2, the experimental "time zero" is not necessarily the same at each wavelength, since it depends on when the kinetic trace reached its maximum at each wavelength; this means that the intensity of each wavelength of the experimental spectrum is produced by a slightly different group of molecules. Even when the time discrepancy is within the IRF (as in the case of ref. 27), it is still another unavoidable element of difference between the experimental spectrum and the simulation;

3. The resolution of the experimental FlUC spectrum is limited by the IRF = 300 fs[27] This means that the experimental spectrum recorded within the IRF does not correspond to the starting time of the simulation (which is the Franck-Condon region) and instead captures an already advanced state of the photoinduced dynamics of 5BU - as it is clear even by just comparing its band shape with the low-energy absorption band.

It should be noted that the sources of inaccuracy 1 and 2 could be significantly less impactful if the time resolution was higher. Item 2 represents also one of the reason why the preferred way to compare a simulation with such experiment is via the kinetic trace of the maximum of each band.

2.3.2 Our strategy

Our protocol is based on computing a fit of each band of the simulated spectrum, to be compared with the fit of the experimental kinetic trace.

Let us start with three observations:

1. Our simulation has a time step of 0.5 fs, while the experiment has a time resolution of 300 fs; the soundest and most significant way to compare them is to compute an average of the simulated spectra to be compared with the experimental ones; but how long should such an average be?

2. The fluorescence simulated at the first frame (0 fs) of our NAD correctly captures the emission of our system in the Franck-Condon region, for the transitions included in the initial conditions: it is indeed very similar to the low-energy band of the absorption spectrum and has only one band. Since, instead, the experimental spectrum recorded within the IRF already shows two bands, there should be a time (which we name *t-zero*) were our simulated spectrum sprouts a second band, either by increasing the energy gap among two PES, or by populating new bright states. All the frames before such time should not be taken into account by the fits of the two bands; the question here is how to find such t-zero;

3. Since Valadan *et al.*[27] computed the kinetic trace for each band, we have to distinguish the bands of our simulated spectrum; however, raw data (the transitions from every trajectory at a given time) form a line spectrum; in order to group them in bands, we need to mimick part of the experimental broadening via the dedicated function 1.14; the problem is to find a value for the variance s^2 that well reproduces the experimental broadening: a too large value could coalesce two bands into one; one too small could count as separated two maxima of the same band;

4. another way to avoid counting minor bands is making the fit algorithm ignore bands which do not respect some characteristic, such as a minimum width, height or prominence, and add their intensities to the closest band instead; in our discussion, we will always talk about excluding bands based on the minimum promincence, but the same reasoning can be done, without difference, for any other band property.

We thus have four parameters to set in order to draw the spectrum, recognize the bands and compute the fit:

1. the length of the average of the spectra (hereafter named *len-avg*);

2. the t-zero;

3. the standard deviation s of the Gaussian broadening;

4. the minimum prominence to accept a band for the fit (hereafter named *band-prom*).

Our strategy is based on the *trial-and-error* approach: via a dedicated script written by one of us, it is possible to automatically check every combination of the aforementioned parameters (in a reasonable range), compute a fit for every combination and choose the one giving the best agreement with the time constants and amplitudes of the experimental fit. Here is the workflow:

For each test value of s:

1. All the instantaneous spectra are computed;
2. the simulation time is divided into intervals, based on the values of t-zero and len-avg; for each test value of t-zero:
   (a) The first interval runs from the initial frame to t-zero;
   (b) A list of test values for len-avg is produced, running from a user-defined minimum to the value of *(simulation time - t-zero)/4* (so that there are always at least four points to compute the four parameters of a biexponential fit); for each test value of len-avg:
      i. The rest of the simulation is divided into equally spaced intervals of length equal to len-avg;
      ii. The average of the instantaneous spectra included in each time interval is computed, producing a series of average spectra for the current combination of s, t-zero and len-avg (hereafter named just *spectra series*);
      iii. For each test value of band-prom and for each average spectrum of the current series:
         A. the bands of the average spectrum are individuated;
         B. If at least one spectrum of the series has a number of bands different from the other spectra, the current series is discharged; this is the strongest approximation of the protocol, but it's necessary to avoid the intricate problem of coupling

the bands of two spectra computed at different times, given that both the energy and the intensity can easily change;

C. the first band of each spectrum (in order of energy) is grouped with the first band of the other spectra in the series; the same for the second one and the eventual following bands;

D. For each series of bands:
   - the maximum and the integral of each band are computed;
   - two biexponential fits are computed: one using the maxima and the other one using the integrals.

The combination of s, time-z, len-avg and band-prom which gives the fit with the best agreement with the experimental parameters is chosen; its average spectra are visually inspected, in order to compare their band shape with the experimental one.

We used a biexponential fitting function, as the experiment does; it should be noted, however, that any function could be used, provided it produces two time constants in agreement with the experimental ones. In particular, an alternative to item 2 could be using a tri-exponential function, where one of the components should capture the initial decay linked to the formation of the second band.

A simplified approach

At the time of the writing of this thesis, the testing of the protocol discussed in section 2.3.2 is still in progress; however, we have already analyzed the results of our simulation and compared them with the experiment, via a simplified approach. This simplified approach do not separate the bands of the spectrum and instead use the maximum and the integral of the whole spectrum to build the simulated kinetic trace to be fitted. Here is the workflow:

1. for each test value of s, all the instantaneous spectra are computed;

2. the maximum and the integral of each spectrum are computed and saved as two distinct kinetic traces;

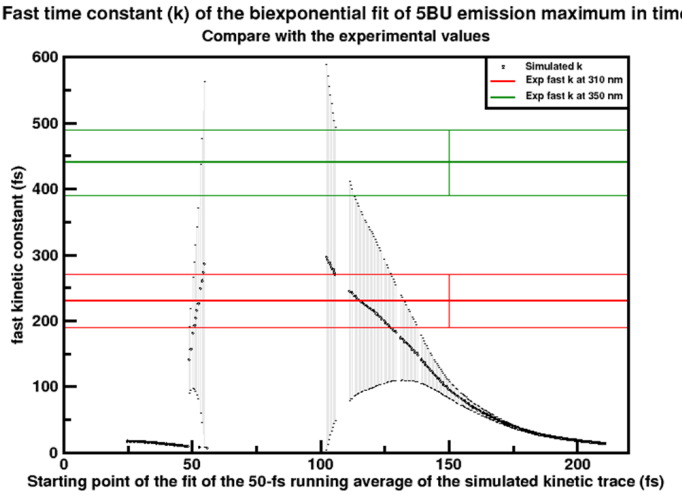


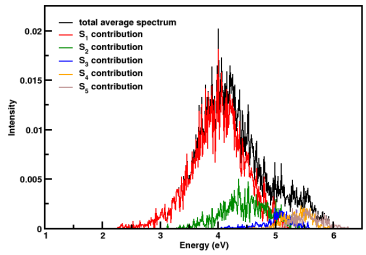
Fig. 2.3: Solid black points: values of the fast kinetic constant obtained from each fit, compared with the starting time of the fit (t-zero); grey bars: error bars of each fit. Thick green line: experimental fast time constant computed from the fit of the kinetic trace of the maximum of the band at 350 nm; thin green lines: its error bar. Thick red line: experimental fast time constant computed from the fit of the kinetic trace of the maximum of the band at 310 nm; thin red lines: its error bar.

3. each kinetic trace is slightly smoothed with an opportune mobile average;

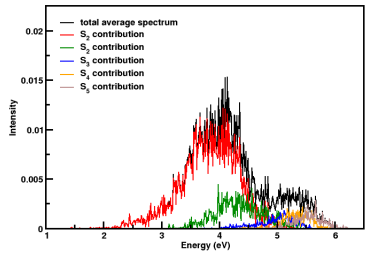
4. for each test value of t-zero, each kinetic trace is fitted, excluding from the fit the times before t-zero.

The results of the best fit computed using this approach is visualized in fig. 2.3.

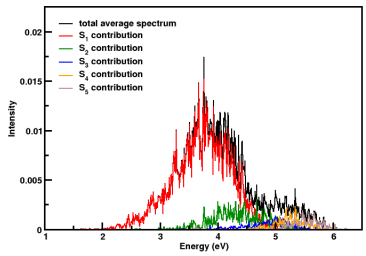
The fast kinetic constant obtained from fitting our simulated transient emission data is in the same order of magnitude as the experimental time constants if we start the fit between 100 and 150 fs, considering the respective errors. Thus, we can tell that our simulation takes less than 200 fs to fully leave the Franck-Condon region, assuming a shape with two bands and start exploring the time interval recorded by the FlUC experiment. The time-resolved spectrum computed from 150 fs is shown and commented in section 2.4.1.



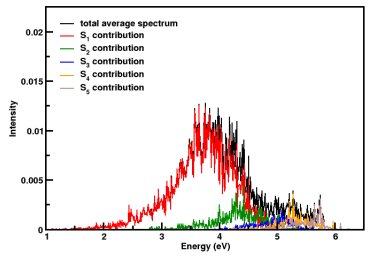
(a) 0-300 fs



(b) 250-550 fs



(c) 500-800 fs



(d) 750-1000 fs

Fig. 2.4: Black: TD-CAM-B3LYP/6-31G(d,p) emission spectra of 5BU in methanol, averaged with windows of 300 fs (figs. 2.4a to 2.4c) or 250 fs (fig. 2.4d), for different time intervals. Variance of the gaussian broadening function: $s^2 = 10 \times 10^6\,\mathrm{eV}^2$. Colored lines: the contributions of the involved transitions: $S_1 \rightarrow S_0$ (red), $S_2 \rightarrow S_0$ (green), $S_3 \rightarrow S_0$ (blue), $S_4 \rightarrow S_0$ (orange), $S_5 \rightarrow S_0$ (brown).

2.4 Results

2.4.1 Transient fluorescence spectrum

The emission spectrum of 5BU in methanol has been computed, according to the procedure explained in section 2.3.2, from the results of the NAD. To complete our comparison with the experimental data, we chose four time windows (which overall cover the whole simulation time) and compute the average of the spectra inside each one. The experimental spectrum in fig. 2.1 is printed at 0, 250, 500 and 1000 fs; given its time resolution of 300 fs (see section 2.1.2) and that

the experimental spectrum recorded within the IRF approximately corresponds with the simulated spectrum at around 150 fs, we selected four windows of width 300 fs each (except for the last one of 250 fs) and centers respectively 150, 400, 650 and 875 fs.

The results are displayed in fig. 2.4. At a first glance to the total spectra, there two bands clearly visible: one at ≈ 4 eV and the other one ≈ 5.5 eV; however, if we look at the contribution of the adiabatic transitions to the spectrum, we see that the spectral region around 4.5 eV is actually a superimposition of the bands of the first and the second excited state. We can thus tell that the emission spectrum of 5BU is composed by three bands:

1. the first one, peaked at ≈ 4 eV, which is by far the most intense and is dominated by S_1(fig. 2.1);

2. the second one, which is covered by the first one, peaks at ≈ 4.46 eV, and is composed by both S_1 and S_2(fig. 2.1);

3. the third one, peaked at ≈ 5.5 eV, contains the few contributions of S_3, S_4 and S_5.

Assuming a sistematical blue shift of 0.46 eV, the low-energy band of the average simulated spectrum can be associated to the experimental band at 350 nm (3.54 eV, see fig. 2.1): they both are the lowest in energy, the more intense and widest band of the respective spectra and have a similar FWHM (0.53 vs 0.69 eV). Assuming the blue shift is constant, the mid energy band could instead be associated with the high-energy experimental band (310 nm or 4.00 eV). The highest-energy band of the average computed spectrum is instead clearly outside of the experimental spectrum range.

The simulated spectrum also shows a clear Stokes shift (the peak energy moves from 4.00 to 3.76 eV) and the intensity decay (the maximum of the average spectrum at 875 ± 125 fs is 0.625 times the original peak, the same ratio between the maxima of the FlUC spectrum at 1 and 0 ps).

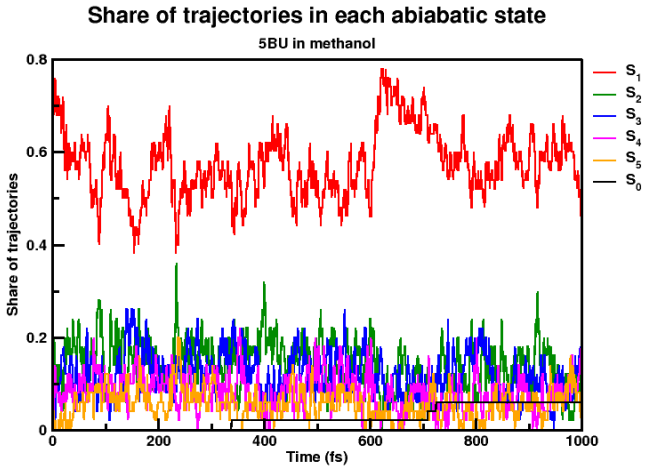


Fig. 2.5: Share of trajectories occupying each adiabatic state at each time: S_0 (black), S_1 (red), S_2 (green), S_3 (blue), S_4 (magenta), S_5 (orange).

2.4.2 Population analysis

The population analysis of the SH trajectories gives us several hints about the molecular processes behind the transient emission profile. Most of the trajectories start in the first excited state, which according to both the absorption spectrum computed from the Wigner distribution (see section 2.2.2) and the one from AIMD sampling (see section 1.2.3) is the brightest state, consistently with its mostly U$\pi\pi^*$ character (see sections 1.2.1 and 1.2.3 and appendices 1.B and 1.C).

As the system leaves the Franck-Condon region, however, so the population of S_1 rapidly falls, covering only 40% of the trajectories already at 100 fs (see fig. 2.5). Such a ultrafast time scale is consistent with the estimated time required to move the emission spectrum from the absorption regime to the early fluorescence one (see the fit of the kinetic trace in section 2.3.2).

The trajectories leaving S_1 populate the upper excited states: we do not observe, indeed, a significant decay to the ground state within the simulation time (1 ps); such an average population of higher excited states is not a surprise,

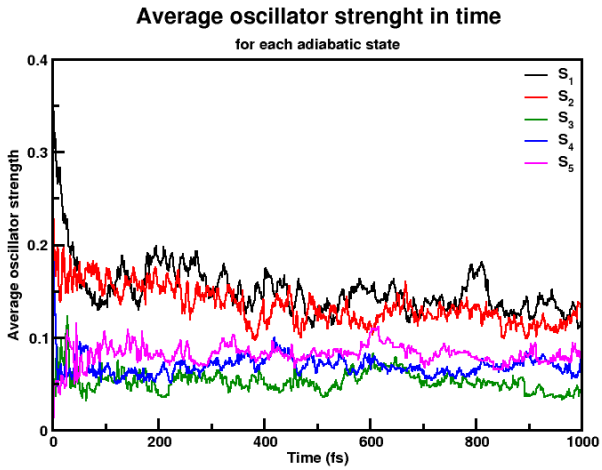


Fig. 2.6: Average oscillator strength of each excited state at each time: S_1 (black), S_2 (red), S_3 (green), S_4 (blue), S_5 (magenta).

since all the PES are, on average, close in energy in the first part of the dynamics ($\approx$ 1.5 eV between S_1 and S_5, see fig. 2.7).

The main recipient of this population exchange is instead the second excited state, which at 100 fs covers almost 1/3 of the trajectories: the rise of S_2 population is thus associated with the raise of the shoulder at 4.5 eV of the emission spectrum (fig. 2.4a), where $S_2 \rightarrow S_0$ is the most important transition. As discussed in section 2.4.1, this shoulder reproduces the high-energy band of the experimental spectrum (fig. 2.1): we can thus rationalize the born of the second experimental band with the population transfer from S_1 to S_2.

The raising emission intensity from S_2 is entirely due to the population increase: indeed, its average oscillator strength (which of course does not depend on the number of occupied trajectories) does not increase at the same time (see fig. 2.6). On the opposite, the average oscillator strength of S_1 sharply decreases in the first 100 fs: it is still a very bright state, but it passes from 0.358 at 0 fs to 0.129 at 87 fs: this informs us that the decay of the fluorescence is not due only the depopulation of the bright U$\pi\pi^*$ state, but also to a change in its character; the exhaustive analysis of this change of character is still in progress.

After 100 fs, the respective net populations (without considering the frequent hoppings) of S_1 and S_2 assume an oscillatory trend: that of S_2 varies roughly between 30 and 5%, that of S_1 between 70 and 40% (see fig. 2.5). At the same time, however, their respective average oscillator strengths (see fig. 2.6) steadily decline: such decay is plausibly associated with the experimental fast component fluorescence decay; the decay of S_1 oscillator strength explains the decay of the low-energy band (dominated by S_1, see fig. 2.4), while that of S_2 explains the decay of its shoulder at $\approx 4\,\text{eV}$. S_1 and S_2 should also have a similar character, judging by their similar average values of oscillator strength from 100 fs onward; this also means the lesser contribution of S_2 to the spectrum is entirely due to its smaller population.

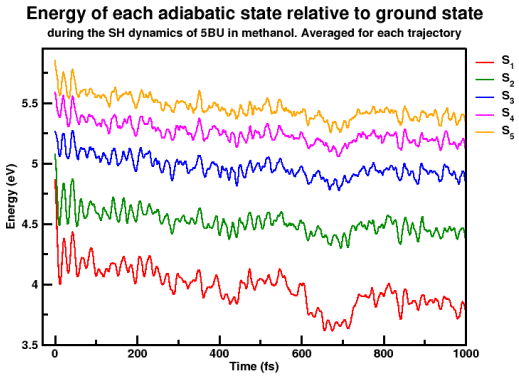


Fig. 2.7: Average energy of each excited state, relative the the ground state: S_1 (red), S_2 (green), S_3 (blue), S_4 (magenta), S_5 (orange).

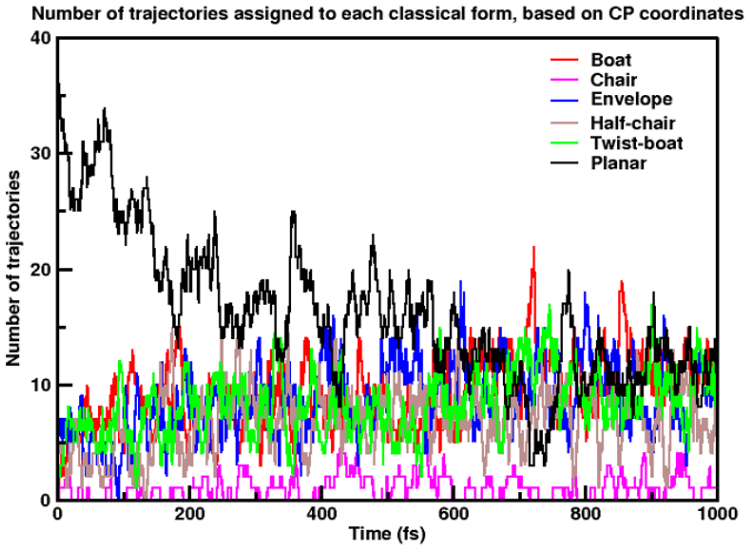


Fig. 2.8: Number of trajectories where the conformation of the uracil ring is assigned to one of the six classical puckering forms, based on Cremer-Pople analysis.

2.4.3 Time-resolved conformational analysis

In order to give a molecular explanation of the electronic features described in sections 2.3.2, 2.4.1 and 2.4.2, we selected a wide number of structural variables able to describe the conformation of 5BU and analyzed their behaviour during the NAD.

The most relevant one, in the context of explaining the fluorescence decay, results to be the out-of-plane puckering motion of the uracil ring. To follow the puckering motion, we computed, for each frame of the NAD, the Cremer-Pople coordinates[153] of each ring.

The Cremer-Pople analysis assign three numbers to each possible conformation of a six-membered molecular ring: the first one, name *total puckering amplitude* q, only tells how close the conformation is to being planar ($q = 0$); the other two, named *puckering angles* θ and ϕ, only tell how close the conformation is to one of the five classical puckered forms (chair, boat, half-chair, envelope and twist-boat). After computing them, we classified as planar all the conformations with $q < 0.15$ and assigned the remaining conformations to the

classical puckered form with the closest values of θ and ϕ.

The results are displayed in fig. 2.8, which shows, for each frame and for each puckering form, the number of trajectories whose uracil is assigned to that puckering form. Almost all the trajectories start with a planar uracil, as it is expected in the ground state; but then, almost all of them lose the planarity to adopt a puckered conformation. The population of puckered conformations is very fast in the first 200 fs (when only 13 trajectories still retain the planarity), then slows down and reaches a minimum of 3 planar conformations at 716 fs. Such timings are compatible with the experimental time constants of the fluorescence decay: in particular, it is plausible to connect the first, ultrafast loss of planarity to the rise of the high-energy fluorescence band (see sections 2.3.2 and 2.4.1, the depopulation of S_1 and the population of S_2.

Thus, our conclusion is that the photoexcitation induces an out-of-plane motion of the uracil ring, which disrupts the symmetry of the bright U$\pi\pi^*$ state, reducing its oscillator strength and favoring the population of the second excited state. After that, in the 200-700 fs range, even more trajectories lose the planarity, even if with a slower trend, while both S_1 and S_2 lose emission intensity: this could be an explanation for the fast component of the fluorescence decay.

Also the phenyl ring presents some puckering, but in its case most of the trajectories remain planar. Other geometrical variables do not show any clear trend common to most of the trajectories; for completeness, we show here the time series of the dihedral angles θ and ϕ (which are defined in section 1.2.1 and are unrelated to the CP θ and ϕ puckering angles).

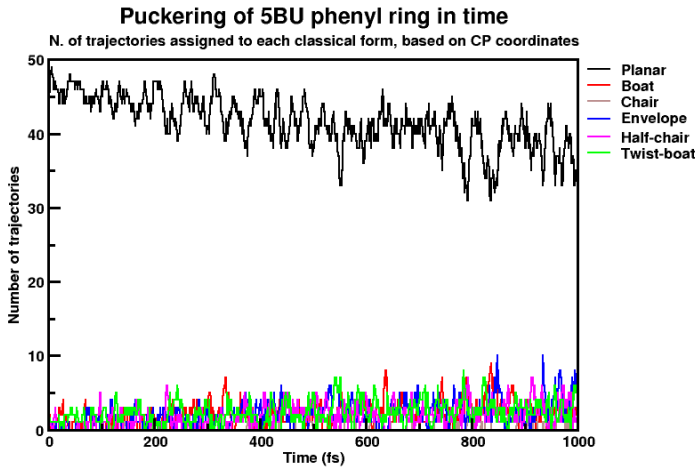


Fig. 2.9: Number of trajectories where the conformation of the phenyl ring is assigned to one of the six classical puckering forms, based on Cremer-Pople analysis.

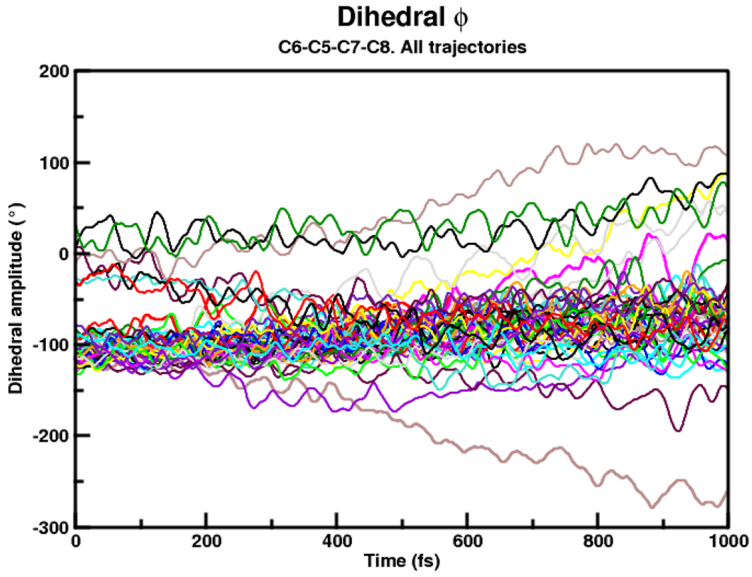


Fig. 2.10: Values of the ϕ dihedral angle for each trajectory in time.

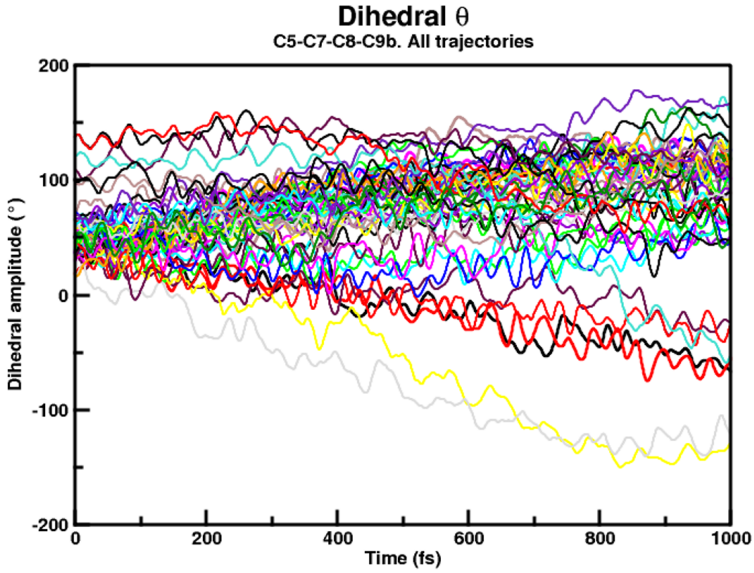


Fig. 2.11: Values of the θ dihedral angle for each trajectory in time.

Chapter 3

High-energy Photoreactivity of 5- and 6-benzyluracil

The absorption of high-energy radiation by molecular systems and the subsequent photoionization usually lead to enhanced reactivity toward unimolecular processes such as rearrangements and fragmentations, compared to the initial neutral, closed-shell molecule.[158–163] From an experimental point of view, photoelectron spectroscopy (hereafter, PES) measurements performed with vacuum-ultraviolet (VUV) light sources can identify the valence ionic molecular states actually populated and their binding energies.[164,165]

Coincidence techniques such as photoelectron-photo-ion coincidence (PEPICO) spectroscopy allow for the investigation of energy-selected unimolecular reaction channels and provide insights into the specific evolution of the molecular ion from well-defined ionic states.[166–169] The interpretation of data collected from PEPICO experiments is often carried out through atomistic *ab initio* simulations to identify the involved molecular species and the main fragmentation pathways,[169–171] although the limited spatial and temporal extent of the potential energy surface exploration by the simulations can prevent a systematic characterization of all transient species.

In this chapter, we present a new computational protocol to characterize the dynamical behavior of ionized systems, aimed at exploring several unimolecular

reaction channels including rearrangements, intramolecular reactions, and fragmentations. In order to interpret PEPICO experiments, we present an extensive exploration of the cation potential energy surface through *ab initio* molecular dynamics (MD) simulations, carried out with efficient semi-empirical potentials for long simulation times and in a highly parallel manner. This can reveal the actual molecular identity of the rearrangement or fragmentation products, as well as the detailed reaction mechanisms.

As challenging test cases, the 5-benzyluracil (5BU) and 6-benzyluracil (6BU) model systems have been employed; their importance as model systems for nucleic acid-protein photo-crosslinking has been described in section I.1. Chapters 1 and 2 explored the photophysics and spontaneous emission, respectively, of 5BU in methanol solution, which are indeed propaedeutic to the study of its photocyclization. Photochemical cross-linking through UV irradiation, based upon the population of nucleobase low-lying electronic excited states, is, however, characterized by rather low yields in both biological systems and model compounds due to fast excited-state deactivation.[172,25] At the same time, possible cross-link mechanisms involving ionized, rather than just excited, states have been much less investigated. Only recently, the employment of pulsed (from ns to fs) high-intensity UV laser sources has been suggested to increase the cross-link efficiency by nucleobase two-photon absorption and subsequent ionization.[173]

The aforementioned theoretical approach has therefore been applied to the investigation of the electronic structure, cross-link, and fragmentation dynamics of 5BU and 6BU. Simulated photoelectron spectra allow for the characterization of binding energies and spatial distribution of valence electrons on both uracil and benzene moieties and help to understand how the position of the benzene ring substituent can influence the electronic properties. PEPICO experiments integrated with molecular dynamics-based exploration of the potential energy surface and further energy refinement, on the other hand, reveal that, at lower energies, ionized 5BU and 6BU, while being resistant to fragmentation, can indeed rearrange to give cross-linked products. At higher energies, instead, fragmentation channels are dominant, leading mainly to the decomposition of

the uracil moiety. Moreover, different fragmentation pathways are observed for the two BU isomers.

The comprehensive investigation of 5BU and 6BU using PES and PEPICO techniques, combined with state-of-the-art theoretical methods, will contribute to a deeper understanding of the electronic structure and photophysical properties of these biomolecules. Hence, it will provide a broader knowledge of the complex photophysical and photochemical processes involved in DNA-protein cross-linking induced by ionization. Such findings, therefore, will have important implications in the field of DNA damage and repair, the development of light-based techniques for protein-nucleic acid cross-linking to isolate biomolecular complexes occurring in cells, as well as the development of therapeutic strategies targeting DNA-protein interactions.

3.1 Methods

3.1.1 Photoionization simulation

5BU and 6BU vertical electron binding energies up to $\approx 13\,\text{eV}$ were predicted using the renormalized partial third-order approximation[174,175] (P3+) with the 6-311++G(d,p) basis set. This method belongs to the electron propagator theory approaches,[176] which require the solution (to different degrees of approximation) of the Dyson equation to evaluate the many-body Green's function, whose poles correspond to the ionization potentials and electron affinities of the N-electron system. Photoelectron spectra of the ground-state neutral species at equilibrium were simulated within a nuclear ensemble approach[39,177] by performing electron binding energy calculations on 500 uncorrelated geometries from previously collected MD trajectories (fig. 3.1; see appendix 3.B for details about MD dynamics). Such an approach, including several initial configurations, can partially recover the experimental band shapes due to vibrational motion.[39,178] Photoionization cross sections for the (N electrons) state $\Phi_i \rightarrow$ (N-1 electrons) state Φ_f transitions were approximately evaluated as $\sigma_{if} = c\|\psi^d_{if}\|^2$, where c collects several constants and ψ^d_{if} is the corresponding Dyson orbital, defined

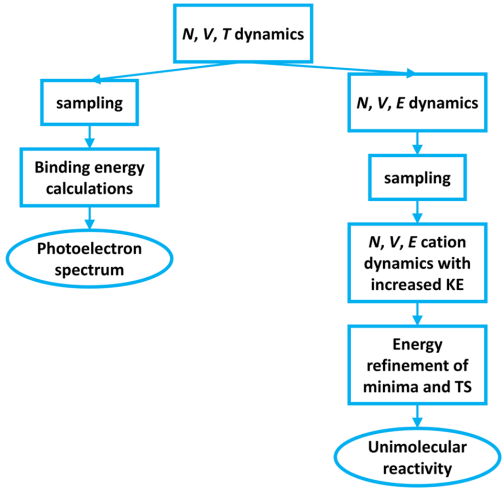


Fig. 3.1: Flowchart describing the computational protocol developed to simulate 5BU and 6BU photoelectron spectra and unimolecular (rearrangements and fragmentations) reactivity.[1]

as[39] $\psi_{if}^{d}(\boldsymbol{r}) = N^{1/2}\langle\Phi_f \mid \Phi_i\rangle_{\{\boldsymbol{r}_{N-1}\}}$, where the integration is performed over the spatial coordinates of the remaining $N-1$ electrons. Total photoelectron spectrum was computed as

$$\Gamma(E) = \sum_f \Gamma_f(E)$$

$$\Gamma_f(E) = \frac{1}{N_p}\sum_{l=1}^{N_p} \sigma_{if}(\boldsymbol{R}_l)\, w\big(E - \Delta V_{if}(\boldsymbol{R}_l)\big) \tag{3.1}$$

where $\Gamma_f(E)$ is the photoelectron spectrum for the adiabatic ion state f, N_p is the number of sampled configurations, $\boldsymbol{R}_l$ is the l-th geometry, $\Delta V_{if}(\boldsymbol{R}_l)$ is the vertical ionization potential, and $w\big(E-\Delta V_{if}(\boldsymbol{R}_l)\big)$ is a broadening function, peaked at $E = \Delta V_{if}$, which is here a Gaussian broadening with a $s^2 = 0.001\,\mathrm{eV}^2$ variance:

$$w\big(E - \Delta V_{if}(\boldsymbol{R}_l)\big) = \exp\Big[-\frac{1}{2}\Big(\frac{E - \Delta V_{if}(\boldsymbol{R}_l)}{s}\Big)^2\Big] \tag{3.2}$$

Photoelectron signals were further characterized through molecular orbitals (MOs) and population analysis of the corresponding removed electrons, cal-

culated for the optimized 5BU and 6BU conformers. No empirical shift was applied to the computed photoelectron spectra to improve the comparison with the experimental ones.

3.1.2 Photofragmentation simulation

5BU and 6BU ionization-induced fragmentation was simulated through high-energy ground-state Born-Oppenheimer molecular dynamics (BOMD) simulations. In particular, the Quantum Chemistry Electron Ionization Mass Spectra (QCxMS) program and protocol were employed.[179–183] After a preliminary BOMD equilibration in the canonical ensemble at the temperature of the PEPICO experiments, a phase-space sampling in the microcanonical ensemble was carried out for the neutral system in its electronic ground state. Then, 1000 uncorrelated snapshots were selected for the subsequent (10 ps) fragmentation dynamics (fig. 3.1). Each sampled snapshot was considered instantaneously ionized, using its coordinates and nuclear velocities as initial conditions for the propagation of the molecular ion on the doublet ground-state potential energy surface. Photoionization leads, in principle, to the ion in an electronic excited state, which then relaxes to a vibrationally excited state of the electronic ground state through internal conversion (IC) and intramolecular vibrational energy redistribution. The internal excess energy (IEE) transfer to the vibrational modes is modeled by increasing nuclear velocities during an initial IC time estimated by an energy-gap law,[182] such as $t_{IC} = \sum_{j>i}(k_h/N_{el}) \exp(\alpha(\varepsilon_i - \varepsilon_j))$, where ε_i are MO energies, $\alpha = 0.5\,\mathrm{eV}^{-1}$, and $k_h = 2\,\mathrm{ps}$. To characterize the structure of the experimentally observed fragments, as well as the main fragmentation mechanisms leading to them, IEE values were stochastically obtained from a distribution according to the original QCxMS protocol.[179]

Since a statistically significant number of fragmentation trajectories must be collected and long simulation times are required to observe fragmentation events, efficient electronic structure approaches must be employed. In this regard, density functional theory (DFT) usually offers a good compromise between cost and accuracy for the description of systems of organic/biological[184–189] interest. Moreover, semi-empirical electronic structure theories (such as den-

sity functional tight binding methods[190–193]) appear even more cost-effective in this context. In particular, the recently developed semi-empirical GFN2-xTB method was employed,[193] as it showed good performance with respect to other semi-empirical methods and was already successfully tested in mass spectra prediction within the QCxMS protocol.[181] Due to the underlying approximations and parametrizations, semi-empirical density functional tight binding methods (such as GFN2-xTB) may indeed estimate bond energies relevant for fragmentation processes with less accuracy than DFT or post-HF methods. Nevertheless, extensive comparisons of experimental mass spectra with the fragments from GFN2-xTB molecular dynamics simulations showed satisfactory agreement, allowing a reliable assignment.[181,182]

In principle, the relative abundances of the fragments could be directly obtained by counting the species observed at the end of each simulation. This approach, however, may lead, for some systems, to large deviations from the experimental values, especially at lower energies,[194] probably due to the limited simulation time (10 ps, compared to the µs experimental observation time) and the inaccuracies of the semi-empirical potential. Therefore, the simulation of relative abundances of PEPICO fragments has not been done here. Energy profiles of the most relevant fragmentation processes were further refined on the cation doublet ground-state potential energy surface at the B2PLYP-GD3BJ/def2-TZVPD//B3LYP-GD3BJ/def2-SVPD level of theory.[195–205] Such calculations were performed with Gaussian16 electronic structure software.[126]

3.2 Results

3.2.1 Simulated photo-electron spectra

Benzyluracil has two distinct torsional degrees of freedom, which describe the relative orientation of the benzene and uracil rings and thus could influence the valence orbital interactions between the two moieties (see appendix 3.B for the definition of the ϑ and φ dihedrals). Therefore, 5BU and 6BU conformational space was extensively explored through long (300 ns) molecular dynamics

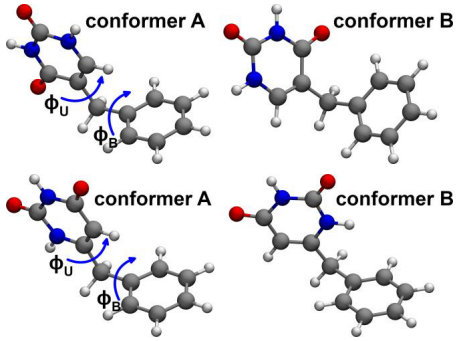


Fig. 3.2: 5BU (up) and 6BU (down) three-dimensional structures for the A (left) and B (right) conformers, found by MD simulations and further optimized at DFT level. ϑ and φ dihedral angles are also shown.[1]

simulations in the gas phase, at the experimental temperature. For both 5- and 6BU, two distinct conformers (hereafter named A and B) were found, both pouplated at the experimental temperature. For both 5- and 6BU, conformer A has the two rings approximately orthogonal, B has them facing each other (see fig. 3.2). The protocol and the results of the conformational analysis are detailed in appendix 3.B.

The simulation of 5BU and 6BU photoelectron spectra, therefore, takes into account such a conformational flexibility and the population of more than one conformer. Photoelectron cross-section was estimated in the 8-13 eV binding energy range from the MD conformational ensemble, allowing for better modeling of the broadening induced by nuclear motion. Band assignment was performed through valence-electron binding energies and spatial distributions calculated from the most populated optimized conformers (A and B for 5BU and 6BU, respectively).

In 5BU, the first experimental band displays a lower peak at 8.94 eV and the most intense one at 9.28 eV (fig. 3.3). From the calculated PES this observation results from three different contributions, that is, ionization from the HOMO, HOMO-1, and HOMO-2 (fig. 3.18). The signal at 8.94 eV likely corresponds to the benzene-localized HOMO of the minor B conformer (see table 3.5 for binding energies and condensed MO-resolved population analysis). The most

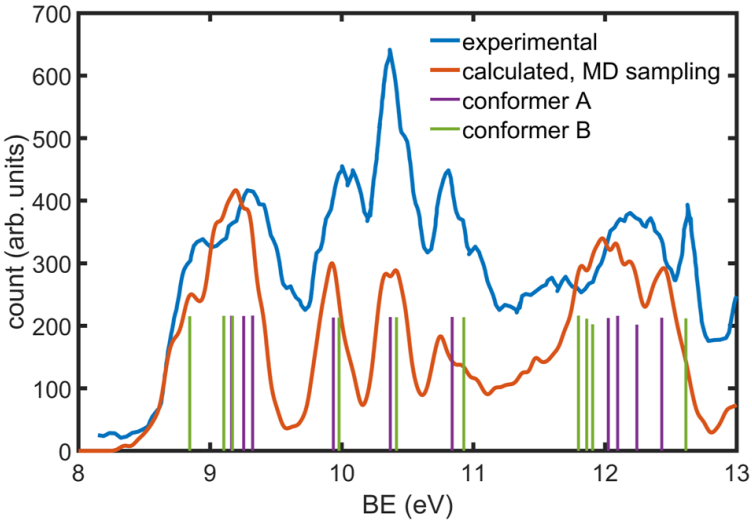


Fig. 3.3: Experimental and simulated 5BU valence photoelectron spectrum in the 8–13 eV binding energy range, the latter evaluated on the conformation ensemble given by the collected MD trajectory. Electron binding energies estimated for 5BU optimized A and B conformers are also shown as vertical purple and green bars, respectively. [1]

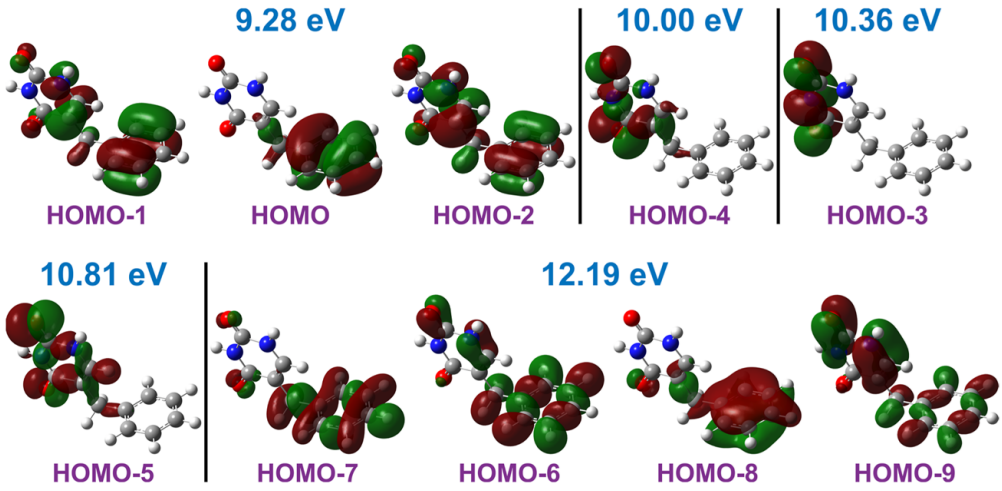


Fig. 3.4: 5BU photoionized electrons in the 8–13 eV binding energy range, calculated on 5BU (most populated) optimized A conformer, reported as MOs isosurfaces (0.02 isovalue). The corresponding BEs from the experimental photoelectron spectrum are reported in blue. [1]

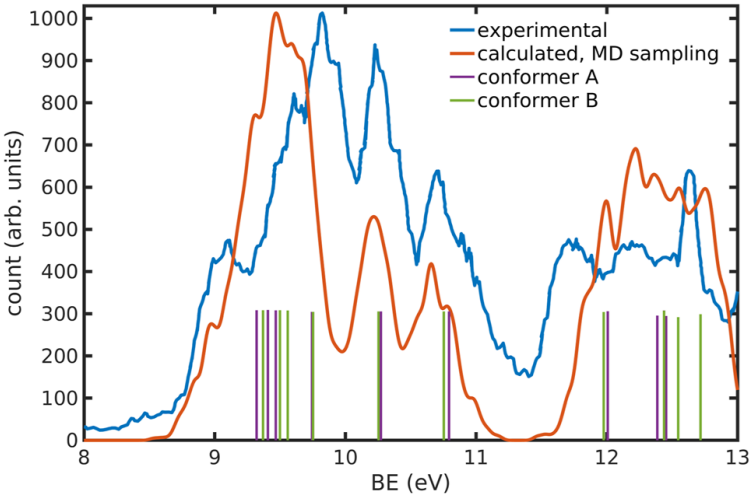


Fig. 3.5: Experimental and simulated 6BU valence photoelectron spectrum in the 8–13 eV binding energy range, the latter evaluated on the conformation ensemble given by the collected MD trajectory. Electron binding energies estimated for 6BU optimized A and B conformers are also shown as vertical purple and green bars, respectively. [1]

intense signal at 9.28 eV can be related to the A conformer HOMO, HOMO-1, and HOMO-2, where the former is a benzene-localized π MO, while the latter two are π MOs with mixed benzene-uracil character (see fig. 3.4 for the MO isosurfaces of the A conformer). The feature at 10.00 eV can be attributed to the HOMO-4 (the n(O4) non-bonding uracil MO), while the HOMO-3 (a uracil π MO) and HOMO-5 (the n(O2) non-bonding uracil MO) are responsible for the bands at 10.36 and 10.81 eV, respectively. The small experimental peak at 11.70 eV could be explained by the ionization of benzene orbitals HOMO-6 and HOMO-7 of the B conformer, although their BEs are slightly blue-shifted. The broad band at 12.19 eV results from the ionization of several MOs: HOMO-6, HOMO-7 (benzene σ MOs), HOMO-8 (benzene π MO), and HOMO-9 (mostly a uracil π MO, but with mixed σ benzene character) of the A conformer.

The 6BU simulated valence photoelectron spectrum (reported in fig. 3.5) shares several features with the 5BU one. The first band again results from contributions of HOMO, HOMO-1, HOMO-2, and HOMO-4 (fig. 3.19). The first peak at 9.11 eV, in particular, could be explained by HOMO-1 and HOMO

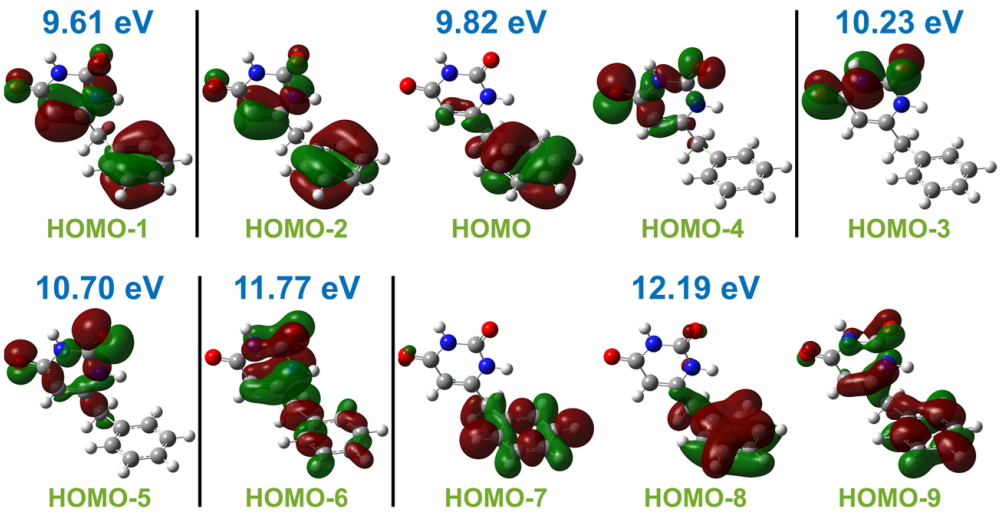


Fig. 3.6: 6BU photoionized electrons in the 8–13 eV binding energy range, calculated on 6BU (most populated) optimized B conformer, reported as MOs isosurfaces (0.02 isovalue). The corresponding BEs from the experimental photoelectron spectrum are reported in blue. [1]

MOs of 6BU structures around the distribution maxima corresponding to the predominant B conformer, which are populated at the experimental temperature (see fig. 3.17). These structures have calculated HOMO-1 and HOMO BEs at about $\approx$8.9 eV, in contrast with the optimized geometry, which instead has 9.37 and 9.56 eV BEs for the HOMO-1 and HOMO, respectively. This leads to double-peaked and broad HOMO and HOMO-1 spectra from the ensemble (see fig. 3.19). Since several 5BU and 6BU valence electrons are quite delocalized on both benzene and uracil moieties (see MO isosurfaces in figs. 3.4 and 3.6), their spatial character and binding energy may be quite sensitive to the relative orientation of the two rings, described by (ϑ, φ) dihedrals. The experimental peak at 9.61 eV can be related to the HOMO-1 of the B conformer (a mixed uracil-benzene π MO, see fig. 3.6), while that at 9.82 eV to the HOMO-2 (another mixed uracil-benzene π MO), HOMO (a benzene-localized π MO), and HOMO-4 (the n(O4) non-bonding uracil MO). It must be remarked that the simulated spectrum from the whole ensemble underestimates the energy of this band by $\approx$0.3 eV, while nevertheless preserving its line shape. The band at 10.23 eV can be assigned to the HOMO-3 (a π MO completely localized on the uracil moiety),

while the HOMO-5 (the n(O2) non-bonding uracil MO) gives the signal at 10.70 eV. The HOMO-6 (a uracil π MO with also a reduced σ benzene character) gives rise to the peak at 11.77 eV. Its BE is slightly overestimated with respect to the experimental value. The signal at 12.19 eV is likely related to the HOMO-7 and HOMO-8 (benzene-localized σ and π MOs, respectively) and the HOMO-9 (a benzene σ MO with also some uracil π character).

3.2.2 Fragmentation pathways

5BU and 6BU GFN2-xTB BOMD trajectories on the radical cation (doublet) potential energy surface allowed us to characterize both the structures and the formation mechanisms of the fragments detected in the PEPICO experiments (see fig. 3.14). In the following, we use AE_{th} to indicate the computed appariting energies, AE_{PEPICO} for the experimental ones.

5BU fragmentation

Considering 5BU fragmentation first (fig. 3.7), an isomer of the molecular ion (M^+, m/z 202) was found, obtained by a H atom migration from uracil N1 to O(C2), followed by uracil N3-C4 bond breaking and H transfer from C6 to N3. It is characterized by $AE_{th} = 9.81$ eV, that is, higher than the estimated adiabatic ionization energy of 8.22 eV. Considering possible ionization-induced uracil-benzene interactions leading to cross-linking and cyclization, no cyclic $5BU^{\cdot+}$ species was found. Further investigation of potential cross-linking indeed predicts a high barrier (1.80 eV, 41.51 kcal mol^{-1}) leading to the syn diastereoisomer of $5BU^{\cdot+}$ cyclic species (fig. 3.20). The anti diastereoisomer, however, appears not stable. The species resulting from a simple hydrogen removal from benzene C3' (m/z 201, $AE_{th} = 12.56$ eV) is not the most stable m/z 201 fragment. An isomer produced by H atom transfer from uracil N3 to O(C4), uracil N1–C2 bond breaking, H migration from the CH2 moiety to N1, and final H–O(C4) dissociation instead has a lower $AE_{th} = 10.63$ eV, very close to the experimental $AE_{PEPICO} = 10.5 \pm 0.2$ eV. An even more stable structural isomer of the m/z 201 fragment is identified at $AE_{th} = 9.57$ eV. This derives from a uracil

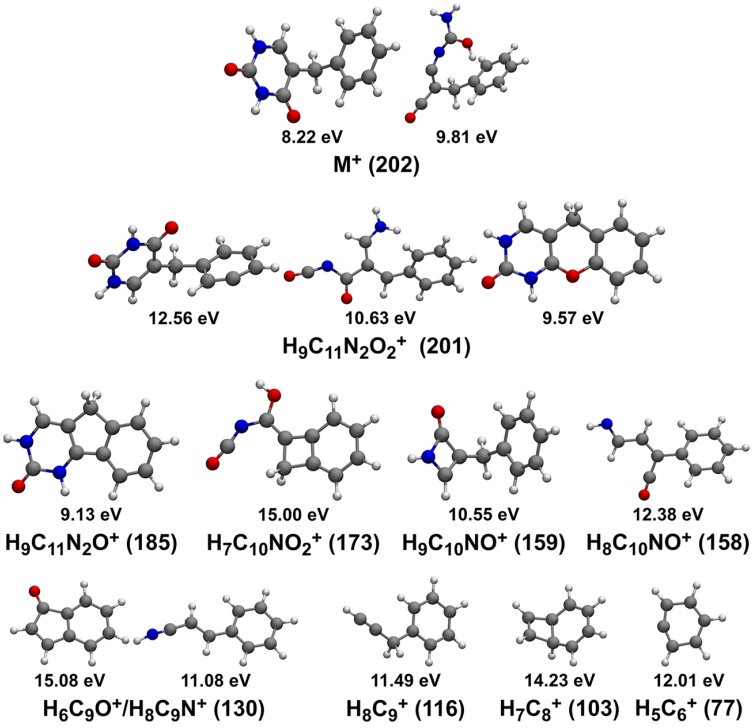


Fig. 3.7: 5BU fragments observed in GFN2-xTB BOMD trajectories. Formula, molecular mass, three-dimensional structure optimized at B3LYP-GD3BJ/def2-SVPD level of theory, and appearance energy at the B2PLYP-GD3BJ/def2-TZVPD level (calculated on final fragmentation products) are reported for each fragment. [1]

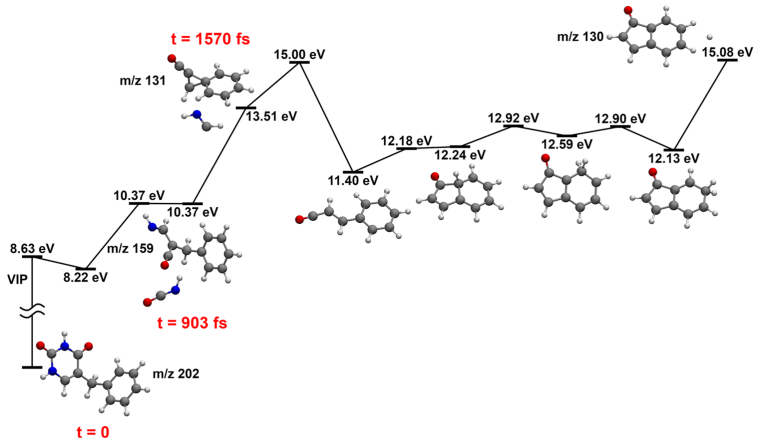


Fig. 3.8: Mechanism of $5BU^{\cdot+}$ decomposition toward m/z 185 cyclic fragment, featuring a uracil C6–benzene C2' cross-link, evaluated at the B2PLYPGD3BJ/def2-TZVPD//B3LYP-GD3BJ/def2-SVPD level of theory. VIP: vertical ionization potential. [1]

O(C4)–benzene C2' bond formation and H atom removal from C2', featuring a central, six-membered ring by cyclization involving the uracil C4 carbonyl oxygen. Nevertheless, kinetic energy barriers could lead to a higher effective AE for this latter, more stable, isomer, thus making it potentially compatible with the experimental AE_{PEPICO}.

The m/z 185 fragment observed in fragmentation MD trajectories is particularly interesting since it shows a uracil C4–benzene C2' ionization-induced cross-link, similar to the uracil C6–benzene C2' cross-link induced in 5BU by photoexcitation,[172,25] which could model potential formation of covalent bonds between nucleic acids and proteins. After a H atom transfer from benzene C2' to uracil O(C4), the uracil C4–benzene C2' bond could form (after 308 fs), followed by OH radical loss from C4 (see fig. 3.8). Calculations on such a fragment predict remarkable stability with $AE_{th} = 9.13$ eV. Taking into account the highest energy kinetic barrier along the mechanism, an $AE_{th,\,kin} = 9.70$ eV is predicted, which is lower than the $AE_{PEPICO} = 10.5 \pm 0.2$ eV.

The m/z 173 fragment ($AE_{th} = 15.00$ eV) is obtained through a rather complex rearrangement, comprising the exchange of uracil C5 between the central CH2 and benzene C2', a H atom moving from C2' to C5, then to C6, to CH2

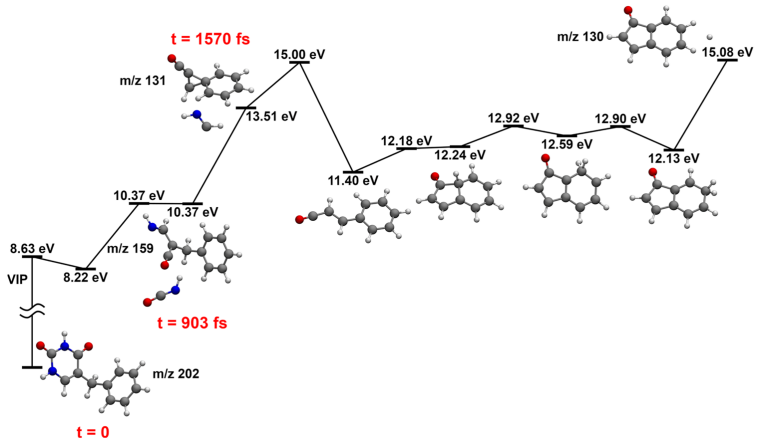


Fig. 3.9: Mechanism of the preferred $5BU^{\cdot+}$ fragmentation process leading to $H_6C_9O^+$ m/z 130 ion. The energy profile is evaluated at the B2PLYP-GD3BJ/def2-TZVPD//B3LYPGD3BJ/ def2-SVPD level of theory. Average dissociation times from GFN2-xTB trajectories are also reported in red. VIP: vertical ionization potential. [1]

and finally to O(C4), another H atom transferred from N3 to O(C2) and then removed, N1–C2 bond dissociation, C5–CH2 bond re-formation (giving a four-membered ring), and final C5–C6 breaking, with a neutral HNCH fragment loss.

The m/z 159 fragment (AE_{th} = 10.55 eV, comparable to the AE_{PEPICO} = 10.7 ± 0.3 eV) results from the loss of a neutral HNCO fragment from the uracil moiety, through N1–C2 and N3–C4 bond breaking and N1–C4 bond formation, giving an N-containing, four-membered ring. After HNCO loss, an alternative pathway involves an H atom loss from central CH2, C6–CH bond formation together with C5–C6 bond breaking and C5 insertion between CH and benzene C1' (through CH–C1' and C5–C1' bond breaking and formation, respectively), yielding an m/z 158 fragment (AE_{th} = 12.38 eV, close to the experimental AE_{PEPICO} = 13.0 ± 0.5 eV).

The m/z 130 fragment appears as the second most abundant one in the 5BU mass spectrum (see fig. 3.14a), as well as in the PEPICO measurements over a wide BE range (13-20 eV, fig. 3.14b). Therefore, the fragmentation pathways leading to such fragment will be analyzed in detail.

Two distinct m/z 130 fragments, $H_6C_9O^+$ and $H_8C_9N^+$ (fig. 3.7), may potentially contribute to the total abundance observed for m/z 130. Considering $H_6C_9O^+$ first, the preferred (94% relative count) route toward such fragment features uracil N1-C2, N3-C4, C5-C6 bond breakings (see fig. 3.21, purple pathway), while a more complex one (4% relative count) is based upon N1-C2, C2-N3, C4-C5, C5-C6 bond breakings and C2-C5 bond formation (see fig. 3.21, green pathway).

Considering in more detail the mechanism of the former fragmentation pathway (fig. 3.9), the concerted N1-C2 and N3-C4 bond breakings in the uracil moiety through a retro Diels-Alder reaction allow the dissociation of a neutral HNCO (isocyanic acid) fragment, as already observed in uracil and uracil derivatives.[206–210] According to semi-empirical BOMD simulations, this occurs, on average, 903 fs after the initial 5BU ionization. The subsequent uracil C5–C6 bond breaking and concerted uracil C5–benzene C1' bond formation instead lead to a spiro bicyclic intermediate and to the removal of a neutral HNCH fragment (t = 1570 fs on average). The labile three-membered ring then opens up through a hydrogen shift from the central CH2 bridge to the uracil C5. A condensed bicyclic intermediate is obtained by bond formation between uracil C4 and benzene C2'. The hydrogen on this latter atom then moves stepwise on benzene C3' and C4' and eventually dissociates, obtaining the m/z 130 $H_6C_9O^+$ fragment. The AE_{th} = 15.08 eV is somewhat high compared with that of the monocyclic m/z 131 precursor along the fragmentation mechanism (AE_{th} = 11.40 eV) as well as to the other $H_8C_9N^+$ m/z 130 fragment (see following discussion). The uracil C4–benzene C2' bond formation along such mechanism, giving the bicyclic m/z 131 species, moreover represents another possible example of an ionization-induced cross-linking process in the 5BU molecular model and so potentially between nucleic acids and proteins.

In contrast with $H_6C_9O^+$, more than one pathway toward the $H_8C_9N^+$ fragment is possible. The most recurrent route (49% relative count, purple pathway in fig. 3.22) corresponds to uracil N1–C2, N3–C4, C4–C5 bond breakings, leading to $H_8C_9N^+$ + HNCO + HCO fragments. A similar pathway (23% relative count, green pathway in fig. 3.22) shows C2–N3 bond breaking instead

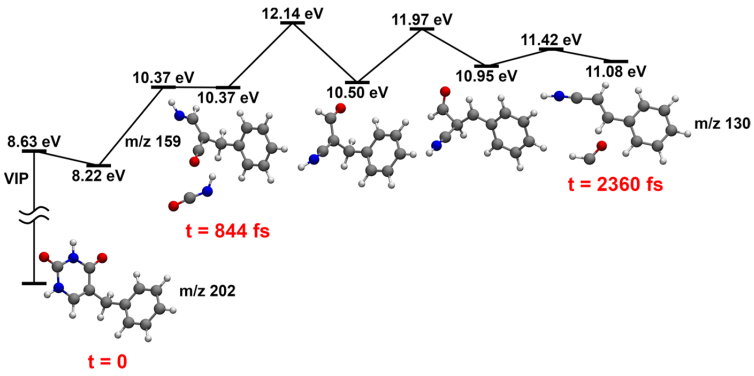


Fig. 3.10: Mechanism of the preferred $5BU^{\cdot+}$ fragmentation process leading to $H_8C_9N^+$ m/z 130 ion. The energy profile is evaluated at the B2PLYP-GD3BJ/def2-TZVPD//B3LYPGD3BJ/def2-SVPD level of theory. Average dissociation times from GFN2-xTB trajectories are also reported in red. VIP: vertical ionization potential. [1]

of N3–C4. A third route (23% relative count, red pathway in fig. 3.22) involves only two bond breakings (N1–C2 and C4–C5), yielding therefore a larger neutral fragment (OCNCOH). Looking in detail at the first fragmentation pathway (fig. 3.10), uracil concerted ring opening at N3–C4 and N1–C2 allows the removal of a HNCO neutral fragment, on average 844 fs after the initial ionization (according to semi-empirical BOMD simulations). Then two hydrogen shifts occur, from uracil C6 to C4 and from the central CH2 to uracil C5. Finally, uracil C4–C5 bond breaking yields a neutral HCO fragment and the m/z 130 $H_8C_9N^+$ ion, at t = 2360 fs on average. Its calculated AE_{th} is 11.08 eV (if considering only the stability of the final fragmentation products) or $AE_{th,\,kin}$ = 12.14 eV, including the contribution of kinetic energy barriers for a better estimate. In the analysis of the experimental yield of 130 shown in fig. 3.15, the yield appears to display two slopes which indicate, consistently with the predictions of the simulations, two different channels leading to the formation of the m/z 130 with AE_{PEPICO} = 12.6 ± 0.5 and 13.4 ± 0.5 eV.

Looking now at the pathway leading to the m/z 116 fragment (fig. 3.7), dissociation at uracil C2–N3, C4–C5 and N1–C6 bonds releases two HNCO neutral fragments along with the charged one, whose AE_{th} = 11.49 eV is in agreement with AE_{PEPICO} = 11.6 ± 0.4 eV. The formation of the m/z 103

fragment involves first an H atom shift from uracil N3 to O(C4), then N1–C2 dissociation, removal of the former H atom from O(C4), transfer of another H atom from the central CH2 to C5, C6–C5 bond breaking with the dissociation of a neutral HNCH fragment, C4–C5 dissociation with OCNCO fragment release, and final bond formation between uracil C5 and benzene C2', leading to a new four-membered cycle and a condensed bicyclic fragment. The higher energy of the latter is confirmed by both AE_{th} = 14.23 eV (which could actually be even higher, if including energy barriers contribution) and AE_{PEPICO} = 16.8 ± 0.5 eV. The dissociation of the CH2–benzene C1' bond instead leads to the 5-methylenyl uracil radical and the m/z 77 phenyl cation.

6BU fragmentation

Two possible structural isomers of the original $6BU^{\cdot+}$ ion, m/z 202 in fig. 3.11, have been found in the MD simulations. The first one (AE_{th} = 11.55 eV) features the opening of the uracil ring, through a H atom shift from benzene C2' to uracil O(C4), C1'–C6' bond breaking and C2'–C6' bond formation, giving a smaller, five-membered ring, H atom transfer from uracil N1 to benzene C1', a H atom shift from the central CH2 to N1 and then to C3', C2–N3 dissociation, and a H atom shift from C1' to C5.

The other isomer is of particular interest, since it is a quite stable (AE_{th} = 7.91 eV), tricyclic form featuring a uracil–benzene crosslink. In fact, a uracil C5–benzene C2' bond can form (after 3.7 ps), followed by two H atoms shifts from benzene C2' to O(C4) and from C5 to C6 (see fig. 3.12). Including energy barrier contributions, an effective $AE_{th,\,kin}$ = 9.70 eV is nevertheless predicted. The approach of the uracil and benzene rings giving the anti-diastereomer appears more favorable than the syn approach (see fig. 3.23), with an energy barrier lower by 0.28 eV. In this regard, 5BU and 6BU isomers show very different behavior with respect to cyclization and cross-link: while in $5BU^{\cdot+}$ a high energy barrier of 1.80 eV (41.51 kcal mol^{-1}) has to be overcome to obtain a cross-linked isomer (see fig. 3.20), the intramolecular bond formation in $6BU^{\cdot+}$ has to overcome a barrier of only 0.76 eV (17.53 kcal mol^{-1}) as shown in fig. 3.12.

The simplest m/z 201 isomer, obtained from H atom removal from benzene

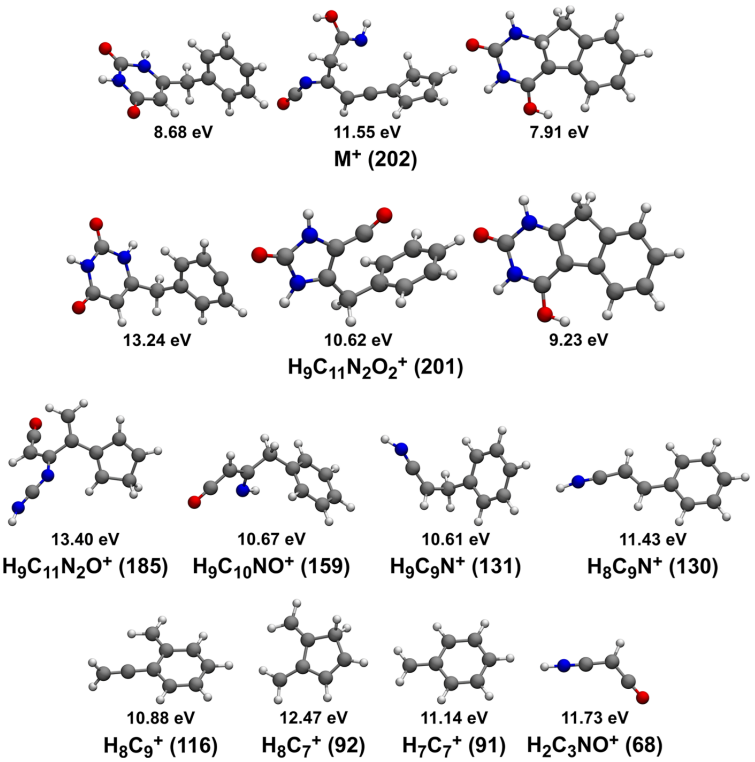


Fig. 3.11: 6BU fragments observed in GFN2-xTB BOMD trajectories. Formula, molecular mass, three-dimensional structure optimized at B3LYP-GD3BJ/def2-SVPD level of theory, and appearance energy at the B2PLYP-GD3BJ/def2-TZVPD level (calculated on final fragmentation products) are reported for each fragment. [1]

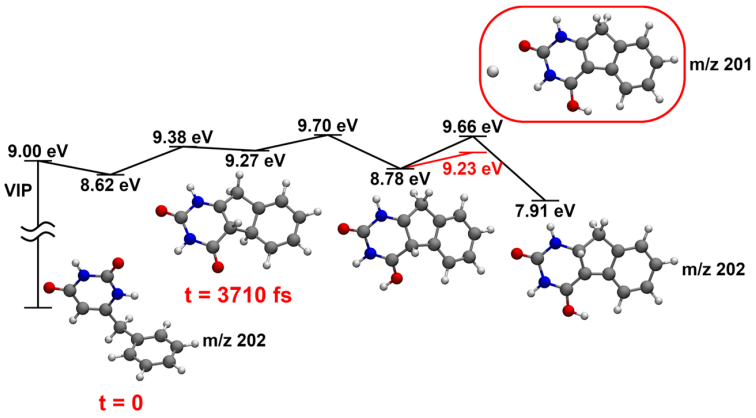


Fig. 3.12: Mechanism of $6BU^{\cdot+}$ isomerization, leading to the cyclic anti isomer with a uracil C5–benzene C2' cross-link. The alternative pathway leading to the m/z 201 cyclic fragment (highlighted in red) is also reported. The energy profile is evaluated at the B2PLYP-GD3BJ/def2-TZVPD//B3LYP-GD3BJ/def2-SVPD level of theory.[1]

C4', is the least stable (AE_{th} = 13.24 eV). Another observed m/z 201 isomer features a rearranged uracil ring, due to H atom dissociation from C5, N3–C4 bond breaking, and N3–C5 bond formation, leading to a five-membered ring. Its AE_{th} = 10.62 eV is close to AE_{PEPICO} = 10.1 ± 0.2 eV. An m/z 201 cyclic isomer (AE_{th} = 9.23 eV) has been also found, whose formation pathway involves a uracil C5–benzene C2' cross-link, a H atom shift from C2' to O(C4), as in the case of the m/z 202 cyclic isomer pathway (fig. 3.12), and a final dissociation step of the H atom on C5. The higher effective $AE_{th,kin}$ = 9.70 eV due to the energy barrier also makes this isomer compatible with the experimental AE.

The m/z 185 fragment, with AE_{th} = 13.40 eV, results from a rather complex mechanism, involving a H atom shift from N1 to O(C2), uracil C5–benzene C1' temporary bond formation, benzene C2'–C6', C1'–C6' bonds formation and cleavage, respectively, while a H atom moves from C2' to C3' and then to C4'. Then the C5–C1' and C6–CH2 bonds break, while the C6–C1' bond is formed; the N3–C4 bond is cleaved, and at last a OH radical is removed from C2. Dissociation of uracil N3–C4 and N1–C2 bonds instead leads to the release of a neutral HNCO fragment and to a charged m/z 159 one, whose AE_{th} = 10.67 eV agrees with the AE_{PEPICO} = 11.0 ± 0.3 eV. C4–C5, C2–N3

and N1–C2 uracil bond breakings (with CO release), N3–C5 bond formation, C6–CH2 dissociation, N3–C5 bond cleavage (with neutral HNCO dissociation), and final C5–CH2 bond formation results in the m/z 131 fragment, with AE_{th} = 10.61 eV, close to $AE_{PEPICO} = 11.2 \pm 0.2$ eV.

The fragmentation toward m/z 130 fragment, $AE_{th} = 11.43$ eV and $AE_{PEPICO} = 12.5 \pm 0.2$ eV, involves a H atom shift from uracil N1 to C6, a H2 molecule release from C5 and C6, another H atom shift from N3 to O(C2), C4–C5 bond cleavage, N1–C5 bond formation, N1–C6 cleavage, H atom transfer from O(C2) to N1, N1–C2 dissociation, releasing a neutral OCNCO fragment, and a H atom shift from the central CH2 to C6.

The m/z 116 fragment results from several bond cleavages at the uracil moiety, involving C4–C5, C2–N3 (releasing a neutral HNCO), N1–C6 (removing a second HNCO fragment) and C6–CH2; then a uracil C6–benzene C2' bond can form and a H atom is transferred from C2' to C6 and then to C5. The $AE_{th} = 10.88$ eV agrees with the $AE_{PEPICO} = 11.2 \pm 0.2$ eV.

The mechanism leading to the m/z 92 fragment, $AE_{th} = 12.47$ eV similar to $AE_{PEPICO} = 12.0 \pm 0.3$ eV, involves a benzene ring rearrangement through C2'–C4' and C3'–C4' bond formation and breaking, respectively, giving a five-membered ring, the transfer of H atoms from C5 to C3' and from C2' to C4', N3–C4 and N1–C2 bond cleavages with HNCO dissociation and final C6–CH2 bond breaking, releasing a neutral HNCCCO fragment.

The m/z 91 fragment observed in MD trajectories is the benzyl cation, obtained by C6–CH2 bond cleavage, also yielding the neutral uracil radical. The $AE_{th} = 11.14$ eV is lower than $AE_{PEPICO} = 13.5 \pm 0.2$ eV, but such underestimation is due to not including kinetic barrier contributions.

Since the m/z 68 ($H_2C_3NO^+$ fragment) is the most abundant fragment in both the 6BU mass spectrum and in the PEPICO measurements over a wide binding energy range (14–25 eV), its formation has been studied in detail. Uracil N1–C2, N3–C4, C6–C7 bond breakings represent the largely preferred (92% relative count) decomposition route (fig. 3.24, purple pathway). A less explored (8% relative count) pathway, however, shows cleavages at uracil C2–N3, C4–C5 and C6–CH2 bonds (fig. 3.24, green pathway).

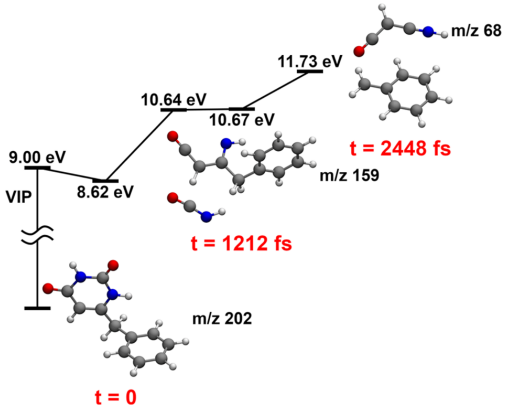


Fig. 3.13: Mechanism of the preferred $6BU^{\cdot+}$ fragmentation process leading to $H_2C_3NO^+$ m/z 68 ion. The energy profile is evaluated at the B2PLYP-GD3BJ/def2-TZVPD//B3LYP-GD3BJ/def2-SVPD level of theory. Average dissociation times from GFN2-xTB trajectories are also reported in red. VIP: vertical ionization potential. [1]

The mechanism of the former pathway has been further characterized (see fig. 3.13). Initial uracil N1–C2 and N3–C4 concerted bond breaking through a retro Diels–Alder step leads to the removal of a neutral HNCO (isocyanic acid) fragment, on average 1212 fs after 6BU ionization (according to semi-empirical BOMD simulations). Further uracil C6–CH2 bond breaking eventually yields the $H_2C_3NO^+$ (m/z 68) fragment ion (AE_{th} = 11.73 eV) and the neutral C_7H_7 benzyl radical, at t = 2448 fs on average. Moreover, it is interesting to note that the corresponding m/z 68 $H_2C_3NO^+$ fragment from 5BU (whose optimized structure is not linear, but cyclic, see fig. 3.25) is instead not found in the experimental mass spectrum and PEPICO measurements. In fact, its energy is predicted to be more than 3 eV higher than 6BU $H_2C_3NO^+$ at the B2PLYP-GD3BJ/def2-TZVPD//B3LYP-GD3BJ/def2-SVPD level.

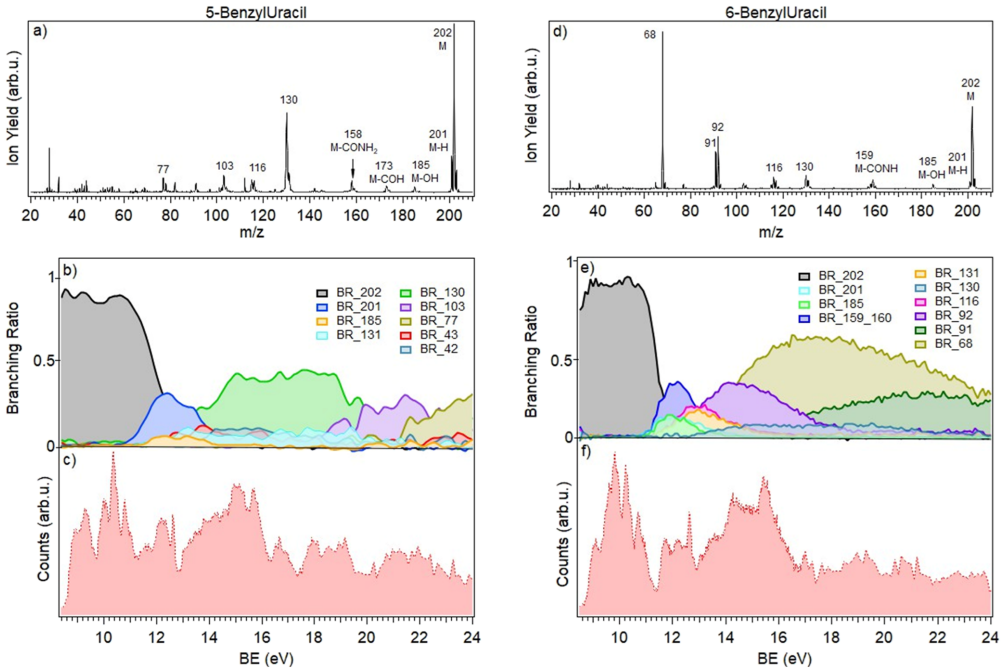


Fig. 3.14: Mass spectra [panels (a) and (d)], branching ratios of the main fragments obtained by the PEPICO measurements [panels (b) and (e)] and photoelectron spectra [panels (c) and (f)] of 5BU and 6BU (left and right panels, respectively). In the photoelectron spectra, the BE region above 18 eV for 5BU and 17 eV for 6BU has been measured with larger energy steps and lower energy resolution. The mass spectra are the sum of all PEPICO measurements up to BE = 24 eV. All measurements are performed at 60 eV incident photon energy. In the case of 5BU, due to the poor statistics of the raw PEPICO data, a three-point smoothing has been applied. [1]

Appendices

3.A PES and PEPICO experimental spectra

Figure 3.14 presents all experimental results for 5BU (left panels) and 6BU (right panels): mass spectra in the top row, branching ratios from the PEPICO measurements in the middle row, and PES spectra in the bottom row.

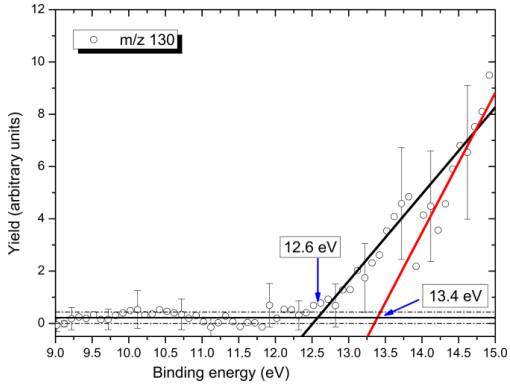


Fig. 3.15: The 5BU PEPICO yield of fragment m/z 130 in the threshold region and the results of the linear fits to extract the AE_{PEPICO} (shown by the arrows). The regions 12.3–13.7 and 13.9–15.0 eV have been considered for the two regions, respectively.[1]

3.B Conformational analysis

In order to perform accurate 5BU and 6BU molecular dynamics simulations, the general AMBER force field (GAFF2)[88] was chosen. GAFF2 atom types were assigned to 5BU and 6BU atoms. Atomic charges were evaluated as restricted electrostatic potential (RESP) charges at the HF/6-31G(d) level, according to the standard AMBER procedure.

The parameters of the relevant 5BU and 6BU torsional degrees of freedom (ϑ and φ, see section 1.2.2 for their definition), as well as those of the central B–CH_2–U (ζ_{BU}) bending term, were re-parametrized against high-level QM energy scans by minimizing the objective function:

$$f(\{p\}) = \sum_{j=1}^{N_{\text{geom}}} \left[(E_{\text{MM}}(\boldsymbol{r}_j, \{p\}) - E_{\text{QM}}(\boldsymbol{r}_j))^2 + K \right]$$

where $\{p\}$ are the MM parameters to be optimized, $\boldsymbol{r}_j$ are N_{geom} geometries (which should span the internal coordinates of interest), E_{MM} are the MM energies, E_{QM} are the QM reference energies, and K is an offset parameter (to be optimized first) accounting for the different "zeros" of the QM and MM energies (so that $f\{p\}$) could ideally become zero).

Table 3.1: Optimized 5BU GAFF2 parameters. Please refer to fig. 1 for atom labels.

Internal coordinate	Atom types	Parameter	Optimized value
C4-C5-C7-C1' (ϑ)	c-cc-c3-ca	$V_1/2$ (kcal mol^{-1})	0.6408
		$V_2/2$ (kcal mol^{-1})	−0.0408
		$V_3/2$ (kcal mol^{-1})	0.8668
		$V_4/2$ (kcal mol^{-1})	0.0010
C5-C7-C1'-C2' (φ)	cc-c3-ca-ca	$V_1/2$ (kcal mol^{-1})	−0.3970
		$V_2/2$ (kcal mol^{-1})	−0.3412
		$V_3/2$ (kcal mol^{-1})	0.0008
		$V_4/2$ (kcal mol^{-1})	0.0136
C5-C7-C1' (ζ_{BU})	cc-c3-ca	k_ζ (kcal mol^{-1} rad^{-2})	64.4028
		ζ_{eq} (°)	114.5572

B2PLYP/6-311++G(d,p) energies were evaluated on 5BU and 6BU structures, relaxed with the original GAFF2 force field, along a two-dimensional (ϑ, φ) grid, spanning the full 360° range, with 10° steps (i.e., [−180°:10°:180°]). Each dihedral term has a Fourier-series functional form:

$$\sum_{n=1}^{4} \frac{V_n}{2}[1 + \cos(n\phi - \gamma)]$$

The four $V_n/2$ torsional energy barriers for each dihedral were simultaneously optimized. Analogously, QM B2PLYP/6-311++G(d,p) energies were calculated on GAFF2-relaxed structures, spanning the [105°:1°:125°] range of the central B–CH_2–U angle. Angle MM terms have a quadratic functional form:

$$k_\zeta(\zeta - \zeta_{\mathrm{eq}})^2$$

Both k_ζ and ζ_{eq} were optimized. Optimized 5BU and 6BU MM parameters are reported in table 3.1 and table 3.2, respectively. The Paramfit program[211], based upon a genetic-type, global optimization, algorithm, was employed for this re-parametrization.

A system composed of 5BU (or 6BU) and 40 Ar atoms (non-bonding parameters from ref. 212) in a $\approx 55 \times 54 \times 54$ °cell was heated up to the experimental

Table 3.2: Optimized 6BU GAFF2 parameters. Please refer to Figure 1 in main text for atom labels.

Internal coordinate	Atom types	Parameter	Optimized value
N1-C6-C7-C1' (ϑ)	ns-cc-c3-ca	$V_1/2$ (kcal mol^{-1})	0.1111
		$V_2/2$ (kcal mol^{-1})	−0.1813
		$V_3/2$ (kcal mol^{-1})	0.8893
		$V_4/2$ (kcal mol^{-1})	−0.0006
C6-C7-C1'-C2' (φ)	cc-c3-ca-ca	$V_1/2$ (kcal mol^{-1})	−0.3540
		$V_2/2$ (kcal mol^{-1})	0.0803
		$V_3/2$ (kcal mol^{-1})	−0.1437
		$V_4/2$ (kcal mol^{-1})	0.0194
C6-C7-C1' (ζ_{BU})	cc-c3-ca	k_ζ (kcal mol^{-1} rad^{-2})	58.6604
		ζ_{eq} (°)	112.2741

temperature (5BU: 406 K, 6BU: 378 K) in 425 ps and 380 ps, respectively (i.e., at a $\approx 1\,\mathrm{K\,ps_{-1}}$ rate). Equilibration was carried out for 10 ns and then production trajectories were collected for 300 ns. Langevin dynamics was performed to sample the canonical ensemble, employing a velocity Verlet integrator with a 1.0 fs time step.

5BU and 6BU trajectories analyzed in the (ϑ, φ) space reveal six distribution maxima (see the two-dimensional distribution functions in figs. 3.16 and 3.17). Actually, they can be reduced to two distinct conformers (hereafter called A and B), if enantiomeric or completely superimposable structures are excluded. Geometries corresponding to (ϑ, φ) distribution maxima were further optimized at B3LYP/6-311++G(d,p) level of theory. A and B conformers mainly differ in the relative orientation of the two rings (being more orthogonal in A conformers and more parallel in B ones, see fig. 3.2) and so in the (ϑ, φ) dihedrals values (see tables 3.3 and 3.4): for 5BU, they correspond to the conformers described in methanol at room temperature (see section 1.2.2). A less than 1 kcal mol^{-1} relative stability is estimated (see tables 3.3 and 3.4): although one conformer is preferred (A by 5BU, B by 6BU), the other one is non-negligibly populated and therefore may contribute to the overall (for instance, spectroscopic) observed properties.

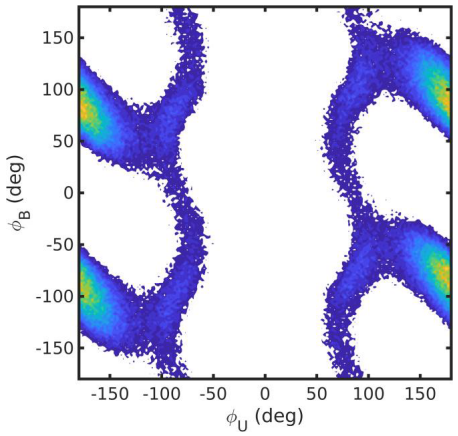


Fig. 3.16: Two-dimensional (ϑ, φ) distribution function from 5BU MD trajectory at 406 K. Populations are reported through a color scale. [1]

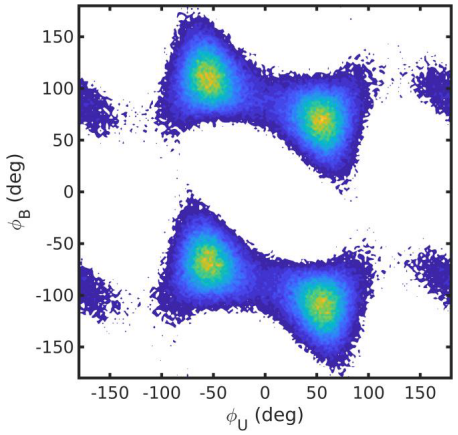


Fig. 3.17: Two-dimensional (ϑ, φ) distribution function from 6BU MD trajectory at 378 K. Populations are reported through a color scale. [1]

Table 3.3: Relevant structural parameters of 5BU conformers found by MD sampling.

	conf. A	conf. B
ϑ (°)	179.96	85.46
φ (°)	−89.68	−79.20
$E_k - E_A$ (kcal mol^{-1})	0	0.8875
N_k/N	0.60	0.40

Table 3.4: Relevant structural parameters of 6BU conformers found by MD sampling.

	conf. A	conf. B
ϑ (°)	180.00	50.07
φ (°)	89.96	67.99
$E_k - E_A$ (kcal mol^{-1})	0	-0.9441
N_k/N	0.12	0.88

3.C Additional PES data

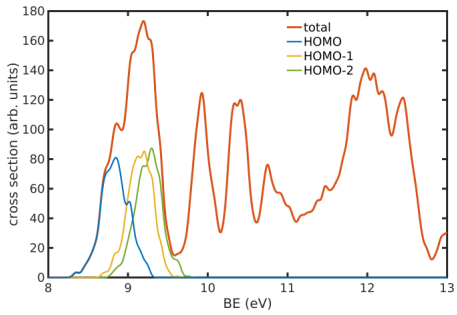


Fig. 3.18: 5BU photo-electron cross-section from HOMO, HOMO-1 and HOMO-2 contributing to the first band of the total photo-electron spectrum. [1]

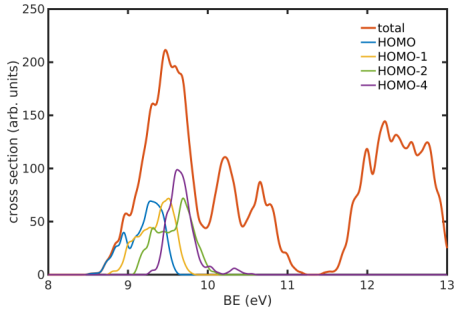


Fig. 3.19: 6BU photo-electron cross-section from HOMO, HOMO-1, HOMO-2 and HOMO-4 contributing to the first band of the total photo-electron spectrum. [1]

Table 3.5: Characterization of the experimental 5BU photo-electron spectrum in the 8–13 eV range. For each band maximum in the experimental spectrum, the corresponding maximum in the calculated spectrum from MD sampling, the corresponding ionized MOs, their calculated binding energies (BEs), and spatial localizations (as MO-resolved Hirshfeld population analysis[213]) are reported. Fragments labels: U = uracil, B = benzene, CH_2 = central CH_2 moiety.

Exp (eV)	Calc (eV)	Ionized MO (conformer)	BE (eV)	U char	B char	CH_2 char
8.94	8.86	h (B)	8.85	0.10	0.87	0.03
9.28	9.19	h-1 (B)	9.10	0.01	0.99	0.00
	9.16	h-1 (A)	9.16	0.33	0.64	0.03
	9.17	h-2 (B)	9.17	0.82	0.11	0.07
	9.26	h (A)	9.26	0.04	0.92	0.04
	9.32	h-2 (A)	9.32	0.58	0.36	0.06
10.00	9.93	h-4 (A)	9.94	0.97	0.01	0.02
	9.98	h-4 (B)	9.98	0.96	0.02	0.02
10.36	10.41	h-3 (A)	10.37	1.00	0.00	0.00
	10.42	h-3 (B)	10.42	0.99	0.01	0.00
10.81	10.76	h-5 (A)	10.84	0.97	0.02	0.01
	10.93	h-5 (B)	10.93	0.94	0.05	0.01
11.70	11.82	h-6 (B)	11.80	0.03	0.90	0.07
	11.86	h-7 (B)	11.86	0.13	0.74	0.13
12.19	11.98-12.44	h-8 (B/A)	11.91-12.24	0.08-0.87	0.05-0.87	0.05-0.17
12.43	12.43–12.61	h–9 (A/B)	12.43–12.61	0.74-0.82	0.16-0.21	0.02-0.05

3.D Additional fragmentation pathways

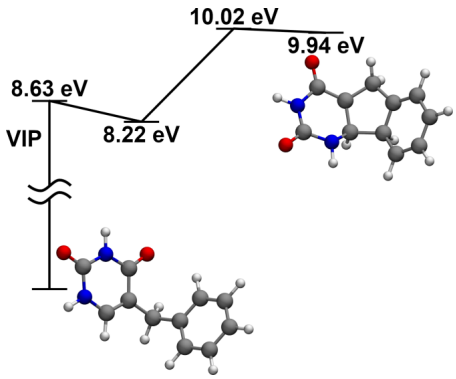


Fig. 3.20: Mechanism of $5BU^{\cdot+}$ isomerization, leading to the cyclic syn isomer with a uracil C6–benzene C2' cross-link.[1]

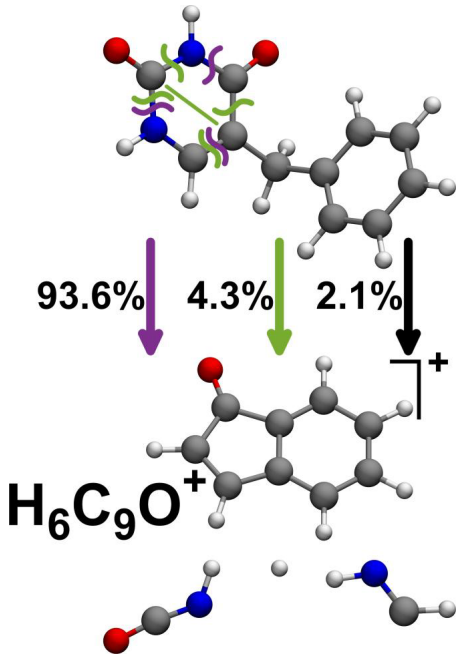


Fig. 3.21: Most common pathways leading to $H_6C_9O^+$ ion from 5BU molecular cation, extracted from GFN2-xTB MD simulations.Wavy lines indicate broken bonds, straight lines formed ones. Pathways are labeled by different colors. Relative counts of each pathway with respect to the total count of all the pathways leading to $H_6C_9O^+$ are reported.[1]

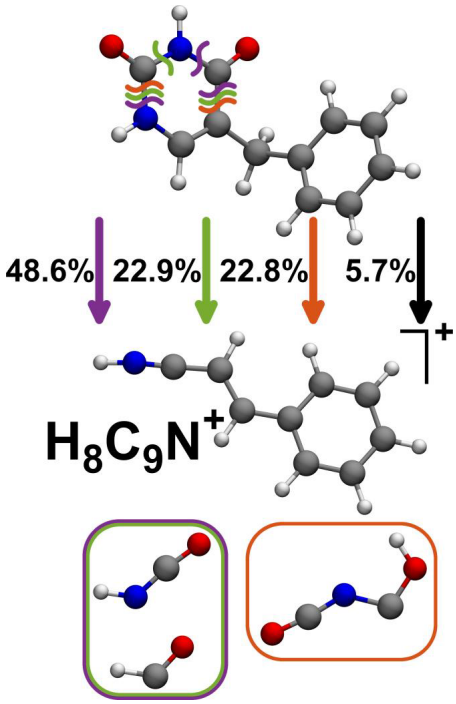


Fig. 3.22: Most common pathways leading to $H_8C_9N^+$ ion from 5BU molecular cation, extracted from GFN2-xTB MD simulations. Wavy lines indicate broken bonds. Pathways are labeled by different colors. Relative counts of each pathway with respect to the total count of all the pathways leading to $H_8C_9N^+$ are reported. Neutral fragments formed in each pathway are labeled with the corresponding color. [1]

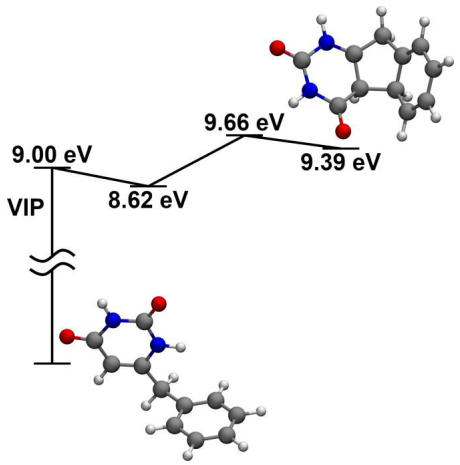


Fig. 3.23: Mechanism of $6BU^{\cdot+}$ isomerization, leading to the cyclic syn isomer with a uracil C5–benzene C2' cross-link.[1]

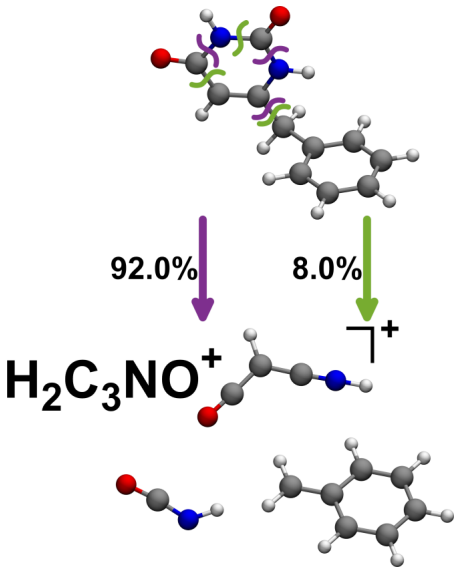


Fig. 3.24: Most common pathways leading to $H_2C_3NO^+$ ion from 6BU molecular cation, extracted from GFN2-xTB MD simulations.Wavy lines indicate broken bonds. Pathways are labeled by different colors. Relative counts of each pathway with respect to the total count of all the pathways leading to $H_2C_3NO^+$ are reported.[1]

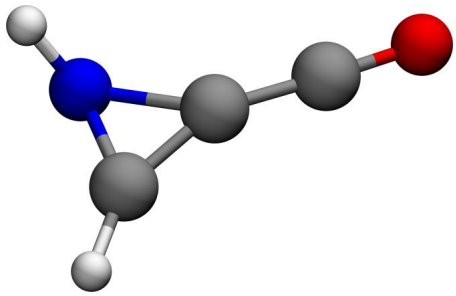


Fig. 3.25: Optimized structure of hypothetical 5BU $H_2C_3NO^+$ m/z 68 fragment, not found in semi-empirical MD simulations. Its energy is 3.06 eV higher than the 6BU corresponding fragment, at the B2PLYP-GD3BJ/def2-TZVPD//B3LYP-GD3BJ/def2-SVPD level of theory.[1]

Chapter 4

Integrating nonadiabatic dynamics with multilayer extrapolative schemes: towards a new Newton-X/Gaussian16 interface

This final chapter brings together two main research lines developed throughout this PhD project: (i) the efficient simulation of complex chemical environments via multilayered partition methods, discussed in chapter 1, and (ii) the simulation of photoinduced nonadiabatic processes via Trajectory Surface Hopping, introduced in chapter 2.

Many chemical systems simultaneously face both challenges - namely, the coupling among multiple potential energy surfaces (PESs) and the interaction between each PES and the surrounding environment.

In detail, this chapter presents the development of a new interface between Gaussian 16 and Newton-X New Series (NX NS),[147] designed to enable trajectory surface hopping nonadiabatic dynamics (NAD) combined with ONIOM

(acronym of "**O**ur own **N**-layered **I**ntegrated molecular **O**rbital and **M**olecular mechanics") extrapolative scheme.[214] The theoretical principles underlying the ONIOM approach were discussed in section 1.1.2, while the fundamentals of Surface Hopping were outlined in section 2.2.1.

Although these methodologies address different aspects of molecular simulations, ONIOM and Newton-X share a common philosophy: both are designed to efficiently perform a specific task while delegating the remaining part of the computations (the evaluation of energy and forces) to the electronic structure program suite with which they are interfaced.

In the case of ONIOM, this means that any computational method and software can be employed to compute the energy of a particular term in the extrapolative eq. 1.10. Conversely, NX NS, is responsible exclusively for propagating the molecular dynamics and to compute the electronic state hoppings, and can be combined with any external program able to compute energy and forces, once a suitable interface has been developed.

The NX NS program represents the latest evolution of the Newton-X package. Compared to the previous NX CS version, NX NS is entirely written in Fortran programming language, which significantly reduces disk-writing operations and thereby enhances computational efficiency. Moreover it is devoted to the propagation of trajectories and the computation of nonadiabatic couplings, while the generation of initial conditions is delegated to NX CS and the post-processing analysis is handled by ULaMDyn, a Python-based, open-source package designed to automate the unsupervised analysis of large datasets generated by NAD simulations.[215]

Regarding ONIOM, a major advantage of the extrapolative scheme lies in its flexibility and modularity. Gaussian 16 supports both two- and three-layer ONIOM calculations, enabling the combination of diverse theoretical methods (*e.g.*, QM/MM, QM/QM', QM/QM'/MM) to achieve an optimal balance between accuracy and computational cost. Unlike additive QM/MM approaches, which require an explicit treatment of QM–MM interaction terms, ONIOM implicitly includes the coupling between the layers through the low-level calculation on the entire system. In the case of ONIOM3 (*model*/*intermediate*/*real*

system), each layer is coupled with the other two. This strategy simplifies the implementation and minimizes the need for empirical parameterization of interaction potentials. Furthermore, the subtractive nature of the method mitigates the problems of to over- and under-counting of interactions between the layers.

From a practical standpoint, Gaussian 16 offers a mature and user-friendly implementation of the ONIOM methodology. [214,73,216–218] The software provides automated link-atom insertion, flexible layer assignment, and analytic gradients and Hessians, thus enabling geometry optimizations, transition-state searches, and vibrational frequency analyses directly at the ONIOM level. Micro-iterative optimization protocols are also available to reduce the number of high-level evaluations required during structural relaxation. In addition, Gaussian 16 supports multiple embedding options, including mechanical and electrostatic embedding, thereby allowing the high-level region to be polarized by the surrounding environment when desired.

The interface developed in this work represents a preliminary proof of concept for enabling Newton-X to perform QM/MM simulations with a commercially available electronic structure software, and outlines a possible strategy for extending such coupling scheme to QM/QM' and multilayered ONIOM computations.

4.1 The Newton-X New Series / Gaussian16 ONIOM interface

The present interface represents an upgrade of the previous NewtonX/Gaussian interface. As in the earlier implementation, Newton-X New Series manages the time propagation and the surface hopping algorithm. At each step of the dynamics, NX NS invokes the Gaussian program to compute the potential energies and corresponding gradients.

The main modifications concern the subroutines dedicated to parsing the Gaussian input, updating the atomic coordinates at each step, and reading energies and forces from the Gaussian output.

The Gaussian input parser has been reformulated to accept any properly

formatted Gaussian input file, avoiding the previous additional formatting requirements(for instance, the previous version did not parse correctly the input file if there was a blank line before the command line). In the command line, the parser now recognizes both the ONIOM keyword and the electronic structure method for the excited-state calculations. It rewrites the route section at each step changing exclusively the specification of the current root state. Within the molecule specification, the parser recognizes and saves separately the atomic coordinates, element symbols, options, and the ONIOM layer definition. This new design allows the program to update the coordinates at each step while preserving all the additional information provided to Gaussian, including user-defined sections such as those containing custom molecular mechanics (MM) parameters. Although these improvements were introduced for ONIOM-based applications, they generally enhance the usability of Gaussian within the Newton-X framework.

On the other end, the NX NS subroutine dedicated to reading energies and gradients from the Gaussian output had already the ability to sum the ground-state energy with the VEE of each transition yielding the excited state potential energies. The new version extend this functionality to ONIOM outputs, it can recognize the ground-state energy of the *model system* and, for each transition, sum the corresponding vertical excitation energy with the energies of the other addenda of the extrapolative eq. 1.10.

For each ONIOM layer, the interface supports any excited-state method compatible with ONIOM framework including CIS, TD-DFT (also in Tamm-Dancoff Approximation, TDA), in both restricted and unrestricted forms as well as all ground- and excited-state methods previously implemented in the Gaussian/NX CS interface. The implemented flexibility allows the execution of QM/MM, QM/QM', and even QM/QM'/MM simulations. In addition, the interface has also been successfully tested in combination with implicit solvation models, thus extending its compatibility with continuum embedding approaches.

Despite these advantages, the interface naturally presents the intrinsic limitations of the ONIOM methodology. Reliably results cannot be expected for

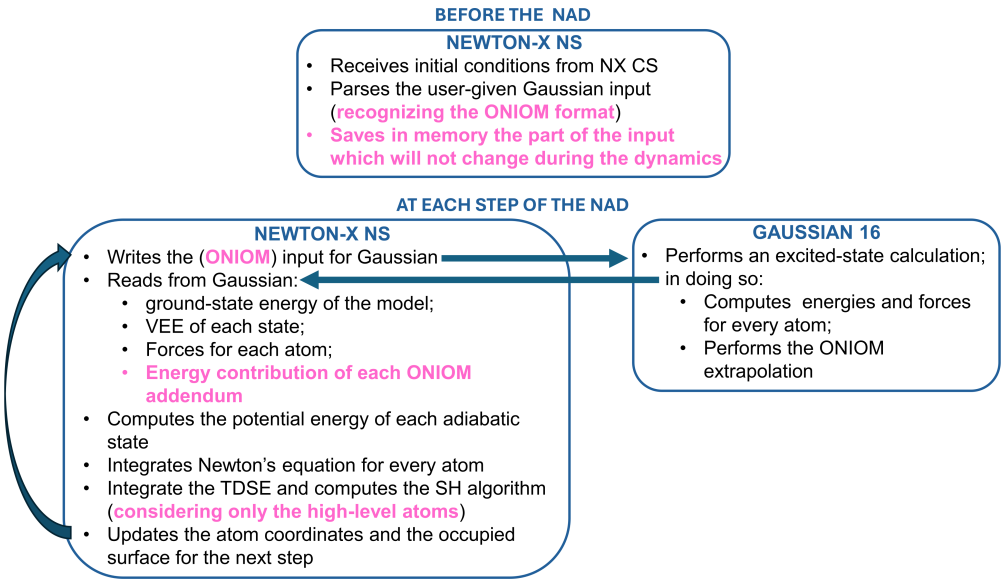


Fig. 4.1: Worflow of the NX NS-Gaussian 16 ONIOM interface.

charge-transfer excitations that involved different layers. Moreover, multi-layer TD calculations has not been implemented here, since managing multiple excited states across layers would be impractical within a surface-hopping framework. Like any other NX interface, moreover, intersystem crossing processes are not currently treated.

Finally, the interface is entirely integrated in the NX NS and requires no additional script or auxiliary programs, only NX NS, Gaussian 16 and their standard dependencies. However, some applications require an additional equilibration step(s) via classical MD to generate initial conditions: in this case it is highly advisable to use a dedicated suite; more information are given in section 4.2.

4.2 Generation of initial conditions in the explicit solvent

The most important and challenging application of this new interface will concern the possibility to collect swarm of non-adiabatic molecular dynamics trajectories of molecular systems embedded in large solvent clusters.

However, this brings with it a new problem: generating meaningful initial conditions (in terms of coordinates and momenta) for a solute in a solvent cluster. In the research group that hosted my PhD project, this problem was thoroughly addressed in the context of ab-initio adiabatic molecular dynamics for sampling excited-state PES and study photo-relaxation phenomena. A summary can be found in the Ref. 31. Newton-X CS v. 2.6 allows to generate initial conditions via three different approaches:[219]

1. Generate configurations according to an harmonic oscillator distribution (classical, Wigner, or quantum correlated), based on normal modes;
2. Generate random velocities from a single structure;[220]
3. Sample random configuration from a previous molecular dynamics.

For molecular systems embedded in large solvent clusters, any approach relying on a single, static configuration - such as minimum energy structure - is intrinsically limited, as it fails to capture the extensive configurational degrees of freedom of the solvent environment. In particular, the global minimum of a large solvent cluster, besides being computationally prohibitive to locate, is not representative of the statistical distribution of solvent molecules in the liquid phase. Additionally, geometry optimization algorithms inherently tend to maximize intermolecular solute-solvent and solvent-solvent interactions, thus producing artificially compact structures that poorly resemble the realistic liquid-like arrangement.

An alternative strategy could involve generating solute conformations according to Wigner distribution and subsequently solvating each of the individually. However, such Wigner-generated geometries possess an excess of kinetic

energy, which would be transferred to the surrounding solvent upon solvation, thereby perturbing the natural solute-solvent interaction pattern and compromising the physics of the resulting ensemble.

On the other hand, selecting random points from a thermalized ground-state molecular dynamics can give a physical sampling of the solute-solvent intermolecular interactions; but, since the nuclei are treated as classical particles, the solute will not have enough kinetic energy to populate the higher frequency vibrational modes, altering the distribution of the initial conditions.

We therefore developed a new protocol, specifically designed for a solute solvated by a significant number of solvent molecules, which includes all the steps for the preparation of the entire system with the aim of overcoming the limitations of current approaches. Our approach involves the following stages:

1. *Geometry optimization of the solute*: the ground-state minimum energy structure of the isolated solute computed at QM level of theory will serve as the reference structure around which explicit solvent molecules will be arranged.

2. *Construction of explicit solvent shells*: generate an explicit solvent sphere that solvates the optimized solute. The solvent sphere should include enough solvent shells to adequately represent short- and medium-range solute–solvent interactions for the timescale of the dynamics. Packing tools such as PackMol[221] or molecular-modeling utilities in the AmberTools suite (useful to generate the input within the NAMD program);[222] ensure that there are no severe steric overlaps and that the solvent composition and density are consistent with the intended thermodynamic conditions.

3. *Classical equilibration*: perform a molecular mechanics (MM) equilibration trajectory to thermalize the solute and solvent system, following a standard MM equilibration protocol and employing periodic boundary conditions (PBC). The purpose of this stage is:

   (a) to remove the correlation between the initially generated configuration and the first step of the production dynamics, so that the

following step of MD sampling can benefit from the ergodic principle;

(b) to redistribute thermal energy between the solute (which was optimized) and the solvent;

(c) to relax any packing artifacts.

4. *Classical production and sampling*: run an additional MD trajectory long enough to meaningfully sample the solvent distribution; from this trajectory, extract a set of configurations, chosen to be statistically independent. A good option is to randomly vary the time between sampled frames, to avoid accidentally matching any low-frequency modes.

5. *ONIOM re-equilibration*: from each random frame selected in item 4, a short ground-state ONIOM(QM/MM) trajectory is initiated. The purpose of this stage is threefold:

   (a) to further sample the solute conformational space at the QM level;

   (b) to allow the solvent molecules to reorganize according to the updated solute electron density and to the new QM solute conformations;

   (c) to re-thermalize the solvent environment around the QM subsystem.

   This step requires particular care. Periodic boundary conditions (PBC) are removed to enable the use Gaussian basis functions and hybrid DFT functionals within the ONIOM framework. Unfortunately, the non-periodic boundary conditions (NPBC) described in Section 1.1.2 cannot be applied here, since the NPBC potential is not yet implemented in Gaussian 16 for these purposes. Nevertheless, excited state QM/MM molecular dynamics typically span short time scales (hundreds of femtoseconds). Therefore, provided that the solvent sphere is sufficiently large, boundary artifacts —propagating from the outer surface towards the solute region — will not significantly affect the inner solvation shells during the production runs. For this reason, no explicit confinement or boundary potential is used; instead, the solvent sphere size is simply increased to mitigate edge

effects. Running multiple short QM/MM trajectories rather than a single long one offers two advantages: it allows straightforward parallelization and further reduces the accumulation of boundary artifacts.

6. *Generation of initial conditions*: from the set of trajectories produced in the previous item 5, random frames are selected as candidates for the initial conditions. For each selected configuration, a TD-DFT calculation is performed to initialize them in the chosen excited state(s). If necessary the absorption spectrum can be computed from the sampled geometries. When energy constraints are imposed, only those configurations falling within the desired energy window are retained as valid initial conditions. Finally, before the production phase, a constrained geometry optimization of the solute is carried out to minimize any spurious leakage of thermal energy between solute and solvent.

This multi-stage protocol separates the equilibration of bulk solvent (performed at the computationally inexpensive MM level) from electronic-structure refinement of the solute (performed at the QM level within an explicit solvent environment). By sampling MM production frames and then allowing short ONIOM QM/MM ground-state relaxations, the procedure produces an ensemble in which solvent configurations are both thermally equilibrated and consistent with the instantaneous QM electronic structure of the solute. The subsequent vertical excitation calculations and local optimizations ensure that each trajectory is initiated from a physically meaningful nuclear and electronic configuration suitable for surface-hopping nonadiabatic dynamics.

Publications List

This thesis' work is based upon the following papers:

1. G. Iuzzolino, F. Perrella, M. Valadan, A. Petrone, C. Altucci, and N. Rega, *Photophysics of a nucleic acid–protein crosslinking model strongly depends on solvation dynamics: an experimental and theoretical study.* Physical Chemistry Chemical Physics, 26:11755–11769, **2024**. doi: 10.1039/D3CP06254F;

2. M. Valadan, F. Perrella, L. Carlini, G. Iuzzolino, J. Chiarinelli, F. Coppola, R. Richter, C. Schiano, A. Petrone, P. Bolognesi, L. Avaldi, C. Altucci, and N. Rega, *Dynamics of energy-selected fragments from nucleobase–protein interaction molecular models revealed by combined photoelectron–photoion coincidence spectroscopy and theoretical simulation.* The Journal of Chemical Physics, 163 (3):034305, 07 **2025**. doi: 10.1063/5.0276616;

3. G. Iuzzolino, F. Perrella, and N. Rega, *Paper in preparation*;

4. G. Iuzzolino *et al.*, *Paper in preparation.*

List of Figures

1 Three-dimensional structures of 5- and 6-benzyluracil, with atom labels.[1] 2
2 Photo-cyclization of 5-benzyluracil to 1,2-indaneuracil.[28] 4

1.1 Multi-layer partition scheme of the ground state AIMD 11
1.2 Multi-layer partition scheme of the TD-DFT calculations 16
1.3 The two rotation described by the dihedrals ϑ and φ.[28] 17
1.4 5BU conformers optimized at the B3LYP/6-31G(d,p) level.[28] 19
1.5 5BU conformers optimized at the CAM-B3LYP/6-31+G(d,p) level.[28] . 20
1.6 5BU total energy as a function of its torsional degrees of freedom. . . . 23
1.7 (ϑ, φ) two-dimensional dihedrals distribution from 5BU AIMD. 25
1.8 RDFs (dashed black) of methanol oxygen atoms - 5BU heteroatoms distances 27
1.9 Distributions of the distance of each 5BU heteroatom from the closest solvent oxygen verifying H_{MeOH}-O_{MeOH}-O or H-N-O_{MeOH} $< 40°$. 28
1.10 TD-CAM-B3LYP/6-31+G(d,p) UV-Vis absorption spectrum of 5BU in methanol. 29
1.11 Distributions of S_1 oscillator strength values among groups of frames taken from the AIMD.[28] 34
1.12 Distributions of S_2 oscillator strength values among groups of frames taken from the AIMD.[28] 35
1.13 NTO isodensity plots of S_1 in gas. 37
1.14 NTO isodensity plots of S_2 in gas. 38
1.15 NTO isodensity plots of S_1 in implicit methanol. 38
1.16 NTO isodensity plots of S_2 in implicit methanol. 39
1.17 NTO isodensity plots of S_1 in methanol cluster. 39
1.18 NTO isodensity plots of S_2 in methanol cluster. 40

1.19 Representative configurations extracted from AIMD. 42
1.20 NTO isodensity plots of the α configuration. 43
1.21 NTO isodensity plots of the β configuration. 43
1.22 NTO isodensity plots of the γ configuration. 44
1.23 NTO isodensity plots of the δ configuration. 44
1.24 NTO isodensity plots of the ϵ configuration. 45

2.1 Slices of the FlUC spectrum of 5BU 48
2.2 Experimental kinetic traces of 5BU fluorescence. 48
2.3 Values of the fast kinetic constant obtained from each fit, compared with the starting time of the fit. 61
2.4 TD-CAM-B3LYP/6-31G(d,p) average emission spectra of 5BU. 62
2.5 Share of trajectories occupying each adiabatic state. 64
2.6 Average oscillator strength of each excited state. 65
2.7 Average energy of each excited state. 67
2.8 Classical puckering form of the uracil ring in time. 68
2.9 Classical puckering form of the phenyl ring in time. 70
2.10 Values of the ϕ dihedral angle for each trajectory in time. 70
2.11 Values of the θ dihedral angle for each trajectory in time. 71

3.1 Protocol to simulate photoionization and high-energy photoreactivity . . 75
3.2 A and B conformers 78
3.3 Experimental and simulated 5BU valence photoelectron spectrum 79
3.4 5BU photoionized electrons as MOs isosurfaces 79
3.5 Experimental and simulated 6BU valence photoelectron spectrum 80
3.6 6BU photoionized electrons as MOs isosurfaces 81
3.7 5BU fragments observed in GFN2-xTB BOMD trajectories 83
3.8 Mechanism of $5BU^{\cdot+}$ decomposition toward m/z 185 cyclic fragment . . 84
3.9 Mechanism of $5BU^{\cdot+}$ fragmentation toward $H_6C_9O^+$ m/z 130 ion 85
3.10 Mechanism of $5BU^{\cdot+}$ fragmentation toward $H_8C_9N^+$ m/z 130 ion 87
3.11 6BU fragments observed in GFN2-xTB BOMD trajectories 89
3.12 Mechanism of 6BU·+ isomerization toward the cyclic anti isomer 90
3.13 Mechanism of 6BU·+ fragmentation toward the $H_2C_3NO^+$ m/z 68 ion . 92
3.14 Mass, photoelectron and PEPICO spectra of 5- and 6BU 93
3.15 The 5BU PEPICO yield of fragment m/z 130 94
3.16 (ϑ, φ) distribution from 5BU MD 97
3.17 (ϑ, φ) distribution from 6BU MD 97

3.18 5BU photo-electron cross-section from HOMO, HOMO-1 and HOMO-2 contributing to the first band of the total photo-electron spectrum.[1] . . 99
3.19 6BU photo-electron cross-section from HOMO, HOMO-1, HOMO-2 and HOMO-4 contributing to the first band of the total photo-electron spectrum.[1] 99
3.20 Mechanism of $5BU^{\cdot+}$ isomerization toward the cyclic syn isomer 101
3.21 Most common pathways leading to $H_6C_9O^+$ ion from $5BU^+$ 102
3.22 Most common pathways leading to $H_8C_9N^+$ ion from $5BU^+$ 103
3.23 Mechanism of $6BU^{\cdot+}$ isomerization toward the cyclic syn isomer 104
3.24 Most common pathways leading to $H_2C_3NO^+$ ion from $6BU^+$ 104
3.25 Optimized structure of hypothetical 5BU $H_2C_3NO^+$ m/z 68 fragment . 105

4.1 Worflow of the NX NS-Gaussian 16 ONIOM interface. 110

List of Tables

1.1 Values of the ϑ and φ dihedral angles for 5BU minimum energy structures 21
1.2 VEE and oscillator strengths of the first four electronic states of 5BU, from minimum energy structures 22
1.3 Analysis of S_1 and S_2 for the representative configurations sampled from 5BU AIMD 30
1.4 Average VEE and oscillator strenght of each transition of 5BU in methanol; average of the first computed band and peak of the first experimental band 31
1.5 Average S_1 oscillator strength for every considered cluster of AIMD frames 33
1.6 Average S_2 oscillator strength for every considered cluster of AIMD frames 33
1.7 VEE (eV) and oscillator strengths of the first four transitions of 5BU, for CAM-B3LYP/6-31+G(d,p) minimum energy structures 36

2.1 Parameters of the experimental fit of 5BU fluorescence kinetic trace . . 49
2.2 Number of trajectories initialized in each conformer and excited state . 53

3.1 Optimized 5BU GAFF2 parameters 95
3.2 Optimized 6BU GAFF2 parameters 96
3.3 Relevant structural parameters of 5BU conformers found by MD sampling. 97
3.4 Relevant structural parameters of 6BU conformers found by MD sampling. 98
3.5 Characterization of the experimental 5BU photo-electron spectrum . . . 100

Bibliography

[1] M. Valadan, F. Perrella, L. Carlini, G. Iuzzolino, J. Chiarinelli, F. Coppola, R. Richter, C. Schiano, A. Petrone, P. Bolognesi, L. Avaldi, C. Altucci and N. Rega, *The Journal of Chemical Physics*, 2025, **163**, 034305.

[2] R. Berisio, F. Schluenzen, J. Harms, A. Bashan, T. Auerbach, D. Baram and A. Yonath, *Nat. Struct. Biol.*, 2003, **10**, 366–370.

[3] A. G. Matera, R. M. Terns and M. P. Terns, *Nat. Rev. Mol. Cell Biol.*, 2007, **8**, 209–220.

[4] T. A. Steitz, *Nat. Rev. Mol. Cell Biol.*, 2008, **9**, 242–253.

[5] T. Glisovic, J. L. Bachorik, J. Yong and G. Dreyfuss, *FEBS Lett.*, 2008, **582**, 1977–1986.

[6] L. A. Yates, C. J. Norbury and R. J. Gilbert, *Cell*, 2013, **153**, 516–519.

[7] M. W. Hentze, A. Castello, T. Schwarzl and T. Preiss, *Nat. Rev. Mol. Cell Biol.*, 2018, **19**, 327–341.

[8] G. L. Hager, C. Elbi and M. Becker, *Curr. Opin. Genet. Dev.*, 2002, **12**, 137–141.

[9] D. J. Patel and Z. Wang, *Annu. Rev. Biochem.*, 2013, **82**, 81–118.

[10] T. Siggers and R. Gordân, *Nucleic Acids Res*, 2014, **42**, 2099–2111.

[11] Q. Du, P. L. Luu, C. Stirzaker and S. J. Clark, *Epigenomics*, 2015, **7**, 1051–1073.

[12] C. D. Allis and T. Jenuwein, *Nat Rev Genet*, 2016, **17**, 487–500.

[13] K. Smith, B. Hodgkins and M. O'Leary, *Biochim. Biophys. Acta, Nucleic Acids Protein Synth.*, 1966, **114**, 1–15.

[14] J. G. Peak and M. J. Peak, *Mutat. Res. Fundam. Mol. Mech. Mutagen.*, 1991, **246**, 187–191.

[15] S. Barker, M. Weinfeld and D. Murray, *Mutat. Res. Rev. Mutat. Res.*, 2005, **589**, 111–135.

[16] T. Nakano, X. Xu, A. M. Salem, M. I. Shoulkamy and H. Ide, *Free Radical Biol. Med.*, 2017, **107**, 136–145.

[17] L. Zhang, K. Zhang, R. Prändl and F. Schöffl, *Biochem. Biophys. Res. Commun.*, 2004, **322**, 705–711.

[18] C. Altucci, A. Nebbioso, R. Benedetti, R. Esposito, V. Carafa, M. Conte, M. Micciarelli, L. Altucci and R. Velotta, *Laser Phys. Lett.*, 2012, **9**, 234–239.

[19] A. Nebbioso, R. Benedetti, M. Conte, V. Carafa, F. De Bellis, J. Shaik, F. Matarese, B. Della Ventura, F. Gesuele, R. Velotta, J. H. A. Martens, H. G. Stunnenberg, C. Altucci and L. Altucci, *Sci. Rep.*, 2017, **7**, 11725.

[20] A. Reim, R. Ackermann, J. Font-Mateu, R. Kammel, M. Beato, S. Nolte, M. Mann, C. Russmann and M. Wierer, *Nat. Commun.*, 2020, **11**, 3019.

[21] Stützer, A. and Welp, L. M. and Raabe, M. and Sachsenberg, T. and Kappert, C. and Wulf, A. and Lau, A. M. and David, S. S. and Chernev, A. and Kramer, K. and Politis, A. and Kohlbacher, O. and Fischle, W. and Urlaub, H. , *Nat. Commun.*, 2020, **11**, 5250.

[22] Knörlein, A. and Sarnowski, C. P. and de Vries, T. and Stoltz, M. and Götze, M. and Aebersold, R. and Allain, F. H. and Leitner, A. and Hall, J., *Nat. Commun.*, 2022, **13**, 2719.

[23] H. Morrison, *Bioorganic photochemistry*, J. Wiley and Sons, Ltd, New York, 1990.

[24] M. Shetlar, *International Congress Series*, 1993, 67–72.

[25] G. Sun, C. J. Fecko, R. B. Nicewonger, W. W. Webb and T. P. Begley, *Org. Lett.*, 2006, **8**, 681–683.

[26] M. Micciarelli, M. Valadan, B. Della Ventura, G. Di Fabio, L. De Napoli, S. Bonella, U. Röthlisberger, I. Tavernelli, C. Altucci and R. Velotta, *J. Phys. Chem. B*, 2014, **118**, 4983–4992.

[27] M. Valadan, E. Pomarico, B. Della Ventura, F. Gesuele, R. Velotta, A. Amoresano, G. Pinto, M. Chergui, R. Improta and C. Altucci, *Phys. Chem. Chem. Phys.*, 2019, **21**, 26301–26310.

[28] G. Iuzzolino, F. Perrella, M. Valadan, A. Petrone, C. Altucci and N. Rega, *Physical Chemistry Chemical Physics*, 2024, **26**, 11755 11769.

[29] M. Micciarelli, C. Altucci, B. D. Ventura, R. Velotta, V. Toşa, A. B. G. Pérez, M. P. Rodríguez, A. R. de Lera and A. Bende, *Phys. Chem. Chem. Phys.*, 2013, **15**, 7161–7173.

[30] M. Micciarelli, B. F. E. Curchod, S. Bonella, C. Altucci, M. Valadan, U. Rothlisberger and I. Tavernelli, *J. Phys. Chem. A*, 2017, **121**, 3909–3917.

[31] A. Petrone, F. Perrella, F. Coppola, L. Crisci, G. Donati, P. Cimino and N. Rega, *Chemical Physics Reviews*, 2022, **3**, 021307.

[32] N. Rega, G. Brancato and V. Barone, *Chem. Phys. Lett.*, 2006, **422**, 367–371.

[33] G. Brancato, N. Rega and V. Barone, *J. Chem. Phys.*, 2006, **125**, 164515.

[34] G. Brancato, N. Rega and V. Barone, *J. Phys. Chem.*, 2008, **128**, 144501.

[35] U. Raucci, F. Perrella, G. Donati, M. Zoppi, A. Petrone and N. Rega, *Journal of Computational Chemistry*, 2020, **41**, 2228–2239.

[36] M. E. Casida, *Recent Advances in Density Functional Methods (Part I)*, World Scientific Publishing Co. Pte. Ltd., 1995, ch. 5, pp. 475–516.

[37] R. E. Stratmann, G. E. Scuseria and M. J. Frisch, *J. Chem. Phys.*, 1998, **109**, 8218–8224.

[38] A. Dreuw and M. Head-Gordon, *Chem. Rev.*, 2005, **105**, 4009–4037.

[39] R. Crespo-Otero and M. Barbatti, *Theor. Chem. Acc.*, 2012, **131**, 1237.

[40] A. V. Marenich, C. J. Cramer and D. G. Truhlar, *J. Phys. Chem. B*, 2015, **119**, 958–967.

[41] A. Petrone, J. Cerezo, F. J. A. Ferrer, G. Donati, R. Improta, N. Rega and F. Santoro, *J. Phys. Chem. A*, 2015, **119**, 5426–5438.

[42] J. Cerezo, F. J. Avila Ferrer, G. Prampolini and F. Santoro, *J. Chem. Theory Comput.*, 2015, **11**, 5810–5825.

[43] J. Cerezo, A. Petrone, F. J. Avila Ferrer, G. Donati, F. Santoro, R. Improta and N. Rega, *Theor. Chem. Acc.*, 2016, **135**, 263.

[44] J. Cerezo, G. Mazzeo, G. Longhi, S. Abbate and F. Santoro, *J. Phys. Chem. Lett.*, 2016, **7**, 4891–4897.

[45] D. Padula, J. Cerezo, G. Pescitelli and F. Santoro, *Phys. Chem. Chem. Phys.*, 2017, **19**, 32349–32360.

[46] J. M. Milanese, M. R. Provorse, E. J. Alameda and C. M. Isborn, *J. Chem. Theory Comput.*, 2017, **13**, 2159–2171.

[47] J. Cerezo, D. Aranda, F. J. Avila Ferrer, G. Prampolini, G. Mazzeo, G. Longhi, S. Abbate and F. Santoro, *Chirality*, 2018, **30**, 730–743.

[48] Y. Liu, J. Cerezo, N. Lin, X. Zhao, R. Improta and F. Santoro, *Theor. Chem. Acc.*, 2018, **137**, 40.

[49] G. Donati, A. Petrone, P. Caruso and N. Rega, *Chem. Sci.*, 2018, **9**, 1126–1135.

[50] S. V. Shedge, T. J. Zuehlsdorff, M. J. Servis, A. E. Clark and C. M. Isborn, *J. Phys. Chem. A*, 2019, **123**, 6175–6184.

[51] S. Del Galdo, M. Fusè and V. Barone, *J. Chem. Theory Comput.*, 2020, **16**, 3294–3306.

[52] C. Reichardt, *Chem. Rev.*, 1994, **94**, 2319–2358.

[53] K. M. Solntsev, D. Huppert, L. M. Tolbert and N. Agmon, *J. Am. Chem. Soc.*, 1998, **120**, 7981–7982.

[54] K. M. Solntsev, D. Huppert and N. Agmon, *J. Phys. Chem. A*, 1999, **103**, 6984–6997.

[55] H. A. Frank, J. A. Bautista, J. Josue, Z. Pendon, R. G. Hiller, F. P. Sharples, D. Gosztola and M. R. Wasielewski, *J. Phys. Chem. B*, 2000, **104**, 4569–4577.

[56] C. Adamo, M. Cossi, N. Rega and V. Barone, *Theoretical Biochemistry*, Elsevier, Amsterdam, The Netherlands, 2001, vol. 9, pp. 467–538.

[57] E. Krystkowiak, K. Dobek and A. Maciejewski, *J. Photochem. Photobiol. A*, 2006, **184**, 250–264.

[58] V. Barone, R. Improta and N. Rega, *Acc. Chem. Res.*, 2008, **41**, 605–616.

[59] M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G. A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H. P. Hratchian, A. F. Izmaylov,

J. Bloino, B. G. Janesko, F. Lipparini, G. Zheng, J. L. Sonnenberg, W. Liang, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, J. A. Montgomery, Jr., J. E. Peralta, F. Ogliaro, M. Bearpark, J. J. Heyd, E. Brothers, K. N. Kudin, V. N. Staroverov, T. Keith, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, N. Rega, J. M. Millam, M. Klene, J. E. Knox, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, R. L. Martin, K. Morokuma, V. G. Zakrzewski, G. A. Voth, P. Salvador, J. J. Dannenberg, S. Dapprich, P. V. Parandekar, N. J. Mayhall, A. D. Daniels, O. Farkas, J. B. Foresman, J. V. Ortiz, J. Cioslowski and D. J. Fox, *Gaussian Development Version, Revision H.37+*, 2010, Gaussian Inc. Wallingford CT.

[60] S. S. Iyengar, H. B. Schlegel, J. M. Millam, G. A. Voth, G. E. Scuseria and M. J. Frisch, *J. Chem. Phys.*, 2001, **115**, 10291–10302.

[61] H. B. Schlegel, J. M. Millam, S. S. Iyengar, G. A. Voth, A. D. Daniels, G. E. Scuseria and M. J. Frisch, *J. Chem. Phys.*, 2001, **114**, 9758–9763.

[62] H. B. Schlegel, S. S. Iyengar, X. Li, J. M. Millam, G. A. Voth, G. E. Scuseria and M. J. Frisch, *J. Chem. Phys.*, 2002, **117**, 8694–8704.

[63] S. S. Iyengar, H. B. Schlegel, G. A. Voth, J. M. Millam, G. E. Scuseria and M. J. Frisch, *Isr. J. Chem.*, 2002, **42**, 191–202.

[64] N. Rega, S. S. Iyengar, G. A. Voth, H. B. Schlegel, T. Vreven and M. J. Frisch, *J. Phys. Chem. B*, 2004, **108**, 4210–4220.

[65] J. C. Phillips, R. Braun, W. Wang, J. Gumbart, E. Tajkhorshid, E. Villa, C. Chipot, R. D. Skeel, L. Kalé and K. Schulten, *J. Comput. Chem.*, 2005, **26**, 1781–1802.

[66] G. J. Martyna, D. J. Tobias and M. L. Klein, *J. Phys. Chem.*, 1994, **101**, 4177–4189.

[67] S. E. Feller, Y. Zhang, R. W. Pastor and B. R. Brooks, *J. Phys. Chem.*, 1995, **103**, 4613–4621.

[68] R. Car and M. Parrinello, *Phys. Rev. Lett.*, 1985, **55**, 2471–2474.

[69] D. K. Remler and P. A. Madden, *Molecular Physics*, 1990, **70**, 921–966.

[70] M. C. Payne, M. P. Teter, D. C. Allan, T. A. Arias and J. D. Joannopoulos, *Rev. Mod. Phys.*, 1992, **64**, 1045–1097.

[71] M. E. Tuckerman, P. J. Ungar, T. von Rosenvinge and M. L. Klein, *The Journal of Physical Chemistry*, 1996, **100**, 12878–12887.

[72] M. Svensson, S. Humbel, R. D. J. Froese, T. Matsubara, S. Sieber and K. Morokuma, *J. Phys. Chem.*, 1996, **100**, 19357–19363.

[73] T. Vreven, K. S. Byun, I. Komáromi, S. Dapprich, J. A. Montgomery Jr, K. Morokuma and M. J. Frisch, *J. Chem. Theory Comput.*, 2006, **2**, 815–826.

[74] K. Morokuma, Q. Wang and T. Vreven, *J. Chem. Theory Comput.*, 2006, **2**, 1317–1324.

[75] M. Berkowitz, C. L. Brooks and S. A. Adelman, *J. Phys. Chem.*, 1980, **72**, 3889–3898.

[76] K. Kratky, *J. Comput. Phys.*, 1980, **37**, 205–217.

[77] C. L. Brooks and M. Karplus, *J. Phys. Chem.*, 1983, **79**, 6312–6325.

[78] D. Beglov and B. Roux, *J. Chem. Phys.*, 1994, **100**, 9050–9063.

[79] G. Brancato, A. Di Nola, V. Barone and A. Amadei, *J. Chem. Phys.*, 2005, **122**, 154109.

[80] E.-S. Riihimäki, J. M. Martínez and L. Kloo, *J. Mol. Struct. THEOCHEM*, 2006, **760**, 91–98.

[81] N. Rega, G. Brancato, A. Petrone, P. Caruso and V. Barone, *J. Chem. Phys.*, 2011, **134**, 074504.

[82] A. Petrone, G. Donati, P. Caruso and N. Rega, *J. Am. Chem. Soc.*, 2014, **136**, 14866–14874.

[83] M. G. Chiariello and N. Rega, *J. Phys. Chem. A*, 2018, **122**, 2884–2893.

[84] A. D. Becke, *Phys. Rev. A*, 1988, **38**, 3098.

[85] C. Lee, W. Yang and R. G. Parr, *Phys. Rev. B*, 1988, **37**, 785.

[86] B. Miehlich, A. Savin, H. Stoll and H. Preuss, *Chem. Phys. Lett.*, 1989, **157**, 200–206.

[87] A. D. Becke, *J. Chem. Phys.*, 1993, **98**, 5648.

[88] J. Wang, R. M. Wolf, J. W. Caldwell, P. A. Kollman and D. A. Case, *J. Comput. Chem.*, 2004, **25**, 1157–1174.

[89] L. Goerigk, A. Hansen, C. Bauer, S. Ehrlich, A. Najibi and S. Grimme, *Phys. Chem. Chem. Phys.*, 2017, **19**, 32184–32215.

[90] M. G. Chiariello, G. Donati, U. Raucci, F. Perrella and N. Rega, *J. Phys. Chem. B*, 2021, **125**, 10273–10281.

[91] J. Antony and S. Grimme, *Phys. Chem. Chem. Phys.*, 2006, **8**, 5287–5293.

[92] F. Perrella, F. Coppola, A. Petrone, C. Platella, D. Montesarchio, A. Stringaro, G. Ravagnan, M. P. Fuggetta, N. Rega and D. Musumeci, *Biomolecules*, 2021, **11**, 1048.

[93] S. Grimme, J. Antony, T. Schwabe and C. Mück-Lichtenfeld, *Org. Biomol. Chem.*, 2007, **5**, 741–758.

[94] F. Coppola, F. Perrella, A. Petrone, G. Donati and N. Rega, *Front. Mol. Biosci.*, 2020, **7**, 569990.

[95] D. Jacquemin, V. Wathelet, E. A. Perpète and C. Adamo, *J. Chem. Theory Comput.*, 2009, **5**, 2420–2435.

[96] C. A. Guido, D. Jacquemin, C. Adamo and B. Mennucci, *J. Phys. Chem. A*, 2010, **114**, 13402–13410.

[97] U. Raucci, M. G. Chiariello, F. Coppola, F. Perrella, M. Savarese, I. Ciofini and N. Rega, *J. Comput. Chem.*, 2020, **41**, 1835–1841.

[98] C. A. Guido, S. Knecht, J. Kongsted and B. Mennucci, *J. Chem. Theory Comput.*, 2013, **9**, 2209–2220.

[99] F. Perrella, U. Raucci, M. G. Chiariello, M. Chino, O. Maglio, A. Lombardi and N. Rega, *Biopolymers*, 2018, **109**, e23225.

[100] É. Brémond, M. Savarese, N. Q. Su, Á. J. Pérez-Jiménez, X. Xu, J. C. Sancho-García and C. Adamo, *J. Chem. Theory Comput.*, 2016, **12**, 459–465.

[101] D. Jacquemin, E. A. Perpète, G. E. Scuseria, I. Ciofini and C. Adamo, *J. Chem. Theory Comput.*, 2008, **4**, 123–135.

[102] S. Miertuš, E. Scrocco and J. Tomasi, *Chem. Phys.*, 1981, **55**, 117–129.

[103] V. Barone and M. Cossi, *J. Phys. Chem. A*, 1998, **102**, 1995–2001.

[104] M. Cossi, N. Rega, G. Scalmani and V. Barone, *J. Chem. Phys.*, 2001, **114**, 5691–5701.

[105] M. Cossi, N. Rega, G. Scalmani and V. Barone, *J. Comput. Chem.*, 2003, **24**, 669–681.

[106] T. Vreven, B. Mennucci, C. O. da Silva, K. Morokuma and J. Tomasi, *J. Chem. Phys.*, 2001, **115**, 62–72.

[107] S. J. Mo, T. Vreven, B. Mennucci, K. Morokuma and J. Tomasi, *Theor. Chem. Acc.*, 2004, **111**, 154–161.

[108] A. Petrone, P. Caruso, S. Tenuta and N. Rega, *Phys. Chem. Chem. Phys.*, 2013, **15**, 20536–20544.

[109] A. Ben–Naim, *Solvation Thermodynamics*, Plenum, New York, 1987.

[110] F. Perrella, A. Petrone and N. Rega, *Phys. Chem. Chem. Phys.*, 2021, **23**, 22885–22896.

[111] A. Nakayama, G. Arai, S. Yamazaki and T. Taketsugu, *J. Chem. Phys.*, 2013, **139**, 214304.

[112] S. Komin, C. Gossens, I. Tavernelli, U. Rothlisberger and D. Sebastiani, *J. Phys. Chem. B*, 2007, **111**, 5225–5232.

[113] J. Gao and M. Freindorf, *J. Phys. Chem. A*, 1997, **101**, 3182–3188.

[114] M.-E. Moret, I. Tavernelli and U. Rothlisberger, *J. Phys. Chem. B*, 2009, **113**, 7737–7744.

[115] I. Tavernelli, B. F. Curchod and U. Rothlisberger, *Chem. Phys.*, 2011, **391**, 101–109.

[116] M. Nonella, G. Mathias and P. Tavan, *J. Phys. Chem. A*, 2003, **107**, 8638–8647.

[117] V. Pham, I. Tavernelli, C. Milne, R. van der Veen, P. D'Angelo, C. Bressler and M. Chergui, *Chem. Phys.*, 2010, **371**, 24–29.

[118] R. M. Abolfath, P. K. Biswas, R. Rajnarayanam, T. Brabec, R. Kodym and L. Papiez, *J. Phys. Chem. A*, 2012, **116**, 3940–3945.

[119] J. Wen, S. Mai and L. González, *J. Phys. Chem. A*, 2023, **127**, 9520–9529.

[120] A. Lami and F. Santoro, *Computational Strategies for Spectroscopy: From Small Molecules to Nano Systems*, John Wiley & Sons, Ltd, New York, NY, 2011, ch. 10, pp. 475–516.

[121] T. Yanai, D. P. Tew and N. C. Handy, *Chem. Phys. Lett.*, 2004, **393**, 51–57.

[122] R. Improta and V. Barone, *Top. Curr. Chem.*, 2015, **355**, 329–357.

[123] R. Improta, F. Santoro and L. Blancafort, *Chem. Rev.*, 2016, **116**, 3540–3593.

[124] A. J. Pepino, J. Segarra-Martí, A. Nenov, R. Improta and M. Garavelli, *J. Phys. Chem. Lett.*, 2017, **8**, 1777–1783.

[125] A. J. Pepino, J. Segarra-Martí, A. Nenov, I. Rivalta, R. Improta and M. Garavelli, *Phys. Chem. Chem. Phys.*, 2018, **20**, 6877–6890.

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

[127] R. L. Martin, *J. Chem. Phys.*, 2003, **118**, 4775–4777.

[128] L. M. Ibele, P. A. Sánchez-Murcia, S. Mai, J. J. Nogueira and L. González, *The Journal of Physical Chemistry Letters*, 2020, **11**, 7483–7488.

[129] R. Borrego-Varillas, A. Nenov, P. Kabaciński, I. Conti, L. Ganzer, A. Oriana, V. K. Jaiswal, I. Delfino, O. Weingart, C. Manzoni, I. Rivalta, M. Garavelli and G. Cerullo, *Nature Communications*, 2021, **12**, 7285.

[130] C. E. Crespo-Hernández, B. Cohen, P. M. Hare and B. Kohler, *Chemical Reviews*, 2004, **104**, 1977–2020.

[131] J. Xu and J. R. Knutson, in *Chapter 8 Ultrafast Fluorescence Spectroscopy via Upconversion*, Elsevier, 2008, p. 159–183.

[132] L. Wang, R. Long and O. V. Prezhdo, *Annual Review of Physical Chemistry*, 2015, **66**, 549–579.

[133] S. Hammes-Schiffer, *Journal of the American Chemical Society*, 2015, **137**, 8860–8871.

[134] R. Crespo-Otero and M. Barbatti, *Chemical Reviews*, 2018, **118**, 7026–7068.

[135] T. Nelson, S. Fernandez-Alberti, A. E. Roitberg and S. Tretiak, *Accounts of Chemical Research*, 2014, **47**, 1155–1164.

[136] E. Brunk and U. Rothlisberger, *Chemical Reviews*, 2015, **115**, 6217–6263.

[137] J. C. Tully and R. K. Preston, *The Journal of Chemical Physics*, 1971, **55**, 562–572.

[138] J. C. Tully, *Faraday Discussions*, 1998, **110**, 407–419.

[139] J. C. Tully, *The Journal of Chemical Physics*, 1990, **93**, 1061–1071.

[140] E. Fabiano, T. Keal and W. Thiel, *Chemical Physics*, 2008, **349**, 334–347.

[141] J. Pittner, H. Lischka and M. Barbatti, *Chemical Physics*, 2009, **356**, 147–152.

[142] M. Barbatti and R. Crespo-Otero, in *Surface Hopping Dynamics with DFT Excited States*, Springer International Publishing, 2014, p. 415–444.

[143] B. Klaumünzer, D. Kröner, H. Lischka and P. Saalfrank, *Physical Chemistry Chemical Physics*, 2012, **14**, 8693–8702.

[144] D. Tuna, N. Došlić, M. Malis, A. L. Sobolewski and W. Domcke, *The Journal of Physical Chemistry B*, 2015, **119**, 2112–2124.

[145] L. Verlet, *Physical Review*, 1967, **159**, 98–103.

[146] W. C. Swope, H. C. Andersen, P. H. Berens and K. R. Wilson, *The Journal of Chemical Physics*, 1982, **76**, 637–649.

[147] M. Barbatti, M. Bondanza, R. Crespo-Otero, B. Demoulin, P. O. Dral, G. Granucci, F. Kossoski, H. Lischka, B. Mennucci, S. Mukherjee, M. Pederzoli, M. Persico, M. Pinheiro Jr, J. Pittner, F. Plasser, E. Sangiogo Gil and L. Stojanovic, *Journal of Chemical Theory and Computation*, 2022, **18**, 6851–6865.

[148] M. Barbatti, G. Granucci, M. Ruckenbauer, F. Plasser, R. Crespo-Otero, J. Pittner, M. Persico and H. Lischka, *NEWTON-X: A package for Newtonian dynamics close to the crossing seam. Version 2*, 2016, www.newtonx.org.

[149] F. Plasser, R. Crespo-Otero, M. Pederzoli, J. Pittner, H. Lischka and M. Barbatti, *Journal of Chemical Theory and Computation*, 2014, **10**, 1395–1405.

[150] C. R. Harris, K. J. Millman, S. J. van der Walt, R. Gommers, P. Virtanen, D. Cournapeau, E. Wieser, J. Taylor, S. Berg, N. J. Smith, R. Kern, M. Picus, S. Hoyer, M. H. van Kerkwijk, M. Brett, A. Haldane, J. F. del Río, M. Wiebe, P. Peterson, P. Gérard-Marchant, K. Sheppard, T. Reddy, W. Weckesser, H. Abbasi, C. Gohlke and T. E. Oliphant, *Nature*, 2020, **585**, 357–362.

[151] P. Virtanen, R. Gommers, T. E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau, E. Burovski, P. Peterson, W. Weckesser, J. Bright, S. J. van der Walt, M. Brett, J. Wilson, K. J. Millman, N. Mayorov, A. R. J. Nelson, E. Jones, R. Kern, E. Larson, C. J. Carey, İ. Polat, Y. Feng, E. W.

Moore, J. VanderPlas, D. Laxalde, J. Perktold, R. Cimrman, I. Henriksen, E. A. Quintero, C. R. Harris, A. M. Archibald, A. H. Ribeiro, F. Pedregosa, P. van Mulbregt and SciPy 1.0 Contributors, *Nature Methods*, 2020, **17**, 261–272.

[152] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot and E. Duchesnay, *Journal of Machine Learning Research*, 2011, **12**, 2825–2830.

[153] D. Cremer and J. Pople, *Journal of the American Chemical Society*, 1975, **97**, 1354–1358.

[154] L. Chan, *RING*, 2020, https://github.com/lucianlschan/RING.

[155] L. Chan, G. R. Hutchison and G. M. Morris, *Journal of Chemical Information and Modeling*, 2021, **61**, 743–755.

[156] C. Wiebeler, F. Plasser, G. J. Hedley, A. Ruseckas, I. D. W. Samuel and S. Schumacher, *The Journal of Physical Chemistry Letters*, 2017, **8**, 1086–1092.

[157] Z. Lan, Y. Lu, O. Weingart and W. Thiel, *The Journal of Physical Chemistry A*, 2012, **116**, 1510–1518.

[158] R. G. Cooks, I. Howe and D. H. Williams, *Organic Mass Spectrometry*, 1969, **2**, 137–156.

[159] J. P. Reilly, *Mass Spectrometry Reviews*, 2009, **28**, 425–447.

[160] J. Wu and S. A. McLuckey, *International Journal of Mass Spectrometry*, 2004, **237**, 197–241.

[161] L. Sleno and D. A. Volmer, *Journal of Mass Spectrometry*, 2004, **39**, 1091 1112.

[162] M. Quack and J. Trose, *International Reviews in Physical Chemistry*, 1981, **1**, 97 147.

[163] T. Baer and W. L. Hase, *Unimolecular Reaction Dynamics: Theory and Experiments*, Oxford University Press, USA, 1996.

[164] J. H. D. Eland, *Photoelectron Spectroscopy : an Introduction to Ultraviolet Photoelectron Spectroscopy in the Gas Phase*, Elsevier Science, 2014, p. 286.

[165] E. Molteni, G. Mattioli, P. Alippi, L. Avaldi, P. Bolognesi, L. Carlini, F. Vismarra, Y. Wu, R. B. Varillas, M. Nisoli, M. Singh, M. Valadan, C. Altucci, R. Richter and D. Sangalli, *Physical Chemistry Chemical Physics*, 2021, **23**, 26793 26805.

[166] J. Eland, *International Journal of Mass Spectrometry and Ion Physics*, 1972, **8**, 143–151.

[167] T. Baer, *International Journal of Mass Spectrometry*, 2000, **200**, 443–457.

[168] M. C. Castrovilli, A. Trabattoni, P. Bolognesi, P. O Keeffe, L. Avaldi, M. Nisoli, F. Calegari and R. Cireasa, *The Journal of Physical Chemistry Letters*, 2018, **9**, 6012 6016.

[169] L. Carlini, A. R. Casavola, J. Chiarinelli, F. Porcelli, E. Molteni, G. Mattioli, P. Bolognesi, D. Sangalli, F. Vismarra, Y. Wu, R. Borrego-Varillas, M. Nisoli, M. Singh, M. Valadan, C. Altucci, R. Richter and L. Avaldi, *Journal of Physics B: Atomic, Molecular and Optical Physics*, 2024, **57**, 105401.

[170] D. o. Barreiro-Lage, J. Chiarinelli, P. Bolognesi, R. Richter, H. Zettergren, M. H. Stockett, S. D az Tendero and L. Avaldi, *Physical Chemistry Chemical Physics*, 2023, **25**, 15635 15646.

[171] M. Selvaraj, A. Subramani, K. Ramanathan, R. Richter, N. Pal, P. Bolognesi, L. Avaldi and U. R. Kadhane, *Physical Chemistry Chemical Physics*, 2024, **26**, 16540 16549.

[172] C. J. Fecko, K. M. Munson, A. Saunders, G. Sun, T. P. Begley, J. T. Lis and W. W. Webb, *Photochemistry and Photobiology*, 2007, **83**, 1394 1404.

[173] D. Angelov, R. Boopathi, I. N. Lone, H. Menoni, S. Dimitrov and J. Cadet, *Photochemistry and Photobiology*, 2022, **99**, 296–312.

[174] J. V. Ortiz, *International Journal of Quantum Chemistry*, 2005, **105**, 803–808.

[175] J. V. Ortiz, *WIREs Computational Molecular Science*, 2012, **3**, 123–142.

[176] H. c. H. Corzo and J. V. Ortiz, in *Electron Propagator Theory*, Elsevier, 2017, p. 267 298.

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

[178] A. Petrone, J. Cerezo, F. J. A. Ferrer, G. Donati, R. Improta, N. Rega and F. Santoro, *The Journal of Physical Chemistry A*, 2015, **119**, 5426 5438.

[179] S. Grimme, *Angewandte Chemie International Edition*, 2013, **52**, 6306 6312.

[180] C. A. Bauer and S. Grimme, *The Journal of Physical Chemistry A*, 2016, **120**, 3755 3766.

[181] J. Koopman and S. Grimme, *ACS Omega*, 2019, **4**, 15120 15133.

[182] V. Ásgeirsson, C. A. Bauer and S. Grimme, *Chemical Science*, 2017, **8**, 4879 4895.

[183] C. A. Bauer and S. Grimme, *The Journal of Physical Chemistry A*, 2014, **118**, 11479 11484.

[184] F. Perrella, A. Petrone and N. Rega, *J. Chem. Theory Comput.*, 2023, **19**, 626–639.

[185] U. Raucci, F. Perrella, G. Donati, M. Zoppi, A. Petrone and N. Rega, *Journal of Computational Chemistry*, 2020, **41**, 2228 2239.

[186] U. Raucci, M. G. Chiariello, F. Coppola, F. Perrella, M. Savarese, I. Ciofini and N. Rega, *Journal of Computational Chemistry*, 2020, **41**, 1835 1841.

[187] F. Perrella, U. Raucci, M. G. Chiariello, M. Chino, O. Maglio, A. Lombardi and N. Rega, *Biopolymers*, 2018, **109**, e23225.

[188] F.-A. Korsaye, F. Perrella, A. Petrone, C. Adamo, N. Rega and I. Ciofini, *The Journal of Physical Chemistry A*, 2024, **128**, 4324 4334.

[189] F. Perrella, A. Petrone and N. Rega, *Journal of Chemical Theory and Computation*, 2024, **20**, 8820 8832.

[190] M. Elstner, D. Porezag, G. Jungnickel, J. Elsner, M. Haugk, T. Frauenheim, S. Suhai and G. Seifert, *Physical Review B*, 1998, **58**, 7260 7268.

[191] M. Gaus, Q. Cui and M. Elstner, *Journal of Chemical Theory and Computation*, 2011, **7**, 931 948.

[192] Y. Yang, H. Yu, D. York, Q. Cui and M. Elstner, *The Journal of Physical Chemistry A*, 2007, **111**, 10861 10873.

[193] C. Bannwarth, S. Ehlert and S. Grimme, *Journal of Chemical Theory and Computation*, 2019, **15**, 1652 1671.

[194] J. Gorges and S. Grimme, *Physical Chemistry Chemical Physics*, 2025, **27**, 6899 6911.

[195] A. D. Becke, *Physical Review A*, 1988, **38**, 3098 3100.

[196] C. Lee, W. Yang and R. G. Parr, *Physical Review B*, 1988, **37**, 785 789.

[197] B. Miehlich, A. Savin, H. Stoll and H. Preuss, *Chemical Physics Letters*, 1989, **157**, 200 206.

[198] A. D. Becke, *The Journal of Chemical Physics*, 1993, **98**, 5648 5652.

[199] P. J. Stephens, F. J. Devlin, C. F. Chabalowski and M. J. Frisch, *The Journal of Physical Chemistry*, 1994, **98**, 11623 11627.

[200] S. Grimme, *The Journal of Chemical Physics*, 2006, **124**, 034108.

[201] S. Grimme, J. Antony, S. Ehrlich and H. Krieg, *The Journal of Chemical Physics*, 2010, **132**, 154104.

[202] S. Grimme, S. Ehrlich and L. Goerigk, *Journal of Computational Chemistry*, 2011, **32**, 1456 1465.

[203] A. Schäfer, H. Horn and R. Ahlrichs, *The Journal of Chemical Physics*, 1992, **97**, 2571 2577.

[204] A. Schäfer, C. Huber and R. Ahlrichs, *The Journal of Chemical Physics*, 1994, **100**, 5829 5835.

[205] F. Weigend and R. Ahlrichs, *Physical Chemistry Chemical Physics*, 2005, **7**, 3297.

[206] J. M. Rice, G. O. Dudek and M. Barber, *Journal of the American Chemical Society*, 1965, **87**, 4569 4576.

[207] R. W. Reiser, *Organic Mass Spectrometry*, 1969, **2**, 467 479.

[208] C. Zhou, S. Matsika, M. Kotur and T. C. Weinacht, *The Journal of Physical Chemistry A*, 2012, **116**, 9217–9227.

[209] C. C. Nelson and J. A. McCloskey, *Journal of the American Society for Mass Spectrometry*, 1994, **5**, 339–349.

[210] F. Tureček and V. Hanuš, *Mass Spectrometry Reviews*, 1984, **3**, 85–152.

[211] R. M. Betz and R. C. Walker, *Journal of Computational Chemistry*, 2015, **36**, 79–87.

[212] K. V. Tretiakov and S. Scandolo, *The Journal of Chemical Physics*, 2004, **120**, 3765–3769.

[213] F. L. Hirshfeld, *Theoretica Chimica Acta*, 1977, **44**, 129–138.

[214] S. Dapprich, I. Komáromi, K. S. Byun, K. Morokuma and M. J. Frisch, *Journal of Molecular Structure: THEOCHEM*, 1999, **461**, 1–21.

[215] M. Pinheiro, M. de Oliveira Bispo, R. S. Mattos, M. T. Do Casal, B. C. Garain, J. M. Toldo, S. Mukherjee and M. Barbatti, *Digital Discovery*, 2025, **4**, 666–682.

[216] T. Vreven and K. Morokuma, *Annual reports in computational chemistry*, 2006, **2**, 35–51.

[217] F. R. Clemente, T. Vreven and M. J. Frisch, *Quantum biochemistry*, 2010, 61–83.

[218] M. Orozco, I. Marchán, I. Soteras, T. Vreven, K. Morokuma, K. V. Mikkelsen, A. Milani, M. Tommasini, M. D. Zoppo, C. Castiglioni *et al.*, *Continuum Solvation Models in Chemical Physics: From Theory to Applications*, 2007, 499–605.

[219] M. Barbatti, *Newton-X CS Documentation, version 2.6*, 2023, www.newtonx.org.

[220] B. Sellner, M. Barbatti and H. Lischka, *The Journal of Chemical Physics*, 2009, **131**, 024312.

[221] J. M. Martínez and L. Martínez, *Journal of computational chemistry*, 2003, **24**, 819–825.

[222] D. A. Case, H. M. Aktulga, K. Belfon, D. S. Cerutti, G. A. Cisneros, V. W. D. Cruzeiro, N. Forouzesh, T. J. Giese, A. W. Gotz, H. Gohlke *et al.*, *Journal of chemical information and modeling*, 2023, **63**, 6183–6191.