Quantum Mechanics Simulations of Nucleic Acids Building Blocks and Small DNA/RNA Models

Fernando Romeo Gella

Máster en Química Teórica y Modelización Computacional

MÁSTERES DE LA UAM 2018 - 2019

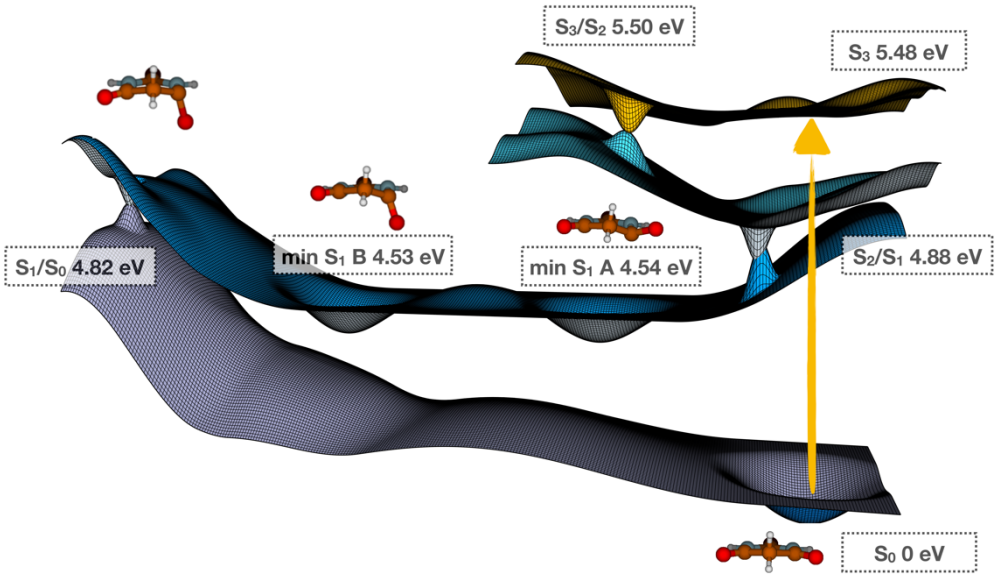
Facultad de Ciencias

UAM Universidad Autónoma de Madrid

Master Erasmus Mundus in

Theoretical Chemistry and Computational Modelling

Quantum Mechanics Simulations of Nucleic Acids Building Blocks and Small DNA/RNA Models



Fernando Romeo Gella

Director: Inés Corral Pérez
Department of Chemistry, Faculty of Science

Master Thesis. Course 2018-2019

Contents

List of Figures

List of Tables

Acknowledgements

It all started as a small project for the "Beca de Colaboración", which ended up as my Master Thesis Project. During these two years, not only was I starting to grow up as a scientist but also as a person, thanks to all the support coming from my friends, new colleagues, supervisors and family. First, I would like to thank the Spanish Government, the European Commission and Cátedra UAM-Fujitsu for funding on the tuition fees, research stays and conference attendance throughout the master. In particular, it is more than relevant to mention the people behind the scenes organising the master and making this experience possible: Wilson, Viviana, Beatriz, Marga, Manuel Alcamí and Sergio, not forgetting about Manuel Yáñez and Otilia Mó for believing in the project and supporting every young motivated student from their own experience.

During these two years, some people stood out from the crowd and have become actual friends and future colleagues. Nacho, Joel, Kilian, Raquel and Judit (a.k.a. MJ) thanks for your help, support, drinks and memories. Also from the master, but this time coming from Euskal Herria, Maria and Maialen helped me, hosted me and taught me some Euskera to survive in Donostia. You all are the key pieces of the master.

To pursue this master I also went to Groningen, where Shirin Faraji welcomed me and was really involved in training me. In her group, everybody made me feel as if I were home during those three months. For all of this, and the future, I am really grateful and I hope to be able to work with you all in some time.

Not to be forgotten is the best Jedi Master one can have. Enrique, without your constant help, scripts, guidelines, suggestions and more scripts, I would have never been able to learn and move forward as much as I did. Thank you for everything and best of luck for your last year of PhD.

In the end, I can just be extremely grateful for finding such a great, warm and helpful supervisor: Inés Corral. Even being aware of my biological background, she has always been on board for helping, supporting and guiding. She worked tirelessly for me to be able to do a PhD. This is the reason why I am so lucky to have you again as a supervisor, but now as a PhD supervisor during four years. Good luck with that!

Abstract

Canonical nucleobases are the building blocks of nucleic acids, and the origin of the genetic identity of any living organism. Despite their simple, but at the same time elegant chemical properties, for instance, their base-pairing properties, considerable pieces of evidence indicate that the family of canonical nucleobases is the result of an evolution process initiated during the prebiotic era. In this context, Barbituric Acid (BA) was proposed as a putative ancestor for current nucleobases, but still its photophysics remains unclear. The aim of this work is to unravel the excited state deactivation mechanism of BA, considering all ranges of pH to better reproduce the extreme environmental conditions of early Earth. CASSCF//CASPT2 calculations were performed in order to explore the potential energy surfaces (PES) of the predominant tautomers of BA in different pH conditions. Absorption spectra were calculated at the Franck-Condon region and comparison with experimental data showed excellent agreement. Minimum energy path calculations were undertaken in order to determine the most plausible deactivation routes. By these means, the principle funnels along the singlet and the triplet manifold were located. Additionally, tentative time resolved information on the deactivation of this system has been extracted from the interpretation of transient absorption spectroscopy data. For this, vertical spectra at regions of the PES where the system is expected to spent some time were carried out. Although a rigorous dynamical study would be necessary to predict excited state lifetimes, the PES for the predominant BA tautomers at acid and basic pH point BA to be a photostable system. For both pH conditions, the molecule showed photostability, what sets a considerable precedent to consider BA as an ancestor of the current nucleic acids.

Resumen

Las nucleobases canónicas son los bloques de construcción de los ácidos nucleicos y el origen de la identidad genética para cualquier organismo vivo. Pese a sus sencillas, pero al mismo tiempo elegantes propiedades, como por ejemplo su patrón de emparejamiento entre ellas, hay importantes evidencias que indican que la familia de las nucleobases canónicas es el resultado de un proceso evolutivo que tuvo comienzo durante le era prebiótica. En este contexto, el Ácido Barbitúrico (BA, por sus siglas en inglés) fue propuesto como un posible precursor de las actuales nucleobases, a pesar de que sus propiedades fotofísicas no están totalmente claras. El objetivo de este trabajo es descifrar los mecanismos de desactivación desde un estado electrónico excitado del BA teniendo en cuenta todos los rangos de pH para reproducir de una manera más fiel las extremas condiciones ambientales de la Tierra primigenia. Se han realizado cálculos CASSCF//CASPT2 para explorar las superficies de energía potencial (PES, por sus siglas en inglés) de aquellas especies predominantes del BA en diferentes condiciones de pH. Así, se han calculado los espectros de absorción en la región Franck-Condon, que coinciden de manera excelente con los datos experimentales disponibles. También se ha procedido a calcular caminos de mínima energía (MEP, por sus siglas en inglés) para determinar las rutas de desactivación más probables. Con todo esto, se ha conseguido localizar las principales rutas disponibles en el conjunto de los diferentes estados singlete y triplete. Además, ha sido obtenida información resuelta en el tiempo sobre la desactivación de este sistema a partir de la interpretación de datos provenientes de espectroscopía de absorción sobre transientes. Para esto, se llevó a cabo el cálculo de los espectros verticales en aquellas regiones de la PES donde se espera que el sistema se acumule más tiempo. Aunque un riguroso estudio dinámico sería necesario para predecir la vida media de los estados excitados, las PES para aquellos tautómeros predominantes del BA a pH ácido y básico dan a pensar que es un sistema fotoestable. Para ambos valores de pH, dado que la molécula muestra fotoestabilidad, hay suficientes indicios que permiten considerar al BA como un posible precursor de los actuales ácidos nucleicos.

Chapter 1

Introduction

1.1 Nucleic Acids: DNA and RNA, the Genetic Information Carriers

Life has been estimated to have started around 4.1 or 4.2 Gyr ago on the Earth [5], and three minimum requirements have been established as necessary for life to start; these are a coexistent atmosphere, water and a landmass with continuous material circulation driven by the sun [6]. Every organism can be characterized as living according to three conditions: (1) compartmentalization with the environment, (2) replication, and (3) metabolism [5]. By fulfilling these three basic requirements, a living organism would be separated from the rest of the environment, would transmit heritable information and would be able to obtain energy from the outside on its own.

Despite the existence of several theories, both scientific and pseudo-scientific, agree on the fundamental role of nucleic acids in terms of using, keeping and passing to offspring the genetic information contained in a living organism. As a matter of fact, nucleic acids (NA) are the basis of life in terms of genetic identity and genetic evolution, which entails function and adaptation, respectively. Current forms of NA polymers are DNA and RNA made of nucleosides linked by phosphodiester bonds. Then, a nucleoside is composed of deoxyribose (DNA) or ribose (RNA) and a nucleobase (either purine- or pyrimidine-like) covalently bonded by a β-N-glycosidic bond arising from $1'$ carbon of the sugar, see picture 1.1.

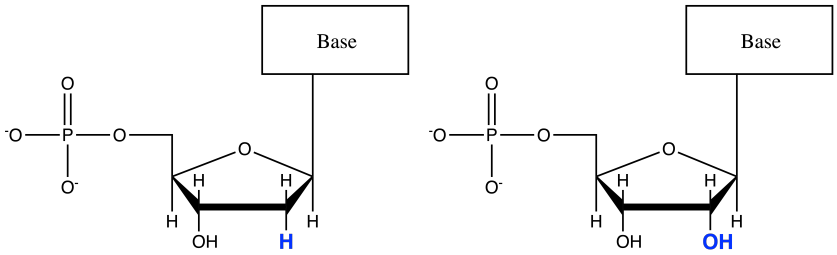


Figure 1.1: Building blocks of DNA an RNA: Deoxyribonucleotide (left) and Ribonucleotide (right). Change in 2' C is highlighted in blue.

1.2 Canonical and Non-Canonical Nucleobases

Canonical nucleobases, those found ubiquitously in naturally-ocurring DNA and RNA, are adenine (A), guanine (G), cytosine (C), thymine (T, DNA-exclusive) and uracil (U, RNA-exclusive). Both A and G belong to the purine family (having a purine ring as a common subunit where any particular property raises from the substituents), while C, T and U belong to the pyrimidine family (having a purine ring as a common subunit where any particular property arises from the substituents); as can be seen in Figure 1.2. Apart from these, DNA and RNA may contain other modified nucleobases like 5-methylcytosine which has been reported to be involved in genetic expression control, in terms of epigenetic and transcriptomic processes [7]. These modified nucleobases, either by methylation, amination or others, give raise to the non-canonical nucleobases family, which can be found along the genome and also in many other biologically-related processes [8].

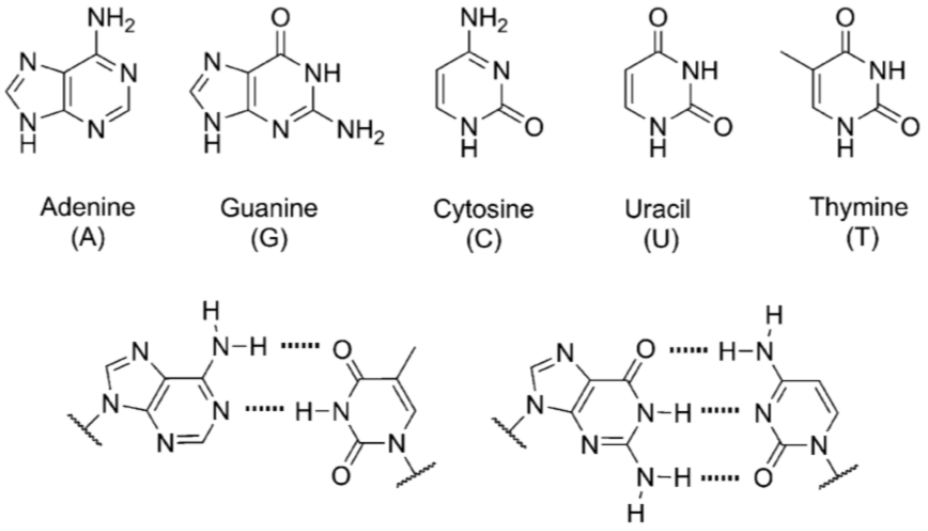


Figure 1.2: Canonical purine (A and G) and pyrimidine (C, U and T) nucleobases present in DNA and RNA (top). Watson-Crick pairing scheme for A-T in DNA, analogous to A-U in RNA, and G-C in both DNA and RNA. Adapted from [1].

These five nucleobases generally pair with each other according to what is called canonical or Watson-Crick scheme: A with T/U, and G with C (see Figure 1.2 bottom panel). Despite this simple and neat arrangement, it is most likely that the family of canonical nucleobases and their elegant chemical properties is rather the result of a continuous evolutive process, rather than these being the original nucleobases and pairing patterns since life first emerged [1]. Apart from the canonical pairing scheme, other arrangements have been reported in nature, in particular for RNA-RNA and RNA-protein interaction, which involve other sides of the nucleobase, such as Hoogsteen edge and Sugar/C-H edge (see Figure 1.3) [2]. As it has been already pointed out, modifications to canonical nucleobases occur on a usual basis; and, in fact, thymine has been reported to be a modification of uracil which remained in our genetic code for functional reasons [9, 10]. Several experiments have shown that canonical nucleobases could have been present in the prebiotic area, but also many others, sharing the core chromophore (pyrimidine and purine rings), but holding different substituents [11–15]. One of the evidences used as an argument on the potential role of a particular nucleobase in primordial DNA/RNA molecules at prebiotic time is its stability, and more specifically its photostability in terms of ultrafast decay to the ground state upon light absorption.

The prebiotic scenario has been depicted as an environment undergoing drastic pH changes, extreme temperatures, and strong high-energy radiation like X-ray and UV, due to the lack of the ozone layer. Taking into account these prebiotic Earth conditions, two processes are thought to have been determinant in the stability of nucleobases ancestors: hydrolytic degradation and UV irradiation. Leaving hydrolytic degradation aside, we will focus on the effect of UV light on these molecules, as photostability seems to be a direct selective pressure during prebiotic times.

Canonical nucleobases strongly absorb UV light in the region comprised between 230 and 280 nm [16]. As a consequence, they might potentially suffer damage upon UV radiation, such as the prevalent photodimerization of uracil or thymine residues [17]. However, in comparison to other heterocycles they are extremely photostable, as they have non-radiative decay pathways which explain their low fluorescence quantum yields [16]. Small modifications on the position and nature of the substituents have great impact on the excited-state lifetime of nucleobases. This is also the case of pH changes (common in the prebiotic area) which are also determinant for the modulation of the decay times; as has been reported by [18], and, in particular, for guanine [1].

Not so recently (1963), it was established that free canonical nucleobases do not spontaneously pair according to the Watson-Crick scheme, as water is well able to satisfy hydrogen-bonding for these heterocycles, and they tend to stack in columnar shapes through their hydrophobic faces [19]. For this reason, canonical nucleobases do not seem to have been adequate ancestors of current nucleic acids, as no selective pressure would

have been applied to them if no recognition pattern arose from pairing. Thus, either a chaperone molecule would have facilitated the pairing, or other non-canonical heterocycles were involved in the prebiotic evolutionary process. Several functionally-equivalent and compatible candidates should have been available with regards to prebiotic chemistry and environment.

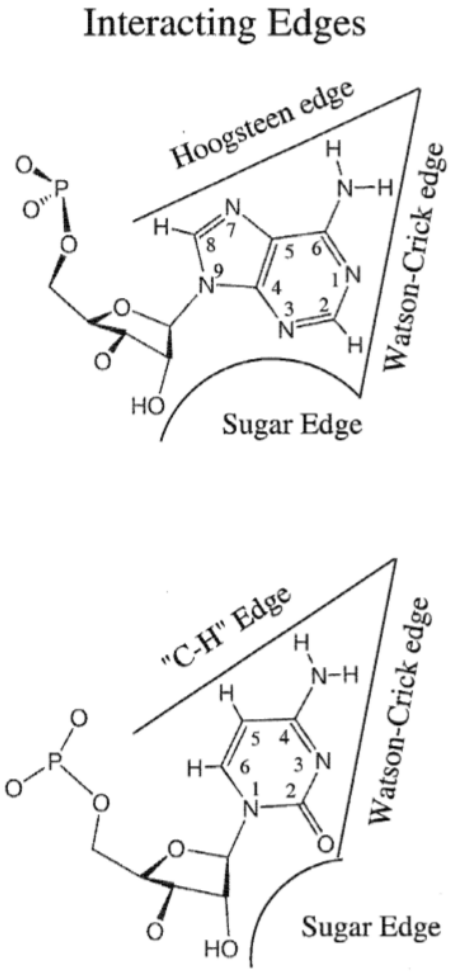


Figure 1.3: Interacting edges on nucleotides: Watson-Crick, Hoogsten and sugar-edge for purines (top) and pyrimidines (bottom). Adapted from [2].

Cafferty and Hud came up with 81 heterocycles (27 purines and 54 pyrimidines) which actually fulfilled the chemical requirements compatible with a prebiotic scenario [10]: (1) chemically compatible with abiotic reactions, (2) water soluble and hydrogen bond patterns analogues to Watson-Crick base pairing, (3) tendency to forming a glycosidic bond with a sugar moiety within prebiotic chemistry, and (4) able to transfer information in a replicative way: principle of continuity. These 81 non-canonical nucleobases were analyzed thoughtfully against those criteria, but the authors did not take into account photostability against UV radiation as a determinant criterion. Purine-pyrimidine base

pairs were unlike to be the original patterns as purine-rich templates have difficulty to incorporate free pyrimidine nucleobases [10]. Purine-purine alternatives are more consistent due to a more favorable assembly process (i.e. G-tetrad rosette as commonly seen in nature), still the insolubility of the purine aromatic rings is a considerable drawback [10]. Following G-tetrad rosette pattern, pyrimidine-pyrimidine pairing seem to be a plausible pattern, mainly motivated by the highly cooperative self-assembly of barbituric acid (BA, Figure 1.4) and 2,4,6-triaminopyrimidine (TAP), and the reactivity of the C2 and C6 sites enabling sugar to enter the game through formation of C-nucleosides [10] (commonly found in non-coding RNA and now used as antiviral drugs [20]).

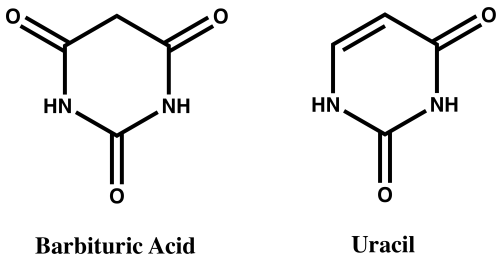


Figure 1.4: Chemical structure of barbituric acid (left) and uracil (right).

As many authors have claimed, photostability against UV-radiation in a prebiotic atmosphere, which is considerably exposed to higher doses of electromagnetic radiation, is a key aspect to consider for putative ancestors of our current nucleic acids. Thus, the aim of this work are to study the photophysics of BA from a theoretical perspective and being able to decipher the intrinsic mechanisms involved in dissipating the absorbed energy. The results of this work will not only shed light of the potential role of BA as an ancestor of DNA and RNA canonical nucleobases, but also will hopefully help establishing the connection between purine and pyrimidine substitution pattern and the origin of the photostability of nucleobases.

1.3 Light-Driven Rotary DNA Molecular Motors

Nowadays, control of biological processes is a key aspect with the increasing research in nanotechnology and biotechnology. In fact, a challenging aspect is to control a particular event, both spatially and temporally, in the most precise manner. In this sense, light, within certain biologically-harmless wavelength range, is thought to be the ideal external agent which can control this process to our convenience, without the use of invasive techniques or the need to contaminate biological samples. For this reason, molecular photoswitches started to raise our interest as they are molecules able to its conformation upon radiation absorption. Some of the most common and studied photoswitches are

azobenzenes and stilbenes. Both molecular families suffer reversible cis-trans isomerization upon UV/Vis or UV/UV radiation, showing high quantum yields and absorption at biologically-friendly frequencies. The photophysics of azobenzene and stilbene has been well studied and is very-well known, despite some aspects are still a matter of discussion between different experts in the field, which raise several questions in such a challenging research field. Interestingly, these families show fast isomerization and low rate of photobleaching; also recently, photoisomerization was reported to be achieved exclusively with visible light, avoiding potential damage to tissue due to UV radiation [21, 22].

A photoswitch molecule can be coupled to a DNA hairpin acting as a bridge between the two DNA complementary oligomeric strands. It has been recently found that the trans form of azobenzene favours double helix stability due to strong $\pi - \pi$ stacking interactions with the adjacent nucleobases, while the cis isomer destabilizes base-pairing, causing the helix to open. Letsinger and Wu observed that bridged complementary oligonucleotide chains show reversibility of the switching event, which is easily monitorable in aqueous solution; also, placing a G-C basepair adjacent to the photoswitch prevents photoisomerization from happening, while A-T basepair favours it [23, 24]. Further research on this phenomena showed that the G-C basepair, but also A-T to a lower level, adjacent to the photoswitch might enhance fluorescence quenching of the photoswitch molecule as the result of a charge transfer between the photoswitch and the mentioned basepair [25, 26]. While fluorescence quenching is still an issue in pure azobenzene-linked hairpins, varying the bridge length between the photoswitch and the phosphate group of the oligomeric chains can solve this aspect, and also the reversibility of the isomerization at physiological conditions, in order to gain more control on the switching event [27]. Also, stilbene-linked hairpins are too stable under physiological conditions that it is difficult to obtain notable photoreversibility, as they undergo cyclation/oxidation of the cis form [28].

Several inconveniences have been stated on two of the most widely studied molecular photoswitches. Due to these and other additional casualties, there is place for new models which are able to overcome these problems. Feringa and collaborators proposed the use of a xylene-based core structure as a novel switch, shown in Figure 1.5. This new switch is symmetric, with limited conformational flexibility, thus maximizing geometrical change in the DNA hairpin duplex structure [3].

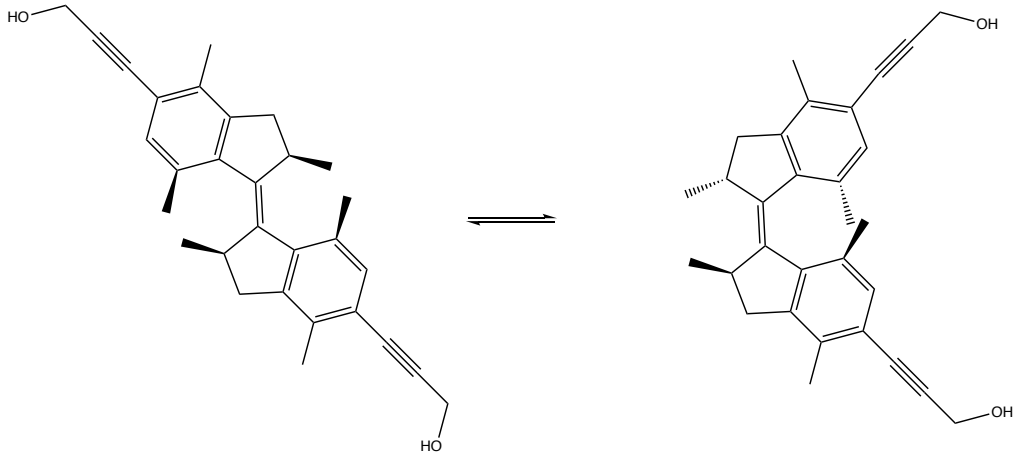


Figure 1.5: Trans (left) and cis (right) isomers of the novel light-driven rotary molecular motor proposed by Feringa and collaborators [3]. Compound name: (3,3'-(2,2',4,4',7,7'-hexamethyl-2,2',3,3'-tetrahydro-[1,1'-biindenylidene]-5,5'-diyl)bis(prop-2-yn-1-ol)).

.

Feringa and coworkers, [3], revealed a two-step isomerization process between the cis and trans isomers, with a different unstable intermediate in each direction of the reaction. The study of the full hybrid molecular motor, 8T-switch-8A, revealed that the hybrid hairpin was more stable than the original one, having a higher melting temperature (T_m). This study also showed that the trans isomer was more stable than the cis one. Furthermore, photoswitching dynamics were experimentally studied showing that: (1) Light absorption occurs at wavelengths longer than 300 nm, preventing potential photodamage; (2) cis-trans isomerization can occur at physiological conditions, although lower rates were observed; (3) and the limiting step is the thermal helix inversion (indicated as Δ in order to obtain the final stable isomer at the end of each side of the cycle, see Figure 1.6).

Additionally to the experimental approaches, the authors performed some preliminary theoretical calculations on the system. First, for the isolated photoswitch, a DFT energy scan was undertaken along a single atom-atom distance coordinate (see [3]), revealing that the cis-isomer is more stable than the trans one. Then, they performed Molecular Dynamics at specific temperatures trying to justify an increased stability of the trans isomer in comparison to the cis isomer for the whole hybrid system. In this case, π-π stacking as described above was found to be responsible for the reported stabilization.

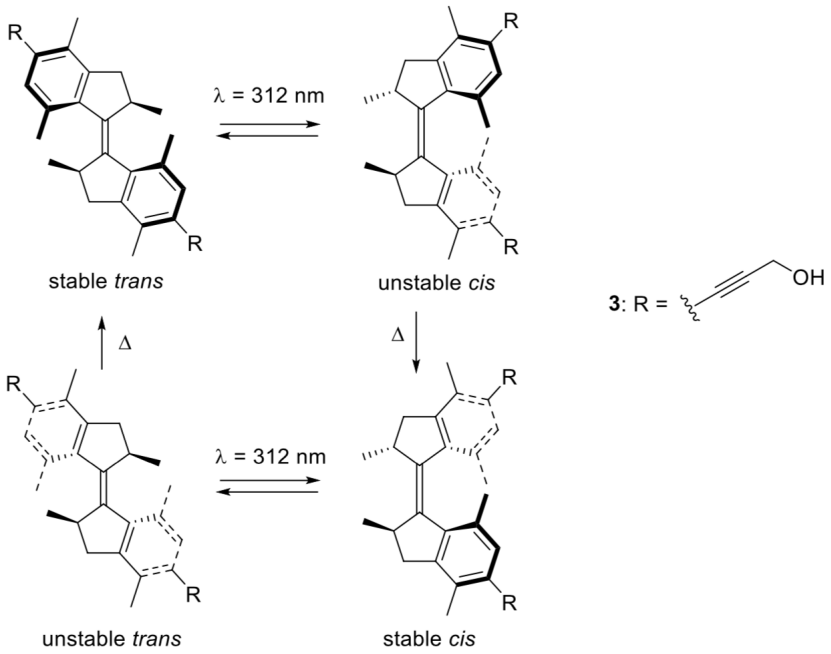


Figure 1.6: Photochemical and thermal isomerization during 360° unidirectional rotary cycle of the light-driven rotary molecular motor shown in Figure 1.5. Adapted from [3].

After this brief introduction, it is evident that a deeper study of the mechanism behind light-driven molecular motors is needed, both to boost basic research on the topic and also for future bio- and nano-technological applications (bionanodevices, drug-delivery systems,...). In particular, a deeper theoretical study describing the molecular and electronic processes following irradiation of the hybrid system will help understanding and realizing how these systems can be improved. Along my master studies, I have undertaken preliminary calculations on the isolated photoswitch, which, however, are not contained in the manuscript due to space constraints. These results include:

- DFT functional benchmarking and comparison of experimental data against CASSCF, TDDFT and SF-TDDFT results on geometry and excitation energies.
- Geometry optimization, absorption spectra and excited state analsysis of both cis- and trans-isomers.
- PES Scan along the reaction coordinate: describing the dihedral between both symmetric subunits of the rotor using TDDFT and SF-TDDFT. This helped us determining a putative energy barrier in the ground state, concluding that this reaction is only feasible if induced by light (see Figure 1.7).

Several other aspects need to be tackled in order to get a more complete picture of the photophysics of this hybrid system. In fact, QM/MM simulations are planned to study the photophysics of the complete hybrid hairpin and its stability. An important

acknowledgement on this part of the project is Prof. Dr. Shirin Faraji, from University of Groningen (The Netherlands), whom hosted me for three months and helped me take the first steps in the topic, in particular thanks to her photoswitch and DFT-based study of excited states experience.

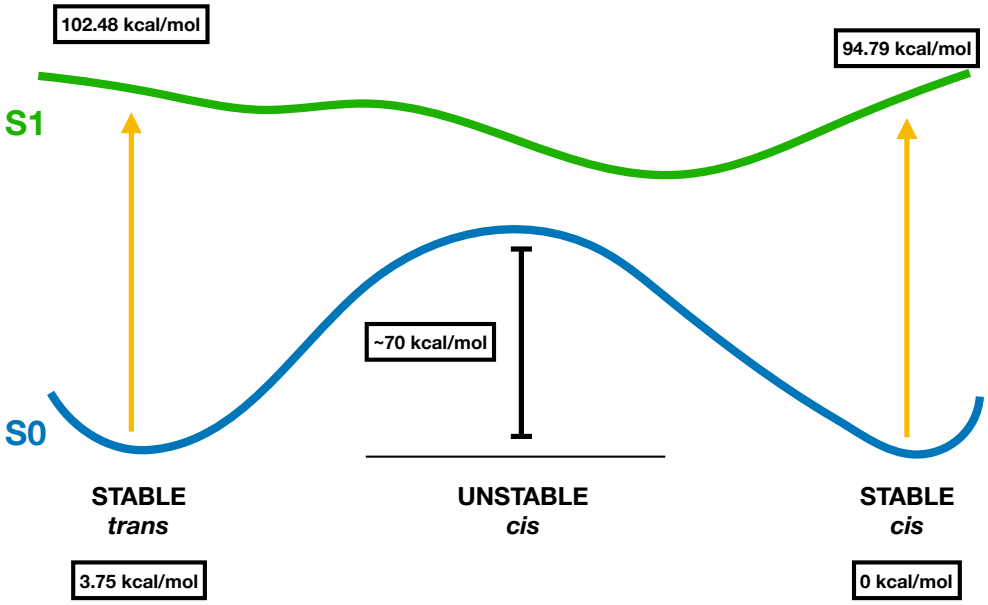


Figure 1.7: Potential Energy Surface obtained after performing a relaxed scan along the dihedral C2-C8-C15-C19 from the trans isomer configuration (∼160°) to the cis isomer configuration (∼6°) on the light-driven rotary molecular motor shown in Figure 1.5.

Chapter 2

Methods

2.1 Hamiltonian

In Quantum Mechanics, each physical system is described by a complex separable Hilbert space, where each pure state of this system corresponds to an unitary ray usually called state vector or ket. The pure state concept arises when the maximum degree of information of the system is considered; and other than a maximal degree information will determine a mixed state. Experimental observables correspond mathematically to linear self-adjoint operators, whose spectrum is real upon acting on state vector, e.g. linear momentum operator ($\hat{p}$), position operator ($\hat{q}$) or angular momentum operator ($\hat{L}$). Now, being Ψ a wavefunction representing a certain state, and $\hat{P}_a$ the projector for a certain value of the discrete eigenvalue spectrum, the probability of obtaining a value "a" for this observable is $P(A = a) = \langle\Psi|\hat{P}_a\Psi\rangle$. The free evolution of a system, represented by Ψ is guided by the time-dependent Schrödinger equation

$$i\hbar\frac{\delta\Psi}{\delta t} = \hat{H}\Psi \tag{2.1}$$

which is a linear partial differential equation that describes the wavefunction of a quantum-mechanical system. The operator involved in this equation (in a.u.) is the Hamiltonian, $\hat{H}$, which for a system with N electrons and M nuclei is defined in atomic units (a.u.) by

$$\hat{H} = -\sum_{A=1}^{M}\frac{1}{2M_A}\nabla_A^2 - \sum_{i=1}^{N}\frac{1}{2}\nabla_i^2 - \sum_{i=1}^{N}\sum_{A=1}^{M}\frac{Z_A}{r_{iA}} + \sum_{i=1}^{N}\sum_{j>i}^{N}\frac{1}{r_{ij}} + \sum_{A=1}^{M}\sum_{B>A}^{M}\frac{Z_A Z_B}{r_{AB}} \tag{2.2}$$

where M_A is the ratio of the mass of nucleus A to the mass of an electron, and Z_A is the atomic number of the nucleus A. The first term corresponds to the kinetic energy of the nuclei; the second term, to the kinetic energy of the electrons; the third term, to the

Coulomb attraction between electrons and nuclei; the fourth and fifth, to the repulsion between electrons and nuclei respectively.

2.2 Born-Oppenheimer Approximation.

For a general molecular system, consisting of several electrons and nuclei, the Hamiltonian (neglecting spin-orbit and spin-spin interactions) can be written as

$$H(r,R) = T_{nuc} + H_{el}(r,R) + V_{MM}(R) \tag{2.3}$$

where r and R represent the colective electronic and nuclear coordinates, respectively, for n electrons and N nuclei. T_{nuc} accounts for the first term of equation 2.2; $H_{el}(r,R) = T_{el}(r) + V_{eN}(r,R) + V_{ee}(r)$, for the three next terms of equation 2.2; and V_{NN}, for the last term of 2.2.

For a fixed nuclear configuration in space (R fixed), the motion of the electrons is described by the electronic Schrödinger equation

$$[H_{el}(r;R) + V_{NN}(R)]\psi_n(r;R) = E_{el,n}(R)\psi_n(r;R) \tag{2.4}$$

where $\psi_n(r;R)$ and $E_{el,n}(R)$ are called adiabatic eigenfunctions and eigenvalues of the electron n, depending parametrically on the nuclear coordinates R. Since the adiabatic eigenfunctions form a complete orthonormal set, the full molecular eigenfunction $\Psi(r;R)$ satisfying the complete Schrödinger equation

$$\hat{H}\Psi(r;R) = E\Psi(r;R)$$

can be expanded in the adiabatic basis $\psi_n(r;R)$

$$\Psi(r;R) = \sum_n \chi_n(R)\psi_n(r;R) \tag{2.5}$$

where $\chi_n(R)$ are the nuclear wavefunctions in the adiabatic representation.

After multiplying by the bra $\langle\psi_m(r;R)|$, substituting the former expansion in the Schrödinger equation, and integrating over the electron coordinates, we obtain the coupled system of equations

$$[T_{nuc}(R) + E_{el,m}(R)]\chi_m(R) + \sum_n \Lambda_{mn}(R)\chi_n(R) = E\chi_m(R) \tag{2.6}$$

Here, $\Lambda_{mn}(R)$ is the non-adiabatic coupling matrix operator, which results from the action

of $T_{nuc}(R)$ on the electronic wavefunctions $\psi_n(r;R)$. Also,

$$\Lambda_{mn}(R) = -\hbar^2 \sum_{\alpha} \frac{1}{M_\alpha}(A^\alpha_{mn}\nabla_{alpha} + \frac{1}{2}B^\alpha_{mn})$$

where A and B matrices are defined as

$$A^\alpha_{mn} = \langle\psi_n(r;R)|\nabla_\alpha|\psi_n(r;R)\rangle$$

$$B^\alpha_{mn} = \langle\psi_n(r;R)|\nabla^2_\alpha|\psi_n(r;R)\rangle$$

The coupled system of equations in matrix form

$$(\mathbf{T}+\mathbf{V})\mathbf{X}(R) = E\mathbf{X}(R) \tag{2.7}$$

where the diagonal matrix $\mathbf{V}_{mn} = E_{el,m}(R)\delta_{mn}$ is the adiabatic potential, and $\mathbf{X}(R)$ represents the vector regarding nuclear coordinates. The non-diagonal kinetic energy matrix is given by $\mathbf{T}_{mn} = \mathbf{T}_{nuc}(R)\delta_{mn} + \mathbf{\Lambda}_{mn}(R)$. Thus, solving the coupled systems of equations is equivalent to solving the problem of nuclear motion in the adiabatic representation. The non adiabatic coupling between different adiabatic states is responsible of non-adiabatic transitions. Nevertheless, the direct calculation of the non-adiabatic coupling terms is usually a difficult task. Additionally, solving the coupled equations is also a challenging task except for some simple systems. Therefore, the adiabatic representation allows introducing the adiabatic approximation which consists of discarding the off-diagonal coupling terms ($\mathbf{\Lambda}_{mn}$ with $m \neq n$). This approximation is based on the rationale that the nuclear mass is much larger than the electron mass, and therefore the nuclei move much slower than the electrons. As a direct consequence of this, nuclear kinetic energies are smaller than those of the electrons and the non-adiabatic coupling terms (A and B) are generally small (as they are a result of nuclear motion).

Let's assume we neglect the non-adiabatic coupling terms, which is equivalent to consider a single term in the adiabatic expansion of the total wavefunction. Then, we obtain the adiabatic approximation for the nuclear wavefunction $\hat{H}_{nuc}\chi_n(R) = E\chi_n(R)$ where the adiabatic Hamiltonian is $\hat{H}_{nuc}(R) = \hat{T}_{nuc}(R) + E_{el,n}(R) + \Lambda_{nn}(R)$, and $E_{el,n}(R)$ represents the adiabatic potential for the nuclear motion obtained from the solution of the electronic Schrödinger equation. The necessary condition that needs to be fulfilled in order to apply the the adiabatic approximation is that the nuclear kinetic energy is much smaller than the energy gap between adiabatic electronic states. In a considerable number of situations, the dependence of B^α_{nn} (coupling terms) on the nuclear coordinates is weak compared to that of the adiabatic potential $E_{el,n}(R)$. Thus, the term $\Lambda_{nn}(R)$ can be neglected and as

a result we obtain the Born-Oppenheimer approximation

$$\hat{T}_{nuc}(R) + E_{el,n}(R)]\chi_n(R) = E\chi_n(R) \tag{2.8}$$

where $E_{el,n}(R)$ is the global multidimensional potential that governs nuclei motion. This $E_{el,n}(R)$, where n denotes the N-th electronic state, is what is generally called potential energy surface (PES). This equation, nuclear Schrödinger equation, can be used for the prediction rovibrational spectra (bound states) or for the solution of scattering problems.

In a few words, in the adiabatic approximation, the nuclear hamiltonian includes the diagional non-adiabatic terms, Λ_{nn}, which are neglected in Born-Oppenheimer approximation. Nevertheless, both approximations rely on the full separation of the electronic and the nuclei motion.

2.3 Approximate Methods in Quantum Chemistry

2.3.1 Variational and Linear Variational Principle

Given a normalized wavefunction $|\Psi\rangle$ for a ground state system that satisfies the appropriate boundary conditions, then the expectation value of the Hamiltonian is an upper bound to the exact ground state energy. This allows an easy measurement of the quality of the wavefunction: the lower the energy, the better the wavefunction.

$$\langle\Psi|\hat{H}|\Psi\rangle \geq \epsilon_0 \quad and \quad \epsilon_0 \leq \epsilon_1 \leq \epsilon_2 \leq \cdots \tag{2.9}$$

Then, the linear variational method is a procedure for finding the best possible approximate solutions to the eigenvalue problem given a fixed set of orthonormal functions (i.e. basis set). This is equivalent to solving the eigenvalue equation in a finite subspace spanned by the basis sets.

2.3.2 Perturbation Theory

Perturbation theory relies on the premise that the actual Hamiltonian of the system does not differ much from a known exact or approximate Hamiltonian of a system, and thus this deviation is treated as a perturbation: $H = H_0 + \lambda V$, where $0 \leq \lambda \leq 1$, being $\lambda = 0$ for an unperturbed system and $\lambda = 1$ the perturbed one. The energy and the wavefunction can be written as a Taylor expansion in terms of powers of the perturbation parameter

λ. Then, the following Schrödinger equation is obtained

$$\begin{aligned}(H_0 + \lambda V)(\Psi^{(0)} + \lambda\Psi^{(1)} + \lambda^2\Psi^{(2)} + \cdots) = \cdots \\ \cdots = (E^{(0)} + \lambda E^{(1)} + \lambda^2 E^{(2)} + \cdots)(\Psi^{(0)} + \lambda\Psi^{(1)} + \lambda^2\Psi^{(2)} + \cdots)\end{aligned} \tag{2.10}$$

By considering different degrees of perturbation, it is possible to obtain the different Moller-Pleset perturbation series. For example, by taking into account all elements of the expansion until λ^2, the second order corrected perturbation series, MP2, which is interesting and accurate towards predicting geometries. This type of calculations are size-consistent, but they are not variational.

2.4 Hartree-Fock Approximation

It is only possible to solve Schrödinger equation in an exact manner for very few systems, such as the hydrogen atom or H_2^+. Nevertheless, most systems, in which quantum and computational chemists are interested in, are multielectronic and must be solved using approximations. The appropriate wavefunction describing a single electron is a spin orbital $\chi(\mathbf{r})$ for $\mathbf{r} = (\mathbf{r}, \omega)$ indicating space and spin coordinates respectively. For a certain number of electrons, the first approximation to the wavefunction is a Hartree product, which is obtained as the product of N spin orbitals wavefunctions

$$\Psi(r_1, r_2, \ldots, r_N) = \chi_i(r_1)\chi_j(r_2)\cdots\chi_k(r_N) \tag{2.11}$$

The Hartree product is an uncorrelated or independent-particle wavefunction. This means that the probability of finding one electron in a certain position is independent of any other electron position. Also, a Hartree Product does not take into account the indistinguishability of fermions (i.e. electrons), and thus, in order to build a well behaved wavefunction, the antisymmetry principle should be now introduced. Thus, for an N-electron system, we define a Slater determinant as

$$\Psi(r_1, r_2, \ldots, r_N) = \frac{1}{\sqrt{N!}} \begin{vmatrix} \chi_i(r_1) & \chi_j(r_1) & \cdots & \chi_k(r_1) \\ \chi_i(r_2) & \chi_j(r_2) & \cdots & \chi_k(r_2) \\ \cdots & \cdots & & \cdots \\ \chi_i(r_N) & \chi_j(r_N) & \cdots & \chi_k(r_N) \end{vmatrix} = |\chi_i\chi_j\cdots\chi_k\rangle \tag{2.12}$$

where $\frac{1}{\sqrt{N!}}$ is a normalization factor. This expression accounts for N electrons occupying N spin orbitals without specifying which electron goes into which orbital, and antisymmetry principle and Pauli exclusion principle (just one electron per spin orbital) are accounted for. It is particularly interesting to highlight that a Slater determinant incor-

porates exchange correlation; this is that the motion of two electrons with parallel spins is correlated, but not that of two electrons with non-parallel spin.

The simplest antisymmetric wavefunction describing the ground state of an N-electron system is a Slater determinant $|\Psi_0\rangle = |\chi_1\chi_2\cdots\chi_N\rangle$. Upon applying the variational principle, the best wavefunction is the one with the lowest possible energy, being $E_0 = \langle\Psi_0|\hat{H}|\Psi_0\rangle$, and $\hat{H}$ the full electronic Hamiltonian. The Hatree-Fock equation is an eigenvalue equation with the form

$$f(i)\chi(r_i) = \epsilon\chi(r_i) \tag{2.13}$$

where $f(i)$ is an effective one-electron operator called the Fock operator

$$f(i) = -\frac{1}{2}\nabla_i^2 - \sum_{A=1}^{M}\frac{Z_A}{r_{iA}} + v^{HF}(i) \tag{2.14}$$

where $v^{HF}(i)$ is an average potential experienced by the ith electron due to the presence of the other electrons. As a result, the many-electron problem is replaced by a one-electron problem treating electron-electron repulsion in an average way. This potential depends on the spin orbitals of the other electrons, so this Hartree-Fock equation is not linear and must be solved in an iterative way: in fact the Hartree-Fock method is sometimes referred as self-consistent-field (SCF).

The SCF method relies on an initial guess of the spin orbitals, which are used to calculate the average field. Then the eigenvalue equation for a new set of spin orbitals is solved until convergence to a minimum energy wavefunction is reached. This solution provides an orthonormal set of spinorbitals to be inserted in a Slater determinant, which is the best variational approximation to the ground state of the system for a single determinant configuration. In theory, there would be an infinite number of basis functions and an infinite number of unoccupied or virtual spinorbitals; while, in practice, a finite set of spatial basis functions is introduced, thus the double of them for spinorbitals. As a result of this, the larger and more complete the basis set is, the greater the flexibility, and the lower the energy will be, until the Hartree-Fock limit is reached.

Basis Sets

Calculations of this kind rely on the approximation of molecular orbtitals by using linear combinations of atomic orbitals (LCAO), which are functions centered on each atom. Thus, a basis set should be specified for each atom and should be chosen in a careful way. John Pople introduced the concept of Minimal Basis Set which refers to the minimum

number of basis functions for a calculation to work. Nevertheless, increasing the flexibility is a must in order to obtained more accurate results, which can be achieved by increasing the number of the basis functions (double-zeta, triplet-zeta,...), using split-valence functions (different contraction schemes for valence and core orbitals), introducing polarization basis functions (having a higher angular momentum which allow to describe better the electron density), or introducing diffusion functions (additional low angular momentum functions which allow to describe broad electron density and long-distance interactions).

In 1989, Dunning realised that it would be a good idea to optimize the basis set with correlated wavefunctions [29], and were designed to converge towards the complete basis set limit, including polarization basis functions. These are called cc-pVDZ for a double-zeta scheme, cc-pVTZ for a triple-zeta scheme, and so on. One set of diffusion functions may be added to the basis set; i.e. aug-cc-pVDZ has diffuse s,p and d functions.

Other type of basis set is the Atomic Natural Orbitals (ANO), which follow a general contraction scheme, different from the Dunning basis set. ANO basis set aim at contracting a large number of primitive of functions to a small number of contracted functions by using natural orbitals from correlated calculations on the isolated atoms. In particular, ANO-L will be used throughout this work [30], which are obtained by averaging the density matrix over several atomic states, ions and the atom in an electric field. Also, ANO-RCC are constructed using CASSCF/CASPT2 density matrices and will be used for studying scalar relativistic effects, such as Spin-Orbit Coupling.

2.5 Post Hartree-Fock Methods

2.5.1 Configuration Interaction and Related Methods

Whitin the frame of Hartree-Fock theory, the ground state Slater determinant is one out of $\binom{2K}{N}$ (combinations obtained with N total electrons and 2K total spin orbitals). Using the Hartree-Fock ground state determinant as a reference, the other determinants can be considered as approximate excited states in which one electron is promoted into another orbital leaving one hole after it. Thus, linear combinations of all posible determinants can be used (including the Hartree-Fock ground state) in order to obtain a more flexible and accurate description of both the ground and the excited states of the system. As a result, the exact wavefunction of any state of a N-electron system can be written as a linear combination of all possible N-electron Slater determinants resulting from a complete set

of spin orbitals

$$|\Phi\rangle = c_0|\Psi_0\rangle + \sum_{ra} c_a^r|\Psi_a^r\rangle + \sum_{a<b,r<s} c_{ab}^{rs}|\Psi_{ab}^{rs}\rangle + \sum_{a<b<c,r<s<t} c_{abc}^{rst}|\Psi_{abc}^{rst}\rangle + \cdots \tag{2.15}$$

being $|\Psi_0\rangle$ the ground state configuration, $|\Psi_a^r\rangle$, the single excitation determinants from the orbital a to the orbital r, $|\Psi_{ab}^{rs}\rangle$ the double excitation determinants from the orbitals a and b to the orbitals r and s, and so on. Within this scheme, the exact energies of the ground and excited states are the eigenvalues of the Hamiltonian, and this method is the so called configuration interaction method (CI). The lowest eigenvalue (ϵ_0) is the exact non-relativistic ground state energy of the system within the Born-Oppenheimer approximation and

$$E_{corr} = \epsilon_0 - E_0 \tag{2.16}$$

is defined as the correlation energy, being E_0 the Hartree-Fock limit energy, since the motion of electrons with non-parallel spin is not correlated within the Hartree-Fock approximation.

This procedure, despite exact, has no practical applications as it is impossible to handle infinite basis sets. Reducing the basis set to a finite set of spin orbitals, the method, which is called full CI, leads to exact solutions within the N-electron subspace spanned by the $\binom{2K}{N}$ determinants. Nevertheless, the number of determinants included in full CI is extremely large, and some truncation must be done: CID, CISD, including only double or single and double excitations respectively. In the end, a compromise between number of basis functions (Hartree-Fock limit) and number of Slater determinants (size of the CI expansion) must be agreed to do accurate, but at the same time feasible, calculations.

Upon definition of the operators that generate single (T_1), double(T_2),... excitations, the corresponding Slater determinants are generated and $T = T_1 + T_2 + ... + T_{2K}$ is the complete operator which would develop the Full CI wavefunction. The Coupled-cluster method (CC) uses the Taylor expansion of e^T, and can be easily truncated. Including T_1 and T_2 operator terms, allow us to have even single excitations to the fourth power, which is the case for CCSD expanded as

$$e^{(T_1+T_2)} = 1 + T_1 + (T_2 + \frac{1}{2!}T_1^2) + (T_2T_1 + \frac{1}{3!}T_1^3) + (\frac{1}{2}T_2^2 + \frac{1}{2}T_2T_1^2 + \frac{1}{4!}T_1^4) \tag{2.17}$$

Size-consistency refers to the idea that the energy of a many-particle systems, even in the presence of interaction, is proportional to the number of particles in the limit (N$\rightarrow$ inf). Full CI is size-consistent, while truncated CI is not. Thus, unlike CISD, CCSD is size-consistent and is able to recover a great fraction of the correlation energy, provided that HF configuration is dominant. If we were able to do CCSDT, the accuracy would be close

to that provided by a Full CI. However, due to its extreme computational cost, CCSD(T) is generally used instead. Here, the triple excitation contribution is taken into account using perturbation theory.

2.5.2 MultiConfigurational SCF, CASSCF and CASPT2

In CI-related methods, only the coefficients expanding the linear combination of Slater determinants are variationally optimized. At difference, multiconfigurational methods also involve the optimization of the coefficients expanding the molecular orbitals in terms of the atomic orbitals. This latter approach is called multiconfigurational self-consistent field method (MCSCF), and, *a priori*, using a smaller number of configuration state functions (CSFs, which are symmetry and spin-adapted linear combinations of Slater determinants) it is possible to obtain more accurate results than single reference CI-related methods. Thus, a good compromise between accuracy and computational cost might be reached upon good selection criteria on how to include or not certain CSFs in the calculation.

CASSCF

One particular MCSCF approach, and also one of the most widely used, is Complete Active Space SCF method (CASSCF). Its efficiency relies on undertaking a Full-CI calculation within a specific set of orbitals. For this, three different orbital regions are defined:

- Inactive space: doubly spatial orbitals.
- Active space: spatial orbitals with occupation numbers between 0 and 2.
- Virtual space: spatial orbitals completely unoccupied.

In fact, the active space, where the Full CI approach will be applied, must be selected by hand and must include chemically-relevant orbitals. This method presents some limitations. One of them is the maximum of orbitals to be involved in the active space (usually $\sim$12-16) depending on computational budget. Even more important is that this method does not include dynamical correlation. There might some problems arising from a state-specific approach such as root flipping (in particular for states which lie close in energy), slow convergence, non-orthogonal states and the computations of some transition properties. A state-average (SA-CASSCF) approach might sort some of these problems, which consists of introducing a weighted average of the energies of each state. The average

energy can be written as

$$E^{average} = N^{-1} \sum_{i}^{N} E_i \tag{2.18}$$

and the SA-CASSCF minimizes this value optimizing the molecular orbitals for N states, which should include our state of interest.

CASPT2

In order to incorporate dynamic correlation and obtain more accurate energies and transition properties (CASSCF tends to overestimate excitation energies), CAS second-order perturbation theory (CASPT2) can be used, as it introduces dynamic correlation via perturbation theory employing the CASSCF wavefunction as reference, defining the Hamiltonian with one-electron terms and expanding. An interesting aspect of CASPT2 is the occurrence of intruder states. An intruder states is characterized when its zeroth-order energy of one or more state functions used to build the first-order wavefunction has a similar value to that of the reference state. Thus, the denominator tends to zero, and singularities may arise from it. It may be tackled by increasing the size of the active space, which can be unpractical in some cases due to computational cost. Thus, level shift CASPT2 [31] is an interesting correction as it introduces a penalty parameter in the denominator in order to control the weight of the intruder states. This approach is controversial and other alternative solutions have been proposed, such as introducing an imaginary level shift which prevents additional new singularities from the application of a shift [32].

Also a correction of single-state CASPT2 is Multistate CASPT2 (MS-CASPT2) [33], which comes in hand when describing nearly degenerated states, avoided crossings or Rydberg states mixing. It solves the non-orthogonality of the wavefunctions; also, it uses as a reference function the SA-CASSCF wavefunction. Another interesting correction is **IPEA-CASPT2**. It was introduced after noticing a widespread underestimation of dissociation energy for open-shell systems, as an empirical correction. In this work, after checking the effect of introducing or not an IPEA shift, it was decided not to use it in CASPT2 calculations; however, MS-CASPT2 and low level shift correction have been used.

2.6 Density Functional Theory

Density Functional Theory (DFT) provides, in most cases, a good compromise between accuracy and computational cost, giving the chance to describe moderately large systems from first principles. DFT is able to describe the electronic states of atoms, molecules and solids in terms of the electronic density ρ, by means of a functional. In 1964, Hohenberg and Kohn published the first paper which led to the development of DFT [34]. They showed that there must exist a functional $F[\rho]$ such that the ground state energy can be expressed as the minimum of the functional

$$E[\rho] = F[\rho] + \int V(r)\rho(r)dr \tag{2.19}$$

where $\rho(r)$ is the electron density and $F[\rho]$ is the functional which is independent of the system. Additionally, the ground state properties of the system can be described in terms of the electron density, instead of using a many-particle wavefunction as in HF-based methods.

DFT relies on the variational principle (section 2.3.1) to determine the energy of the ground state of a given system, which in the end determines the ground state wavefunction Ψ_0 and the energy E_0 for a system with N electrons and a given nuclear potential V_{ext}. Ultimately, the ground state energy is a functional of the total number of electrons and the nuclear potential $E_0 = E[N, V_{ext}]$.

The first Hohenberg-Kohn theorem indicates that the electron density determines the Hamiltonian operator, and, as a consequence, all the properties of the system. Being the Hamiltonian fully determined, except for a constant shift of the energy, it follows that the many-body wavefunctions for all states, ground and excited, are derived. Therefore, all properties of the system can be inferred from the ground state density exclusively.

The second Hohenberg-Kohn Theorem ensures that ground state density exists, and that the ground state energy is obtained provided that the true electron density is used. Also, the functional $E[\rho]$ on its own is enough to determine the exact ground state energy and density.

2.6.1 Kohn-Sham Method

In 1965, Kohn and Sham proposed a revolutionary idea that made available the use of DFT in computational chemistry by introducing the concept of orbitals in this scheme [35]. The idea is to substitute the complex interacting many-body system by an auxiliary system, which can be solved as it is formed by a set of independent non-interacting

particles. The Kohn-Sham ansatz assumes that the ground state density of the original interacting system is equal to that of the non-interacting one and can be built upon two main assumptions: (1) the exact ground state density is equal to the ground state density of the auxiliary system of non-interacting particles, and (2) the auxiliary hamiltonian is chosen in order to have the usual kinetic energy operator and an effective local potential acting on an electron of spin σ at point r.

The electron density $\rho(r)$ and the kinetic energy $T_{KS}[\rho]$ are respectively defined as

$$\rho(r) = \sum_{i=1}^{N} |\psi_i(r)|^2 \tag{2.20}$$

$$T_{KS}[\rho] = -\frac{1}{2}\sum_{i=1}^{N} \langle \psi_i | \nabla^2 | \psi_i \rangle \tag{2.21}$$

where ψ_i represent the occupied molecular orbitals and the sum runs over each occupied molecular orbital N. Similarly, the kinetic energy associated to the complex system is replaced by the kinetic energy associated to the auxiliary one. So, the universal functional is now rewritten as

$$F_{HK}[\rho] = T_{KS}[\rho] + \frac{1}{2}\int \frac{\rho_1(r)\rho_2(r')}{|r-r'|} dr dr' + E_{xc}[\rho] \tag{2.22}$$

The second term is the expression for the Hartree energy in terms of the Hartree density and potential

$$E_{HT}[\rho(r)] = \frac{1}{2}\int v_H(r)\rho(r) dr = \frac{1}{2}\int \frac{\rho_1(r)\rho_2(r')}{|r-r'|} dr dr' \tag{2.23}$$

All the many body effects of exchange and correlation are grouped into the exchange and correlation energy

$$E_{xc}[\rho] = F_{HK}[\rho] - (T_{KS}[\rho] + E_{HT}[\rho]) \tag{2.24}$$

The determination of the orbitals proceeds exactly as in Hartree-Fock theory, by solving the Kohn-Sham equations $H^{KS}\psi_i^{KS} = \epsilon_i^{KS}\psi_i^{KS}$ The Fock operator and the Kohn-Sham operator (H^{KS}) are almost the same, except for the exchange correlation potential part that in DFT is defined by the operator $V_{xc}(r)$, in contrast to the complicated term defined in HF. Finally, this exchange correlation potential operator is such that it contains all the many electron effects, in contrast again to the HF operator, which includes the exchange but not the electron correlation.

The self-consistent solution of the Kohn-Sham equation eventually delivers the ground state density of the system from which, a priori, all the properties of the system can be derived. The critical point is to find a way to describe the exchange-correlation energy

functional, which has brought many flavours along the years, i.e. Local Density Approximation (LDA), Generalized Gradient Approximation (GGA) and hybrid functionals among others.

LDA considers the exact exchange-correlation potential term of a given particle located in r, and only depends on the electron density at that specific point

$$E_{xc}^{LDA} = \int \rho(r)\epsilon_{xc}^{LDA}[\rho(r)]dr \tag{2.25}$$

where ϵ_{xc}^{LDA} is the sum of the exchange and correlation contribution. In particular,

$$\epsilon_{xc}^{LDA} = -\frac{3}{4}(\frac{3}{\pi})^{\frac{1}{3}}\rho(r)^{\frac{4}{3}} \tag{2.26}$$

is the LDA exchange energy, and the correlation terms are inferred from parametric equations fitting perturbative or quantum Monte Carlo calculation results. Some examples are LYP [36] or PW92 [37]. This exchange-correlation functional is exact for the uniform electron gas (for example, s an p blocks of the periodic table, where there are homogeneous electron densities).

GGA uses the gradient of the density as a variable, introducing thus, a semi-local character of the electron density in the formulation.

$$E_{xc}^{GGA}[\rho, \nabla\rho] = \int \rho(r)\epsilon_{xc}^{GGA}[\rho(r), \nabla\rho(r)]dr \tag{2.27}$$

Then, the exchange energy terms is

$$E_{x}^{GGA}[\rho, \nabla\rho] = \int F_x(s(r))\rho(r)^{\frac{3}{4}}dr \tag{2.28}$$

where F_x corresponds to a function of the reduced density gradient defined as $s(r) = \frac{|\nabla\rho(r)|}{6\pi^2\rho(r)^{\frac{3}{4}}}$ GGA functionals provide a better description of s, p, f, d blocks. Among the most popular GGA functionals, we can highlight BLYP [38] and PBE [39].

Hybrid functionals are a class of functionals used to approximate the exchange-correlation, which contain a part of the exact Hartree-Fock exchange and another fraction inferred by interpolation of post-HF and semi-empirical results. The most popular hybrid functional is B3LYP, which is defined as

$$E_{xc}^{B3LYP} = (1-a)E_x^{LDA} + aE_x^{HF} + b\Delta E_x^{B88} + cE_c^{LYP} + (1-c)E_c^{VWN}$$

where a=0.20, b=0.72 and c=0.81, [40, 41]. Another example, is the long-range corrected ωB97X-D [42].

2.6.2 Time-Dependent DFT

Excited states may be described in the frame of DFT using TDDFT, which calculates poles in response of the ground state density to a time-varying applied electric field. These poles are Bohr frequencies (excitation energies). This involves the solution of an eigenvalue equation, with an exchange-correlation correction, iteratively for the lowest excitation energies (within the widespread non-TDA scheme). It is a popular method as it is as expensive as the simple CIS method, which in turn, accounts for electron correlation. This method is only suitable for low-lying valence excited states (below the first ionization potential of the molecule or below the first Rydberg state), and valence excitation energies are usually remarkably better than CIS despite being very functinal-dependent.

Time-dependent DFT can be applied if the molecular system is subject to a time-dependent external potential. Due to this potential, the electron density of the system changes in time. The time-dependent analogue of the first Hohenberg-Kohn theorem is the Runge-Gross theorem [43]. This theorem proves that the exact time-dependent electron density determines the time-dependent external potential up to an additive constant and thus the time-dependent wavefunction up to an additive time-dependent phase-factor. Thus, the time-dependent Kohn-Sham equation takes the form

$$[-\frac{\hbar}{2m_e}\nabla^2 + V_s(r,t)]\phi(r,t) = i\hbar\frac{\delta}{dt}\phi(r,t)$$

where V_s is the time-dependent one-electron local potential chosen, such that, the density of the non-interacting n-electron system is the same as the density of the interacting n-electron system.

Often the adiabatic approximation is applied, which assumes that the functional dependence of the time-dependent exchange-correlation potential on the exact time-dependent electron density is the same as in the time-independent case. Thus, existing functionals for ground state DFT can be used, in principle, for TDDFT calculations.

2.7 Methods for Describing Excited States

Experimental methods used for studying excited states in molecules include conventional, stationary spectroscopy and also time-resolved spectroscopy. This later approach emerges to obtain information of the fast electronic motion and also on the slower nuclear movement which may become strongly coupled in the case of potential energy surfaces (PES) approaching each other energetically.

Born-Oppenheimer and Franck-Condon (FC) approximations are the fundamental foundations used when calculating electronic spectra and for electronic states description. FC is a consequence of BO, which says that the probability of a transition between two vibrational levels is greater when their wavefunctions are similar. In the framework of FC (vertical) excitations, several methods based on closed-shell Hartee-Fock method can be used as a first guess and high-accuracy is expected. Franck-Condon relies on the dominance of a single reference configuration (closed-shell HF) using it as a reference in the ground state upon which excited states are produced and described. Even in these cases, a strong variation of the character of electronic states, such as Rydberg, charge transfer, local or delocalized, is generally found. Thus, choosing the appropriate method and basis set is a tricky question to answer when working with excited states.

In a broad sense, there are two main groups of methods: those based on a single reference (SR) wavefunction and those which use more than one reference to construct the wavefunction (multireference-MR). SR reference methods may divide the molecular orbitals in three different kinds: frozen, reference doubly occupied and unoccupied. Then, electronic excitations are generated from the reference doubly occupied set to the virtual unoccupied, leaving the frozen set doubly occupied always. Examples of SR methods are those based on CI,such as CIS and CISD; CC, such as EOM-CCSD; and TDDFT. [44]

In many situations, specifically away from the FC region where the vertical excitation takes place, SR method might lack of accuracy when describing the wavefunction. In order to add more flexibility to the wavefunction, MR methods are introduced, which generally divide the molecular orbital space in four different sets: frozen, reference doubly occupied, active and unoccupied virtual orbitals. Electronic excitations arise within the active space. In particular, CASSCF, see Section 2.5.2, includes all possible excitations that may arise from the active orbitals for a certain number of electrons. Thus, MR methods are able to use as a reference several wavefunctions, including those from the ground and the excited states, being able to describe in a more exact manner regions in which two states are degenerated, such as conical intersections (see next section). Apart from CASSCF, other MR methods are available such as MR-CI, MR-CC or CASPT2. [44]

In this work, and many others in the field of computational photochemistry, CASSCF and CASPT2 are used as reference methods. In particular, an interesting and succesful protocol is CASPT2//CASSCF. It consists in doing geometry optimization at CASSCF level and energies are calculated at CASPT2, which usually works fine. Nevertheless, this approach is not always valid. For instance, the states in which dynamical correlation effects are important in determining the geometry of the molecule, a CASPT2//CASSCF approach might be misleading and CASPT2//CASPT2 would result in a better option. Despite a much greater computational cost, it may help to describe those problematic regions and obtain better geometries.

2.8 Study and Description of Excited states

2.8.1 Light-Matter Interaction

Upon light-matter interaction, several processes can potentially occur. Very frequently, these processes are of photophysical nature, which occur leaving the chemical structure intact. On the other hand, photochemical processes involve a change in the chemical structure of the matter following light interaction. It is important to recall that light is technically electromagnetic radiation which can be classified according to its wavelength or frequency. In particular, the energy of monochromatic light - this is of one specific wavelength- is proportional to its frequency and inversely proportional to its wavelength according to $E = h\nu = \frac{hc}{\lambda}$. In this work, we will focus on the description of UV light induced processes, whose energy range usually involves electronic transitions, resulting in an excited electronic state of the molecule. Visible light is also able to induce electronic transitions in certain systems, though. Shorter wavelengths (higher frequencies), such as X- and γ-rays, will cause ionization processes or core and inner valence excitations; while longer wavelengths (lower frequencies), such as infrared and microwave radiation, will lead to rovibrational excitation. Thus, in this work, we will focus on UV light, within the range 10-400 nm, and visible light, within the range 400-700 nm.

Generally, when referring to a chemical or physical process, it is assumed that it takes place in the same electronic state, which is frequently the ground state of the molecule. Within the Born-Oppenheimer approximation (see Section 2.2), this means that the molecule stays in the same potential energy surface (PES), which is the energy profile accounted for in most of processes. Upon electronic excitation, the molecule is promoted to a higher energy PES, describing an excited state, which might change the shape of the energy profile along a particular reaction coordinate. As a result, different minima, reaction pathways and molecular properties are accesible.

2.8.2 Light Absorption

Generally, except for some particular species, the ground state of an organic molecule is a closed-shell singlet state (S_0). Upon photon absorption with appropriate energy, the molecule is excited to a higher electronic state. Nevertheless, this transition is restricted to: (1) spatial symmetry, where the dipole moment of the transition must be different from zero; (2) orbital overlap, as a direct consequence of conservation of spatial symmetry, leading to more intense transitions, the greater is the orbital overlap of both states; (3) spin conservation, as generally transitions with high intensities occur from a singlet to a singlet, or from a triplet to a triplet, etc. As a consequence of the Franck-Condon

principle, an electronic transition takes place in fs timescale, without nuclear movement, but with a change in the electric charge distribution. After a vertical excitation, the molecule may relax to a minimum geometry in the excited state PES. [45]

The interaction of a system with the electromagnetic field is determined by the polarization of the transition, which can be measured by the Transition Dipole Moment (TDM)

$$TDM = \langle \Psi_i | \hat{\mu} | \Psi_f \rangle \tag{2.29}$$

where $\hat{\mu}$ is the dipole moment operator. In particular, the intensity of an electronic transition, either for emission or absorption, is measured by the oscillator strength, *f*, proportional to the square of the TDM and defined as

$$f = \frac{2}{3}|TDM|^2 \tag{2.30}$$

being the energy difference between both states. Experimental absorption spectra can be compared with theoretical calculations by comparing f with the absorption intensity.

A molecular energy level comprises many more levels than an atomic one, where energy can only be deposited in the form of electronic excitations. In molecules, atoms can vibrate and rotate, which generates a set of rovibrational states for each electronic level. The band structure of molecular absorption spectra shows, that upon absorption of a photon, the transitions to any of the rovibronic states are not equally probable, leading to wide bands. The reason of this phenomenon is that an electronic excitation is much faster than the period of molecular vibration. Also, after excitation, a rearrangement of the electronic distribution takes place, leading to a change in vibrational modes. [45]

In organic UV/visible spectroscopy most of excitations depicted are of character $\pi\pi^*$ (an electron from a π bonding orbitalis promoted to a π antibonding orbital) or $n\pi^*$ (an electron from a lone pair is promoted to a π antibonding orbital). Generally, $\pi\pi^*$ transitions are more intense than $n\pi^*$, as the latter are spatially forbidden by breaking spatial symmetry and orbital overlap, resulting in smaller TDM. Other types of excitations involve σ orbitals, which are more energy-requiring, and are not in the typical energy range in which experiments are registered.

2.8.3 Deactivation Mechanisms for Excited States

Figure 2.1 shows the deactivation from an electronic excited state to the ground state.

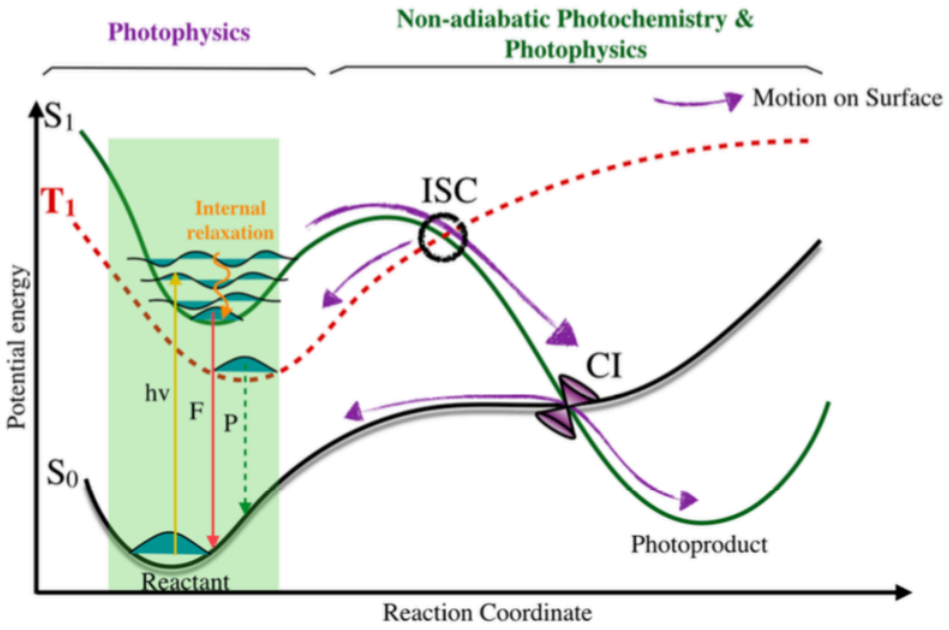


Figure 2.1: Jablonski diagram showing non-radiative and radiative deactivation pathways between electronic states. Legend: (F) Fluorescence, (P) Phosphorescence, (IC) Internal Conversion, (ISC) Intersystem Crossing. Taken from PhD Thesis by Serra Arslancan (May 2018).

Vibrational Relaxation

Vibrational or internal relaxation is the process where the molecule looses energy until the vibrational ground state of the excited state PES is reached. Nevertheless, it does not allow to reach back the ground state on its own, and need any of the mechanisms described next.

Internal Conversion

Internal conversion (IC) involves the transfer of population from an excited PES to the ground state PES or another lower-lying excited PES. In this case, the key topographical feature of the PES is a conical intersection (CI) region in which both states involved are nearly degenerate, and the population transfer is allowed, provided that both states have the same spin. At any point nearby a surface crossing, BO approximation is no longer valid as the non-adiabatic terms become large enough not to be neglected (see Section 2.2, which requires a non-adiabatic approach to describe the region. In particular,

ultrafast photophysical processes are thought to happen generally through CIs between two degenerate states of the same spin.

The topography of the two degenerate PES allows a classification of the CI involved, according to Ruedengerg and collaborators [46]: peaked and sloped, see Figure 2.2. A peaked CI is characterized by an increase of energy in all the directions of the upper PES and a decrease of energy in all the directions of the lower PES. The crossing points will be at the minimum energy point of the upper surface. A sloped CI is found when both PES have similar slopes, and generally only one minimum is reached after the CI (contrarily to peaked CI, where two different pathways to minima can be followed).

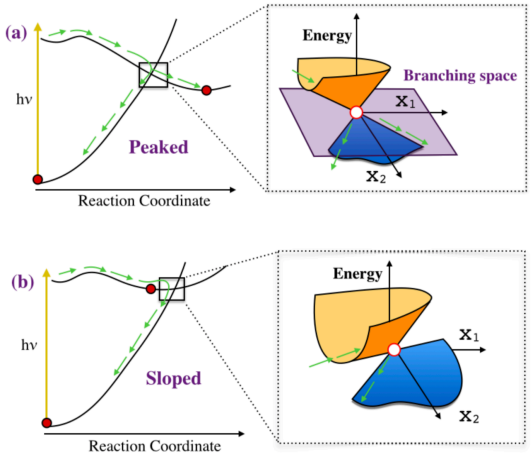


Figure 2.2: Peaked (top) and sloped (bottom) conical intersections. Taken from PhD Thesis by Serra Arslancan (May 2018).

Intersystem Crossing

Intersystem crossing (ISC) which is analogous to internal conversion processes but here states of different spin are involved. As a consequence, this process requires spin-flipping, which is usually not a very fast event, but it is still possible in the fs timescale. Despite being quantum mechanically forbidden, due to the orthonormality of singlet and triplet wavefunctions, spin-orbit coupling (SOC) allows ISC to happen and even compete with IC. SOC arises as a result of an electron creating both spin and orbital angular momenta, and its interaction with the charge of the nucleus, reason why it is particularly favorable in heavy atoms. Also, El-Sayed rules [47] state that, provided that the ISC involves a change in state character, such as from a $\pi\pi^*$ state to a $n\pi^*$ state of different multiplicity and viceversa, the process may be fast.

Fluorescence

Fluorescence takes place by the emission of a photon, simultaneous to the retrieval of the population from the excited to the ground state. It can be characterized by its emission spectrum, which generally differs from the absorption spectrum by a shift to the red known as the Stokes shift.

Phosphorescence

Phosphorescence occurs by the emission of a photon, simultaneous to the return of the population from the excited state of different multiplicity to the ground state. Commonly it is also shifted to longer wavelengths, which, due to its spin-forbidden nature, takes place slower than fluorescence.

Chapter 3

Results and Discussion

3.1 Characterization of the Predominant Species of Barbituric Acid in the Ground State

Barbituric acid (BA), shown in Figure 1.4 left, is an organic molecule of particular interest for biological media. In particular, its derivatives (called barbiturates) are involved in several pharmacological applications as hypnotic, anaesthetic, antitumoral drugs, etc [48]. Due to its interesting pharmacological properties, BA has been extensively studied and highly precise chemical and physical parameters, such as structural and spectroscopic data, are available in several conditions. According to Briggs and colleagues [49], barbituric has been reported to have a pK_a of 3.99 at room temperature, showing a more acidic character in comparison to other substituted barbiturates. For this reason, at pH $<$ pK_a we would expect BA to be neutral, while for pH $>$ pK_a it should be deprotonated. In particular, at physiological pH, BA is assumed to be deprotonated.

In section 1.2, the most generally accepted conditions of prebiotic Earth were discussed. At this time, the Earth is believed to have been subject of abrupt pH changes and gradients along different media even within a reduced amount of time and physical space, most probably ranging throughout the whole pH ladder. As a consequence of these harsh pH variations, BA is thought to have been present in both neutral and deprotonated forms at prebiotic era. Therefore, all those derived species should be taken into account in order to have a reliable study on the stability of this non-canonical nucleobase. Not surprisingly, due to existence of double C-O bonds and α hydrogens, several tautomeric forms might occur in aqueous media. Thus, first step is to determine which of these are the predominant species in the selected media, so that we can focus on the study of their photostability. Note that photophysics of a chemical entity might change dramatically depending on its tautomeric form.

For determining the most abundant species in aqueous solution, all tautomers of BA, both deprotonated and neutral, were considered. Altogether, 30 species (see Figure A.2) were determined for acidic media pH $<$ pK_a; while 18 species (see Figure A.1) occurred for neutral/alkali media, pH $>$ pK_a (i.e. physiological pH). The structures of these species were optimized in the gas phase, and were energetically ordered according to their Gibbs Free Energy. Geometry optimizations were performed at MP2/cc-pVTZ level of theory, while single point energy calculations were done at MP2/aug-cc-pVTZ including water as a continuum with PCM as implemented in Gaussian09. Results were collected in Table A.2 and Table A.1. Neutral tautomers were grouped into 10 families[1], and the most stable form of each family was selected for further analysis, reducing the number of candidates from 30 to 10. All the deprotonated species were, however, taken into account, due to the more reduced and, thus, more reasonable number. Nevertheless, the most stable structures for both neutral and deprotonated species were those containing three keto substituents (see Figure 3.1).

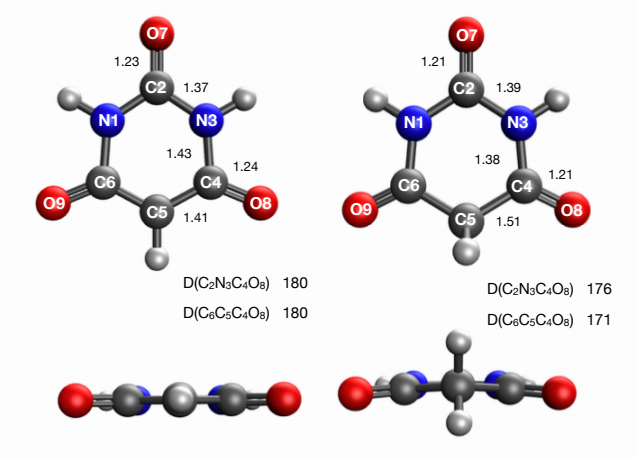


Figure 3.1: Most stable structure of deprotonated (left) and neutral (right) BA species at their minimum energy geometry at MP2/cc-pVTZ level of theory. Color code: blue-N, red-O, black-C and white-H. Distances reported in $\AA$ and angles in degrees.

In BA, carbonyl, amine and hydroxy groups present lone pair electrons, acting as potential hydrogen bond acceptors, but also amine and hydroxi groups might act as hydrogen bond donors, when interacting with water molecules from the media. In order to obtain a more realistic picture of solvent solute interactions, and thus of neutral and deprotonated BA, three explicit water molecules were placed around the molecule in specific positions (see Figure A.3). BA with 3 water molecules clusters were reoptimized for the previously selected structures in gas phase at MP2/cc-pVTZ level of theory and single point energy

[1]In each family group, the tautomers have the same bonding pattern differing on the hydrogens relative orientation. Also, each family is assigned a letter, followed by a number in the case that more than one tautomer belongs to a family.

calculations at the minimum energy geometries were done at MP2/aug-cc-pVTZ. Then, species were ordered according to Gibbs Free Energy as shown in Tables A.1 and A.2.

The Boltzmann distribution has been widely used as a probability distribution which indicates the population distribution for certain states as a function of their energy and the system temperature. According to this distribution, the fraction of population characterized by a particular energy value (i.e."A") is proportional to the exponential assuming a canonical partition function as specified below

$$P(''A'') \quad \alpha \quad e^{\frac{-\Delta G}{k_B T}}$$

being k_B Boltzmann constant, T the temperature of the system and ΔG the relative energy. The distribution is usually normalized so that the sum of the occurrence for all accessible states adds up to 1.

According to a Boltzmann distribution at room temperature (T = 298 K), the predominant species for the cluster of deprotonated BA including three water molecules falls into one single tautomer, the one containing three carbonyl groups (tautomer A) as reported in Table 3.1 and Figure A.1. Our results are consistent with the literature on the deprotonated tautomer, which also focuses on this species [50]. Other deprotonated forms are expected to contribute to the total population with smaller weights than $\sim 0.01\%$, so they were not considered.

On the other hand, a single neutral tautomer is thought to occur as a predominant species ($\sim$98%), with residual contribution from other species, 4-hydroxy (C2) ($\sim$2%), see Table 3.2 and Figure A.2. Interestingly, the fraction of the second most abundant tautomer increases to $\sim$17%, competing with the triketo species (A), when considering the electronic potential energy instead.

Figure 3.1 shows the geometries of both gas phase neutral and deprotonated predominant tautomers. Both geometries are almost planar and symmetric along the plane containing C_2-O_7 and CH/CH_2 groups. As a remark, the geometries of the BA water-containing clusters, see Figure A.3, are extremely similar compared to gas phase BA species in terms of planarity of the ring and its substituents. In fact, these planar minima are in agreement with crystallographic [51] and theoretical data [52, 53] for neutral and deprotonated [53] forms.

Table 3.1: Gibbs Free Energy for each of the deprotonated BA·3H_2O clusters at room temperature (T=298K) in a water continuum under PCM scheme at MP2/aug-cc-pVTZ level of theory and population distribution according to a Boltzmann distribution. Labelling of tautomers corresponds to Figure A.1.

Protonated Species	Gibbs Free Energy (Hartree)	ΔG° (kcal/mol)	Population (%)
A	-717.765586	0.00	99.99
G	-717.756463	5.73	0.01
D1	-717.750813	9.27	0.00
I1	-717.745431	12.65	0.00
H	-717.743398	13.92	0.00
F3	-717.740198	15.93	0.00
F2	-717.739841	16.16	0.00
C2	-717.738026	17.29	0.00
E1	-717.737612	17.55	0.00
E3	-717.736285	18.39	0.00
F1	-717.735559	18.84	0.00
D2	-717.734706	19.38	0.00
F4	-717.733725	19.99	0.00
B2	-717.732154	20.98	0.00
B1	-717.730052	22.30	0.00
C1	-717.728429	23.32	0.00
I1	-717.725370	25.24	0.00
E2	-717.654656	69.61	0.00

Table 3.2: Gibbs Free Energy for each of the neutral BA·3H_2O clusters at room temperature (T=298K) in a water continuum under PCM scheme at MP2/aug-cc-pVTZ level of theory and population distribution according to a Boltzmann distribution. Labelling of tautomers corresponds to Figure A.2.

Protonated Species	Gibbs Free Energy (Hartree)	ΔG° (kcal/mol)	Population (%)
A	-718.204485	0.00	97.82
C2	-718.200896	2.25	2.18
F3	-718.189074	9.67	0.00
C3	-718.187067	10.93	0.00
D2	-718.183358	13.26	0.00
B1	-718.182172	14.00	0.00
I3	-718.181062	14.70	0.00
H3	-718.179550	15.65	0.00
G3	-718.165821	24.26	0.00
E3	-718.162796	26.16	0.00

The concluding remarks of this section which are relevant for the following undertaken studies are:

1. The deprotonated form of BA appears to be represented by a single tautomer, the triketo (A), independently of the thermodynamic scheme considered: potential vs. Gibbs Free energy.

2. The neutral form of BA seems to be represented by a predominant species (triketo - A), and a secondary one (4-hydroxy - C2) which might also be present in solution in different proportions depending on the scheme used. Subsequent spectroscopic studies were, however, undertaken on the predominant species of each form, although they could be easily expanded to cover the 4-hydroxy species of the neutral form.

The next sections of this work will focus on getting insights into the properties of the electronic excited states of BA, using a theoretical approach and contributing to the interpretation of experimental results [4].

3.2 Absorption Spectra Calculation of BA

In this work, most of the calculations involving excited states description have been performed in Molcas 7.8 under the MS-CASPT2//SA-CASSCF scheme, see section 2.5.2, where single point calculations at MS-CASPT2/ANO-L-VTZ were undertaken on geometries optimized at the SA-CASSCF/ANO-L-VTZ level of theory. These methods require an active space selection and this was carefully selected by analyzing the low-lying excited states character of preliminary calculations done at the Franck-Condon geometry. At first, the active space includes the complete π system and a lone pair of each of the C-O, which correspond to chemically-relevant and generally involved in the low-lying electronic transitions of small organic molecules.

For the neutral species, 5 π orbitals, 3 lone pairs and also 3 antibonding π^* were included in the preliminary calculations. Out of these orbitals, one lone pair was later excluded from the active space as its occupation was too high, letting other non-chemically relevant orbitals for the description of the low lying excited states in. Thus, in a second stage, calculations were performed using an active space of 14 electrons and 10 molecular orbitals made up of 5 π, 2 lone pairs and 3 π^* as depicted in Figure 3.2. Also, four states were considred in our calculations; specifically the ground state (S_0), two dark states (S_1 and S_2) and the bright or spectroscopic state (S_3), as can be seen in Table 3.3.

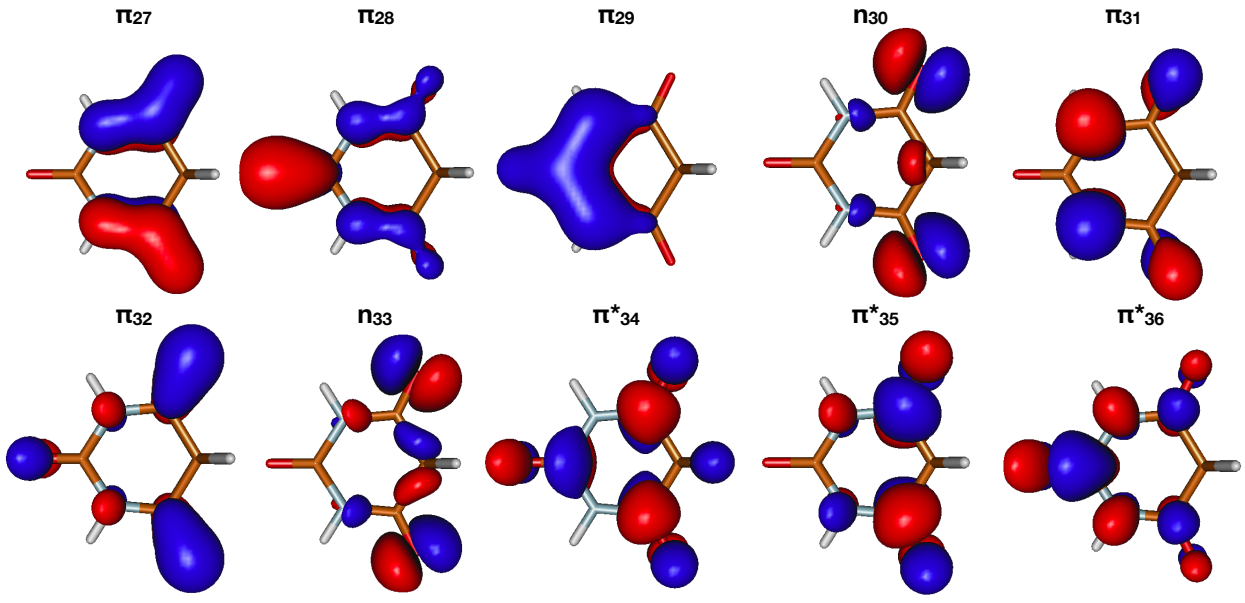


Figure 3.2: CASSCF molecular orbitals included in active space (27-36) for calculations of the neutral BA ground state minimum at MS-CASPT2(14,10)//ANO-L-VTZ level of theory with 4 roots.

Table 3.3: Excitation character, energy and absorption wavelength at the Franck-Condon geometry of the neutral predominant tautomer at MS-CASPT2(14,10)//ANO-L-VTZ level of theory with 4 roots.

Character	Excitation Energy (eV)	λ_{abs} (nm)	Osc. strength
$S_0 \rightarrow S_1$ $(n_{33}\pi^*_{34})/(n_{30}\pi^*_{35})$	4.65	266	0.000006
$S_0 \rightarrow S_2$ $(n_{33}\pi^*_{35})/(n_{30}\pi^*_{34})$	4.88	255	0.000741
$S_0 \rightarrow S_3$ $(\pi_{31}\pi^*_{34})$	5.48	226	0.294367

On the other hand, for the deprotonated form of BA, all 6 π orbitals (as C_5 is now sp^2, instead of sp^3 as in the neutral form), 3 lone pairs and 3 antibonding π^* were included at first. After the analysis of the character of the lowest lying electronic transitions, all lone pairs were excluded from the active space as their occupation was quite high. Furthermore, the first excited state (S_1) was the spectroscopic one and transitions involving lone pairs were much higher in energy. One of the most stable excitations at the FC region was found to be a Rydberg excitation that quickly raises in energy along the relaxation coordinate of the spectroscopic state. Thus, calculations were performed using an active space of 12 electrons and 10 molecular orbitals (including the Rydberg as seen in Figure 3.3), and three states for all the calculations: the ground state (S_0), the dark Rydberg state (S_1) and the spectroscopic state (S_2), as reported in Table 3.4. protonated

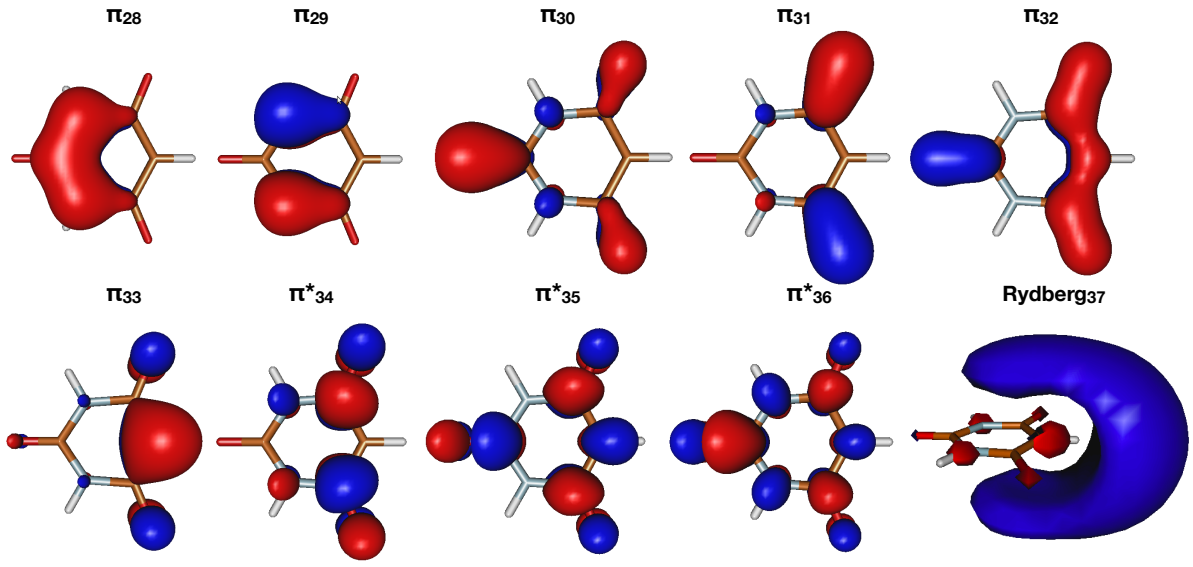


Figure 3.3: CASSCF molecular orbitals (28-37) included in active space for the calculation of the deprotonated BA ground state minimum at MS-CASPT2//ANO-L-VTZ level of theor with 3 roots. Orbital 37 is a Rydberg state which changed in orientation for a better representation.

Table 3.4: Excitation character, energy and absorption wavelength at the Franck-Condon geometry of the deprotonated predominant tautomer at MS-CASPT2(12,10)//ANO-L-VTZ level of theory with 3 roots.

Character	Excitation Energy (eV)	λ_{abs} (nm)	Osc. strength
$S_0 \rightarrow S_1(\pi_{33}Ryd_{37})$	4.07	304	0.004988
$S_0 \rightarrow S_2(\pi_{33}\pi^*_{34})$	4.59	270	0.601026

Absorption spectra were calculated in gas phase using the optimized geometries obtained as described in the previous section excluding the three explicit water molecules, and the active space as commented in the previous paragraphs. Table 3.3 shows excitation energies at the Franck-Condon region for the neutral form, while table 3.4 shows those for the deprotonated species. For the neutral species, the lowest lying bright electronic state lies at 226 nm and shows $\pi\pi^*$ character, being localized at $C_4C_5C_6$. For the deprotonated species, the spectroscopic state is stabilized and absorbs at 270 nm, showing a $\pi\pi^*$ character.

Figure 3.4 shows the absorption spectra, both for the neutral and deprotonated species, and they mostly coincide with the experimental spectra at the corresponding pH. Calculations of absorption spectra were done at MS-CASPT2//ANO-L-VTZ level of theory with 15 roots. The line spectrum was convoluted using gaussians and a FWHM (Full Width at Half Maximum) of 10 nm to better reproduce the experimental absorption spectrum [4].

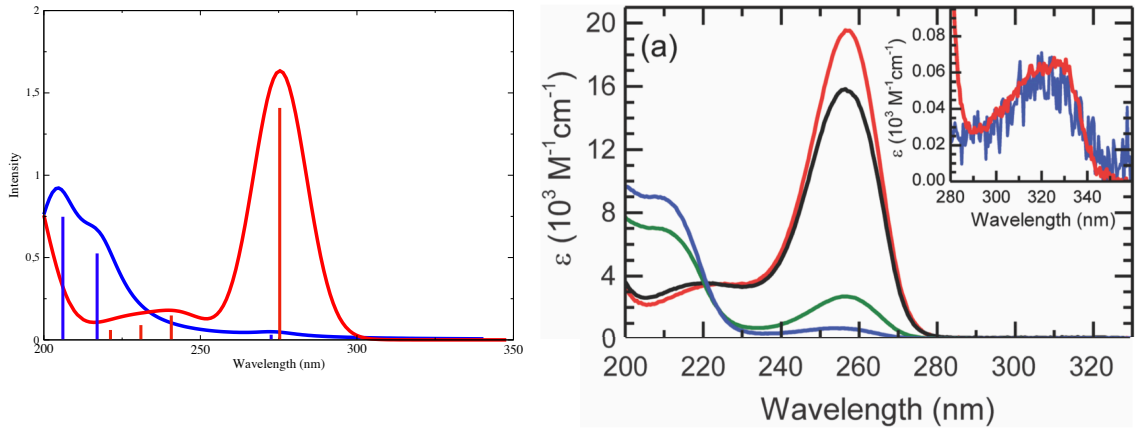


Figure 3.4: Calculated (left) and experimental (right) absorption spectra of the predominant neutral (blue) and deprotonated (red) tautomeric forms of BA within 200-350 nm range in gas phase. Experimental data was taken from [4].

3.3 Exploration of the Excited State Potential Energy Surfaces of BA at Different pH Values

In order to explore the potential energy surfaces (PES) of the different species of interest, we are going to use several techniques. A minimum energy pathway (MEP) is defined as the path determined by a sequence of geometries obtained by subsequent optimizations on a hypersphere [54]. In these calculations, the minimum energy path is found by minimizing the energy of an hypersphere with a fixed radius in mass-weighted coordinates, where the origin of the sphere is the geometry of the reference structure. This implies tracing a minimum steepest-descent energy path to a final minimum energy geometry or a degeneracy in a downward manner. As matter of fact, this was the main approach used for the exploration of the PES in combination with CASSCF level of theory. Then, for the critical points of the PES such as minima and crossing points, CASPT2 was used for the calculation of the final energies after incorporating dynamic correlation in the wavefunction (see 2.5.2).

Starting from the Franck-Condon geometry, which corresponds to the minimum energy geometry of the ground state PES, a MEP calculation along the spectroscopic state of interest was launched. This leads either to a crossing point with a lower lying state, or to a minimum. Starting from a two state crossing point, and assuming a peaked topography according to Ruedenberg classification [55], two MEPs are in principle possible, each of which would follow the gradient of one of the two electronic states involved in the crossing. On the other hand, if a minimum point is found, the first step is to perform a geometry optimization and frequency calculation in order to check whether it is a real minimum or other kind of stationary point. To further explore the PES, starting from the minimum position we searched for a minimum energy crossing point (MECP) with the algorithm implemented in Molpro [56–58], which are based on the calculation of the gradient of both states involved in the crossing and their non-adiabatic coupling. This new MECP would act as a crossroad from where two new MEPs would start following the gradients of the 2 new electronic states involved. Any of these MEP can be lead to either towards an already explored minimum or a yet unexplored one. Provided that a new minimum is found, a connection with other minima should be established using linear interpolation between both minima geometries, looking for an intrinsic reaction pathway with a transition state, etc.

3.3.1 Deprotonated Tautomer Potential Energy Profile

The scheme for the potential energy profile of BA at pH > pK_a (3.99) is depicted in Figure 3.5. From the Franck-Condon geometry, a minimum of the spectroscopic state $^1(\pi\pi^*)$ is found at 4.34 eV, relatively close in energy from the initial FC energy 4.59 eV. Nevertheless, geometry changes are dramatic, mainly due to an out-of plane deviation of the carbonyl group at C4 position, which can be seen in Figure 3.6. In this particular region of the PES that connects the FC point to the minimum of the S_1 state, a Rydberg orbital suddenly gets into the active space, displacing the bright state to higher potential (S_2) for a few steps of the MEP. Nevertheless, a few points further along the relaxation pathway of the spectroscopic state, this state disappears. Thus, due to the ephemeral occurrence of the Rydberg state along this route, not to be found again during the rest of the calculations, this orbital was omitted from the active space for the sake of simplicity.

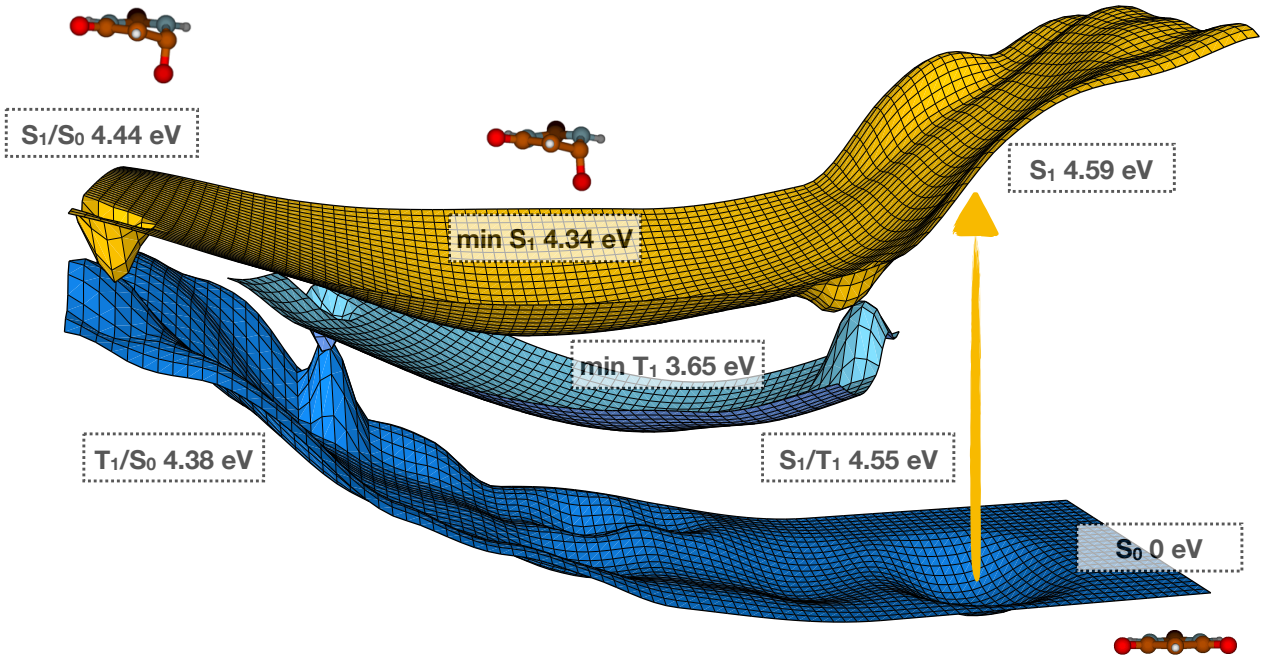


Figure 3.5: Scheme for the potential energy profile along the coordinates relevant to the deactivation of excited BA in basic pH conditions. Energies in eV are set relative to the GS.

Assuming the minimum of the $^1(\pi\pi^*)$ is reached in a direct manner, population would quickly accumulate in it. From this point, an energy-accessible sloped MECP was found with the ground state at 4.44 eV, showing a similar geometry to the $^1(\pi\pi^*)$ minimum, with a more twisted carbonyl group (C_4O_8). Around the MECP region, the population might eventually reach the S_0 PES and come back to the Franck-Condon geometry in the GS, as a direct smooth pathway was found. As a remark, uracil canonical nucleobase has been

extensively reported to have a conical intersection of the same structural nature [59, 60], as the one found for deprotonated BA.

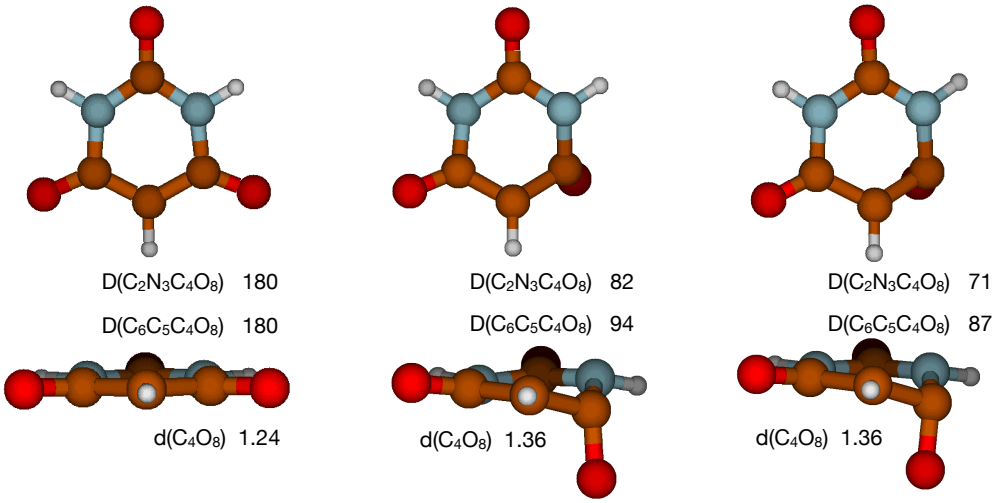


Figure 3.6: Geometries of the Franck-Condon S_0 minimum (left), minimum of the S_1 state (center) and S_1/S_0 conical intersection (right). Distances reported in $\mathring{A}$ and angles in degrees.

The described relaxation pathway, $S_1 \rightarrow S_1$ min $\rightarrow S_1/S_0$, seems to be the most probable path to follow for photo-deactivation, since singlet-singlet crossings are energetically accessible and they are more efficient than other decay funnels transferring population to PES of other multiplicities. Nevertheless, in order to explore other possible decay routes and analyze a putative singlet-triplet pathway, we have estimated the position of the triplets at the minimum of the $^1(\pi\pi^*)$ state. In fact, T_1 $^3(\pi\pi^*)$ was found to be relatively close in energy to the S_1 at this minimum, and a crossing point between both states was localized at 4.55 eV. This crossing directly connects the S_1 PES with the T_1 minimum located at a considerably lower energy, 3.65 eV, which in turn crosses with S_0 at 4.38 eV. This point would work as an Intersystem Crossing (ISC) funnel to potentially return the population to the Franck-Condon geometry. Taking into account the moderate Spin-Orbit Couplings (SOCs) calculated at the S_1/T_1 and T_1/S_0 ISC funnels ($\sim$10 cm^{-1}), the $S_1 \rightarrow S_1$ min $\rightarrow S_1/T_1 \rightarrow T_1$ min $\rightarrow T_1/S_0$ pathway is expected to be a less efficient route and, thus, to play a less important role in the decay of the excited population. Also, geometric changes along the triplet route are more dramatic as illustrated in Figure 3.7.

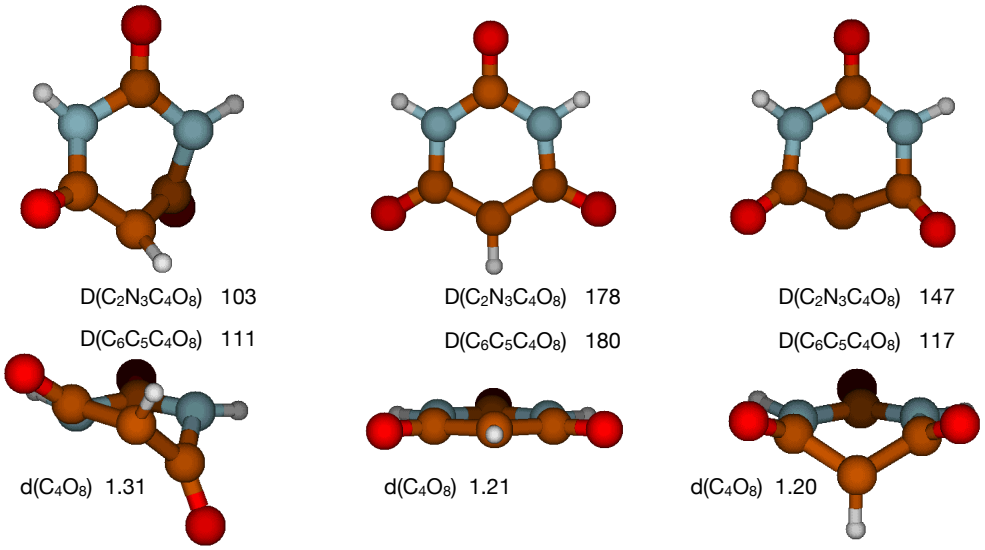


Figure 3.7: Geometries of the S_1/T_1 crossing point (left), T_1 $^3(\pi\pi^*)$ minimum (center) and T_1/S_0 crossing point (right). Distances reported in $\mathring{A}$ and angles in degrees.

According to our calculations, the completely singlet mediated path for the deprotonated tautomer, occurring at basic pH (pH 7.4 - physiological), is compatible with a photostable system in agreement with experimental results [4]. However, the singlet-triplet mixed deactivation path is not as straightforward as the latter. In order to be able to get a more reasonable judgement, transient absorption spectra determination and dynamical calculations should be performed to observe whether T_1 is populated or not.

3.3.2 Neutral Tautomer Potential Energy Profile

The scheme for the potential energy profile along the coordinate relevant to the deactivation of neutral BA is depicted in Figure 3.8. In this case, as reported in Table 3.3, three excited states were considered, the bright one is S_3 $^1(\pi\pi^*)$ and the two dark S_1 and S_2 $^1(n\pi^*)$ states. Upon UV light irradiation, the system would predominantly populate the S_3 state. From this state, the system is expected to relax to the lower lying S_2 $^1(n\pi^*)$ via the IC funnel S_3/S_2 located at around 5.50 eV connected with the FC region through a really flat PES. Once in the S_2 state, further relaxation will proceed through a crossing with the S_1 $^1(n\pi^*)$ state. Geometries of these crossing points are reported in Figure 3.9 top-center and top-right.

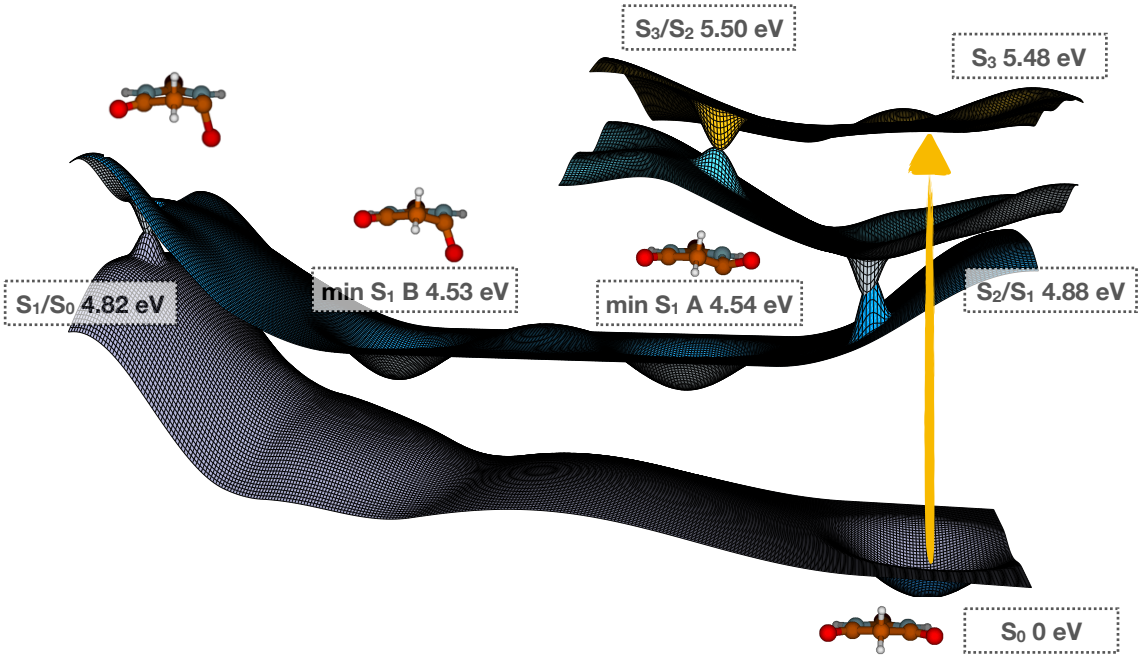


Figure 3.8: Scheme for the potential energy profile describing the deactivation of neutral excited BA, through the singlet states. Energies in eV are set relative to the GS.

This sloped S_2/S_1 crossing point found at 4.88 eV, is directly connected to A S_1 minimum state, which shows an out of plane deformation which affects mainly the carbonyl at C_4O_8, as shown in Figure 3.9 (bottom left pannel). Further decay is possible through the S_1/S_0 funnel shown in Figure 3.9 (bottom righ panel) at around 4.82 eV, which strongly resembles the crossing points found for the deprotonated species. This crossing directly connects the S_1 PES with the starting FC S_0 geometry.

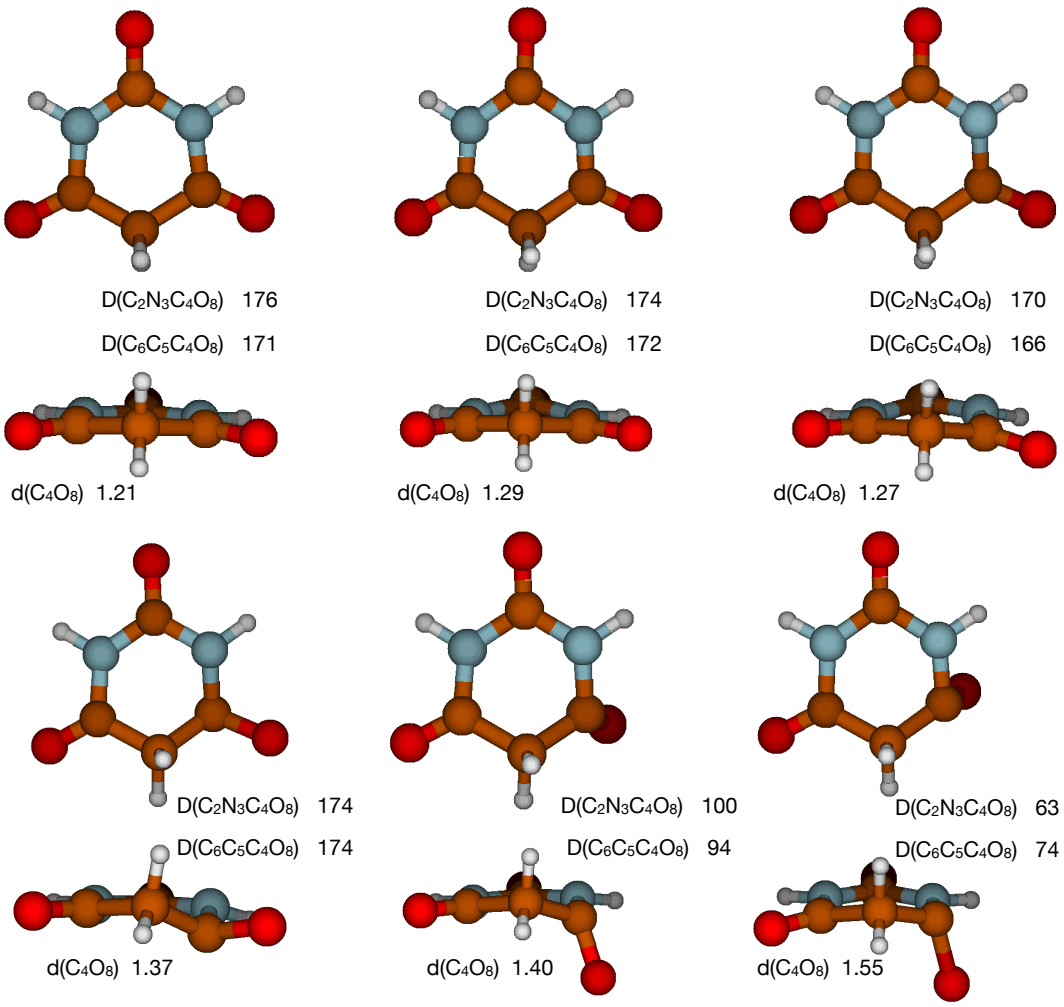


Figure 3.9: Geometries for the S_0 minimum (top-left), S_3/S_2 conical intersection (top-center) and S_2/S_1 conical intersection (top-right), S_1 (A) minimum (bottom-left), S_1 (B) minimum (bottom-center) and S_1/S_0 conical intersection (bottom-right). Distances reported in $\mathring{A}$ and angles in degrees.

Importantly, this S_1/S_0 CI is not directly accessible from the S_1(A) minimum reported above (Figure 3.9 - bottom left panel), but from an additional minimum of the S_1 PES S_1(B) (Figure 3.9 bottom-center panel) and bottom-right, which is really close in geometry to the S_1/S_0 CI (Figure 3.9 bottom-right panel). This new minimum S_1(B) has an energy of 4.53 eV, very similar to that of the minimum S_1(A) (4.54 eV). The connection between both minima was mapped through a relaxed scan and an energy barrier of less than 0.1 eV was found. Thus, both minima seem to be potentially accessible after photoirradiation, the population eventually evolving to the S_1/S_0 conical intersection and returning back to the Franck-Condon geometry in the ground state.

For this tautomer, we also found a possible deactivation route through the triplet manifold in the proximity of the S_1(A) minimum, as can be seen in Figure 3.10. In fact, the S_1/T_2 ($^1(n\pi^*)/^3(\pi\pi^*)$) crossing point shows considerable SOC values (>100 cm^{-1}), indicating

potential population transfer towards the T_2 $^3(\pi\pi^*)$ minimum . Then, the T_2/T_1 CI ($^3(\pi\pi^*)/^3(n\pi^*)$) is located at 5.1 eV, which would allow the population of the T_1 minimum localized at 4.3 eV relative to the GS. Eventually, deactivation to the GS through the T_1/S_0 crossing point at 4.9 eV would occur, as again the ISC is characterized by very large SOC values (> 100 cm^{-1}), from where the FC geometry can be easily reached for the GS. Geometries of the triplet-mediated deactivation pathway are reported in Figure 3.11. An interesting highlight is that similar geometries are shown for the the triplet minima, T_2 and T_1, in comparison to minimum S_1(A), and most of the change occurs in the bond C_4O_8 length.

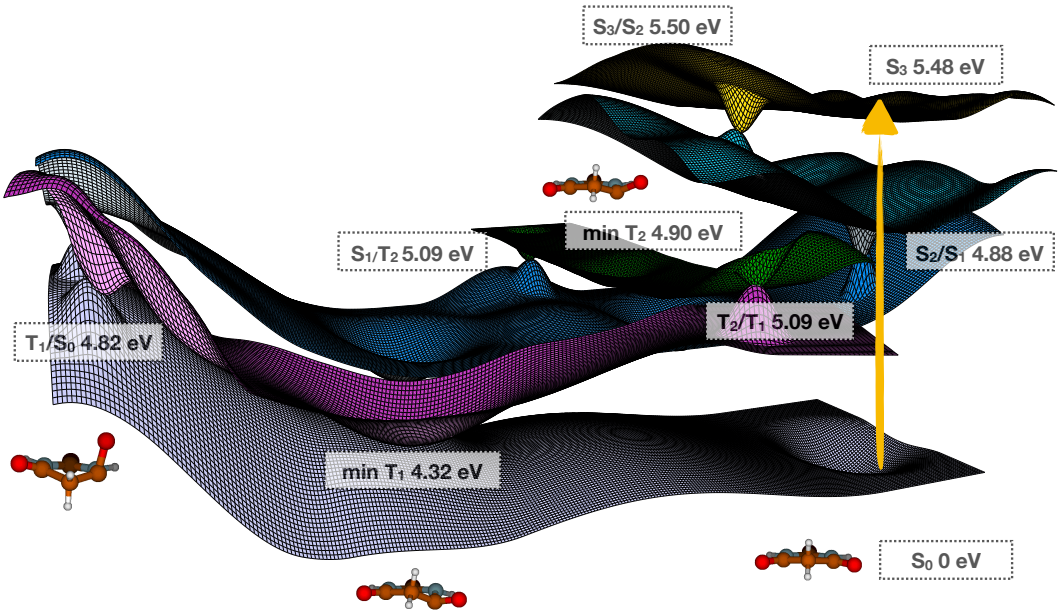


Figure 3.10: Scheme for the potential energy profile along the triplet manifold and the coordinates relevant to the deactivation of excited BA in acidic pH conditions. Energies in eV are set relative to the GS.

Although higher in energy, the triplet pathway is also actually feasible. In fact, both routes (exclusive singlet and mixed singlet-triplet pathways) are likely to compete, although the higher energies barriers that characterize the mixed singlet-triplet pathway might indicate that this pathway is minoritary. The possible competition between the exclusive singlet and the mixed singlet-triplet routes has been considered through the interpretation of experimental Transient Absorption Spectra, which will be tackled in the next section.

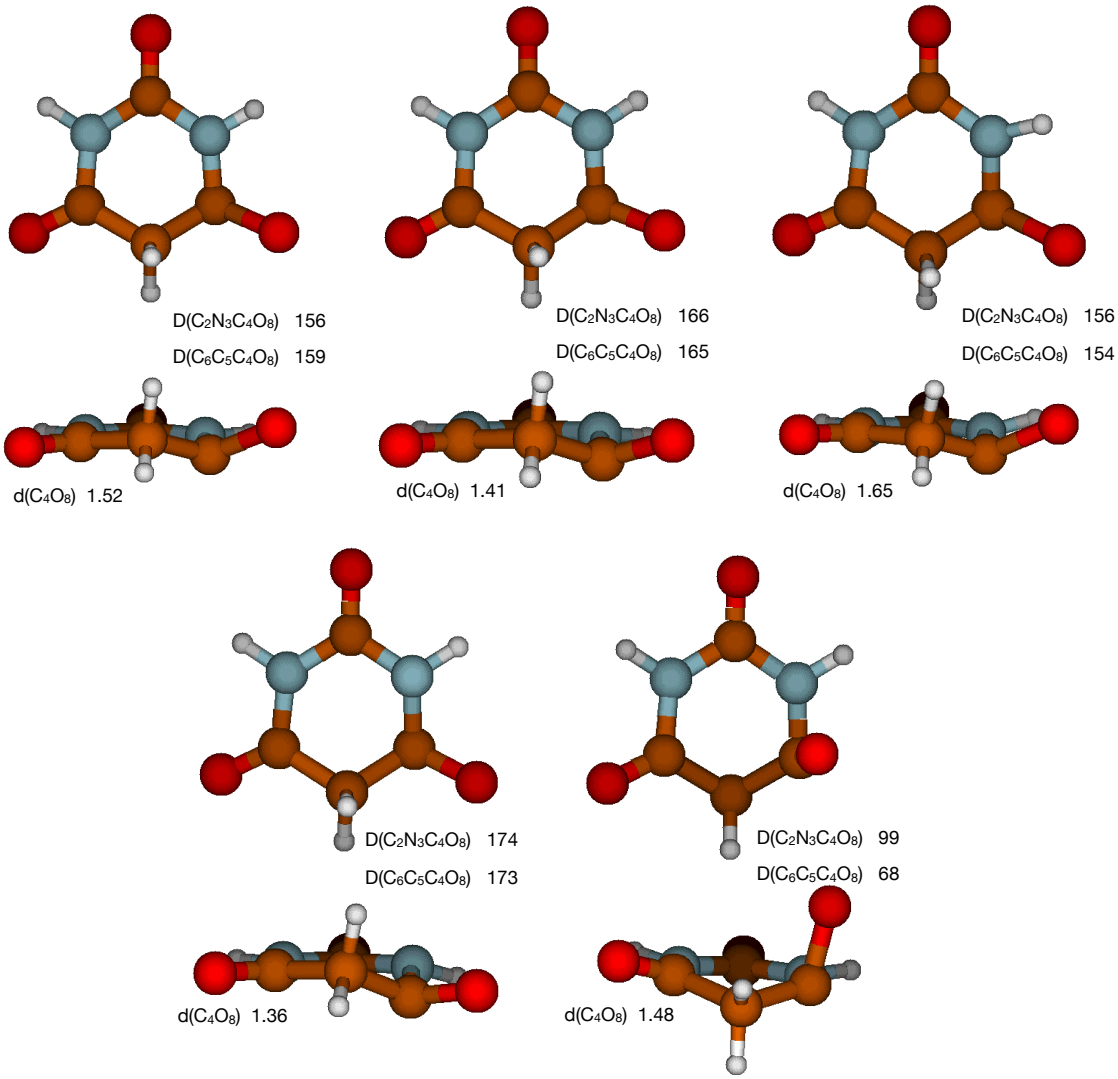


Figure 3.11: Geometries for the S_1/T_2 crossing point (top-left), T_2 minimum (top-center) and T_2/T_1 conical intersection (top-right), T_1 minimum (bottom-left), T_1/S_0 crossing point (bottom-right). Distances reported in $\mathring{A}$ and angles in degrees.

3.4 Transient Absorption Spectra as an Static Approach to Get a Dynamic Picture

Transient absorption spectroscopy measures the change in absorbance of the sample with time following photo-excitation. Its main application is the study of intermediate states in photophysical or photochemical processes as this technique is able to characterize different electronically excited states species provided that their absorption footprint exists and differs one from each other. In this technique, two pulses are used: the first one excites the ground state population to the excited state using a particular wavelength, and the second pulse excites the excited state population to a higher-energy state. This way it is possible to obtain the time-resolved picture of the transient absorption spectra. This principle was used by Crespo and colleagues [4] for studying the photophysics of BA in an experimental manner. Thus, it is the aim of this project to help in the interpretation of these spectra and thus to unravel the dynamics of this system.

Figure 3.12 shows experimental TAS results of BA in PBS at pH 7.4 (Figure 3.12 top panel) corresponding to the deprotonated species and pH 2.5 (Figure 3.12 bottom panel) corresponding to the neutral species, following 268 nm irradiation. For pH 7.4, after irradiation, two main absorption peaks are observed centered at ∼375 nm and ∼525 nm at 0.0 ps. Later on, at ∼0.28 ps time delay, the latter peak decreases its intensity, while that of the former increases. For larger time delays, both signals decrease their intensities, until no further signal is registered (at 1.08 ps time delay), and it could be interpreted as the molecule coming back to the ground state.

At pH 2.5 (Figure 3.12 bottom panel), immediately after irradiation, the TAS spectra show peaks at ∼375 nm and ∼550 nm, analogous to the peaks described observed at pH 7.4. Despite this similarity, for 0.29 ps time delay, it seems that the peak at ∼550 nm decreases in intensity, and blue shifts to ∼475 nm for the maximum of absorption, as well as increasing slightly intensity at ∼375 nm. In this case, deactivation to the ground state occurs in slightly longer time scales, but, still, at 1.63 ps no further absorption takes place.

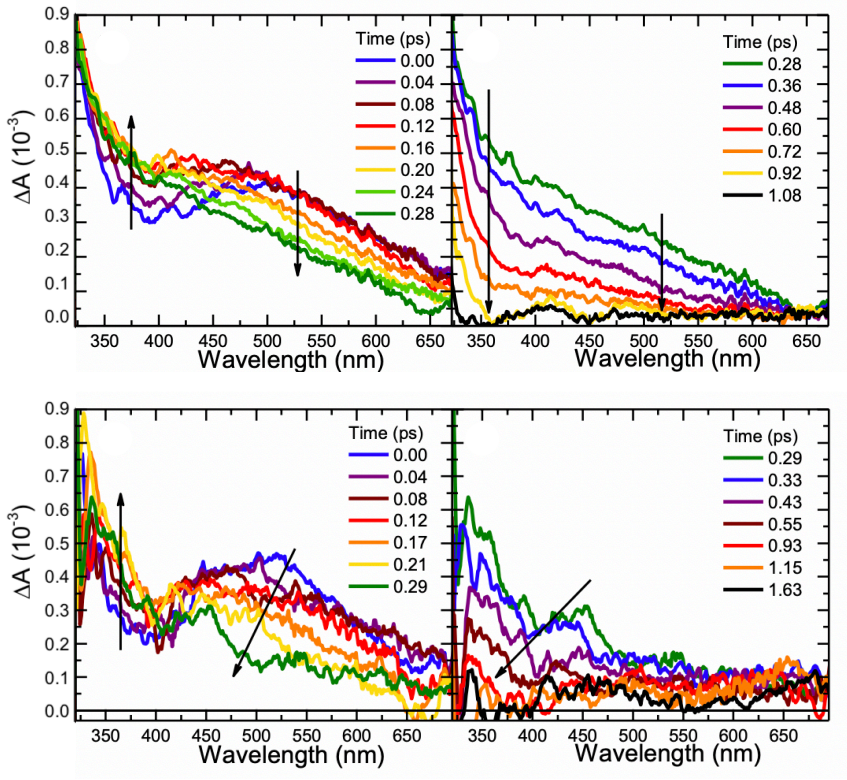


Figure 3.12: Evolution of the transient absorption spectra (350-650 nm) of BA in PBS at pH 7.4 (top) and pH 2.5 (bottom), following 268 nm irradiation. Times correspond to the delay between both pulses. Adapted from [4].

For modelling the steady state absorption spectra of molecules, it is necessary to calculate the energies of the different electronic states. Thus, the absorption energy is calculated as the difference between the excited state and the ground state energies of the molecule. The calculation of the transition dipole moment between the ground and the excited state is also necessary so that we can obtain the oscillator strength which translates into the intensity of absorption. Keeping in mind this scheme, we can reproduce these calculations not using as a reference the ground state of the molecule, but instead considering the excitation of an excited species. This was the principle used for calculating the TAS in gas phase (note that to the experimental TAS were obtained in aqueous PBS buffer) and, in analogy to the GS absorption spectra, the MS-CASPT2/ANO-L-VTZ was used as a protocol.

The spectroscopic footprints at those stationary points, such as minima of the calculated PES in the previous section where the population is expected to accumulate for a longer time, were calculated. To assist in the interpretation of the experimental results, TAS signals were calculated at the position of all the singlet and triplet minima, considering 15 roots.

3.4.1 Deprotonated BA Transient Absorption Spectra

For this species, besides the Franck-Condon ground state minimum, two other excited state minima were calculated: S_1 $^1(\pi\pi^*)$ and T_1 $^3(n\pi^*)$. The absorption spectra for these species are collected in Figure 3.13. S_1 mainly absorbs at ∼400 nm and ∼540 nm, while T_1 absorbs mainly at ∼380 nm showing a weak absorption at ∼600 nm. Unfortunately, the similarity between both absorption spectra prevents distinguishing both species in the experimental spectra. Thus, it is difficult to say whether T_1 is populated or not. Nevertheless, taking into account the PES in Figure 3.5, the low SOC values obtained for the S_1/T_1 crossing point and the calculated TAS shown in Figure 3.13, it can be concluded that the S_1 state is mainly populated, while T_1 might be slightly populated or even not populated at all.

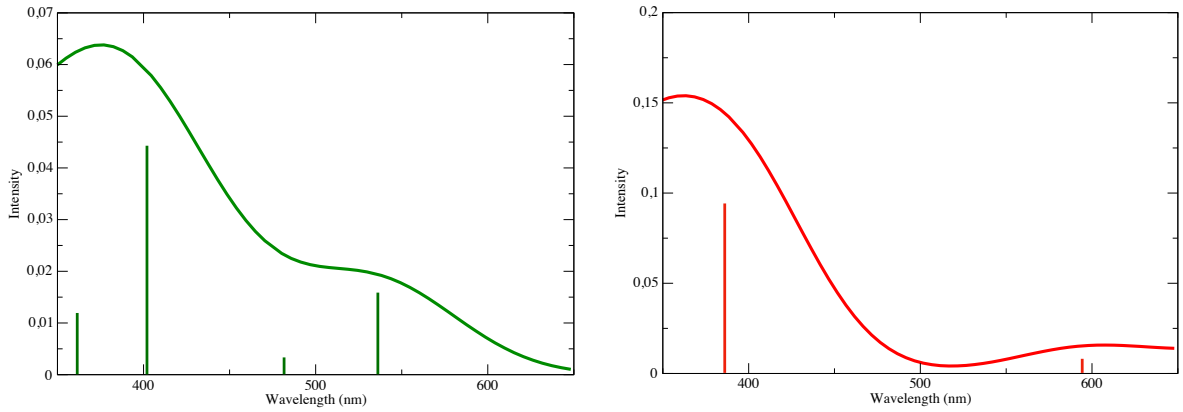


Figure 3.13: Simulated transient absorption spectra for S_1 (left-green) and T_1 (right-red) minima of deprotonated BA. Vertical lines indicate vertical excitation absorption wavelength and their heights correspond to the oscillator strength. Note that the Y-axix scale is different for the two species.

All in all, calculated TAS spectra shows great similarity to the experimental one, despite being slightly red-shifted, most probably due to the solvent effects which were not accounted for in the theoretical calculations. This was also observed in the calculation of the absorption spectra shown in Figure 3.4, where the main peak at around 260 nm in experiment was a peak at around 275 nm in the one obtained using theoretical approaches. In particular, S_1 minimum calcualted TAS spectrum was peaking with a red-shift of around 15-25 nm in comparison to experimental values; and also, the more intense peak at 400 nm is the one showing further shifting to the red.

3.4.2 Neutral BA Transient Absorption Spectra

For this species, besides the Franck-Condon ground state minimum, four other other excited state minima were calculated: S_1(A) $^1(n\pi^*)$, S_1(B) $^1(n\pi^*)$, T_1 $^3(n\pi^*)$ and T_2 $^3(\pi\pi^*)$. The left panel of Figure 3.14 shows calculated TAS for both singlet minima. S_1(A) shows a strong absorption in the region between 420 nm and 450 nm, while S_1(B) absorbs at $\sim$ 510 nm. The presence of both species is somehow in agreement to experimental TAS reported in Figure 3.12 at 0.00 ps. Despite being slightly deviated, it should be taken into account again that gas phase TAS were calculated, while experimental were done in aqueous PBS buffer. Contrarily to the deprotonated absorption spectra commented in the previous section, no significant red-shift is shown between theoretical and experimental values shown in Figure 3.4. Nevertheless some experimental spectral change occurs at around 0.29 ps time delay, and this is hardly reproduced from any of the studied singlet transient absorption spectra.

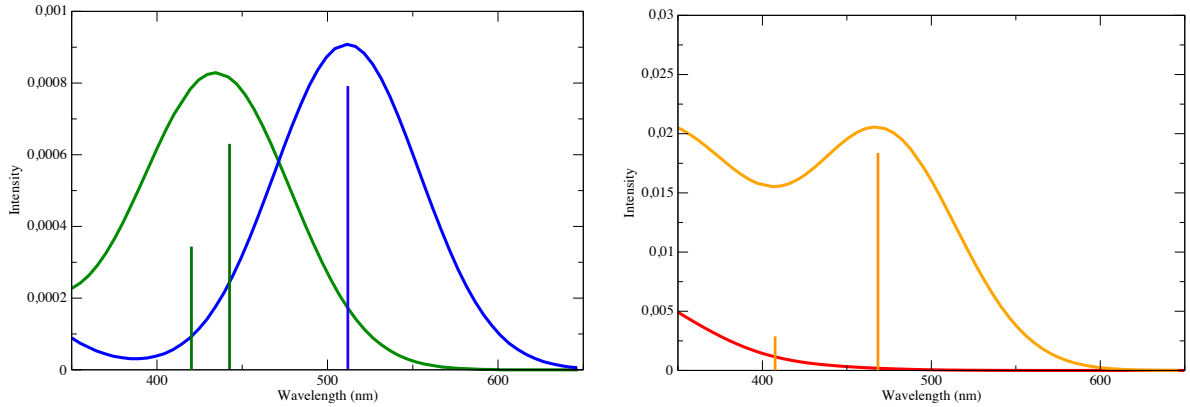


Figure 3.14: Simulated transient absorption spectra for S_1(A) (left-green), S_1(B) (left-blue), T_1 (right-red) and T_2 (right-orange) minima of neutral BA. Vertical lines indicate vertical excitation absorption wavelength and their height corresponds to the oscillator strength. Note that the Y-axix scale is different for the two species.

Also, TAS were calculated for T_1 and T_2 and are shown in Figure 3.14. Unfortunately, T_1 shows very little absorption within 350-650 nm, so it is hardly possible to detect this species from TAS interpretation. Contrarily, T_2 does have a clear footprint absorbing mainly at $\sim$470 nm with weak absorption at $\sim$410 nm. Nevertheless, it may be able to explain the increase in intensity at $\sim$ 350 nm for time delay of 0.29 ps, as T_1 strongly absorbs around that region. This would compatible with the spectral evolution for time delays $>$ 0.29 ps, resulting in transfer of population from the S_1 state to the T_2, and eventually decaying to the ground state. From this data, it may be concluded that deactivation through the triplet manifold might take place, and this idea is supported by the high SOC values obtained for the S_1/T_2 and T_1/S_0 crossing points. Despite the high

SOC values, these crossing points have been reported to be high in energy (see Section 3.3.2, which might be difficult to reach after 268 nm irradiation. Other aspect to take into account is the speed of the deactivation process (< 2 ps), which makes hard to believe that the triplet manifold is involved. In order to gain insight in this point, further studies must be done trying to get the dynamical picture of the process.

Chapter 4

Conclusions

Throughout this work, we explored the photophysics of Barbituric Acid in different pH conditions, acidic and neutral-alkali, aiming at deciphering the possible role of this molecule as a ancestor of current nucleobases. First, the tautomeric species were determined in terms of stability in water environment. Neutral and deprotonated species show one predominant triketo form in solution, which were used in subsequent calculations. Considering these tautomers, the absorption spectrum of BA was computed, showing an excellent agreement with the experiments. In particular, the neutral form shows a maximum of absorption at 210-220 nm, while the deprotonaed form strongly absorbs at 260 nm.

The excited state PES were explored in gas phase for both species. From the S_1, the deprotonated form shows, at physiological pH, a potential decay mechanism which follows the following sequence of steps: $S_1^* \rightarrow S_1$ min $\rightarrow S_1/S_0 \rightarrow S_0$. Also, an alternative pathway along the triplet manifold, involving the following steps: $S_1^* \rightarrow S_1$ min $\rightarrow S_1/T_1 \rightarrow T_1$ min $\rightarrow T_1/S_0 \rightarrow S_0$ was studied. After considering experimental decay times, the SOC values at the triplet-singlet crossing points and the comparison of calculated and experimental TAS, the most plausible pathway for the deactivation to the GS seems to be the one proceeding exclusively along the singlet manifold.

The neutral form of BA, present at acidic pH, was a characterized by a more complex PES landscape. On the one hand, the exclusively singlet-mediated decay pathway, $S_3 \rightarrow S_3/S_2 \rightarrow S_2/S_1 \rightarrow S_1(A)$ min $\rightarrow S_1(B)$ min $\rightarrow S_1/S_0 \rightarrow S_0$, is characterized by two minima on the S_1 PES showing TAS fingerprints consistent with the experimental results at the first stages of the measurements. For longer delay times, the experimental TAS shows a change in the absorption signal shape which was not coincident with any of the calculated singlet TAS. Thus, an additional decay pathway through the triplet manifold was studied:

$S_3 \rightarrow S_3/S_2 \rightarrow S_2/S_1 \rightarrow (S_1(A)\ \text{min} \leftrightarrow S_1(B))\ \text{min} \rightarrow S_1/T_2 \rightarrow T_2\ \text{min} \rightarrow T_2/T_1 \rightarrow T_1$ min $\rightarrow T_1/S_0 \rightarrow S_0$. This singlet-triplet alternative pathway is supported by high SOC values at ISC points, and a good compromise between experimental and simulated TAS.

Although molecular dynamics simulations would be necessary to establish the actual deactivation mechanisms at different pH conditions and to determine the excited state lifetimes, based on our static calculation it is possible to classify Barbituric Acid as a photostable system and, thus, a probable nucleobase ancestor.

Bibliography

[1] A. C. Rios and Y. Tor, "On the origin of the canonical nucleobases: An assessment of selection pressures across chemical and early biological evolution," *Israel Journal of Chemistry*, vol. 53, no. 6-7, pp. 469–483, 2013.

[2] N. B. Leontis and E. Westhof, "Geometric nomenclature and classification of rna base pairs," *RNA*, vol. 7, no. 4, p. 499–512, 2001.

[3] A. S. Lubbe, Q. Liu, S. J. Smith, J. W. de Vries, J. C. M. Kistemaker, A. H. de Vries, I. Faustino, Z. Meng, W. Szymanski, A. Herrmann, and B. L. Feringa, "Photoswitching of DNA Hybridization Using a Molecular Motor, journal = Journal of the American Chemical Society," vol. 140, no. 15, pp. 5069–5076, 2018. PMID: 29551069.

[4] M. M. Brister, M. Pollum, and C. E. Crespo-Hernandez, "Photochemical etiology of promising ancestors of the RNA nucleobases," *Physical Chemistry Chemical Physics*, vol. 18, pp. 20097–20103, Aug 14 2016.

[5] N. Kitadai and S. Maruyama, "Origins of building blocks of life: A review," *Geoscience Frontiers*, vol. 9, no. 4, pp. 1117 – 1153, 2018.

[6] J. M. Dohm and S. Maruyama, "Habitable trinity," *Geoscience Frontiers*, vol. 6, no. 1, pp. 95 – 101, 2015. Special Issue: Plate and plume tectonics: Numerical simulation and seismic tomography.

[7] S. Hussain, J. Aleksic, S. Blanco, S. Dietmann, and M. Frye, "Characterizing 5-methylcytosine in the mammalian epitranscriptome," *Genome Biology*, vol. 14, p. 215, Nov 2013.

[8] H. Grosjean, "Nucleic Acids Are Not Boring Long Polymers of Only Four Types of Nucleotides: A Guided Tour," *See Ref.*, vol. 52, 01 2000.

[9] B. Poole A., Penny and S. B., "Cofounded cytosine! tinkering and the evolution of dna," *Nature Reviews Molecular Cell Biology*, vol. 2, p. 147, 2001.

[10] B. J. Cafferty and N. V. Hud, "Was a Pyrimidine-Pyrimidine Base Pair the Ancestor of Watson-Crick Base Pairs? Insights from a Systematic Approach to the Origin of RNA," *Israel Journal of Chemistry*, vol. 55, no. 8, pp. 891–905, 2015.

[11] E. Borquez, H. J. Cleaves, A. Lazcano, and S. L. Miller, "An Investigation of Prebiotic Purine Synthesis from the Hydrolysis of HCN Polymers," *Origins of Life and Evolution of Biospheres*, vol. 35, pp. 79–90, Apr 2005.

[12] O. L. E., "Prebiotic Chemistry and the Origin of the RNA World," *Critical Reviews in Biochemistry and Molecular Biology*, vol. 39, no. 2, pp. 99–123, 2004.

[13] H. J. Cleaves II, K. E. Nelson, and S. L. Miller, "The prebiotic synthesis of pyrimidines in frozen solution," *Naturwissenschaften*, vol. 93, pp. 228–231, May 2006.

[14] H. L. Barks, R. Buckley, G. A. Grieves, E. Di Mauro, N. V. Hud, and T. M. Orlando, "Guanine, Adenine, and Hypoxanthine Production in UV-Irradiated Formamide Solutions: Relaxation of the Requirements for Prebiotic Purine Nucleobase Formation," *ChemBioChem*, vol. 11, no. 9, pp. 1240–1243, 2010.

[15] R. Saladino, G. Botta, S. Pino, G. Costanzo, and E. Di Mauro, "Genetics first or metabolism first? The formamide clue," *Chem. Soc. Rev.*, vol. 41, pp. 5526–5565, 2012.

[16] C. T. Middleton, K. de La Harpe, C. Su, Y. K. Law, C. E. Crespo-Hernández, and B. Kohler, "DNA Excited-State Dynamics: From Single Bases to the Double Helix," *Annual Review of Physical Chemistry*, vol. 60, no. 1, pp. 217–239, 2009. PMID: 19012538.

[17] F. E., "DNA damage and repair," *Nature*, vol. 421, pp. 436–440, 2003.

[18] R. W. Sinkeldam, N. J. Greco, and Y. Tor, "Fluorescent analogs of biomolecular building blocks: Design, properties, and applications," *Chemical Reviews*, vol. 110, no. 5, pp. 2579–2619, 2010. PMID: 20205430.

[19] P. O. P. Ts'o, I. S. Melvin, and A. C. Olson, "Interaction and association of bases and nucleosides in aqueous solutions," *Journal of the American Chemical Society*, vol. 85, no. 9, pp. 1289–1296, 1963.

[20] K. Wellington and S. Benner, "A Review: Synthesis of Aryl C -Glycosides Via the Heck Coupling Reaction," *Nucleosides, nucleotides nucleic acids*, vol. 25, pp. 1309–33, 02 2006.

[21] R. Siewertsen, H. Neumann, B. Buchheim-Stehn, R. Herges, C. Näther, F. Renth, and F. Temps, "Highly Efficient Reversible ZE Photoisomerization of a Bridged Azobenzene with Visible Light through Resolved S1($n\pi^*$) Absorption Bands," *Jour-*

nal of the American Chemical Society, vol. 131, no. 43, pp. 15594–15595, 2009. PMID: 19827776.

[22] A. A. Beharry, O. Sadovski, and G. A. Woolley, "Azobenzene Photoswitching without Ultraviolet Light," *Journal of the American Chemical Society*, vol. 133, no. 49, pp. 19684–19687, 2011. PMID: 22082305.

[23] R. L. Letsinger and T. Wu, "Control of Excimer Emission and Photochemistry of Stilbene Units by Oligonucleotide Hybridization," *Journal of the American Chemical Society*, vol. 116, no. 2, pp. 811–812, 1994.

[24] R. L. Letsinger and T. Wu, "Use of a Stilbenedicarboxamide Bridge in Stabilizing, Monitoring, and Photochemically Altering Folded Conformations of Oligonucleotides," *Journal of the American Chemical Society*, vol. 117, no. 28, pp. 7323–7328, 1995.

[25] F. D. Lewis, T. Wu, Y. Zhang, R. L. Letsinger, S. R. Greenfield, and M. R. Wasielewski, "Distance-Dependent Electron Transfer in DNA Hairpins," *Science*, vol. 277, no. 5326, pp. 673–676, 1997.

[26] J. C. Genereux and J. K. Barton, "Mechanisms for DNA Charge Transport," *Chemical Reviews*, vol. 110, no. 3, pp. 1642–1662, 2010. PMID: 20214403.

[27] L. Wu, K. Koumoto, and N. Sugimoto, "Reversible stability switching of a hairpin DNA via a photo-responsive linker unit," *Chem. Commun.*, pp. 1915–1917, 2009.

[28] H. K. Cammenga, V. N. Emel'yanenko, and S. P. Verevkin, "Re-investigation and Data Assessment of the Isomerization and 2,2′-Cyclization of Stilbenes and Azobenzenes," *Industrial & Engineering Chemistry Research*, vol. 48, no. 22, pp. 10120–10128, 2009.

[29] T. H. Dunning, "Gaussian basis sets for use in correlated molecular calculations. I. The atoms boron through neon and hydrogen," *The Journal of Chemical Physics*, vol. 90, no. 2, pp. 1007–1023, 1989.

[30] K. Pierloot, B. Dumez, P.-O. Widmark, and B. O. Roos, "Density matrix averaged atomic natural orbital (ANO) basis sets for correlated molecular wave functions," *Theoretica chimica acta*, vol. 90, pp. 87–114, Jan 1995.

[31] B. O. Roos, K. Andersson, M. P. Fülscher, L. Serrano-Andrés, K. Pierloot, M. Merchán, and V. Molina, "Applications of level shift corrected perturbation theory in electronic spectroscopy," *Journal of Molecular Structure: THEOCHEM*, vol. 388, pp. 257 – 276, 1996.

[32] N. Forsberg and P. Åke Malmqvist, "Multiconfiguration perturbation theory with imaginary level shift," *Chemical Physics Letters*, vol. 274, no. 1, pp. 196 – 204, 1997.

[33] J. Finley, P. Åke Malmqvist, B. O. Roos, and L. Serrano-Andrés, "The multi-state CASPT2 method," *Chemical Physics Letters*, vol. 288, no. 2, pp. 299 – 306, 1998.

[34] P. Hohenberg and W. Kohn, "Inhomogeneous Electron Gas," *Phys. Rev.*, vol. 136, pp. B864–B871, Nov 1964.

[35] W. Kohn and L. J. Sham, "Self-Consistent Equations Including Exchange and Correlation Effects," *Phys. Rev.*, vol. 140, pp. A1133–A1138, Nov 1965.

[36] C. Lee, W. Yang, and R. G. Parr, "Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density," *Phys. Rev. B*, vol. 37, pp. 785–789, Jan 1988.

[37] J. P. Perdew, J. A. Chevary, S. H. Vosko, K. A. Jackson, M. R. Pederson, D. J. Singh, and C. Fiolhais, "Atoms, molecules, solids, and surfaces: Applications of the generalized gradient approximation for exchange and correlation," *Phys. Rev. B*, vol. 46, pp. 6671–6687, Sep 1992.

[38] A. D. Becke, "Density-functional exchange-energy approximation with correct asymptotic behavior," *Phys. Rev. A*, vol. 38, pp. 3098–3100, Sep 1988.

[39] J. P. Perdew, K. Burke, and M. Ernzerhof, "Generalized Gradient Approximation Made Simple," *Phys. Rev. Lett.*, vol. 77, pp. 3865–3868, Oct 1996.

[40] A. D. Becke, "Density-functional thermochemistry. III. The role of exact exchange," *The Journal of Chemical Physics*, vol. 98, no. 7, pp. 5648–5652, 1993.

[41] P. J. Stephens, F. J. Devlin, C. F. Chabalowski, and M. J. Frisch, "Ab Initio Calculation of Vibrational Absorption and Circular Dichroism Spectra Using Density Functional Force Fields," *The Journal of Physical Chemistry*, vol. 98, no. 45, pp. 11623–11627, 1994.

[42] J.-D. Chai and M. Head-Gordon, "Long-range corrected hybrid density functionals with damped atom–atom dispersion corrections," *Phys. Chem. Chem. Phys.*, vol. 10, pp. 6615–6620, 2008.

[43] E. Runge and E. K. U. Gross, "Density-Functional Theory for Time-Dependent Systems," *Phys. Rev. Lett.*, vol. 52, pp. 997–1000, Mar 1984.

[44] H. Lischka, D. Nachtigallová, A. J. A. Aquino, P. G. Szalay, F. Plasser, F. B. C. Machado, and M. Barbatti, "Multireference Approaches for Excited States of Molecules," *Chemical Reviews*, vol. 118, no. 15, pp. 7293–7361, 2018. PMID: 30040389.

[45] E. Anslyn, D. Dougherty, E. Dougherty, and U. S. Books, *Modern Physical Organic Chemistry*. University Science Books, 2006.

[46] G. J. Atchity, S. S. Xantheas, and K. Ruedenberg, "Potential energy surfaces near intersections," *The Journal of Chemical Physics*, vol. 95, no. 3, pp. 1862–1876, 1991.

[47] M. A. El-Sayed, "Spin—Orbit Coupling and the Radiationless Processes in Nitrogen Heterocyclics," *The Journal of Chemical Physics*, vol. 38, no. 12, pp. 2834–2838, 1963.

[48] G. Mohammadi Ziarani, F. Aleali, and N. Lashgari, "Recent applications of barbituric acid in multicomponent reactions," *RSC Adv.*, vol. 6, pp. 50895–50922, 2016.

[49] T. P. Briggs AG., Sawbrige JE. and W. JM., "Thermodynamics of dissociation of some barbituric acids in aqueous solution," *Journal of the Chemical Society B-Physical Organic*, no. 7, pp. 802–&, 1969.

[50] D. Villemin, I. Charif, S. M. Mekelleche, and N. Bar, "Natural Bond Orbital Analysis of Cyclic and Acyclic "C-H" acids.," *Eur.Chem Bull., 5 (6) 274-279 (2016)*, 01 2016.

[51] W. Bolton, "The crystal structure of anhydrous barbituric acid," *Acta Crystallographica*, vol. 16, pp. 166–173, Mar 1963.

[52] I. G. Shterev and V. B. Delchev, "Solvent influence on the excited states of the oxo form of barbituric acid and the mechanisms of the out-of-plane non-radiative elongation of the nh bond: A comparative theoretical and experimental study," *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, vol. 125, pp. 384 – 390, 2014.

[53] F. Zuccarello, G. Buemi, C. Gandolfo, and A. Contino, "Barbituric and thiobarbituric acids: a conformational and spectroscopic study," *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, vol. 59, no. 1, pp. 139 – 151, 2003.

[54] F. Aquilante, L. De Vico, N. Ferré, G. Ghigo, P.-a. Malmqvist, P. Neogrády, T. B. Pedersen, M. Pitoňák, M. Reiher, B. O. Roos, L. Serrano-Andrés, M. Urban, V. Veryazov, and R. Lindh, "MOLCAS 7: The Next Generation," *Journal of Computational Chemistry*, vol. 31, no. 1, pp. 224–247, 2010.

[55] T. J. Martinez, "Seaming is believing," *Nature*, vol. 467, pp. 412 EP –, 09 2010.

[56] H.-J. Werner, P. J. Knowles, G. Knizia, F. R. Manby, and M. Schütz, "Molpro: a general-purpose quantum chemistry program package," *Wiley Interdisciplinary Reviews: Computational Molecular Science*, vol. 2, no. 2, pp. 242–253, 2012.

[57] H. Werner and P. J. Knowles, "A second order multiconfiguration SCF procedure with optimum convergence," *The Journal of Chemical Physics*, vol. 82, no. 11, pp. 5053–5063, 1985.

[58] P. J. Knowles and H.-J. Werner, "An efficient second-order MC SCF method for long configuration expansions," *Chemical Physics Letters*, vol. 115, no. 3, pp. 259 – 267, 1985.

[59] Z. Lan, E. Fabiano, and W. Thiel, "Photoinduced Nonadiabatic Dynamics of Pyrimidine Nucleobases: On-the-Fly Surface-Hopping Study with Semiempirical Methods," *The Journal of Physical Chemistry B*, vol. 113, no. 11, pp. 3548–3555, 2009. PMID: 19239209.

[60] R. Improta, F. Santoro, and L. Blancafort, "Quantum Mechanical Studies on the Photophysics and the Photochemistry of Nucleic Acids and Nucleobases," *Chemical Reviews*, vol. 116, no. 6, pp. 3540–3593, 2016. PMID: 26928320.

Appendix A

Additional data

A.1 Tautomers of BA: neutral and deprotonated

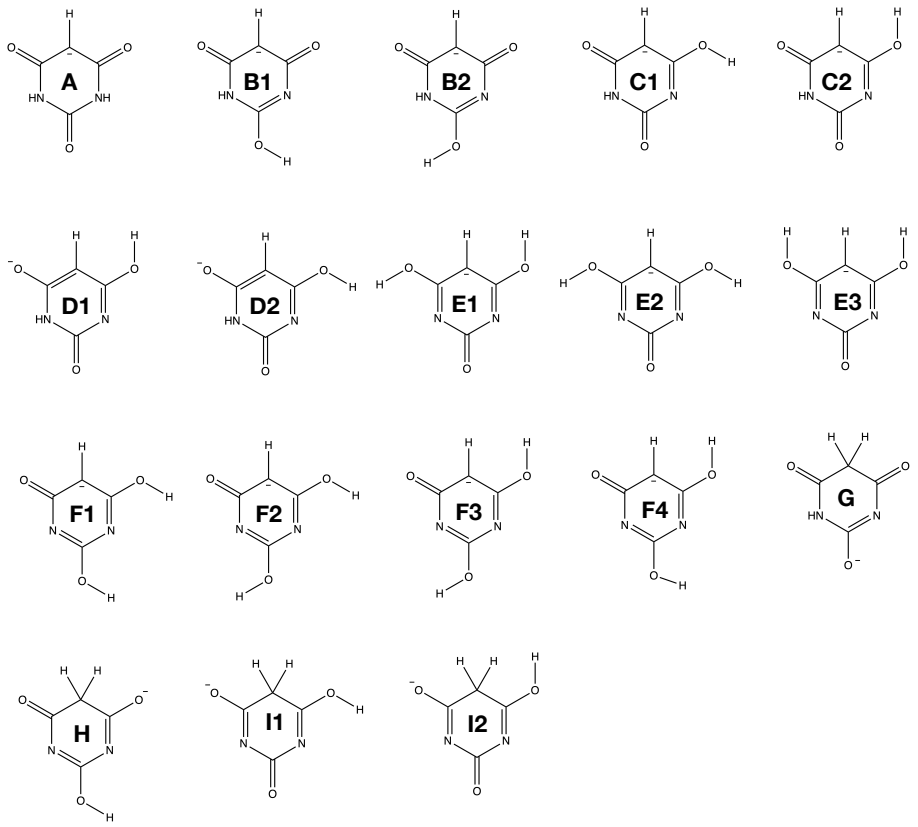


Figure A.1: Tautomers for deprotonated BA.

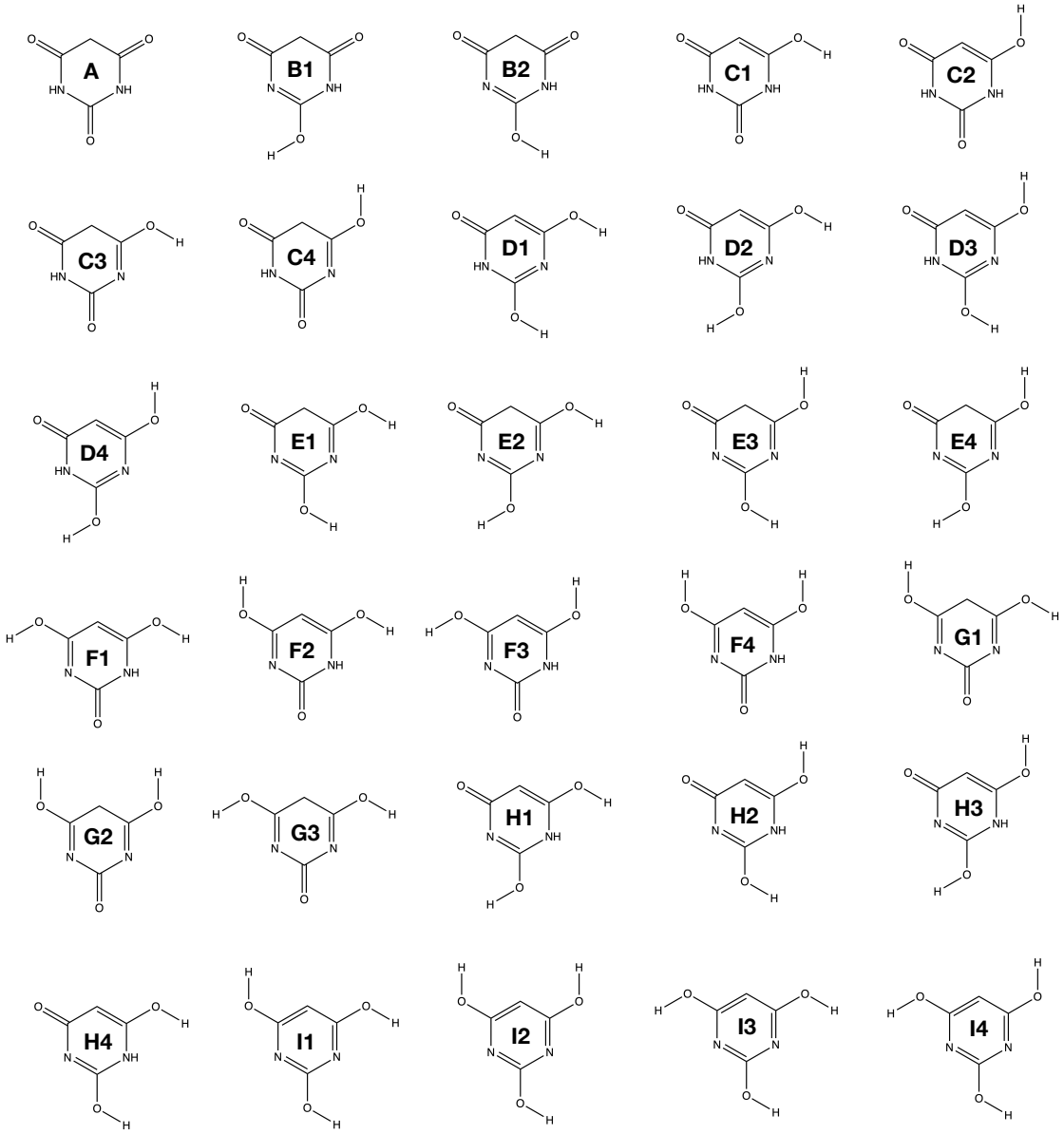


Figure A.2: Tautomers for neutral BA.

A.2 Gibbs Free Energy for Deprotonated Tautomers.

Table A.1: Gibbs Free Energy for each of the deprotonated BA at room temperature (T=298K) in a water continuum under PCM scheme at MP2/aug-cc-pVTZ level of theory. Labelling of tautomers corresponds to Figure A.1

Deprotonated Species	Gibbs Free Energy (Hartree)	ΔG° (kcal/mol)
A	-488.789735	0.00
E3	-488.781271	5.31
H	-488.777227	7.85
D2	-488.775213	9.11
I2	-488.761927	17.45
B1	-488.761705	17.59
D1	-488.759712	18.84
B2	-488.759053	19.25
F1	-488.758735	19.45
I1	-488.757437	20.27
E1	-488.757270	20.37
F3	-488.756808	20.66
C2	-488.753973	22.44
E2	-488.753451	22.77
G	-488.753246	22.90
F2	-488.752869	23.13
F4	-488.749874	25.01
C1	-488.746910	26.87

A.3 Gibbs Free Energy for Protonated Tautomers.

Table A.2: Gibbs Free Energy for each of the neutral BA at room temperature (T=298K) in a water continuum under PCM scheme at MP2/aug-cc-pVTZ level of theory. Labelling of tautomers corresponds to Figure A.2.

Protonated Species	Gibbs Free Energy (Hartree)	ΔG° (kcal/mol)
A	-489.234894	0.00
C2	-489.221559	8.37
C1	-489.217355	11.01
C3	-489.213085	13.69
C4	-489.209237	16.10
D1	-489.208269	16.71
B1	-489.206794	17.63
D3	-489.206750	17.66
I3	-489.206441	17.85
F3	-489.206253	17.97
I2	-489.204312	19.19
I4	-489.204208	19.26
F1	-489.203798	19.51
D2	-489.203669	19.59
F4	-489.202977	20.03
D4	-489.202019	20.63
I1	-489.201957	20.67
F2	-489.200433	21.62
B2	-489.199860	21.98
H3	-489.197131	23.70
H1	-489.191825	27.03
G3	-489.190752	27.70
H4	-489.190531	27.84
E2	-489.188184	29.31
G2	-489.185297	31.12
E4	-489.184680	31.51
E3	-489.183571	32.21
H2	-489.183351	32.34
G1	-489.180723	33.99

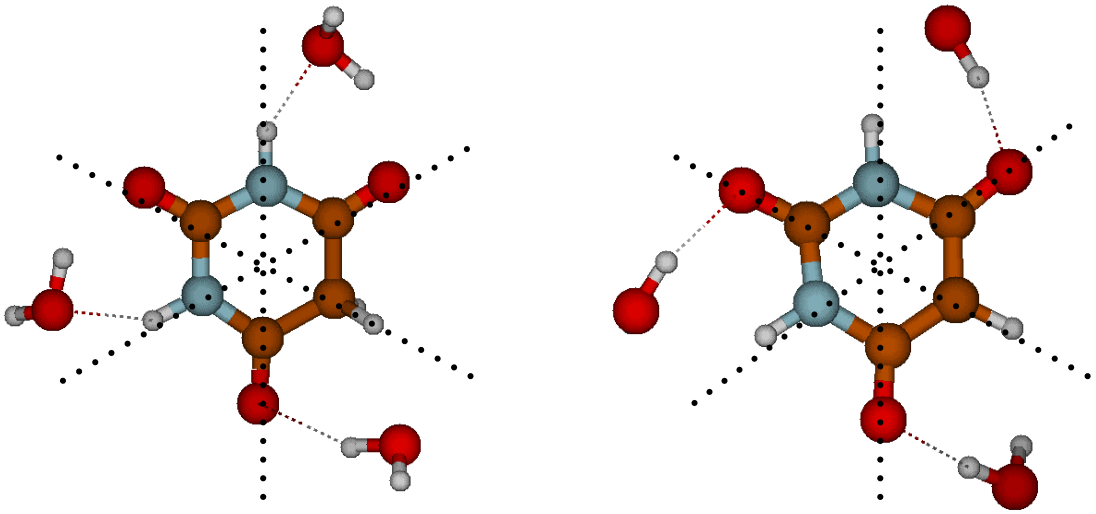


Figure A.3: Ground state minimum energy geometries of BA surrounded by three water molecules for the neutral (left) and deprotonated (right) predominant forms. Dotted lines separate the positions where the three water molecules were placed for the solvation of the species.