Alma Mater Studiorum – Università di Bologna

DOTTORATO DI RICERCA IN

Scienze Chimiche

Ciclo XXVII

Settore Concorsuale di afferenza: 03/A2

Settore Scientifico disciplinare: CHIM/02

MODELLING PHOTOINDUCED EVENTS IN SOLVATED BIO-CROMOPHORES BY HYBRID QM/MM APPROACHES

Presentata da: **Salvatore Flavio Altavilla**

Coordinatore Dottorato **Relatore**

Prof. Aldo Roda **Prof. Marco Garavelli**

Esame finale anno 2016

Contents

1   Introduction

Biological systems such as nucleic acid and proteins have been widely studied
in the past 50 years and will most likely continue to be a topic of interest in
the forefronts of natural science for many years to come. For such purpose,
computational modelling can provide useful information and help us un-
derstand biochemical mechanisms at molecular, atomic and even electronic
levels. The interaction between light and matter is one of the most inter-
esting phenomena in nature; such interaction can be considered responsible
of life on earth. Besides, light absorption, several photophysical or photo-
chemical deactivation mechanisms can compete depending on their relative
efficiencies [1]. The aim of this introduction is to provide an overview about
the photochemical concepts, theoretical methods and molecular systems that
have been studied.

1.1   The photochemical process

The purpose of this thesis is to investigate, through high-level computations,
some of the interesting photo-biochemical processes involved in biological
systems -such as retinal or DNA- in solvent. The investigated processes dis-
play an ultrafast decay from their excited states back to the ground state.
Novel computational techniques have been exploited in order to study the
mentioned systems. Such techniques allows determining the Franck-Condon
point following the steepest descent paths departing from it or from the real
crossing point. The techniques are based on the study of the system de-
cay from an excited-state to the ground-state through the so-called Conical
Intersection (CI) [2], which is a locus where ground-state and excited-state
potential layers cross. The idea that a conical intersection could play a key
role in the decay mechanisms for excited state reactions was first proposed
by Teller [3] more than 30 years ago. Before the decay mechanism through
the CI was recognized, radiationless decay was deemed to occur in the ex-
cited state at a minimum, where the ground state and excited state potential
surfaces are close to each other but do not cross. However, suchprocess could

not account for the fast femtosecond decay experimentally ob- served. The CI provides a fast decay path from an excited state to the ground one and accounts for the femtosecond decay that characterizes such processes; in fact, ultrafast radiationless decay is the main indicator of a mechanism involving CI [4]. Existence of the CI has been experimentally verified and confirmed by means of sophisticated instrumentation; moreover such decays have been thoroughly characterized in terms of decay times, excited state lifetimes and reaction triggering energies. Besides, quantum yield measurements provide information concerning the possible ground state paths accessible from the CI. Such experimental data can be rationalized according to suitable computational models in order to provide a deeper insight in the mechanisms underlying the photochemistry of diverse systems and to provide a better understanding of the photochemical properties of such systems.

1.2   Hybrid QM/MM methods

Hybrid quantum mechanics/molecular mechanics (QM/MM) simulations are a viable method to investigate chemical reactions taking place in condensed phase [5]. According to such methods, the region where the investigated process occurs is studied at a suitable quantum mechanical (QM) level while the environment surrounding that region is modelled as a molecular mechanics (MM) force field. The boundary between the QM and MM subsystems has to be treated carefully in order for the simulation to provide sensible results. This approach is suitable for the study of broad systems such as proteins, DNA and, more in general, solvated molecules. QM/MM techniques will be discussed in greater detail later on in this thesis, moreover QM/MM practical applications for simulating photobiological systems will be also presented. In fact, we have performed QM/MM calculations for the chemical systems described below in the solvent. Moreover, in the research activity carried out in the Ph.D. activity underlying this thesis, algorithms and software scripts have been developed and used to carry out QM/MM computations.

1.3   The COBRAMM suite

COBRAMM (Computational Bridge between Ab-initio and Molecular mechanic) is a hybrid QM/MM calculations software acting as an interface between different tools which execute the various simulation steps. One of the aims of the experimental activity carried out was to locate the conical

intersection (CI) of complex systems. To do that, the gradient projection method has been used, which includes the effect of the MM atoms on the structure and energy of the CI both implicitly via the one electron Hamiltonian (electrostatic embedding) and explicitly through the force on the movable MM atoms. It has been implemented also a routine aimed at executing the Intrinsic Reaction Coordinate (IRC) on QM/MM systems within the COBRAMM package. Such scripts have been verified and validated on the GMP system in solvent. A program for simulating transient spectra has been implemented. The script describes the time-evolution of the oscillator strengths for different state transitions as a superposition of Gaussian functions. A first script fits the time-evolution of the oscillator strengths to a time-sequence of Gaussian curves, for each possible state transition. Then, another one provides a graphical representation of the produced spectra. The program allows using several states for the simulation of the spectra.

1.4   Retinal

Many biological systems can and do interact with light. The study of photoinduced reactions is valuable for the understanding of the biochemical mechanisms that characterize such systems. Usually the interaction with light occurs by means of proteins, whose reactivity is mainly due to the non-proteic moiety of the cromophores that are responsible for light absorption and for the execution of the photochemical processes which characterize the protein function. Processes such as photoisomerization of the retinal cromophores, which induce the conformational changes that are at the basis of the activity of rhodopsin proteins, involve conjugated π-systems. The process of vision, which involves the transduction of the light impinging the eye into neural signals that carry visual information to the brain where they are processed is one of such processes, and in particular one that everybody is familiar with. Light is first collected by the eye front lens (the crystalline lens), which focusses the photon beam on the retina. The retina consists of a large number of photoreceptors, which are cells that are capable of transducing relative light intensities into neurochemical signals. In particular, the photoreceptor outer segments absorb light thanks to chromophores that are covalently bound to an *opsin*, which is a single peptide transmembrane protein. Rhodopsin is a photoreceptor protein which is extremely sensitive to light, and it is the pigment that is involved in the scotopic vision. It is located in the rod outer segment and is characterized by a peak absorption wavelength at about 500nm. Since it is easily prepared, rhodopsin has been widely studied. The chromophore of rhodopsin is the 11-cis-retinal, which

forms rhodopsin when bonded to opsin, which alone would not absorb light in the visible spectrum. When the cis-retinal absorbs a photon in the visible spectrum, it can isomerize to all-trans-retinal. As a consequence the shape of the molecule changes.
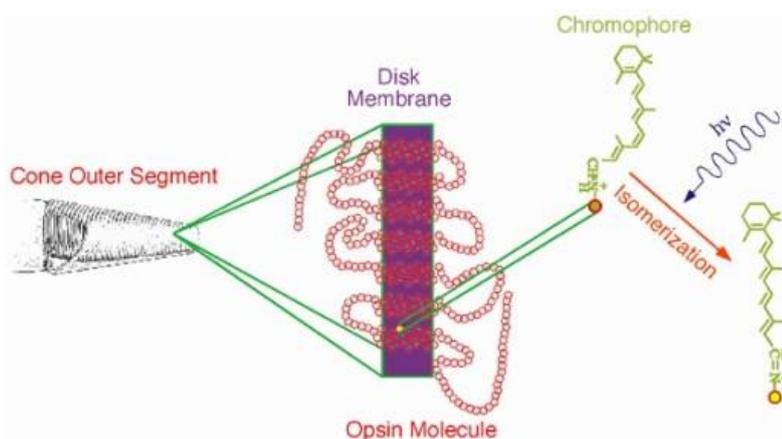
Figure 1.1: Scheme summarizing the phototransduction process. The central part of the figure depicts how a photopigment is embedded within a disk membrane of a photoreceptor outer segment. The filled yellow circle is a lysine which it is attached to the retinal cromophore. The absorption of a photon may cause the chromophore to change from the 11-cis form to the all trans form as shown by the chemical reaction. This conformational change in the chromophore causes a change in the shape of the opsin molecule. The opsin change in shape changes it from an inactive to an activated enzyme, and thus light absorption is transduced into a biochemical signal within a photoreceptor.

Bacteriorhodopsin (bR) is an integral membrane protein usually found in two-dimensional crystalline patches known as *purple membrane*, which can occupy up to nearly 50% of the surface area of the archaeal cell. It is the retinal molecule inside bR that changes its conformation when absorbing a photon, resulting in a conformational change of the surrounding protein and the proton pumping action.

Understanding the fundamental intrinsic photophysical and photochemical properties of retinal chromophores and environmental effects of the surroundings (proteins, solvent, organic and inorganic supports,etc.) is crucial for the design of tailored photosensitive devices. The discrepancy between solution and protein behavior has been largely attributed to the complexity of the protein pocket with its three-dimensional arrangement of amino acids resulting in a unique steric and dielectric environment [6, 7]. The tunability of the absorption spectra, isomerization yield, and reaction speed make RPSB an bR the ideal candidate for investigations aimed at disclos-

Figure 1.2: The absorption of a photon may cause the chromophore to change from the all-trans-retinal to the 13-cis-retinal. The retinal is covalently linked to Lys216 in the chromophore by Schiff base action. After photoisomerization of the retinal molecule, Asp85 becomes a proton acceptor of the donor proton from the retinal molecule. This releases a proton from a "holding site" into the extracellular side (EC) of the membrane. Reprotonation of the retinal molecule by Asp96 restores its original isomerized form.

ing the molecular and structural origins of efficient photochemistry [6, 7, 8, 9, 10, 11, 12, 13, 14, 15]. As a consequence, RPSB and bR has become a paradigm for understanding the origins of activation and suppression of ultrafast relaxation processes, which is essential for the rational engineering of photoreactivity.

1.5 DNA

Deoxyribonucleic acid (DNA), the hereditary basis of the genetic identity of life, has always been a major topic of discussions since its structure was discovered in 1953. The genomic material and the mechanisms that process the disposal of the excess energy attained upon UV-light irradiation are of paramount importance from both biomedical and biotechnological standpoints. The initially populated excited states and their fate along the distinct deactivation routes present in the DNA/RNA double helix chains relate simultaneously to the intrinsic photostability of the genomic material [16, 17], as well as to the damaging photo-reactions that ultimately yield mutations and single and double-strand breaks that have been associated to increasingly occurring diseases like skin cancer [18, 19, 20]. Besides the biological relevance, the intricate photophysical properties of nucleobases can be also exploited to design optical photoresponsive nanodevices (among a long list of prospective applications [21]), yet it is their biological relevance that drags our attention. From a biophysical and chemical point of view, knowledge of the fundamental properties of the deactivation processes, including the associated underlying molecular motions, is essential to rationalize the intrinsic photostability of the genomic material and to characterize

the aforementioned routes leading to detrimental photochemical reactions [18]. A proper characterization of monomeric chromophores is therefore essential to understand the photoinduced events occurring in DNA/RNA from a bottom-up approach that can yield definitive answers regarding their role in photostability and photodamage [22]. In this thesis we focus on water-
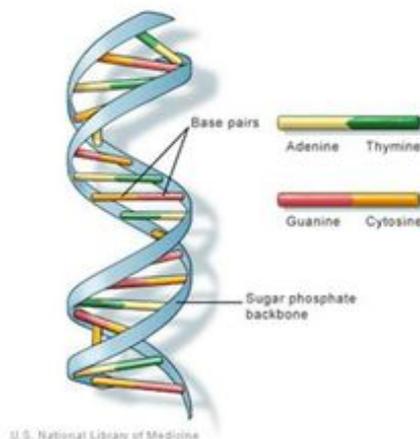
Figure 1.3: Schematic representation of the DNA structure

solvated guanine monophosphate (GMP) an on Cytdine (Cyt), two of the canonical DNA/RNA nucleobases. GMP has also been recently proposed as an interesting compound for nanotechnological applications due to its outstanding self-assembly capabilities [23], especially in its quadruplex form, which has even been located in DNA/RNA chains [24]. Whereas guanine in vacuum has been extensively studied at a high-level multireference ab initio level [25, 26], in the present work a theoretical assessment of the deactivation routes embodying the main photophysical and photochemical features of GMP by employing high-level ab initio multireference perturbation theory methods coupled with a quantum-mechanical/molecular mechanics (QM/MM) approach is proposed, in order to ascertain the role of the environmental perturbations in these type of systems as they remain relatively unknown [26, 27]. For what concern cytidine, formed by a cytosine attached to a ribose ring and component of RNA, is here presented only an initial study. Further knowledge on the environmental effects affecting the photoprocesses occur- ring in the DNA/RNA chromophores upon UV-light irradiation will provide essential information that can be properly referred to water-solvated DNA/RNA systems such as those found in cells.

2 QM, MM and hybrid QM/MM methods: essential concepts and techniques

2.1 Introduction

This chapter briefly overviews the basic concepts that are the basis of the quantum mechanics. Unlike classical mechanics where quantities can vary continuously, in systems obeying quantum mechanics quantities can have only some well-defined discrete values: they are *quantized*. The development of quantum mechanical theories was promoted by the apparent inconsistency of the experimental results obtained when studying microscopic systems and those predicted using the classical mechanics models that had been developed to account for macroscopic systems behaviour. Quantum mechanics is based on the postulate of the existence of a wave function Ψ for any system and of operators that, applied to Ψ, yield the observable properties of such system [28]. Let ϑ be an operator and e a scalar value of a system property, then the concept can be expressed as:

$$\vartheta\Psi = e\Psi \qquad (2.1)$$

where in general Ψ is a $(1xn)$ column vector, ϑ is a (nxn) matrix, and the following constraint must be met: $|\Psi^*\Psi|$ must have unit probability density. According to matrix algebra, ϑ is an Eigenfunction and e is an Eigenvalue of the equation. Very roughly speaking, according to this approach Ψ is much like a system status database from which individual property values can be obtained by applying the appropriate operator. For example if the Hamiltonian operator H is used for ϑ, then the system energy E is returned, and the resulting equation is the Schroedinger equation 2.2.

$$H\Psi = E\Psi \qquad (2.2)$$

The Hamiltonian operator takes into account a number of contributions to the total energy depending on the structure of the system itself (e.g. an

atom, a molecule, etc...) and the interactions with the surroundings it is subjected to. In the case of molecules, the Hamiltonian operator accounts for no less than five different contributions that include the kinetic components, the attraction between the nuclei and electrons, and energies related to interactions among atoms. Other terms may add in the case the observed system is subjected to external electric and/or magnetic fields . In general for a given molecular system the wave function is not easily expressed due to the interaction between the movements of the particles which influence each other. The analysis of such systems can be greatly simplified by taking into account the fact that the nuclei move much slower than the electrons. So it is practical to consider the nuclei at a fixed position when computing electronic energies. This approximation is known as the *Born-Oppenheimer approximation*, and it simplifies expressing the Hamiltonian in that the internuclear potential energy term becomes a constant for the geometry and the correlation terms in the electrons-to-nucleus potential energies are eliminated: the Schroedinger equation becomes an electronic Schroedinger equation, whose Eigenvalue is the electronic energy. Such approximation holds true in most applications, and the relevant equation is equation n 2.3.

$$\left(H_{electronic} + V_N\right)\Psi_{electronic} = E_{electronic}\Psi_{electronic} \tag{2.3}$$

In such equation the term V_N is the inter-nuclear potential energy term, which is a geometry-dependent constant. Due to the invariance of the wave function to constant values, equation 2.3 can be solved neglecting V_N, thus obtaining the so-called *pure electronic energy* and then adding the neglected term to obtain the electronic energy. Moreover, equation 2.2 may have many Eigenfunctions Ψ_i, each returning an energy Eigenvalue E_i. Such Ψ_i constitute a set of wave functions. One assumption that can be made is that such functions Ψ_i are orthonormal, i.e. equation 2.4 holds true:

$$\int \Psi_i\Psi_j d\mathbf{r} = \delta_{ij} \tag{2.4}$$

where δ_{ij} is equal to one if and only if $i = j$ else it is equal to zero. So considering equation 2.2 for the wave function Ψ_i , by multiplying both its right- and left-hand sides times Ψ_i and integrating them in dr, applying equation 2.4 we obtain equation 2.5.

$$\int \Psi_i H\Psi_i d\mathbf{r} = E_i \int \Psi_i\Psi_i d\mathbf{r} = E_i\delta_{ii} = E_i \tag{2.5}$$

that provides a method for computing molecular energy given a wave function. Given an orthonormal set of wave functions Ψ_i, a generic wave function

can be expressed as a linear combination of these, as expressed in equation 2.6.

$$\Phi = \sum c_i \Psi_i \tag{2.6}$$

As a consequence of applying the condition expressed by equation 2.5 to a wave function Φ in the form of equation 2.6, the energy associated to the wave function Φ is a linear combination of the energies E_i associated to the single orthonormal set of wave functions Ψ_i according to equation 2.7.

$$E_\Phi = \int \Phi H \Phi d\mathbf{r} = \sum_i c_i{}^2 E_i \tag{2.7}$$

The quantum mechanical assumption that quantities can assume only discrete values implies that among the energies E_i of the orthonormal set of wave functions Ψ_i there will exist one (or more than one) that is minimal. That value, labeled E_0, is the energy associated to the "ground state" of the system, i.e. the state of the system at its lowest energy state which is a condition towards which systems tend at rest. Thus, assuming an orthonormal set of wave functions, equation 2.8 holds:

$$E_\Phi = \int \Phi H \Phi dr \geq E_0 \tag{2.8}$$

If the basis set is not orthonormal, an analogous principle holds thus a lowest energy bound still corresponding to the ground state can again be found. In this case however the non orthonormality of the basis set must be taken into account, therefore the expression of the energy E_Φ in equation 2.8 must take into account the fact that the basis set is no longer orthonormal. This leads to the energy form of equation 2.9:

$$E_\Phi = \frac{\int \Phi H \Phi dr}{\int \Phi^2 dr}. \tag{2.9}$$

It can be easily observed how equation 2.9 is obtained simply by reorganizing the first two members of equation 2.5 and that the energy term in equation 2.8 is obtained from equation 2.9 when the denominator of the latter is equal to 1 (which is the orthonormality condition). So, for the purpose of choosing (or iteratively adjusting) a basis set suitable to constructing a wave function as in equation 2.6, the figure of merit of the currently chosen basis set is the magnitude of its associated E_i values, and accordingly also of E_0. Such energy magnitudes should be minimized. According to the Born-Oppenheimer approximation, nuclei can be considered to be occupying fixed positions if compared to electrons, so that electronic wave functions can be considered instead of atomic and molecular system wave functions. For such systems

the Schroedinger equation simplifies to that shown as equation 2.3. Further simplification can be attained considering electronic wave functions for systems where there is only one nucleus and one electron: in such case each wave function describes an atomic orbital, and the associated pure electron energy is the magnitude of the energy of the electron when it is in that atomic orbital. Due to the possibility of neglecting the electron-to-electron interaction in the Hamiltonian associated to the single electron Schoedinger function, it is useful to use as basis set for equation 2.6 a set of single electron atomic orbitals to build more complex molecular orbitals. This principle is the *Linear-combination of atomic orbitals, (LCAO)* technique. Considering the Ψ_i wave function in equation 2.6 as atomic orbitals, then the construction of any wave function $\Phi = \sum c_i \Psi_i$ (whose quality is described by the aforementioned figure of merit) requires determining the c_i coefficients that minimize the energies of any linear combination of the chosen atomic orbitals. This involves computing the partial derivatives of the energy (in the form of equation 2.9) with respect to c_i $\forall i \in [1, N]$ and equating them to zero. By substituting equation 2.6 into equation 2.9 two integral terms can be identified in the resulting expression. One is called the *resonance integral* expressed in equation 2.10

$$H_{ij} = \int \Psi_i H \Psi_j dr \qquad (2.10)$$

and the other is called *overlap integral*, which quantitatively describes the spatial overlap of given couples of basis functions. Its expression is as in equation 2.11.

$$S_{ij} = \int \Psi_i \Psi_j dr \qquad (2.11)$$

With this formalism, the equation that must be solved to find the c_i set that minimize energy is equation 2.12.

$$\sum_{i=1}^{N} c_i \left(H_{ji} - E S_{ji} \right) = 0 \qquad (2.12)$$

To ensure that the set of the aforementioned derivatives has nontrivial solution, its characteristic (also known as *secular*) equation must be satisfied. Such condition is displayed in equation 2.13.

$$\begin{vmatrix} H_{11} - E S_{11} & H_{12} - E S_{12} & \cdots & H_{1N} - E S_{1N} \\ H_{21} - E S_{21} & H_{22} - E S_{22} & \cdots & H_{2N} - E S_{2N} \\ \vdots & \vdots & \ddots & \cdots \\ H_{N1} - E S_{N1} & H_{N2} - E S_{N2} & \cdots & H_{NN} - E S_{NN} \end{vmatrix} = 0 \qquad (2.13)$$

The solution of equation 2.13 returns N eigenvalues E, which - used in the energy minimization equation 2.12 - allows obtaining the basis set coefficients c_i that characterize the considered molecular orbital. The discussion carried on so far concerning a one-electron system can be extended to a multi-electron system by assuming that the energy associated to the latter is equal to the sum of the energies of the occupied one-electron molecular orbitals, each being characterized by its Hamiltonian h_i expressed as per equation 2.14, where M is the number of the nuclei of the system and where the interelectronic repulsion term has been neglected.

$$h_i = -\frac{\hbar}{2m_e}\nabla_i^2 - \sum_{k=1}^{M} e^2 \frac{Z_k}{r_{ik}} \tag{2.14}$$

The multi-electron Hamiltonian may then be expressed as per equation 2.15 i.e. as the sum of the one-electron Hamiltonians outlined so far.

$$H = \sum_{i=1}^{N} h_i \tag{2.15}$$

Each h_i, applied to a one-electron wave function ψ_i will return the energy ϵ_i associated to that wave function, i.e. the one-electron Schroedinger equation 2.16 holds.

$$h_i\psi_i = \epsilon_i\psi_i \tag{2.16}$$

Note that the one-electron ψ_i are not known. It can be demonstrated that the multi-electron wave function can be constructed as the product of the one-electron wave functions (*Hartree-product*) and that the respective overall energy is the sum of the one-electron energies, as shown in equation 2.17

$$\begin{cases} \Psi_{HP} = \prod_{i=1}^{N} \psi_i \\ E_{HP} = \sum_{i=1}^{N} \epsilon_i \end{cases} \tag{2.17}$$

The construction of the Hartree-product wave function requires knowledge of the individual ψ_i one-electron wave functions, which are actually not known. One possibility is to undertake an iterative process based on an initial-guess ψ_i set, which is used to compute the h_i operators to use in the one-electron Schroedinger equation. The solution to such equation provides a new set of ψ_i. The process outlined so far is iterated until the difference between two subsequent sets of h_i is below an acceptance threshold. Such process is called *self-consistent field* (or *SCF*) method[29]. A wave function as expressed by 2.17 does not satisfy the Pauli exclusion principle according to which any molecular orbital can be populated with two electron having

opposite spin number. Let α be the spin quantum number, a multi-electron wave function that satisfies the Pauli exclusion principle can be expressed as a *Slater determinant*[30] (equation 2.18).

$$
\Psi_{SD} = \frac{1}{\sqrt{N!}} \begin{vmatrix} \psi_1(1)\,\alpha(1) & \psi_2(1)\,\alpha(1) & \cdots & \psi_N(1)\,\alpha(1) \\ \psi_1(2)\,\alpha(2) & \psi_2(2)\,\alpha(2) & \cdots & \psi_N(2)\,\alpha(2) \\ \vdots & \vdots & \ddots & \cdots \\ \psi_1(N)\,\alpha(N) & \psi_2(N)\,\alpha(N) & \cdots & \psi_N(N)\,\alpha(N) \end{vmatrix} \tag{2.18}
$$

2.2 The Hartree-Fock SCF method

The Hartree-Fock SCF method is an iterative method for finding molecular orbitals according to the multi-electron wave function theory outlined so far. Therefore Hartree-Fock molecular orbitals can be found as Eigenfunctions of a set of one-electron operators, but in this case in the electronic repulsion effects are taken into account. The one-electron Fock operator can be defined for the $i - th$ electron as expressed in equation 2.19:

$$
f_i = -\frac{1}{2}\nabla_i^2 - \sum_k^{nuclei} \frac{Z_k}{r_{ik}} + V_i^{HF}\{j\} \tag{2.19}
$$

where the first term is related to the one-electron kinetic energy, the second term is the nuclear attraction contribution and finally V_i^{HF} is the Hartree-Fock potential. In order to determine the molecular orbitals, the method requires the availability of N basis functions and solving the secular equation 2.20 for E, the values of F_{ij} and S_{ij} being computed explicitly [31, 32].

$$
\begin{vmatrix} F_{11} - ES_{11} & F_{12} - ES_{12} & \cdots & F_{1N} - ES_{1N} \\ F_{21} - ES_{21} & F_{22} - ES_{22} & \cdots & F_{2N} - ES_{2N} \\ \vdots & \vdots & \ddots & \cdots \\ F_{N1} - ES_{N1} & F_{N2} - ES_{N2} & \cdots & F_{NN} - ES_{NN} \end{vmatrix} = 0 \tag{2.20}
$$

The S_{ij} elements are overlap integrals, analogous to those previously expressed by equation 2.11. The F elements depend on both the chosen basis set and the molecular orbitals. Given the couple of basis functions μ and ν, $F_{\mu\nu}$ is computed as in equation 2.21.

$$
F_{\mu\nu} = \left\langle \mu \left| -\frac{1}{2}\nabla^2 \right| \nu \right\rangle - \sum_k Z_k \left\langle \mu \left| \frac{1}{r_k} \right| \nu \right\rangle + \sum_{\lambda\sigma} P_{\lambda\sigma} \left[(\mu\nu|\lambda\sigma) - \frac{1}{2}(\mu\lambda|\nu\sigma) \right],
$$
$$
\tag{2.21}
$$

where the first and second term of equation 2.21 consist in *one-electron integrals* as expressed by 2.22:

$$\langle \mu \left| g \right| \nu \rangle = \int \phi_\mu \left(g\phi_\nu \right) d\mathbf{r} \tag{2.22}$$

and the third term involves the *exchange integrals* in the form expressed by equation 2.23:

$$(\mu\nu|\lambda\sigma) = \int \int \phi_\mu \left(1 \right) \phi_\nu \left(1 \right) \frac{1}{r_{12}} \phi_\lambda \left(2 \right) \phi_\sigma \left(2 \right) d\mathbf{r} \left(1 \right) d\mathbf{r} \left(2 \right) \tag{2.23}$$

where ϕ_μ, ϕ_ν, ϕ_λ and ϕ_σ are electron probability densities. The term $P_{\lambda\sigma}$ in equation 2.21 accounts for the contributions of base functions to a molecular orbitals.

2.3 Configuration interaction

In order to account for interaction between electrons in a multi-electron system the *Configuration Interaction (CI)* and the *Moeller-Plesset (MP)* [33] methods have been developed. Both methods rely on the Hartree-Fock wave reference function and add a correction term which is computed according to the variational principle in the case *CI* method and on the perturbation theory in the case of the *Moeller-Plesset* [33] method. In the case of the Configuration Interaction, a basis set consisting of Slater determinants is linearly combined to obtain multi electron wave functions. Coefficients for such linear combination are the Hartree-Fock coefficients. The ground state wave function can be expressed on the basis of the determinants of the ground state and all the possible electronic excitations as:

$$\Psi = \sum_L c_L \phi_L. \tag{2.24}$$

The relevant Schroedinger equation is:

$$\sum_L c_{LI} \langle \phi_K | \hat{H} | \phi_L \rangle = E \sum_L c_{LI} \langle \phi_L | \phi_L \rangle. \tag{2.25}$$

The Slater determinants constitute an orthonormal basis set and the overlap matrix between the various Slater determinants is $\langle \phi_L | \phi_L \rangle = \delta_{KL}$. Rewriting Eq. 2.25 in a more compact fashion we obtain the usual Schroedinger equation:

$$\mathbf{HC_1} = E\mathbf{C_1}, \tag{2.26}$$

where the $\mathbf{C_1}$ coefficients weigh the contribution of each Slater determinants. Among them ,the $\mathbf{C_{11}}$ coefficient is associated to the aforementioned

Hartree-Fock reference wave function and the remaining are associated to the excited state determinants and are the correction terms that characterize the technique. The procedure outlined would return the value of the exact non relativistic energy associated to a system that satisfies the Born-Oppenheimer approximation, but it is impractical in that it requires an infinite basis set. To overcome this issue, one possibility would be to exploit a non-complete basis set consisiting of a finite set of spinorbitals and then diagonalize the obtained Hamiltonian. Such technique is the *Full CI* and returns the exact solution, but still requires an impractically large basis set. In practice the basis set used is truncated and limited in size, and purposely restricted only to the wave functions pertaining to a given excitation type, such as the single excitation (CIS), or double excitation (CID) or both single and double excitation terms (CISD). From a performance standpoint the CIS method is fast but characterized by poor accuracy, while CISD is better but much more demanding in terms of CPU performance. Equation 2.26 can be solved for any of the Eigenvalues each one representing the energy of an excited state.

2.4   Multiconfigurational self consistent field

The methods outlined in section 2.3 are based on a linear combination of Hartree-Fock wavefunctions, with the drawback of requiring a large number of terms to obtain satisfactory results. This in turn implies a heavy CPU usage and long convergence time. The multiconfigurational self consistent field (MCSCF) [34] method overcomes such limitations that render such methods impractical by optimizing both the CI coefficients and the orbitals coefficients. This allows using a limited number of wave functions as linear combination terms thus limiting the CPU usage and convergence time. Several techniques can be exploited to carry out the MCSCF method. The most straightforward one optimizes the CI coefficients and the orbitals coefficients separately. The CI coefficients can be optimized as per equation 2.26 using the procedure outline in section 2.3. The molecular orbitals coefficients optimization is done through a unitary transformation of the orbitals coefficients. Let Φ be an orthonormal basis set of orbitals and U the unitary transformation matrix that, applied to the basis set gives the transformed orbitals.

$$\Phi' = U\Phi \qquad (2.27)$$

where U can be constructed as:

$$U = e^A = \begin{pmatrix} e^{A_{11}} & \cdots & e^{A_{1n}} \\ \vdots & \ddots & \vdots \\ e^{A_{n1}} & \cdots & e^{A_{nn}} \end{pmatrix}. \tag{2.28}$$

that is based on a matrix (A) that is a skew-symmetric $(A^T = -A)$ in order to ensure that U is unitary. Therefore A can be expressed as a linear combination of E_{ij} matrices the elements of which are equal to 1 only at the $i-th$ row and $j-th$ column, whereas the other elements are zero. This allows rewriting equation 2.28 as:

$$U = e^{\sum_{ij} A_{ij} E_{ij}} \tag{2.29}$$

For what concerns the wave functions, an analogous transformation is applied:.

$$|\Psi'\rangle = e^A |\Psi\rangle \tag{2.30}$$

having energy E' as per equation 2.31.

$$E' = \langle \Psi' | H | \Psi \rangle = \langle e^A \Psi | H | e^A \Psi \rangle. \tag{2.31}$$

Such equation can be rewritten in the form of a Taylor expansion truncated to the second-order derivative term as shown in equation 2.32.

$$E' = E0 + \sum_{ij} A_{ij} \left(\frac{\partial E}{\partial A_{ij}} \right)_0 + \frac{1}{2} \sum_{ijkl} A_{ij} Akl \left(\frac{\partial^2 E}{\partial A_{ij} \partial A_{kl}} \right) \tag{2.32}$$

The same equation can in turn be expressed in matrix form, assuming that A is a column vector containing the parameters and defining F as a row vector containing the first- order derivatives of the energies with respect to the skew-symmetric matrix elements, and defining G as the energy second-order derivatives. According to such assumptions, equation 2.32 can be expressed as:

$$E' = E0 + FA + \frac{1}{2} A^T G A \tag{2.33}$$

The E' energy can be minimized if $\frac{\partial E\hat{a}}{\partial A} = 0$ that is true if $F + GA = 0$. That is:

$$A = -G^{-1} F \tag{2.34}$$

Since, according to equation 2.27, $U = e^A$ we can obtain U again as a Taylor expansion of the matrix exponential truncated to the second order term:

$$U = 1 + A + \frac{1}{2} A^2 \tag{2.35}$$

the results of such sum of matrices, in the case A is skew-symmetric as this case, has the same form a rotation matrix has. This means that the orbitals optimization process is carried out by means of rotations which are iterated until convergence. Simulation softwares often implement the *Complete Active Space Self Consistent Field, CASSCF* or the *Restricted Active Space Self Consistent Field, RASSCF*, both of them considering excitation as taking place only between a defined number of orbitals. According to such methods the orbitals are subdivided into three groups: the core orbitals group, which is treated as a set of Hartree-Fock wave functions, the group of the occupied and empty active orbitals, where excitations are considered explicitly and eventually the virtual orbitals, treated as empty Hartree-Fock wave functions. These two methods display significant differences in the excitations that are introduced. In particular in the CASSCF case all the possible excitations are taken into account. On the contrary, in the RASSCF case the excitations can be selected and the active space is further subdivided into three sections Section $RAS1$ contains occupied orbitals where only a defined number of electron can be excited. This is similar to a truncated CI. Section $RAS2$ contains both occupied and empty orbitals processed with CASSCF technique. Section $RAS3$ are orbitals that are empty and can be occupied only by a defined number of electrons. The choice of the orbitals basis set for simulation has to be done taking into account the fact that the CPU usage dramatically increases with the number of the active orbitals introduced, so the use of the optimum basis set is not always viable. Therefore experience and intuition are an asset in the choice of a suitable limited set of active orbitals.

2.5   The CASPT2 method

The CASPT2 method is a post MCSCF method which allows computing the correlation energy of the rest of the system after the active part of the system has been treated with MCSCF. CASPT2 method [34] is based on correcting the wave functions applying the perturbation theory. Given a CASSCF wave function, the 2nd order perturbation can be written on the basis of the $i-th$ perturbation Hamiltonians $\hat{H}_i$ and the perturbation weight λ:

$$\hat{H} = \hat{H}_0 + \lambda \hat{H}_1. \tag{2.36}$$

The energies can be expressed in a similar fashion:

$$E_0 = \langle \Psi_0 | \hat{H}_0 | \Psi_0 \rangle$$
$$E_1 = \langle \Psi_0 | \hat{H}_1 | \Psi_0 \rangle \tag{2.37}$$
$$E_2 = \langle \Psi_0 | \hat{H}_1 | \Psi_1 \rangle$$

In equation 2.37 the orthogonality is implied between the $0 - th$ order wave function and the perturbation. The $\hat{H}_0$ term is the $\hat{F}$ Fock operator and since the CASSCF wave function is invariant with respect to rotations in the core, active and virtual orbitals, the Fock operator can be expressed as a superposition of diagonal subsetes f_i:

$$\hat{H}_0 = \hat{F} = f_{cc} + f_{aa} + f_{vv} + f_{ac} + f_{vc} + f_{va} \tag{2.38}$$

in which the subscript c refers to *core*, a refers to *active* and v refers to *virtual* orbital spaces. The Hamiltonian of the perturbation is the difference between the whole Hamiltonian and $\hat{H}_0$ so that the wave function must be corrected by a term Ψ_1 that is the linear combination of the remaining configurations ϕ_μ in the CI space, i.e. except those of the active space. From a mathematical point of view:

$$\Psi_1 = \sum_\mu C_\mu \phi_\mu \tag{2.39}$$

where the C_μ are the linear combination coefficients:

$$C_\mu = -\frac{\langle \phi_\mu | \hat{H}_1 | \Psi_0 \rangle}{E_\mu - E_0}. \tag{2.40}$$

From the CPU burden standpoint the CASPT2 is very demanding.

2.6 Molecular mechanics basics

Molecular mechanics is a method for the modeling of molecular systems (either small or large) using classical mechanics. Unlike quantum chemical models (where no reference is made to chemical bonding), molecular mechanics describes molecules in terms of bonded atoms, which have been distorted from some idealized geometry due to nonbonded van der Waals and Coulombic interactions [35, 36]. This concept is reflected in the molecular mechanics molecule energy which is a *strain* energy represented as a sum of terms which are contributions caused by deviations from ideal conditions in terms of bond distances, bond angles, torsion angles, along with contributions due to non-bonded (van der Waals and Coulombic) interactions. Force

fields, sets consisting of a functional form and a group of parameters, are used to compute the potential energy of the systems. Such force fields can be obtained either from experiments or by quantum-mechanical simulations or can even be guessed; their quality and derivation can be suitable for energy but not for geometry, for example, or vice versa. Anyway, the system has to be defined in terms of the atom types and their hybridization, and bond location and types have to be guessed. The molecular mechanics molecule energy can be expressed as in the following equation 2.41 :

$$
\begin{aligned}
E_{strain} \quad = \quad & \sum_{bonds} \frac{1}{2} K_b \left(b - b_{eq} \right)^2 + \sum_{angles} \frac{1}{2} K_\vartheta \left(\vartheta - \vartheta_{eq} \right)^2 + \\
& + \sum_{dihedrals} K_\phi [1 + \cos\left(n\phi - \delta \right)] + \\
& + \sum_{non-bonded} \left[\frac{A_{ij}}{R_{ij}^{12}} - \frac{B_{ij}}{R_{ij}^6} + \frac{q_i q_j}{\epsilon R_{ij}} \right] + \\
& + \sum_{H-bonded} \left[\frac{C_{ij}}{R_{ij}^{12}} - \frac{D_{ij}}{R_{ij}^{10}} \right] \qquad (2.41)
\end{aligned}
$$

where the first term accounts for the bond lengths, with force constants K_b that are specific to the atom-type pair considered; the second term accounts for the bond-angle bending and is characterized by the force constant K_ϑ . The first two terms display a quadratic dependency to distances and angles, respectively. The third term describe the torsion potential with the dihedral angle ϕ; its functional form reflects its inherent periodicity. The fourth one is related to non-bonded interactions (van der Waals contribution plus dispersion energy and the electrostatic term); eventually, the fifth term accounts for bonded interactions such as hydrogen bondings, $\pi-$conjugation or anomeric effects, etc.. Generally speaking, the first three terms tend to zero for unstrained and unhindered molecules. The fourth term can attain negative values especially if the molecule is a globular one characterized by a large amount of non-bonded attractive terms. The outlined technique has both advantages and drawbacks. Among the pros are certainly the fact that it can tackle large molecules and simple energy functions are quickly solved, and often returns reliable results for what concerns the van der Waals interactions at non-bonded distances for larger molecules, and for computing vibrational properties. Among the cons are the non-uniqueness of the functional forms, often proprietary to the make of the simulation tool and that it cannot tackle unusual or unpredicted bondings or interactions since bonding information has to be passed as a known information.

2.7   QM/MM hybrid methods

The understanding of the role of biomolecules (polypeptides, enzymes, nucleic acids) usually requires the knowledge of their electronic structure, of their geometry, and often of their reactivity. This knowledge can only be reached by means of quantum chemical computations. The main challenge, from a quantum chemistâs point of view, is their size: often thousands of atoms and their treatments by means of standard quantum chemical codes are still unrealistic. The knowledge of the electronic properties of the whole molecule is not always required to understand the chemical behavior of a large molecular system. The chemically interesting part can be a solute in a solution, a local defect, the active site of an enzyme, or a local chromophore. The rest of the system cannot be discarded, as it interacts with the center of interest by means of mechanical constraints, electrostatic, and induction interactions. The simplest idea is to perform an embedding of the interesting chemical part into its surroundings treated by means of a less demanding method, that is, molecular mechanics. This gives rise to a family of treatments known under the generic name of Quantum Mechanical/Molecular Mechanical (QM/MM) methods. MM (Molecular Mechanics) methods are computational techniques aimed at the study of large molecular systems. Such techniques are based on the assumption that atoms are as rigid as charged spheres, bonded to each other. Bonds, angle bendings and torsions are described in terms of harmonic potentials , unbonded interactions are described as charge dipoles. Such methods are computationally fast but cannot model chemical reactions since the latter imply a change in the structure of the system. On the contrary, QM (Quantum Mechanics) are based on the calculation of electronic wave functions, therefore they are able to handle and describe system changes as those occurring in chemical reactions. QM techniques are however computationally demanding, moreover they are not suitable to study very large systems. Hybrid QM/MM methods [37, 38, 39, 40, 41]have been developed recently to overcome the drawbacks of the previously mentioned methods. It is based on the subdivision of the whole system into two smaller subsystems: the part containing the atoms involved in the reaction, and the remaining atoms which constitute the larger environment surrounding the reaction. The former is treated by means of QM methods, while the latter is treated at MM level. The boundary region between the two is very critical in obtaining sensible results under a physical point of view [40, 38, 42].

COBRAMM (Computational Bridge between Ab-initio and Molecular mechanic) is a software that has been developed at the Department of Chem-
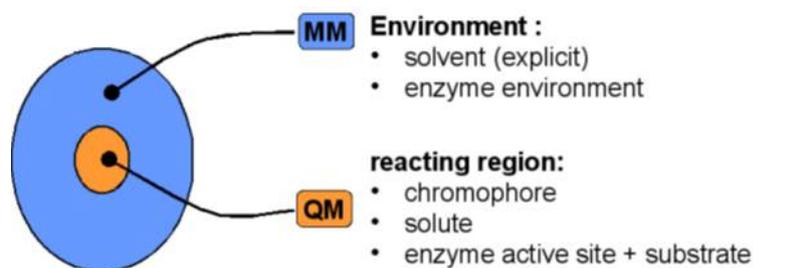
Figure 2.1: Schematic representation of the hybrid QM/MM approach.

istry at the University of Bologna to perform hybrid QM/MM calculations. Such program is a modular system that acts as an interface between different programs that execute single phases of the whole calcula- tion and exchange data by means of the interface itself. The used programs can be chosen to tailor the computational steps to the requirements of the specific computational process required. This way the most efficient software tools can be chosen and run together to return the results. QM and MM calculations are performed separately and the respective results are processed together to return the QM/MM energies and the complete force sets. Such data are then used to generate a new geometry. In doing this, the system is subdivided into two independent regions: one is the external one (*Opt1*) which is handled by a fast but rough algorithm; the other is smaller and it is called *Opt2/MD* and includes the QM region, which is treated using sophisticated algorithms such as BFGS [43]. In the case the QM and Opt2/MD regions coincide, then the system is partitioned in the high and low layer, the first one being treated at QM level using accurate algorithms and the second one being treated at MM level with fast optimization algorithms. On the contrary, if the QM region does not coincide with the Opt2/MD region, then an intermediate layer, called *medium* layer is included and consists of MM atoms that undergo accurate geometry optimizations or MD along with QM atoms. This layer is treated at QM level, and it improves the simulation of the electrostatic interaction between the QM and MM regions. The Electrostatic Embedding approach [38, 44, 39, 41] is used to model the electrostatic influence of the MM region to the QM layer. Several type of calculations are currently supported and allow tailoring the calculation to the requirements dictated by the investigated system. COBRAMM uses one to three of the mentioned layers according to the selected calculation type. For example the HML calculation uses the low, medium and high layers, the HM uses the high and the medium, and so on. The boundary region is handled by means of the *atom-link* approach. This is useful in the cases where at least one

covalent bond crosses the QM/MM boundary. Such crossing implies that the valence orbitals of the QM atom whose bond is cut remains unfilled: the obtained QM structure is called *model*. The atom-link approach fills such vacancy by adding an hydrogen atom as link-atom in the QM region as a bridge between the latter and the MM region. The structure thus obtained is called *model-H* and allows performing the QM calculation. Such approach is widely adopted and gives good results even though artifacts might be obtained, since the H atoms actually alter the structure of the system and the MM point charge distribution at the boundary has to be altered accordingly. In order to avoid that, the position of the hydrogen atom-links are moved each step in a position which is related to the position of the QM and MM atoms whose bond crossed the boundary but keeping fixed the distance of the H atom with respect to the position of the mentioned QM frontier atoms. As far as the non-bonding cross terms are concerned, the Van der Waals terms are treated at MM level whereas the electrostatic interactions across the regions are treated at QM level using an Electrostatic Embedding scheme according to which the QM computations are carried out on the model-H surrounded by atomic point charges of the MM layers (the *emb* charges). Instead, *Pod* point charges are the unmodified point charges from the MM force field. Differences between the *emb* and *pod* sets are limited to the boundary region.Charge distribution in the *emb* scheme is arranged so that the charge associated to the MM atoms bonded to the atom-link hydrogens are zero since such charge is assumed to be redistributed towards the inner atoms so to preserve the total charge of the system. The total energy of the QM/MM system can be expressed basically as the sum of the QM and MM contributions according to a subtractive scheme [39].

$$E^{tot} = E_{QM}^{model-H} + E_{MM}^{real} + E_{QM}^{el_{model-H/emb}} - E_{MM}^{el_{model/pod}} - E_{MM}^{model-H} \quad (2.42)$$

In equation 2.42 the first and third term are computed at QM level, the first one being the QM energy of the model-H in vacuum, and the third term describes the electrostatic interaction between the model-H and the *emb* subsystem. This ensures the wave function will be perturbed by the MM charges so that the QM subsystem is polarized by the MM environment. The remaining terms of equation 2.42 are MM energies. Their contribution can be rewritten as:

$$E_{MM} = E_{MM}^{pod} + E_{MM}^{VdW_{model/pod}} + E_{MM}^{bond,bending,torsion_{model/pod}} \quad (2.43)$$

where the first term are the real MM atoms, while the latter two are the contributions both from the MM and QM atoms that are treated at MM

level. The assumptions behind the overall QM/MM energy scheme reported
is that the atom-link is an hydrogen atom and that charge redistribution has
been taken into account as previously described. In order to perform the
computation outlined so far COBRAMM can interface to commercial pack-
ages for QM and MM calculations including MOLCAS55, Gaussian03 [45],
and others for the QM part; Amber [46] and Tinker [47] for the MM part.
COBRAMM not only interfaces to the third-party QM and MM packages ac-
cording to the required calculation configuration but also gathers data from
the the results to return the total energy and gradient for the current step
and eventually call an external application for geometry optimization or for
molecular dynamics to produce the geometry for the subsequent step until
the convergence criterion has been met. Figure 2.2 displays a COBRAMM
computation flow chart [48].

The computation of the second-derivative allows evaluating the curva-
ture of the Potential-Energy Surface or the zero-point energy or other quan-
tities. The Hessian matrix is computed by means of numerical techniques
based on the energy evaluation and first derivative calculation around the
point of interest in the PES, keeping unattached the QM wave function when
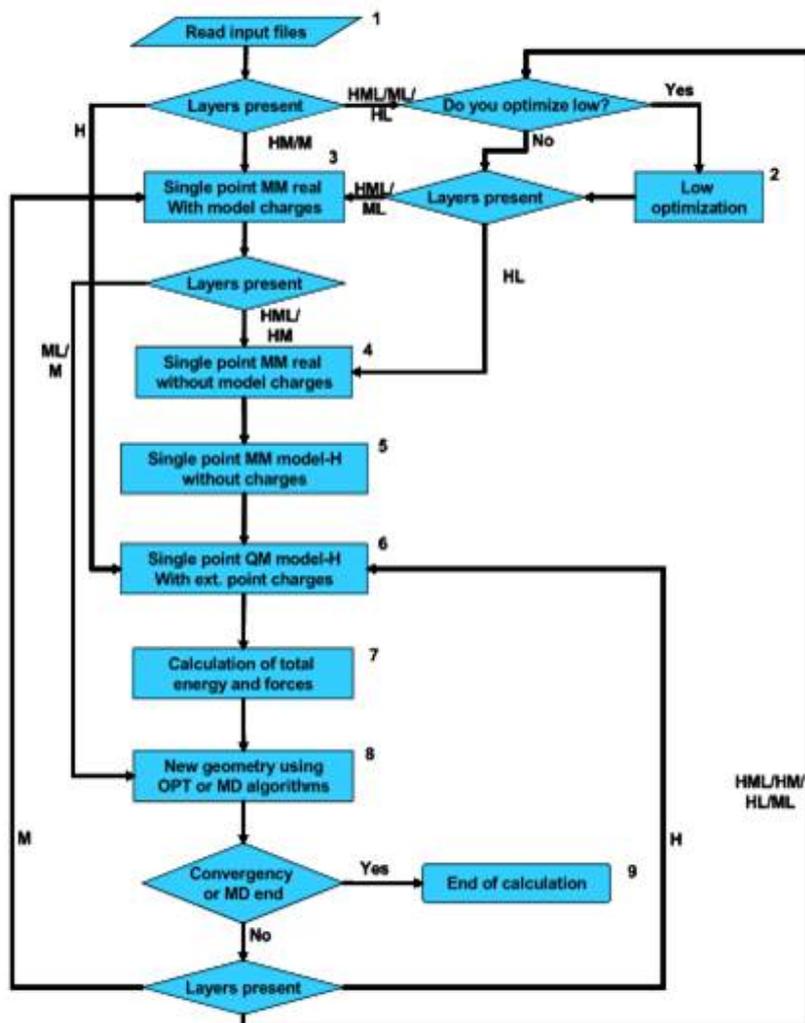only MM atoms are moving in order to decrease the CPU burden.

Figure 2.2: COBRAMM workflow

3 Potential energy surfaces and optimization techniques: stationary points, reaction paths and conical intersections

The potential energy surface (PES) is a central concept in computational chemistry. A PES is the relationship (mathematical or graphical) between the energy of a molecule (or a collection of molecules) and its geometry. The Born-Oppenheimer approximation says that in a molecule the nuclei are essentially stationary compared to the electrons. This is one of the cornerstones of computational chemistry because it makes the concept of molecular shape (geometry) meaningful, makes possible the concept of a PES, and simplifies the application of the Schroedinger equation to molecules by allowing us to focus on the electronic energy and add in the nuclear repulsion energy later. Potential energy surfaces are important because they aid us in visualizing and understanding the relationship between potential energy and molecular geometry (and in understanding how computational chemistry programs locate and characterize structures). The identification of stationary points on the potential energy surface (PES), which have a chemical meaning, remain a troublesome issue [49, 50]. Two are the most important reasons of this difficulty: first is due to the big number of internal coordinates of molecules ($6N-6$ where N is the number of atoms), second the algorithms often fail locating minima and most often fail locating transition states (TS). If we are interested to study excited electronic states we have also to consider the possibility of a crossings between states, such region are called conical intersection, in the second part of the chapter summarize their the most important properties.

3.1 Stationary points

For a thermally activated reaction, the energy of the TS and the shape of the potential energy surface around the TS can be used to estimate the reaction rate (see other reviews in this series). The steepest descent reaction path

(SDP) from the TS down to the reactants and to the products is termed the minimum energy path (MEP) or the intrinsic reaction coordinate (IRC; the MEP in mass-weighted coordinates). The reaction path from reactants through intermediates (if any) to products describes the reaction mechanism [51]. A more detailed description of a reaction can be obtained by classical trajectory calculations [52, 53, 54] that simulate molecular dynamics by integrating the classical equations of motion for a molecule moving on a potential energy surface. Photochemistry involves motion on multiple potential energy surfaces and transitions between them. The first and second derivatives of the energy with respect to the geomet- rical parameters can be used to construct a local quadratic approximation to the potential energy surface:

$$E\left(x\right) = E\left(x_0\right) + g_0^T \Delta x + \frac{1}{2}\Delta x^T H_0 \Delta x \qquad (3.1)$$

where g_0 is the gradient $\left(\frac{dE}{dx}\right)$ at x_0, H_0 is the Hessian $\left(\frac{d^2E}{dx^2}\right)$ at x_0 and $\Delta x = x - x_0$. The gradient and the Hessian can be used to confirm the character of minima and TSs. The negative of the gradient is the vector of forces on the atoms in the molecule. Because the forces are zero for minima, TSs, and higher- order saddle points, these structures are also termed stationary points. The Hessian or matrix of second derivatives of the energy is also known as the force constant matrix. The eigenvectors of the mass-weighted Hessian in Cartesian coordinates correspond to the normal modes of vibration (plus five or six modes for translation and rotation) [55]. For a structure to be characterized as a minimum, the gradient must be zero and all of the eigenvalues of the Hessian corresponding to molecular vibrations must be positive; equivalently, the vibrational frequencies must be real (the vibrational frequencies are proportional to the square root of the eigenvalues of the mass-weighted Hessian). For a TS, the potential energy surface is a maximum in one direction (along the reaction path) and a minimum in all other perpendicular directions. Therefore, a TS is characterized by a zero gradient and a Hessian that has one (and only one) negative eigenvalue; correspondingly, a TS has one and only one imaginary vibrational frequency. An $n - th$ order saddle point (also called a stationary point of index n) has a zero gradient and is a maximum in n orthogonal directions and hence has n imaginary frequencies. For a TS, the vibrational mode corresponding to the imaginary frequency is also known as the transition vector. At the TS, the transition vector is tangent to the reaction path in mass-weighted coordinates.

3.2 Coordinates

In principle, any complete set of coordinates can be used to represent a molecule and its potential energy surface. However, choosing a good coordinate system can significantly improve the performance of geometry optimizations. Inspection of the Hessian used in the local quadratic approximation to the potential energy surface, in equation 3.1, can reveal some favourable aspects of a good coordinate system. For example, an optimization will be less efficient if there are some very stiff coordinates and some very flexible coordinates. This corresponds to a mixture of very large and very small eigenvalues of the Hessian, (i.e., the Hessian is an ill-conditioned matrix). Strong coupling between coordi- nates can also slow down an optimization. This corresponds to off-diagonal Hessian matrix elements that are comparable in magnitude to the diagonal elements. Strong anharmonicity can seriously degrade the performance of an optimization. If the Hessian changes rapidly when the geometry of the molecule is changed, or if the valley around a minimum is strongly curved, then the quadratic expression in equation 3.1 is a poor approximation to the potential energy surface and the optimization will be slow to converge. The nature of the Hessian and the anharmonicity of the potential energy surface will be directly affected by the choice of the coordinate system. There are a number of coordinate systems that are typically used for geometry optimization. Cartesian coordinates are perhaps the most universal and the least ambiguous. An advantage is that most energy and derivative calculations are carried out in Cartesian coordinates. However, they are not well suited for geometry optimization because they do not reflect the "hemical structure" and bonding of a molecule. The x, y, and z coordinates of an atom are strongly coupled to each other and to the coordinates of neighboring atoms. Internal coordinates such as bond lengths and valence angles are more descriptive of the molecular structure and are more useful for geometry optimization. Bond stretching requires more energy than angle bending or torsion about single bonds. More importantly, the coupling be- tween stretches, bends, and torsions are usually much smaller than between Cartesian coordinates. In addition, internal coordinates are much better than Cartesians for representing curvilinear motions such as valence angle bending and rotation about single bonds. For an acyclic molecule with N atoms, it is easy to select set of $3N - 6$ internal coordinates to represent the molecule ($3N - 5$ coordinates for a linear molecule). Z-matrix coordinates are an example of such a coordinate system [56]. It is straightforward to convert geometries and derivatives between Cartesian and Z-matrix internal coordinates [57].

For acyclic molecules, the set of all bonds, angles, and torsions represents the intrinsic connectivity and flexibility of the molecule. However, for a cyclic molecule, this introduces more than the $3N - 6$ coordinates required to define the geometry of the molecule. Such a coordinate system has a certain amount of redundancy in the geometric parameters [57, 58, 59, 60, 61, 62]. Because only $3N - 6$ of these redundant internal coordinates can be transformed back to Cartesian coordinates in three dimensions, certain combinations of the redundant internals must be constrained during the optimization.

3.3   Newton and quasi-Newton methods

As described in standard texts on optimization [63, 64, 65, 66] most nonlinear optimization algorithms are based on a local quadratic approximation of the potential energy surface; equation 3.1. Differentiation with respect to the coordinates yields an approximation for the gradient, given by:

$$g\left(x\right) = g_0 + H_0 \Delta x \tag{3.2}$$

At a stationary point, the gradient is zero, $g(x) = 0$; thus, in the local quadratic approximation to the potential energy surface, the displacement to the minimum is given by:

$$\Delta x = -H_0^{-1} g_0 \tag{3.3}$$

This is known as the Newton or Newton-Raphson step. Newton or quasi-Newton steps are required to reach a stationary point. For minimization, the Hessian must have all positive eigenvalues (i.e., positive definite). If one or more eigenvalues are negative, the step will be toward a first or higher-order saddle point. Thus, without some means of controlling the step size and direction, simple Newton steps are not robust. Similarly, if the aim is to optimize to a TS, the Hessian must have one and only one negative eigenvalue, and the corresponding eigenvector (i.e., the transition vector) must be roughly parallel to the reaction path. At each step, Newton methods require the explicit calculation of the Hessian, which can be rather costly. Quasi-Newton methods start with an inexpensive approximation to the Hessian. The difference between the calculated change in the gradient and the change predicted with the approximate Hessian is used to improve the Hessian at each step in the optimization [63, 64, 65, 66].

$$H^{new} = H^{old} + \Delta H \tag{3.4}$$

For a quadratic surface, the updated Hessian must fulfil the Newton condition:

$$\Delta g = H^{new}\Delta x, \tag{3.5}$$

where $\Delta g = g\left(x^{new}\right) - g\left(x^{old}\right)$ and $\Delta x = \left(x^{new} - x^{old}\right)$. However, there are an infinite number of ways to update the Hessian and fulfill the Newton condition. For TS optimization, it is important that the Hessian has one and only one negative eigenvalue. This should be checked at every step of the optimization. If the Hessian does not have the correct number of negative eigenvalues, the eigenvalues need to be shifted or one of the methods for step size control Newton and quasi-Newton methods are the most efficient and widely used procedures for optimizing equilibrium geometries and can also be used effectively to find TSs. For each step in the Newton method, the Hessian in Eq. 3.3 is calculated at the current point. For quasi-Newton methods, Eq. 3.3 is used with an approximate Hessian that is updated at each step of the optimization (see below). Because actual potential energy surfaces are rarely quadratic, several needs to be used (see Step Size Control). The initial estimate of the Hessian for TS optimizations must have one negative eigenvalue and the associated eigenvector should be approximately parallel to the reaction path. The Hessian update should not be forced to be positive definite. The Powell-symmetric-Broyden (PSB) update [66] fulfils this role: this may be a bottleneck. The updating methods can be reformulated to update the inverse of the Hessian. For example, the BFGS formula for the update of the inverse Hessian is:

$$\Delta B^{BFGS} = \frac{\Delta x \Delta x^T}{\Delta x^T \Delta g} - \frac{B^{old}\Delta g \Delta g^T B^{old}}{\Delta g^T B^{old}\Delta g}, \tag{3.6}$$

where $B = H^{-1}$, and the updated inverse Hessian obeys $\Delta x = B\Delta g$. Limited memory quasi-Newton methods such as *L-BFGS* avoid the storage of the full Hessian or its inverse which would require $O(n^2)$ memory for n variables. Instead, they start with a diagonal inverse Hessian, and store only the Δx and Δg vectors from a limited number of previous steps; thus, the storage is only $O(n)$. The inverse Hessian is written as a diagonal Hessian plus the updates using the stored vectors. For the Newton step, $x^{new} = x^{old} - Bg^{old}$, the product of the updated inverse Hessian and the gradient involves only $O(n)$ work because it can be expressed in terms of dot products between vectors.

3.4   Step size control

The quadratic approximation to the potential energy surface is satisfactory only for a small local region, usually specified by a trust radius, τ . Steps outside this region are risky and optimizations are more robust if the step size does not exceed τ . An initial estimate of τ can be updated during the course of an optimization based on how well the potential energy surface can be fit by a quadratic expression. A typical updating recipe is as follows [63]:

$$\rho = \frac{\Delta E}{g^T \Delta x + \frac{1}{2}\Delta x^T H_0 \Delta x} \tag{3.7}$$

if $\rho > 0.75$ and $\frac{5}{4}|\Delta x| > \tau^{old}$, then $\tau^{new} = 2\tau^{old}$. If $\rho < 0.25$ then $\tau^{new} = \frac{1}{4}|\Delta x|$; otherwise $\tau^{new} = \tau^{old}$ [65]. The simplest approach to step size control is to scale the Newton step back if τ is exceeded. A better approach is to minimize the energy under the constraint that the step is not larger than τ. In the trust radius method (TRM), this is done by using a Lagrangian multiplier, λ, and corresponds to minimizing $E(x) - \frac{1}{2}\lambda(\Delta x^2 - \tau^2)$. With the usual quadratic approximation for $E(x)$, this yields:

$$g_0 + H_0\Delta x - \lambda\Delta x = 0$$

or

$$\Delta x = -\left(H_0 - \lambda I\right)^{-1} g_0 \tag{3.8}$$

where I is the identity matrix. For minimizations, λ must be chosen so that all the eigenvalues of the shifted Hessian, $H - \lambda I$, are positive, i.e., λ must be smaller (more negative) than the lowest eigenvalue of H.

3.5   Constrained optimizations

Under a variety of circumstances, it may be necessary to apply constraints while optimizing the geometry (e.g., scanning potential energy surfaces, coordinate driving, reaction path following, etc.). For nonredundant coordinate systems and simple constraints, the coordinates being held constant can be easily removed from the space of variables being optimized. For more general constraints and/or redundant internal coordinate systems, constraints can be applied by penalty functions, projection methods, or Lagrangian multipliers. In the penalty function method, the constraints $C_i(x) = 0$ are imposed by adding an extra term, $\frac{1}{2}\sum \alpha_i C_i(x)^2$, to the energy in Eq. 3.1 and the energy is minimized as usual. Because the α_i need to have suitably

large values so that the constraints are approximately satisfied at the minimum, the optimization may converge much slower than the corresponding unconstrained optimization. The preferred method for including constraints in an optimization is by using Lagrangian multipliers.

$$Pg_0 + PH_0P\Delta x + \alpha\left(I - P\right) = 0$$
$$P = I - \sum_i c_i c_i^T / \left|c_i\right|^2 ,$$

(3.9)

where the c_i are a set of orthogonal constraint vectors and $\alpha > 0$. For redundant internal coordinates, the projector needs to remove the coordinate redundancies as well as the constraint directions [67].

3.6 Constrained optimizations: the IRC

The concept of the reaction pathway has become important in the study of potential energy surfaces for chemical reactions. In general, the reaction path can be defined as the curve on the potential energy surface connecting the reactants and products through the transition state. This curve can be found by following the steepest descent path or minimum energy path (MEP) from the transition state toward reactants and products. When mass-weighted Cartesian coordinates are used, the path becomes the intrinsic reaction coordinate (IRC) discussed by Fukui [68]. The steepest-descent path from the saddle point to the minima can be defined easily, but depends on the particular choice of coordinate system. Cartesian coordinates would yield a different path than internal coordinates. Furthermore, internal coordinates are not unique, since a number of different sets of bond lengths, angles and torsions can represent the same structure. An intrinsic reaction path can be defined independently of the coordinate system by appealing to classical mechanics. For a given energy surface, the movement of a classical particle must be the same regardless of whether Cartesian coordinates or any of a number of different sets of internal coordinates are used.

The IRC is defined by the following differential equation:

$$\frac{dx}{ds} = \frac{g}{|g|} = v$$

(3.10)

where x is the (mass-weighted) coordinates, s is the path length and v is the (negative) normalized gradient. To solve this the starting geometry needs to be slightly displaced from the TS structure, this can be done by using the normal mode of the imaginary frequency to displace the structure along the coordinate that leads to the associated minima. Once this is done then an optimization technique is used to follow the path of highest negative

gradient to a minimum. Several algorithms have been proposed for the calculation of the IRC. One of the most widely used is the one proposed by Schlegel which is useful for following also curved reaction paths (as in many reactions) [69]. Firstly the gradient is calculated at a point x_k , an expansion point is generated from here by taking a step of $\Delta s/2$ along the direction of the gradient, where s is a step of predefined size. The point x_{k+1} is chosen so that the reaction path between x_k and x_{k+1} is an arc of a circle and so that the gradients g_k and g_{k+1} are tangent to this path. This algorithm requires an $n-1$-dimensional optimization on a hypersphere about the point $x\left(k+1\right)=x\left(k\right)-\frac{1}{2}sgk/|gk|$ where s is the stepsize. In this method quite large step sizes $(20-40)$ bohr amu can be used and also IRC with large curvature are followed quite closely. The energy is then minimized on a hypersphere with radius $\frac{1}{2}\Delta s$, located at the expansion point [69].
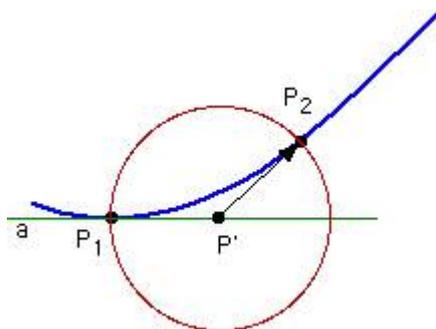

Figure 3.1: Starting from point P_1 on the path (shown in blue) construct auxiliary point P' located a distance of $n/2$ away from P_1 along tangent a (shown in green). The construction of P' does not involve any energy or gradient calculations. On a (hyper)sphere of radius $n/2$ centered at P' search for the point of lowest energy P_2. This latter point is the new point on the IRC path. This sequence is repeated until the geometry convergence criteria are fulfilled in direction along the pathway.

3.7   Conical intersection and non-crossing rule

In the case of diatomic molecules, the potential energy surfaces of two states, such as the ground state and the lowest- energy among the excited states will intersect only if the states are characterized by a different spatial or spin simmetry. On the contrary, in the case of polyatomic systems [3] two PES of a polyatomic molecule can in principle intersect even if they belong to states of the same symmetry and spin multiplicity. In order to determine the crossing conditions for polyatomic molecules [70] we can assume that the solutions of the Schroedinger equation for the electronic wave function

have been found except two, ϕ_1 and ϕ_2. The latter, in conjunction with the already found solutions constitute a complete orthonormal set. We also hypothesize that ϕ_1 and ϕ_2 correspond to the two states (whose energy is E_1 and E_2) whose crossings we are studying. Then it must be possible to express each of the two remaining electronic Eigenfunctions (which describe the states we want to examine) in the form:

$$\Psi = c_1\phi_1 + c_2\phi_2 \tag{3.11}$$

The determination of the energies associated to the two states requires computing the characteristic equation:

$$\begin{bmatrix} H_{11} - E & H_{12} \\ H_{21} & H_{22} - E \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \end{bmatrix} = 0 \tag{3.12}$$

knowing that $H_{ij} = <\phi_i|H|\phi_j>$ and $H_{12} = H_{21}$. Solving 3.12 returns the following energy values

$$E_{1,2} = \frac{H_{11} + H_{22} \pm \sqrt{(H_{11} - H_{22})^2 + 4H_{12}^2}}{2} \tag{3.13}$$

The energy of the two states match if the two energy solutions do coincide, i.e. the following equalities hold:

$$\begin{aligned} H_{11} &= H_{22} \\ H_{12} &= H_{21} = 0. \end{aligned} \tag{3.14}$$

For this condition to be met, there must exist at least two nuclear coordinates which vary independently. In the case of a diatomic molecule the only variable nuclear coordinate is the interatomic distance; the two states can intersect only if their spatial or spin symmetries are different. In the case of a molecule consisting of three or more atoms, the crossing is possible due to the abundance of degrees of freedom. By choosing two independent variable values that satisfy Eq. 3.14, the remaining degrees of freedom can be varied in the crossing region. Let x_1 and x_2 be the two independent coordinates and let their origin satisfy Eq. 3.14: the characteristic equations may be rewritten as:

$$\begin{bmatrix} W + h_1 x_1 - E & l x_2 \\ l x_2 & W + h_2 x_1 - E \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \end{bmatrix} = 0 \tag{3.15}$$

or

$$\begin{bmatrix} W + (m+k)\,x_1 - E & l x_2 \\ l x_2 & W + (m-k)\,x_1 - E \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \end{bmatrix} = 0 \tag{3.16}$$

where the following substitution has been made:

$$m = \tfrac{1}{2}\left(h_1 + h_2\right)$$
$$k = \tfrac{1}{4}\left(h_1 - h_2\right) \tag{3.17}$$

The characteristic equations return the energy values E as:

$$E_{12} = W + mx_1 \pm \sqrt{k^2 + x_1^2 l^2 x_2^2} \tag{3.18}$$

which is the equation a double cone with the same vertex, at the x_1 and x_2 origin. If ϕ_1 and ϕ_2 are the diabatic components of an adiabatic electronic configuration, then the crossing takes place if the two diabatic components do cross and for this to happen H_{11} must be equal to H_{22}, which means that the energies for the two diabatic functions must be the same. Indeed, at the point where $H_{11} = H_{22}$, then $E_1 = H_{11} - H_{12}$, and $E_2 = H_{11} + H_{12}$, that means that an energy difference exists between the two states.

$$\Delta E = E_2 - E_1 = 2H_{12} \tag{3.19}$$

This establishes a condition to determine whether crossing would happen or not, that is if the exchange term H_{12} is zero the two state energies are the same therefore the crossing takes place. In general, this condition holds when the two electronic states have different symmetry, be it spatial or spin. On the contrary, if $H_{12} \neq 0$ then the energies are different therefore no crossing takes place. This happens generally when the electronic states have the same symmetry. The exchange H_{12} value provides information concerning also the depth of the minimum where the avoided crossing (the higher the value the shallower the minimum and vice-versa). In the case of polyatomic system real crossing happens due to the higher number of degrees of freedom irrespective of the symmetry relation between the two states: suitable values of two independent coordinates ensure crossing, then the remaining $n - 2$ can span so that a $n - 2$ dimensional hyperline is found along which the two states intersect.

3.8   The "physical chemistry" of conical intersections

Conical intersections are key part of the mechanisms underlying the photochemical reactivity processes. Indeed, a CI point is like a funnel that connects a ground state potential surface to an excited state one; when a reactant in its excited state decays to its ground state, two or more products
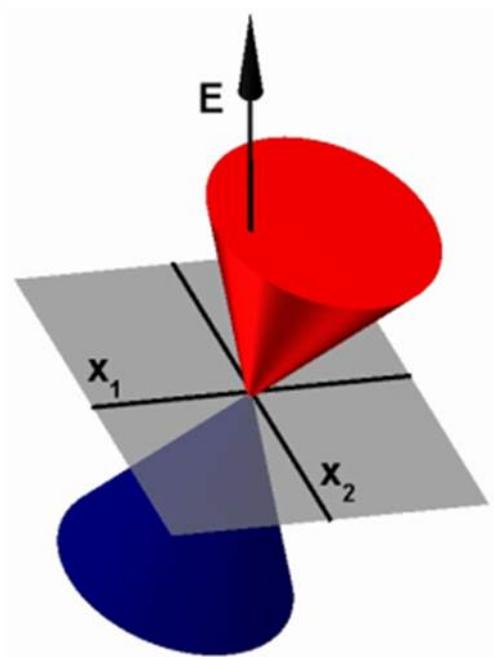
Figure 3.2: Schematization of the conical intersection concept between the ground state (blue) and the excited state (red) .
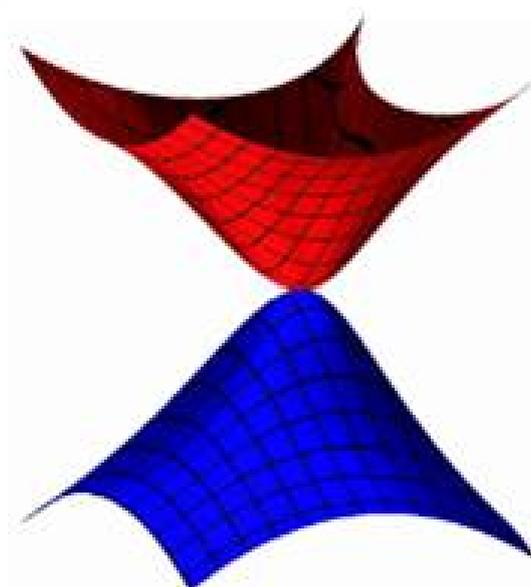

Figure 3.3: Two-dimension conical intersection .

can be obtained according to the relaxation path followed. From a mathematical standpoint the double cone shape is observable plotting the energies as a function of the gradient difference vector $\mathbf{x}_1$ and the gradient of the interstate coupling vector $\mathbf{x}_2$:

$$\begin{aligned} \mathbf{x}_1 &= \frac{\partial(E_1 - E_2)}{\partial \mathbf{q}} \\ \mathbf{x}_2 &= \langle \mathbf{C}_1 \frac{\partial \hat{H}}{\partial \mathbf{q}} \mathbf{C}_2 \rangle \end{aligned} \tag{3.20}$$

$\mathbf{C}_1$ and $\mathbf{C}_2$ being the configuration interaction eigenvectors (the latter being parallel to the non-adiabatic coupling vector) and $\hat{H}$ is the CI Hamiltonian. $\mathbf{x}_2$ is parallel to the non-adiabatic coupling vector $\mathbf{g}\left(\mathbf{q}\right)$. By spanning along these geometric coordinates, the *branching space* is identified, which is the plane on which the system energetic degeneracy is lifted. The remaining $n-2$ dimensions identify the *intersection space* (or *seam*), where there is an infinite number of conical intersections: by moving along the seam space the molecule is taken from a point of conical intersection to another one, still of conical intersection.
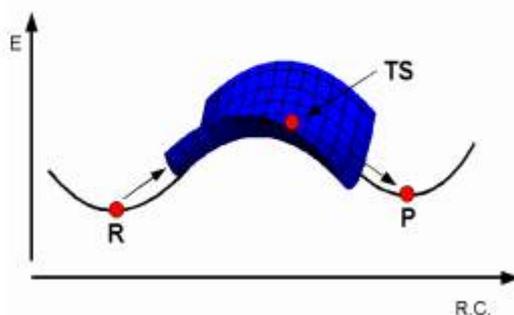
Figure 3.4: â Representative behavior of a thermal reaction.

The probability of a surface hop to occur is:

$$P = e^{-\frac{\pi}{4}\zeta} \tag{3.21}$$

where the value of ζ is:

$$\zeta = \frac{\Delta E\left(\mathbf{q}\right)}{\hbar |\dot{\mathbf{q}}| g\left(\mathbf{q}\right)} \tag{3.22}$$

that means that the lower the energy gap the more likely is the transition, and also that the latter is favoured by a large value of the non-adiabatic coupling g. $\Delta E < 2kcal/mol$ is required for the hop probability not to be negligible.
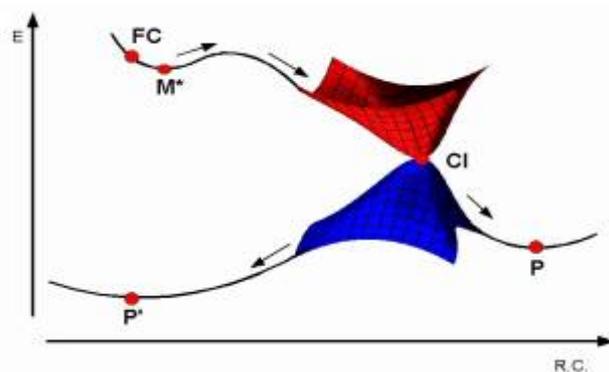
Figure 3.5: Representative behavior of a photochemical reaction.

Often the conical intersection point of chemical interest is located along a valley on the excited state potential energy surface, with the system moving downwards from the Franck-Condon region and reaching the excited state minimum M^*, which is connected to the conical intersection point trough a transition state. In the case no transition state exists to connect the excited state to the conical intersection point, the lowest-lying intersection point must be found within the intersection space.

3.9 Conical intersection optimization

In the case no intermediate or transition point is between the excited state and the conical intersection point, constrained geometry optimization techniques have to be implemented in order to find out the position of the low-energy stationary points having the same energy on two potential energy surfaces. Optimization is to be carried out along directions that are orthogonal to the hyperline spun by the $\mathbf{x}_1$ and $\mathbf{x}_2$ vectors, i.e. . the intersection space. In this case, the excited state potential energy surface gradient will not be zero at the optimized conical intersection point since that wouldnât be the vertex of a double cone. Instead, the projection of the excited state potential energy surface projection on the intersection space goes to zero at the optimized conical intersection. Let $\mathbf{x}_3 \ldots \mathbf{x}_n$ the vectors that identifty the intersection space , then the two conditions that have to be met at a optimized conical intersection point are:

$$E_2 - E_1 = 0$$
$$\frac{\partial E}{\partial \mathbf{x}_3} = \ldots = \frac{\partial E}{\partial \mathbf{x}_n} = 0. \tag{3.23}$$

Another viable method has been implemented in *Gaussian03* and is based on the fact that in the intersection space the following expression holds:

$$\frac{\partial (E_2 - E_1)^2}{\partial \mathbf{q}} = 2 (E_2 - E_1) \mathbf{x}_1 \qquad (3.24)$$

Since the size of the step should depend only on the energy difference and not on the $\mathbf{x}_1$ magnitude, the gradient along the step to the minimum of the energy difference $E_2 - E_1$ should be:

$$\mathbf{f} = \frac{(E_1 - E_2) \mathbf{x}_1}{\sqrt{\mathbf{x}_1 \mathbf{x}_1}} \qquad (3.25)$$

which will be zero when the two energies have the same value, irrespective of the $\mathbf{x}_1$ magnitude. By defining $\mathbf{P}$ as the projection of the gradient of E_2 on the intersection space we obtain a gradient value:

$$\mathbf{g} = \mathbf{P}\frac{E_1}{\partial \mathbf{q}} \qquad (3.26)$$

then we obtain the next-step gradient value as

$$\mathbf{g} = \mathbf{g} + \mathbf{f} \qquad (3.27)$$

so that starting from an initial point the technique finds out the position of the closest intersection point then it moves down the intersection space. The set of all bonds, valence angles, and torsions (if necessary, augmented by out-of-plane bends and linear bends) constitutes a primitive redundant coordinate system [71, 72] In some cases, it may be advantageous to form linear combinations of the primitive redundant internals to form natural or delocalized redundant internal coordinates [73, 74, 75, 76] or symmetry-adapted redundant internal coordinates. For periodic systems such as solids or surfaces, unit cell parameters need to be added [77, 78, 79] (either explicitly or implicitly via coordinates that cross the boundaries of the unit cell). For molecules in nonisotropic media, additional coordinates are needed to specify the orientation of the molecule. For systems containing more than one fragment, additional coordinates are required to specify the positions of the fragments relative to each other. The union of the redundant internal coordinates for the reactants and products is usually a good coordinate system for TS optimization [72]. The transformation of Cartesian coordinates and derivatives to redundant internals is straightforward, but the back transformation of a finite displacement of redundant internals to Cartesian usually is solved iteratively [71, 73, 74, 75, 76, 72, 46, 47, 80, 81, 82, 83, 84, 77, 78, 79]. Most methods for efficient geometry optimization rely on first deriva-

tives of the energy; some also require second derivatives. For most levels of theory used routinely for geometry optimization, the first derivatives can be calculated analytically at a cost comparable to that for the energy. Analytic second derivatives are also available for several levels of theory, but the cost is usually considerably higher than for first derivatives. With the possible exception of optimization of diatomic molecules, derivative-based geometry optimization methods are significantly more efficient than energy-only algorithms. If analytic first derivatives are not available, it is possible to use simplex and pattern search methods,[85, 86, 87] but these become less efficient as the number of degree of freedom increases [88]- Thus, it may be more efficient to compute gradients numerically and to use a gradient-based optimization algorithm than to use an energy-only algorithm. The first and second derivatives of the energy with respect to the geometrical parameters can be used to construct a local quadratic approximation to the potential energy surface:

$$E\left(x\right) = E\left(x_0\right) + gT_0x + \frac{1}{2}xTH_0x \qquad (3.28)$$

where g_0 is the gradient (dE/dx) at x0, H_0 is the Hessian $\left(d^2E/dx^2\right)$ at x_0, and $x = x - x_0$. The gradient and Hessian can be used to confirm the character of minima and TSs. The negative of the gradient is the vector of forces on the atoms in the molecule. Because the forces are zero for minima, TSs, and higher order saddle points, these structures are also termed *stationary points*. The Hessian or matrix of second derivatives of the energy is also known as the force constant matrix. For a function of several variables, the first derivatives with respect to each of the variables form a vector termed the gradient [48]. At a critical point, the first derivative of the potential energy function with respect to all coordinates will be zero (i.e. $\partial E/\partial R_i = 0$, $\forall i$). This is true for all the critical points, so to distinguish them different criteria have to be defined. These criteria arise from a requisition relative to the second derivatives of potential energy surface, defining the Hessian matrix, for a function of several variables. If a second derivate of PES is a negative value it is called imaginary frequency. In classical mechanics, the first derivative of the potential energy for a particle is the opposite of the force on the particle, and the second derivative (for a quadratic potential) is the force constant. Minima occur if the second derivatives (the force constants) with respect to all coordinates are positive, and correspond to the stable products, reactants, or intermediate structures. Graphically they correspond to points localized at the bottom of the valley (fig. 3.1. If there is one or more negative eigenvalues (imaginary frequencies), the point is a $n - th$ saddle point (also called a local maximum), depending on the

number of imaginary frequencies. The first order saddle point (having only one imaginary frequency) corresponds to a transition state, corresponding to a maximum with respect to only one coordinate and a minimum to all the rest of coordinates; higher order saddle points on PES have no chemical interest[3]. At the end conical intersections are critical points characterized from the degenerate value of two electronic state. Actually, there should be defined other two coordinates to guarantee the condition of degeneracy, as it will be widely discussed next.

4 Adiabatic and non-adiabatic molecular dynamics

4.1 Classical molecular dynamics

Molecular Dynamics (MD) is a computational chemistry technique which allows the determination of the behaviors of atomic or molecular systems at a macroscopic level, starting from interactions at the microscopic level. A molecular dynamics simulation is basically characterized by the sampling algorithm in the phase space, and the choice of the interaction potential, V(r), between the particles of the system. The way in which the phase space in sampled distinguishes the various simulative techniques. Classical molecular dynamics extracts trajectories of the atoms (i.e., a set of configurations) integrating the Newtonâs motion equations. Other methods utilize different algorithms to generate the set of configurations, and can be divided into two groups, one that collects the systematic search techniques, while the other includes the random search methods. At the beginning of the informatic era, Alder and Wrainwright carried out the first molecular dynamics simulations [89], studying simple fluids using models which represent atoms as hard spheres. As the potential functions used for the calculations got more complex, thus allowing to obtain simulated data directly comparable with experimental results, it became immediately evident that MD could be used as a powerful and reliable tool for the investigation of the dynamic properties of real system. Over the years, the development of more powerful computers and the appearance of high-level programming languages made possible to simulate always more complex system, from water to small protein. At the end of the 70s, the first simulations of complex biomolecule were performed. The expansion of MD in studies of biochemistry is due to the successful use of it in the reproduction of experimental data of proteins and macromolecules[90, 91, 92, 93, 94]. Nowadays, MD simulations allow to study a broad spectrum of chemical and biological processes, such as structural and functional changes as a result of mutation in proteins, thermodynamic and kinetic properties of various systems[95, 96], investigations and development

47

of new drugs, etc. The exponential growth of the computers power, in conjunction with the development of parallel machines, allowed in a few years to increase the simulation time, the size of the simulated systems and the improvement of the obtained results.

4.1.1 Equation of motion

Classical MD merely describe a system of particles through the motion of their nuclei. In fact, the Born-Oppenheimer approximation allows the separation of the motion of the nuclei from the motion of the electrons, therefore treating them implicitly. Thus, the motion of the molecular system appears to be described as a system of punctual masses that move thanks to an effective potential field produced by the electrons. It is reasonable to not use a quantum treatment: if you exclude the H and He atoms, the motion of the nuclei at room temperature can be described with reasonable accuracy by classical mechanics. In classical MD, the trajectory of a molecular system (i.e., its configurations as a function of time) is generated by the simultaneous integration of the Newton's equations of motion, for all the atoms of the system:

$$\frac{d^2 \mathbf{r}_i}{dt^2} = m_i^{-1} \mathbf{F}_i \tag{4.1}$$

$$\mathbf{F}_i = -\frac{\partial V(\mathbf{r}_1, \ldots, \mathbf{r}_N)}{\partial \mathbf{r}_i} \tag{4.2}$$

The force acting on the i-th atom is $\mathbf{F}_i$, the mass is m_i, and t is time, while $V(\mathbf{r}_1, \ldots, \mathbf{r}_N)$ is the potential energy function, the so-called force field, which represents the forces with which nuclei interact. This function is an effective interaction, since the effect of the (omitted) electronic degrees of freedom is mediated and embedded in the effect of the atomic degrees of freedom expressed in the model.

4.1.2 Molecular dynamics with MM force fields

In MD it is possible to use different force field, mostly depending on the system under investigation. A typical effective potential for systems consisting of N atoms with molecular mass M_i (i=1,2, ..., N) and positive cartesian vectors r_i, has the following form:

$$V(\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_N) = \sum_{bonds} \frac{1}{2} K_b (b - b_{eq})^2 + \sum_{angles} \frac{1}{2} K_\theta (\theta - \theta_{eq})^2$$

$$+ \sum_{dihedrals} K_\phi [1 + cos(n\phi - \delta)] + \sum_{improper\, d.a.} \frac{1}{2} K_\xi (\xi - \xi_{eq})^2$$

$$+ \sum_{VdW} 4\epsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^6 \right] + \sum_{Coul} \frac{q_i q_j}{4\pi\epsilon_0 r_{ij}} \qquad (4.3)$$

The first term is the vibration energy of the covalent bond between two atoms linked by a totally harmonic potential, where b_{eq} is the minimum bond energy, and K_b is the force constant which depends on the bond type; both are derived from experimental data and/or quantum calculations. The second term (three bodies interaction) represents the energy due to the deformation of the valence angle, θ, expressed as a harmonic potential; θ_{eq} is the reference valence angle and K_θ id the force constant. The third and fourth terms represent the four bodies interactions: the third is a sine term that refers to the dihedral angles ϕ; the fourth, ξ, is of harmonic type, and it is used to describe the improper dihedral angles. The last two terms in the force field equation are the effective non-bonding interactions, expressed as van der Waals and Coulomb interactions between the i-th and the j-th atom at distance r$_{ij}$. The van der Waals term is expressed in different ways, depending on the force field. In the example it is expressed as classical Lennard-Jones function (constant ϵ_{ij} and σ_{ij}. The force field parameters can be determined in several ways, the most elegant method is to calibrate these parameters with the results of *ab initio* quantum calculations on small molecular aggregates. However, in this way the results are often not satisfactory, because of the many approximations which must be taken into account with this type of procedure. Alternatively, or in order to optimize a set of parameters obtained from *ab initio* calculations, experimental data can be used (crystal structures, lattice energies and dynamics, spectroscopic data, density, vaporization enthalpy, solvation free energy, NMR data, etc.)[97, 98, 99, 100]. The obtained parameters are valid only in the physical-chemical conditions in which they were determined and this implies that the force fields used in MD programs are specifically optimized for certain physical systems (organic molecules, inorganic molecules, biomolecules, etc.).

4.1.3 Long range interactions

In a MD simulation, most of the computation time is occupied by the calculation of non-bonding interactions. Therefore, it is appropriate to look for a method which allows to increase the computational efficiency of the simulation of the simulation acting on the van der Waals and Coulomb components. Various techniques have been developed to reduce the computational cost. The most commonly used are the cut-off radius and the Ewald summation. The first method severs the non-bonding interactions when the distance between the interacting particles is larger than a specific value, the so called cut-off radius, r_c. Therefore, the i.th particle of the system can interact only with the ones which are contained in a sphere of r_c radius, centered in r_i. Obviously, this method introduces approximations, especially in the electrostatic calculations, which can remarkably affect dynamic and structural properties of the system. To overcome this problem, in addition to increasing the cut-off radius, a switch function can be introduced, which ensures the continuity of dynamic and structural properties beyond the cut-off distance. Despite this method suffers from the aforementioned limitations, it remains among the most common for the calculation of intermolecular forces. The more recent second method proposes the use of a periodic lattice, in which all the interactions between the central box and its infinite replicas are considered. The Coulomb energy in a periodic system formed by N particles is obtained by summing over all the pairs of atoms. One is in the central box, and the other one is in a replica:

$$E = \frac{1}{8\pi\epsilon_0} \sum_{|\mathbf{n}|=0}^{\infty} \left(\sum_{i=1}^{N} \sum_{j=1}^{N} \frac{q_i q_j}{|\mathbf{r}_{ij} + \mathbf{n}|} \right) \tag{4.4}$$

where n are the particles and i, j are the periodic imagines. It can be shown that this summation over n, for potentials varying with r^{-1}, can not be completely convergent. Namely, its limit can vary or diverge if the order of the terms is changed. Therefore, this expression can not be used in the calculation of the electrostatic energy. To solve this problem, each charge can be considered as surrounded by a Gaussian charge distribution ρ^G of equal intensity and opposite sign[101]. This Gaussian distribution will have the effect of shielding the interactions between neighbor charge, and we could consider the interaction energy of short range type. The total charge distribution is:

$$\rho_i(\mathbf{r}) = \rho_i^q(\mathbf{r}) + \rho_i^G(\mathbf{r}) \tag{4.5}$$

where $\rho_i^q(\mathbf{r})$ is the charge distribution of the i-th particle and $\rho_i^G(\mathbf{r})$ is the corresponding Gaussian distribution:

$$\rho_i(\mathbf{r}) = q_i \delta(\mathbf{r} - \mathbf{r}_i) \tag{4.6}$$

After the calculation of the interaction energy in real space, to return to he original distribution function $\rho_i(\mathbf{r})$, a function equal to $\rho_i^G(\mathbf{r})$ must be subtracted in reciprocal space, using the Fourier transformation. The final expression of the total interaction energy will be:

$$
\begin{aligned}
E \;=\; & \frac{1}{8\pi\epsilon_0} \sum_{i=1}^{N} \sum_{j=1}^{N} \left(\sum_{|\mathbf{n}|=0}^{\infty} \frac{q_i q_j \, erfc(\alpha|\mathbf{r}_{ij} + \mathbf{n}|)}{|\mathbf{r}_{ij} + \mathbf{n}|} \right. \\
& + \frac{1}{\pi L^3} \sum_{\mathbf{k}\neq 0} \frac{4\pi^2 q_i q_j}{k^2} \exp(-k^2/4\alpha^2) \cos(\mathbf{k} \cdot \mathbf{r}_{ij}) \bigg) \\
& - \frac{\alpha}{4\pi^{3/2}\epsilon_0} \sum_{i=1}^{N} q_i^2 + \frac{|\sum_{i=1}^{N} q_i \mathbf{r}_i|^2}{2\epsilon_0 L^3 (2\epsilon' + 1)}
\end{aligned}
\tag{4.7}
$$

where $erfc(x)$ is the complementary error function, which tends to zero for $x \to \infty$. With regard to the first term, for α large enough, the sum is reduced to a single term with $\mathbf{n} = 0$; the second term is a sum over the reciprocal vectors $k = 2\pi n/L$. The last two terms are, respectively, the correction function due to the presence of a self-interaction term in $\rho_i^G(\mathbf{r})$, and the contribution of the depolarizing field to the energy, which is affect by the outer dielectric effect.

4.1.4 Boundary conditions

An important feature of the MD simulations is the way in which the boundary conditions are treated. In general, because of the computational limits, a simulated molecular system consists of 10^4-10^5 atoms, significantly less than the corresponding real system. Because of this, the molecules of a simulated system will be more frequently subject to the so-called surface effects. For example, if they are contained in a cubic box, those which are close to the surface will suffer the effects of very different forces with respect those ones present in the bulk. The study of a simulated system of limited an finite size, can lead to the creation of artifacts compared to the actual behavior, especially in the case of homogeneous liquids or solutions. Usually, the periodic boundary conditions (PBC)[101] are introduced to reduce the surface effects. In this way the central box is surrounded by identical

replicas that create a condition of infinite periodicity. During a simulation, when a molecule comes out from the original box, its periodic image enters the box from the opposite side, maintaining the same velocity and direction, and allowing the reproduction of periodic systems, such as crystals or fluids. Interactions within the PBC are treated with the convention of minimum image (MI), where an atom can interact only with the first neighbors. To avoid the anisotropy due to the cubic form of MI and artifacts arising from the interaction of an atom i of the box with an atom j and its periodic image j' simultaneously, a spherical cut-off with radius r_c is used, such that the shorter side of the box is greater than twice the value of cut-off ($L \geq 2r_c$, with L is the box side). This condition is necessary, but not enough, to prevent artifacts. For example, when you want to simulate a solute in a solvent, it is also necessary that a solute atom does not interact with solute atoms belonging to periodic images. To avoid this, it is necessary that the distance of each atom of the solute from each face of the box is greater than the half of cut-off radius ($d \geq 0.5r_c$, with d the distance between solute and box face).

4.1.5 Constraints in MD

In MD simulations different types of constraints can be applied. The most common are those that block the positions of certain atoms, the so-called *Position Constraints*. The use of constraints in MD is necessary to eliminate the high frequency vibrations of bonds, which require short integration steps. In the case of biological macromolecules in solution, the vibrations involving hydrogen atoms determine the higher frequency vibrations. For the O-H bond, the stretching frequency is in the order of 10^{14} Hz, so an average period is of the order of 10 fs[102]. This limits the integration step (time step) that can be used in a simulation to about 0.5 fs, considering that to properly reproduce the trend of a periodic function is necessary to sample at least 20 times per period. With the introduction of a method which binds these bonds, or, in practice, all the covalent bonds, one can arrive at an integration step of 2 fs, that is the commonly used value, also in this work. As the binding vibrations are virtually decoupled from the other system vibrations, this constraints does not alter the rest of the dynamics, while this does not occur for the vibrations of valence angles. The most commonly used method is the so-called SHAKE[103], which was introduced in 1977. The SHAKE method consists in an iterative procedure, for which after each integration step the initial positions of the atoms $\mathbf{r}'_i(t + \Delta t)$ vary until the constraints conditions are not fulfilled in new positions $\mathbf{r}_i(t + \Delta t)$. The use

of SHAKE, despite some disadvantages, allow to increase the integration step of 2-4 times.

4.1.6   Initial conditions

Once the force field, the boundary conditions and the constraints are defined, the initial coordinates and velocities of the system must be set. If the observation time were sufficiently long, the results should be independent from the initial conditions, but since most physical systems reach equilibrium in a very longer time than those simulated, it is appropriate to perform simulations starting from a configuration which is as close as possible to the equilibrium one. The initial velocities can be achieved with a simple Maxwell distribution, which depends on the initial temperature of the simulation:

$$\rho_{x_i} = \sqrt{\frac{m_i}{2\pi k_B T}} exp\left(-\frac{\frac{1}{2}m_i v_{x_i}^2}{k_B T}\right) \tag{4.8}$$

where ρ_{x_i} is the probability density of the component v_{x_i}.

4.1.7   Integration of the equation of motion

Given the complexity of the force field expression (4.3), the integration of the equations of motion (in order to obtain the trajectory of the N atoms of the system) does not have an analytical solution. For this reason, several approximate methods have been developed. They numerically extract the trajectories from an MD simulation. The characteristics of a good integration algorithm can be summarized as follows:

- it must allow the use of a long time step: it is evident that the higher the integration step, the smaller the number of integrations required in a simulation. Also, if the Δt is too large, the risk is to fail to follow certain components of the motion of the system, such as the vibrations, thus leading to misleading results. Therefore, it is appropriate to adjust the integration step on the higher frequency motions;

- it must be reversible in time and preserve constant the quantities in the system. Newton's equations are reversible in time and therefore it is desirable that the algorithm used faithfully reproduce this data;

The algorithms differ in the accuracy of calculation, the memory occupation and the type of representation. The nature of the algorithms derives

from the expansion of the Taylor series of the quantities $\mathbf{r}_i(t + \Delta t)$ and $\mathbf{v}_i(t + \Delta t)$. In MD, algorithms that reach up to the third order of the Taylor expansion are generally used, as algorithms of lower order are inaccurate, while those of higher order are too expensive from a computational point of view. Tipically, the Verlet algorithm [104, 105] is used. The algorithm is fast, requires little memory and has a good conservation of energy for short times. It is obtained by additions and subtractions of Taylor's expansions of the coordinates $\mathbf{r}_i(t + \Delta t)$ and $\mathbf{r}_i(t - \Delta t)$ of particles with mass m_i:

$$\mathbf{r}_i(t + \Delta t) = -\mathbf{r}_i(t - \Delta t) + 2\mathbf{r}_i(t) + \frac{(\Delta t^2)}{m_i}\mathbf{F}_i(t) \tag{4.9}$$

where $\mathbf{F}_i(t)$ is the force exerted on the particle i at time t and where terms higher than third order ones have been neglected. This pattern of integration is therefore correct to fourth order in Δt. The velocities are given by:

$$\mathbf{v}_i(t) = \dot{\mathbf{r}}_i(t) = \frac{1}{2\Delta t}[\mathbf{r}_i(t + \Delta t) - \mathbf{r}_i(t - \Delta t)] \tag{4.10}$$

The error on velocity is on the $(\Delta t)^3$ order.

4.1.8   Temperature and pressure control methods

Different types of statistical ensembles can be used in MD, and they are characterized according to thermodynamic changes, restricted or let free to equilibrate with the external environment:

- micro canonical ensemble (NVE), characterized by constant numbers of particles (N), volume (V) and total energy (E), i.e. an isolated system;

- canonical ensemble (NVT), in which the internal energy of the system can vary, while the temperature is held constant by heat exchange with an external thermal bath, thus an isothermal and isochore system ;

- NPT ensemble, in which the volume can vary in order to maintain constant pressure, equal to the outside pressure, thus an isothermal and isobaric system.

In order to compare the results of an MD simulation with experimental data, the simulated system must be put in the same conditions in which these data were obtained, and this often forces to run the simulation at constant

pressure and temperature. Over the years, various algorithms have been developed.

4.1.9   Structural analysis: RMSD

One of the first analysis that generally are carried out is the calculation of the Root Mean Square Deviation (RMSD) compared to a reference structure in function of time. The purpose of this analysis is to obtain a first qualitative information on the stability of the simulation and the achievement of the equilibrium. It can be expressed through the following relation:

$$RMSD(t) = \left[\frac{1}{N} \sum_{i=1}^{N} (\mathbf{r}_i(t) - \mathbf{r}_i^{ref})^2 \right]^{\frac{1}{2}} \tag{4.11}$$

where $\mathbf{r}_i(t)$ represents the position of the i-th atom at time t, $\mathbf{r}_i^{ref}$ represents its position in the reference structure and N represents the total number of the generated structures.

4.2   Excited state non-adiabatic molecular dynamics

Molecular dynamic simulations have become an established technique for the understanding of fundamental processes that occur on the ultrafast, femtosecond time scale, where simulations are required to help interpret the data. The approach based on molecular dynamics (MD) QM/MM simulations is especially suitable for complex systems, and it allows accounting for finite temperature effects. When the forces are computed from a QM potential, the procedure is known as *ab initio molecular dynamics* (AIMD). Assuming the Born–Oppenheimer approximation valid, the forces may be computed after optimizing the wave function at each step during the dynamics. In the field of molecular modelling of complex systems, MD QM/MM simulations are aimed at describing biochemical processes with realistic model systems and at following their time-evolution at finite temperature. The QM/MM approach allows the system of interest to be investigated by taking fully into account environmental effects. Unfortunately, the time-scale which can be explored by QM/MM MD simulations is limited by the costly evaluation of forces from electronic structure calculations (the QM part of the QM/MM potential). Many popular quantum chemical programs now include the possibility of performing QM/MM calculations [42, 106, 107,

108]. Because of the need to iterate the equations of motion several thousand times, highly optimized coding is required to tackle QM/MM-based MD simulations in an efficient way. The COBRAMM program package is designed to conduct simulations of large systems in the photochemical field.

4.2.1   Excited-state dynamics

Excited-state dynamics is the branch of the theoretical and physical chemistry devoted to simulating molecular processes triggered by UV-visible light absorption. This involves the implementation of nuclear dynamics methods to determine the time evolution of the molecular geometry used, along with electronic structure methods capable of computing electronic excited-state potential energy surfaces. Applications to such techniques include photochemistry and electronic spectroscopy. The goal of a computational approach to the simulation of photo-induced processes is the complete description of the mechanisms at the molecular level from the promotion to the excited electronic state to the formation of products or regeneration of reactants back in the electronic ground state. Following light absorption, several photophysical or photochemical deactivation mechanisms can compete depending on their relative efficiencies. They are summarized in figure 4.1, which depicts a typical Jablonski diagram. The excess energy can be released through light emission (i.e. fluorescence from a singlet or phosphorescence from a triplet) or through radiationless decay processes (i.e. an internal conversion take place between singlets or intersystem crossing between a singlet and a triplet) that convert it into vibrational energy in S0. The spectral position, shape, and width of the absorption or emission spectra are influenced by the competing events that dominate the short-timescale dynamics of the molecule (changes in the molecular geometry through nuclear motion coupled to changes in the electronic density through transfer of electronic population). Moreover, a photochemical reaction (photo-isomerization or photodissociation) may occur during the deactivation time, which leads to the formation of photoproducts. However, the Jablonski diagram does not provide any description of the course of a photochemical reaction at the microscopic level [109, 110, 111]. From a computational chemistry standpoint, the concept of intrinsic reaction coordinate (IRC) [112, 51] gives a stationary/average description of the most significant molecular geometries encountered along the reaction path from reactants to products. In practice, an IRC is derived as a classical trajectory for the motion of the nuclei whereby the kinetic energy is removed at each step. This requires knowledge of the local force field that drives the motion of the nu-
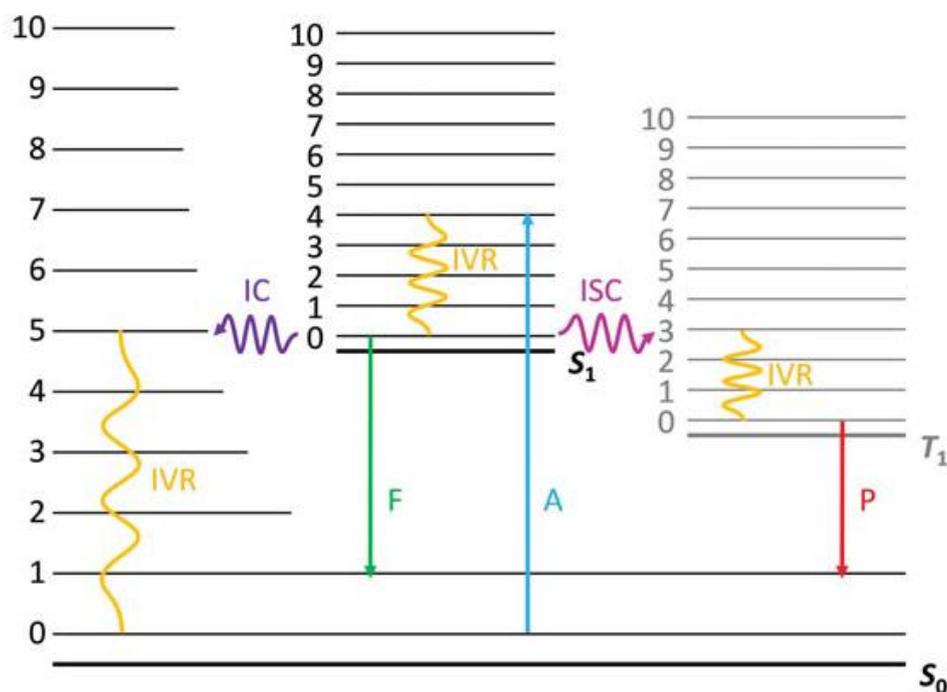
Figure 4.1: Jablonski diagram (A: absorption; F: fluorescence; P: phosphorescence; IVR: intramolecular vibrational redistribution; IC: internal conversion; ISC: intersystem crossing). The numbers from 0 to 10 indicate the quantum number of the photoactive vibrational mode in each of the three electronic states involved.

clei, that is the negative gradient of the potential energy field. According to the Born-Oppenheimer approximation, the electronic and nuclear motions are separated and the nuclear potential energy is given as the electronic adiabatic mean field calculated at each relative arrangement of the nuclei. Such energy is the electronic energy obtained by solving the time-independent electronic Schroedinger equation with quantum chemistry (electronic structure) methods. Two kinds of photochemical reactions must be identified: adiabatic ones and non-adiabatic ones. An adiabatic reaction corresponds to a reaction path on the same electronic excited state. Subsequent excited photoproducts eventually return to their electronic ground state after light emission. On the contrary, a non-adiabatic reaction returns hot products in the electronic ground state directly, with a reaction path that jumps from the excited state to the ground state without emitting light. Figure 4.2 illustrates adiabatic and non-adiabatic reaction paths on two potential energy surfaces. Describing a thermal chemical reaction with an IRC that connects reactants and products through a transition state is a well-established com-

putational chemistry technique. Following the bottom of a potential energy valley is significant because the system has time to regain equilibrium and redistribute the nuclear kinetic energy or transfer it to the surrounding environment. On the contrary, most photochemical reactions are ultrafast, they start from a highly excited initial condition and involve effects for the description of which it is essential to take into account explicitly the nuclear kinetic energy. Indeed, the aim of molecular dynamics simulations is to work out the time evolution of the molecular geometry following the system excitation. Ground-state dynamic simulations describe the nuclear motion
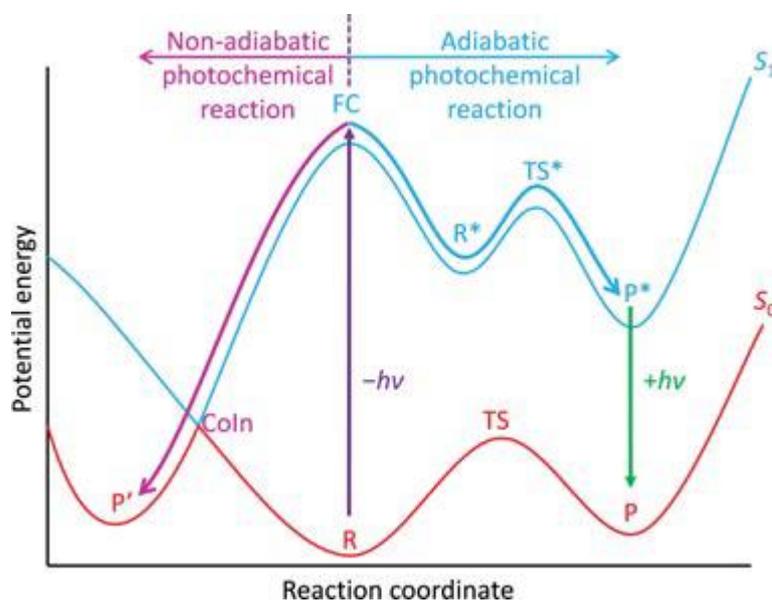
Figure 4.2: One-dimensional representation of nonadiabatic and adiabatic reaction paths involving two potential energy surfaces (R: reactant; TS: transition state; P: product; CoIn: conical intersection).

on the potential energy surface that corresponds to the adiabatic electronic ground state [113, 114, 115]. From an operational point of view, excited-state dynamics and ground-state dynamics differ only when several coupled electronic states must be considered simultaneously and can transfer population, that is when the Born–Oppenheimer approximation does not hold. Thus, simulations of adiabatic photochemical reactions, or computations of absorption/emission spectra corresponding to electronic transitions between a pair of noninteracting adiabatic electronic states, use similar techniques to ground-state dynamics. In contrast, simulating radiationless decay processes requires a non-Born-Oppenheimer description according to which the nuclear motion on several potential energy surfaces accounts for the transfer

of electronic population through vibronic interactions called non-adiabatic couplings. Non-adiabatic couplings were studied by Zener and Teller. The latter, when extending the 1932 Zener paper found that non-adiabatic couplings around molecular geometries where two electronic states are degenerate are characterized by a strong magnitude. Moreover, Teller demonstrated the possible existence of real crossings (to be later known as conical intersections) between potential energy surfaces in polyatomic molecules. Zimmerman [116] and Michl [117] were the first to suggest, independently, that internal conversion occurring at a conical intersection was crucial for the understanding of certain photochemical mechanisms in organic molecules. In order to account for the effect of non-adiabatic couplings, a quantum mechanical treatment of the time evolution of the molecular geometry (relative nuclear positions) and of the electronic populations (occupancies of the electronic states) are mandatory [118, 119, 120, 121, 122, 123, 124, 125, 126]. This is achieved by solving the time-dependent Schroedinger equation for the molecular Hamiltonian in order to propagate the molecular wavepacket. Starting from a given initial condition this process returns the probability of being in a given molecular geometry in a given electronic state at a particular time (see Figure 4.3). This is the domain of quantum dynamics (QD). Quantum dynamics methods are established techniques. Their main computational burden is working out accurate potential energy surfaces and their mutual couplings in non-adiabatic systems. More rough treatments such as semiclassical dynamics are aimed at more or less the same but are based on the solution of Newton classical equations of motion to drive a statistically meaningful set of trajectories (see Figure 4.3) while accounting for possible electronic population transfers through various approximations. Simulation techniques can be subdivided into two groups: grid-based methods (for QD) and trajectory-based methods (for semiclassical dynamics). In traditional methods, a global representation of the potential functions and non-adiabatic couplings must first be computed, often by means of data fitting to mathematical models. Trajectory-based methods also produce direct dynamics, characterized by quantum chemistry calculations which are executed along trajectories in order to return a moving local representation of the surfaces. The two main drawbacks for the techniques of the two groups are the high computational complexity which is exponentially increasing with the size of the system in the case of grid-based methods and the poor accuracy of the trajectory-based methods. In the present PhD work, the second approach based on semiclassical QM/MM trajectories has been employed and will be documented in details below.
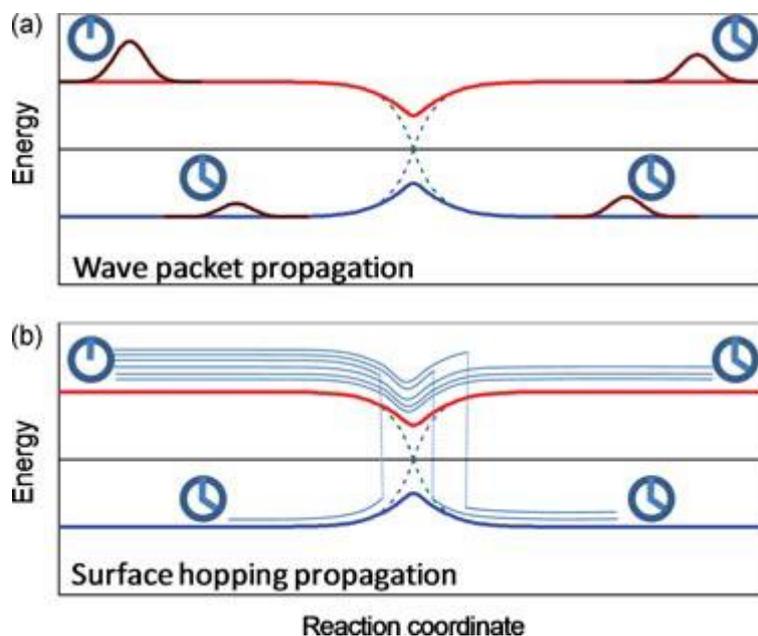
Figure 4.3: Adiabatic (solid) and diabatic (dashed) potential energy curves for the ground and first excited states along the reaction coordinate of a hypothetical molecule, with an avoided crossing on the middle way. (a) Wave packet propagation. (b) Trajectory surface hopping trajectories.

4.2.2 Non-adiabatic dynamics with trajectory surface hopping methods

The trajectory surface hopping (TSH) method is a general methodology for dynamics propagation of nonadiabatic systems. It is based on the hypothesis that the time evolution of a wave packet through a potential-energy branching region can be approximated by a set of independent semiclassical trajectories stochastically distributed among the branched surfaces. As it was proposed about 40 years ago, the TSH methodology has become one of the main techniques for non-adiabatic dynamics propagation in molecular physics and chemistry. One reason for this success lies on its intuitive conceptual background together with its high computational efficiency when compared to full quantum mechanical propagation.

General concepts

In the TSH method, the molecular evolution is approached by a mixed quantum-classical treatment, which separates the adiabatic and the non-adiabatic processes. The adiabatic dynamics of the nuclei is propagated in

a classical fashion on a single Born-Oppenheimer surface at any given time, whereas the branching of the population due to non-adiabatic effects is introduced by a stochastic algorithm allowing the exchange of electronic state during the dynamics propagation. This partition gives rise to a semiclassical trajectory always propagated on a single surface, but with the possibility of exchanging the surface. The statistical character of the wave packet propagation is recovered by preparing an ensemble of such semiclassical trajectories. So formulated, the TSH approach owes much of its success to its intuitive conceptual background, practicability, and low computational costs as compared to full quantum wave packet propagation. The advantages of the method can be summarized as follows:

- Simple propagation of Newton equations for the nuclei, usually effected in Cartesian coordinates.

- The local character of the method does not require building multidimensional potential energy surfaces.

- The local character allows straightforward on the fly implementations, where energies, energy gradients, and nonadiabatic couplings are computed along and simultaneously to the trajectory only for the classical positions.

- The evolution of the dynamics can include all the nuclear degrees of freedom.

- The connection to different quantum chemical methods, including QM/MM partitioning, is straightforward, provided these methods can provide excited-state energies, energy gradients (preferentially by analytical methods), and nonadiabatic couplings.

- The independent trajectory character allows for straightforward parallelization of the computational work.

The formulation of the TSH, as proposed by Tully and Preston [127] is carried out according to the semiclassical approach of Nikitin. First, it is assumed that the nuclei evolve along a general trajectory $\mathbf{R}^c(t)$. Such trajectory will be identified as a classical trajectory. Then, the time dependent wave function for the electrons is written in an electronic basis $(\{\Phi_k\}|\langle\Phi_k|\Phi_l\rangle_r = \delta_{kl})$ as:

$$\varphi(\mathbf{r}, \mathbf{R}^c, t) = \sum_j c_j(t)\Phi_j(\mathbf{r}; \mathbf{R}^c(t)), \tag{4.12}$$

where r identifies the electronic coordinates. By substituting such assumption in the time-dependent electronic Schroedinger equation

$$(i\hbar\frac{\partial}{\partial t} - H_e)\varphi(\mathbf{r}, \mathbf{R}, t) = 0 \tag{4.13}$$

it is possible to work out the following set of differential equations for the coefficients c (t):

$$i\hbar\frac{dc_k}{d_t} + \sum_j (-H_{kj}^c + i\hbar\mathbf{F}_{kj}^c \cdot \mathbf{v}^c)c_j = 0 \tag{4.14}$$

where H_{kj}^c are the matrix elements $\langle\Phi_k|H_e|\Phi_j\rangle_r$, and $\mathbf{F}_{kj}^c$ is the non-adiabatic coupling vector between the two states k and j. Given an atom m, $\mathbf{F}_{kj}^c$ can be expressed as:

$$\mathbf{F}_{kj}^{c,m} \equiv \langle\Phi_k|\nabla_{\mathbf{R}_m}|\Phi_j\rangle_r. \tag{4.15}$$

For the derivation of Eq. 4.14 , which is referred to as semiclassical time-dependent Schroedinger equation ($SC-TDSE$), the following relation holds:

$$\langle\Phi_k|\frac{\partial}{\partial t}|\Phi_j\rangle_r = \mathbf{F}_{kj}^c \cdot \mathbf{v}^c \tag{4.16}$$

where $\mathbf{v}^c$ is the nuclear velocity vector, the superscript c indicating that the quantity is evaluated at a fixed nuclei position. If either an adiabatic representation ($\Phi_k^a|H_{kj} = V_k\delta_{ki}$) or a diabatic representation ($\Phi_k^d|H_{kj} = W_{kj}, \mathbf{F}_{kj} = 0$) is adopted then Eq. 4.14 can be further simplified. In TSH, the nuclear motion is described by Newton equations propagated in a single electronic state. Given the coordinates of nucleus m with mass Mm , such equations can be expressed as:

$$\frac{d^2\mathbf{R}_m^c}{dt^2} - \frac{f_m^c}{M_m} = 0, \tag{4.17}$$

where the force is proportional to the gradient of the potential energy:

$$f_m^c = -\nabla_{\mathbf{R}_m}H_{ll}^c \tag{4.18}$$

Moreover, when integrating the classical equations, the transition probabilities between the current state l and any of other state j are evaluated. Then, a probabilistic algorithm is exploited to determine the state where the dynamics is going to proceed. The propagation of the $SC - TDSE$ is usually done in adiabatic representation, which provides better results if compared to those attained with a diabatic representation [128] and requires a reasonable computational effort. However, local diabatization methods in

the intermediary steps of the propagation of the $SC-TDSE$ yield very stable results as well. The way the transition probability is computed is the leading distinguishing feature among the various TSH models. The simplest hopping algorithms just assume that the probability is one if the energy gap between two states is smaller than some predefined energy threshold [129]. More refined algorithms take into account variations of wave function coefficients as a measurement of the non-adiabatic coupling [130]. Most of the early applications of the TSH approach were based on Landau-Zener or related transition probability formulae [131, 132, 133]. There are several techniques that allow computing the transition probabilities for the TSH approach. The most popular one is the fewest-switches method proposed by Tully, which minimizes the number of hopping events per Δt time step. According to this condition, the probability of hopping between the states l and k is:

$$P_{l \to k} = \frac{Population\ increment\ in\ k\ due\ to\ flux\ from\ l\ during\ \delta t}{Population\ of\ l} \quad (4.19)$$

The population of each electronic state k is expressed based on the coefficients $c_k(t)$ by the diagonal elements of the density matrix defined as:

$$\rho_{lk}(t) = c_l c_k^* \quad (4.20)$$

On the basis of equation 4.14 the hopping probability can be expressed as:

$$P_{l \to k} = max \left[0, \frac{2\Delta t}{\rho_{ll}} (\hbar^{-1} Im(\rho_{kl}) H_{lk}^c - Re(\rho_{kl}) \mathbf{F}_{kl}^c \cdot v^c) \right] \quad (4.21)$$

A significant difference does exist between the fewest-switches approach and the Landau-Zener one, in that while the latter returns the total non-adiabatic event probability and as such it must be applied only once every time the crossing region is approached, the former returns an instantaneous probability value for the non-adiabatic event, therefore it should be applied at every step of the simulation along the trajectory. In the case of a hopping event, the conditions that have to be simultaneously met for a transition from surface l to surface k to take place at time t are that a uniformly selected random number r_t in the [0,1] interval satisfies the following condition:

$$\sum_{n=1}^{k-1} P_{l \to n}(t) < r_t \leq \sum_{n=1}^{k} P_{l \to n}(t); \quad (4.22)$$

and that the energy gap between the two states is:

$$V_k(\mathbf{R}^c(t)) - V_l \mathbf{R}^c(t)) \leq \frac{(\sum_m^{N_{at}} \mathbf{v}_m^c \cdot \mathbf{F}_{kl}^{c,m})^2}{2 \sum_m^{N_{at}} M_m^{-1} (\mathbf{F}_{kc}^{c,m})^2} \quad (4.23)$$

The meaning of the latter condition is that after a hop energy must have not increased. Such relation has been obtained assuming that total energy has conserved, and that a correction term has been added to ensure such purpose. Such correction term is a linear momentum in the direction of $F_{kl}^{c,m}$. In the cases where F is not explicitly computed, the correction can be applied along the direction of $\mathbf{v}_c$, thus obtaining the following condition:

$$V_k(\mathbf{R}^c(t)) - V_l(\mathbf{R}^c(t)) \leq E_{kin}(\mathbf{v}^c). \tag{4.24}$$

Instead, in the case where only the first condition is met, the hopping is said to be *frustrated*. Decoherence issues that may jeopardize self-consistency in TSH can occur because the $SC - TDSE$ is propagated along a single trajectory $\mathbf{R}^c$ which is determined by the gradients for a specific electronic state l. Such issues are due to the effect of the amplitudes of the remaining states which are artificially restricted to propagating in the same direction as the SC-TDSE. A correction can be effected by means of the *nonlinear decay of mixing* model [134]. Such correction is applied at each time step by expressing the solutions of equation 4.14 in the form [135]:

$$
\begin{aligned}
c_k' &= c_k exp(-\tfrac{\Delta t}{\tau_{kl}}) \quad \forall k \neq l \\
c_l' &= c_l \left[\tfrac{1 - \sum_{k \neq l} |c_k'|}{|c_l|^2} \right]^{\frac{1}{2}} \\
\tau_{kl} &= \tfrac{\hbar}{|V_{kk} - V_{ll}|} \left(1 + \tfrac{\alpha}{E_{kin}} \right)
\end{aligned}
\tag{4.25}
$$

where l is the current state, E_{kin} is the nuclear kinetic energy, Δ_t is the integration interval, α is an empirical parameter whose recommended value is 0.1 Hartree. Ideally, the TSH method should be self-consistent, i.e. for statistically converged ensembles the occupation and the average population should coincide. Actually, self-consistency is jeopardized by the missing decoherence, and it can be restored with the mentioned correction. Recent studies have demonstrated that occupation computed by means of the TSH technique implementing the fewest switches algorithm in an adiabatic representation approximately satisfies the detailed balance condition, which means that the trajectories are distributed among the quantum states according to the Boltzmann quantum populations.

Application in molecular photochemistry and photophysics

The largest field of applications of TSH method so far has been investigations of photoinduced physical and chemical processes in molecules. The basic concept is illustrated in Figure 4.4. The molecule is initially considered in its ground state when it is photoexcited. After the photo-absorption,

the molecule relaxes through the excited states (only one in the figure) and can return to the initial conformation in the ground state (trajectory T1) or form a new chemical product (trajectory T2). The scheme in Figure 4.2 fails to capture the intrinsically multidimensional character of this process. Far from following a single reaction coordinate, the non-adiabatic dynamics of molecules, as revealed by TSH simulations, deeply depends on collective motions of all nuclear degrees of freedom. A main class of such investigations involves photoinduced cis-trans isomerization of molecules [136, 137]. These investigations have also revealed details of the photophysical processes, including the possibility of activation of many different reaction pathways [138, 139]. Due to their importance in biology, the non-adiabatic dynamics of nucleic acid fragments [140, 141] and of retinal [142, 143], the chromophore of the visual protein Rhodopsin, have been widely studied in this PhD work.
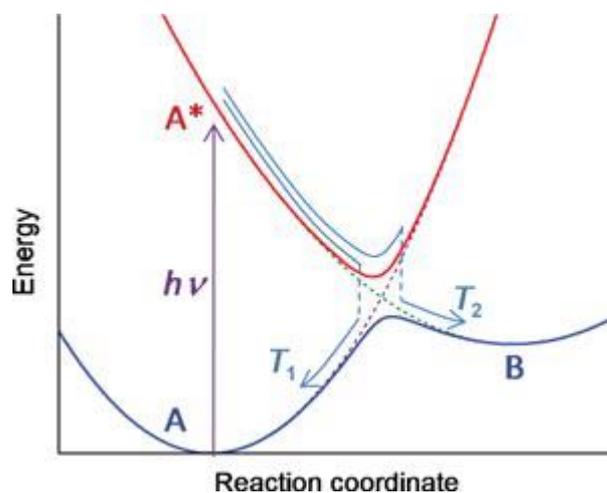

Figure 4.4: Adiabatic (solid) and diabatic (dashed) potential energy curves for the ground and first excited states along the isomerization reaction coordinate of a hypothetical molecule A. Two trajectories (T1 and T2) hopping near the crossing are illustrated. Trajectory T1 is a photophysical process ending in A. Trajectory T2 leads to isomerization into the photochemical product B.

5 Solvated nucleobasis: the guanine monophosphate case

In this chapter the photophysics and photochemistry of water-solvated guanine monophosphate (GMP) is characterized in order to elucidate the main photo-processes occurring upon UV-light irradiation. The effect of the solvent and of the phosphate group on the energetic and structural features of this system are evaluated for the first time employing high-level *ab initio* methods and they are here thoroughly compared to those *in vacuo* previously reported in literature and to the experimental evidence to assess the extent to which they influence the photoinduced mechanisms. Solvated electronic excitation energies of solvated GMP at the Franck-Condon (FC) region show a red shift for the $\pi\pi^* L_a$ and L_b states, whereas the energy of the oxygen lone-pair $n\pi^*$ state is blue-shifted. The main photoinduced decay route is promoted through a ring-puckering motion along the bright lowest-lying L_a state toward a conical intersection (CI) with the ground state, involving a very shallow stationary point along the minimum energy pathway in contrast to the barrierless profile found in gas-phase, the point being placed at the end of the minimum energy path (MEP) thus endorsing its ultrafast deactivation in accordance with time-resolved transient and photo-electron spectroscopy experiments. The role of the $n\pi^*$ state in the solvated system is severely diminished as the crossings with the initially populated L_a state and also with the L_b state are placed too high energetically to partake prominently in the deactivation photo-process. The proposed mechanism present in solvated and *in vacuo* DNA/RNA chromophores validates the intrinsic photostability mechanism through CI-mediated non-radiative processes accompanying the bright excited-state population toward the ground state and subsequent relaxation back to the FC region.

5.1 Introduction

A proper characterization of the monomeric chromophores is essential to understand the photoinduced events occurring in DNA/RNA from a bottom-up

approach that can yield definitive answers regarding their role in photosta-
bility and photodamage [22]. This line of research embodies the main efforts
carried out by the research community over the last two decades , and even
though plenty of advances have been made there is still no definitive con-
sensus regarding the main deactivation routes present in the nucleobases .
Several decay paths have been identified both *in vacuo* and in solution and
assigned to a number of processes, ranging from a barrierless deactivation
through a ring-puckering motion of the bright $\pi\pi^*$ state to a long-lived sig-
nal arising due to a possible crossing with a dark $n\pi\ddot{I}^*$ state [144], as well as
an even longer- lived pathway widely attributed to the role of triplet states,
specially prominent in non-canonical nucleobases with heavy-atom substi-
tutions [145, 146, 147, 148]. In this study we focus our attention to water-
solvated guanine monophosphate (GMP), one of the canonical DNA/RNA
nucleobases. GMP has also been recently proposed to be an interesting
compound for nanotechnological applications due to its outstanding self-
assembly capability [23], especially in its quadruplex form, which has even
been located in DNA/RNA chains [24, 149]. Whereas, guanine *in vacuo*
has been extensively studied at a high-level multireference *ab initio* level
[25, 140, 26] as well as experimentally [150, 151], a lesser degree of scrutiny,
mainly at a density functional theory level [152, 153, 154, 155, 156], has been
considered on the effect of the phosphate group and the solvent on the pho-
toinduced processes occurring in this compound [157]. In the present work
we propose a theoretical assessment of the deactivation routes embody- ing
the main photophysical and photochemical features of GMP by employing
high-level *ab initio* multireference perturbation theory methods coupled with
a quantum-mechanical/molecular mechanics (QM/MM) approach, in order
to ascertain the role of the environmental perturbations in these type of sys-
tems as they remain relatively unknown [158, 159, 160]. Further knowledge
on the environmental effects affecting the photo-processes occurring in the
DNA/RNA chromophores upon UV-light irradiation will provide essential
information that can be properly translated to water-solvated DNA/RNA
systems such as those found in cells.

5.2   Computational details

5.2.1   MM dynamics and sampling

MM simulations were performed for GMP in water to obtain a represen-
tative starting geometry to be employed for all subsequent computations
and analyses. The MM dynamics calculations were carried out using the

Amber-11 suite of programs exploiting the Parm99 force field [46, 161]. Initially, a cubic solvent box comprising 700 water molecules of explicit TIP3P [97] with one Na^+ counterion were considered. The entire system was then heated from 0 to 300 K for 1 ns at constant volume and constant pressure (1 atm), and then finally a simulation run for 100 ns recording the values time-evolution (hereafter named *snapshots*) every 200 fs was performed. In order to select the initial geometry we performed a cluster analysis based on a Root Mean Square (RMS) coordinate deviation analysis on the guanine moiety over all snapshots recorded along the MM dynamics run within a $2.0\overset{\circ}{A}$ difference using the MMTSB toolbox. We obtained three different clusters, denoted a, b, and c in 5.1, with populations relating to the 93, 6, and 1% of the total number of structures obtained along the dynamics run, respectively. The selected initial structure was therefore chosen as the snapshot with the closest geometrical parameters to the centroid of the average structures obtained in cluster A, being the most representative.

5.2.2 QM/MM calculations

QM/MM calculations were performed using the COBRAMM interface developed in our group [48, 162]. The boundary between the QM and MM regions has been set so that all the guanine atoms are included in the QM region, placing the link atom between the N9 of guanine and the carbon of the sugar ring directly attached to the nucleobase (see 5.2). Such choice of the boundary between the QM and MM regions is justified in terms of charge redistribution on the nucleobase and its π-system in order to better reflect the covalent link between guanine and the monophosphate group. A three-layer approach (high, medium, and low) was used throughout ([48]): guanine was included in the QM region (high layer, ball, and stick representation in figure 5.2), whereas the medium layer comprises the movable MM atoms within a 10 $\overset{\circ}{A}$ radius surrounding the GMP moiety, the remainder of the MM system being kept frozen during all optimization procedures in the low layer. Equilibrium geometries and photoreaction paths [163] were determined by using fully unconstrained optimizations and minimum energy path (MEP) computations on the relevant potential energy hypersurfaces by employing the intrinsic reaction coordinate (IRC) and optimization algorithms as implemented in the Gaussian 09 program package [164] interfaced with COBRAMM. CI optimizations were performed with the gradient projection algorithm of Bearpark et al. (1994) [165] as implemented in COBRAMM at the QM/MM level. Further details can be found in [166]. Energies and gradients in the QM region were computed making use of the complete active

space self-consistent field (CASSCF) and complete active space second-order perturbation theory (CASPT2) methods [167, 168], as implemented in the MOLPRO- 2010 [169, 170] and MOLCAS-7 [171, 172] suite of programs, respectively. All gradient and non-adiabatic coupling calculations have been performed with the CASSCF implementation of the MOLPRO-2010 code. Subsequent CASPT2 calculations on the key structures obtained along the optimizations and MEPs have been carried out as implemented in the MOLCAS package in order to correct the energy values due to the lack of dynamic correlation present in the CASSCF method. This procedure is referred to as the CASSCF/CASPT2 protocol, and has been successfully employed over the years to tackle a variety of photoinduced phenomena from a theoretical standpoint [163, 173, 174, 175]. The active space selected comprised the full π space with the exception of the lowest occupied π orbital plus the three lone-pair orbitals to provide a proper description of the $n\pi^*$ states, thus making 18 electrons in 13 orbitals. The removal of the lowest occupied π orbital is carried out due to its occupation number being very close to two and given that its removal does not affect the energy values obtained while speeding up the computations. An imaginary level shift of 0.2 a.u. was employed in the perturbation step to avoid intruder states. Two different basis sets were employed throughout: $6 - 31G^*$ was used in order to map the hypersurfaces at the CASSCF level whereas atomic natural orbital (ANO) type [176] basis set with the large (ANO- L) primitive set C,N,O(14s9p4d3f) / H(8s4p3d) contracted to O[3s2p1d]/C[3s2p1d]/N[3s2p1d]/H[2s1p] (ANO-L 321/21 hereafter) was employed in order to refine the single-point CASPT2 energies.

5.3   Results and discussion

The results are subdivided into three different sections: first, the main geometrical parameters of the ground-state Franck-Condon minimum of 9H-guanine both *in vacuo* and in solvated GMP are presented, together with its corresponding vertical spectra, drawing some conclusions on the influence the solvent and phosphate group have on the relative position of the excited states.

Next, the different minima and conical intersections describing the photoprocess are presented. Finally, a rationale of the photoinduced mechanisms in GMP is drawn in conjunction with prev ous results and experimental data acquired from literature, yielding concluding notes on the photophysical and photochemical decay channels featuring in GMP and leading to its intrinsic photostability.
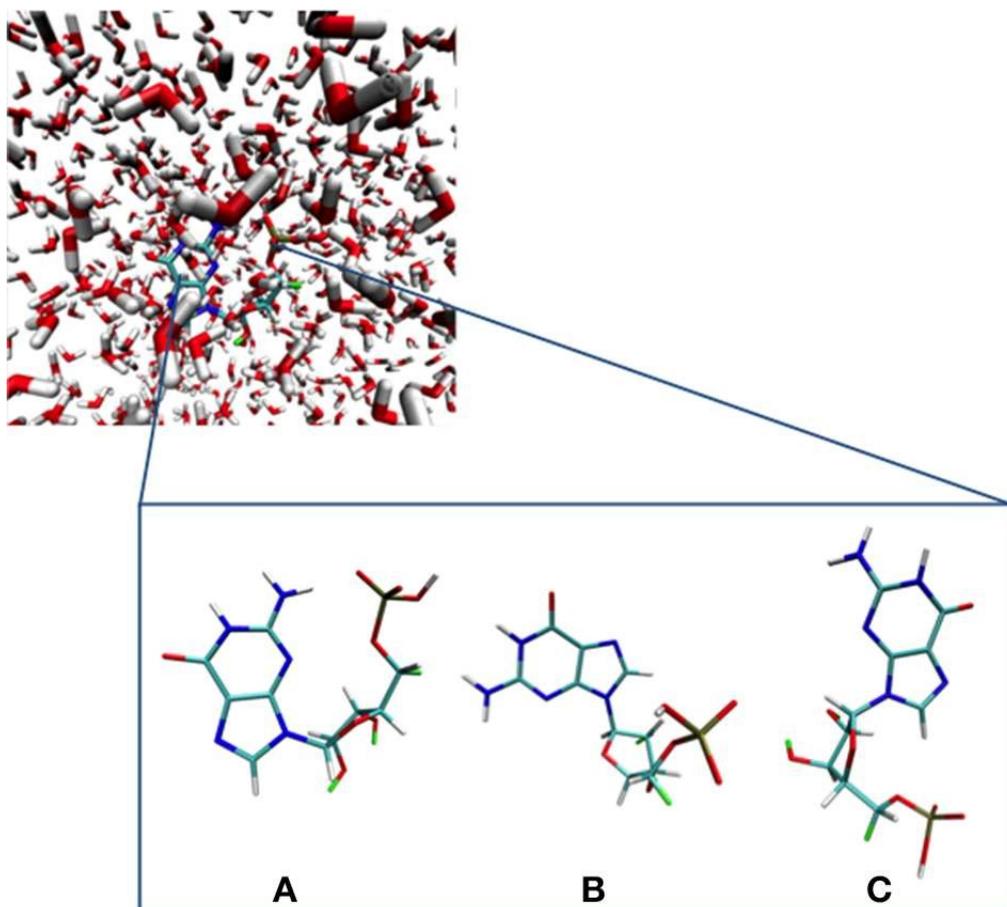
Figure 5.1: GMP and its different main conformations along the MM dynamics run obtained through a RMS deviation cluster analysis. (AâC) depict the most important conformations extracted from the molecular dynamics simulation (see text).

5.3.1 UV absorption at the Franck-Condon region

The optimized FC structure of GMP in solution (5.3) shows a strong resemblance with the *in vacuo* 9H-Guanine CASSCF/6- $31G^*$ structure previously reported in literature [25]. The table reported in figure 5.4 presents the main geometrical parameters of both structures. As can be seen, bond lengths and dihedral angles are analogous in giving rise to a planar structure. Small solvation effects are noticed in the GS structure mainly due to its relatively small dipole moment, providing slight shortenings in the C6-N1 bond distance in the presence of the solvent. A very shallow stationary point $(L_a)_{sp}$, which is not present in the gas phase, has been located at the end of the L_a MEP featuring a pronounced elongation of the C2-N3 and a shortening of the N3-C4 distances, similar to those featured by the conical intersections
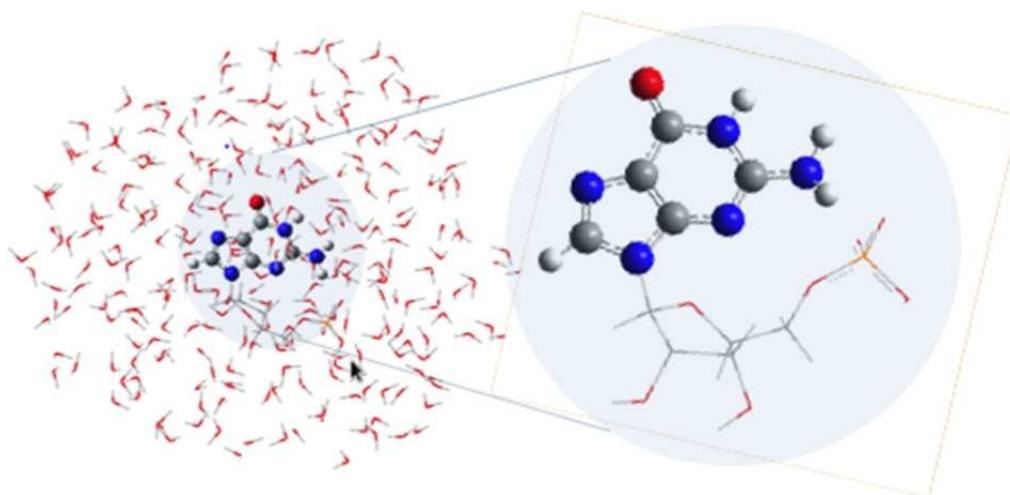
Figure 5.2: GMP and the three different QM/MM partitions defined in the present study. The high layer treating the guanine moiety at the QM level (ball and stick representation), the medium layer encompassing the movable MM water molecules and the phosphate group enclosed within a 10 Åradius distance from GMP, and the low layer containing the remainder of the MM region that is kept frozen throughout the calculations (see text).

between the L_a and GS states both *in vacuo* and in solution. The two CIs located between the L_a and GS states do show pronounced differences due to the solvent, yielding large elongations in the C2-N3 bond compared to its *in vacuo* counterparts. This is mainly due to the large dipole moment featured by the L_a state, which makes it a more influenced state upon solvation. A similar behavior can be seen in the CIs between the L_a and L_b states, inducing large deviations in the C4-C5, C5-C6 and especially pronounced in the C6-O distance, where water-solvated GMP experiences an elongation of 0.15 Å. The $n_O\pi^*$ state minimum does not show significant differences highlighting the scarce influence of the polar solvation on these types of states. The CIs between $n_O\pi^*$ and the polar $\pi\pi^*$ L_a and L_b states are profoundly influenced by the solvent comparing them to their *in vacuo* counterparts, featuring large differences in the C4-C5 bond distance. These differences are mainly attributed to the polar character of the $\pi\pi^*$ L_a/L_b states than to the effects on the $n_O\pi^*$ state. Overall, it can be concluded that solvation has an important effect on the polar $\pi\pi^*$ L_a /L_b states, while being negligible for the $n_O\pi*$ state. These states feature excitations that are prominently placed on the six-member ring of guanine featuring noticeable changes in the structure upon solvation, whereas the five-member ring remains relatively unchanged. A possible cause for this effect could lie on the presence of the phosphate group, which is tied to the five-ring member
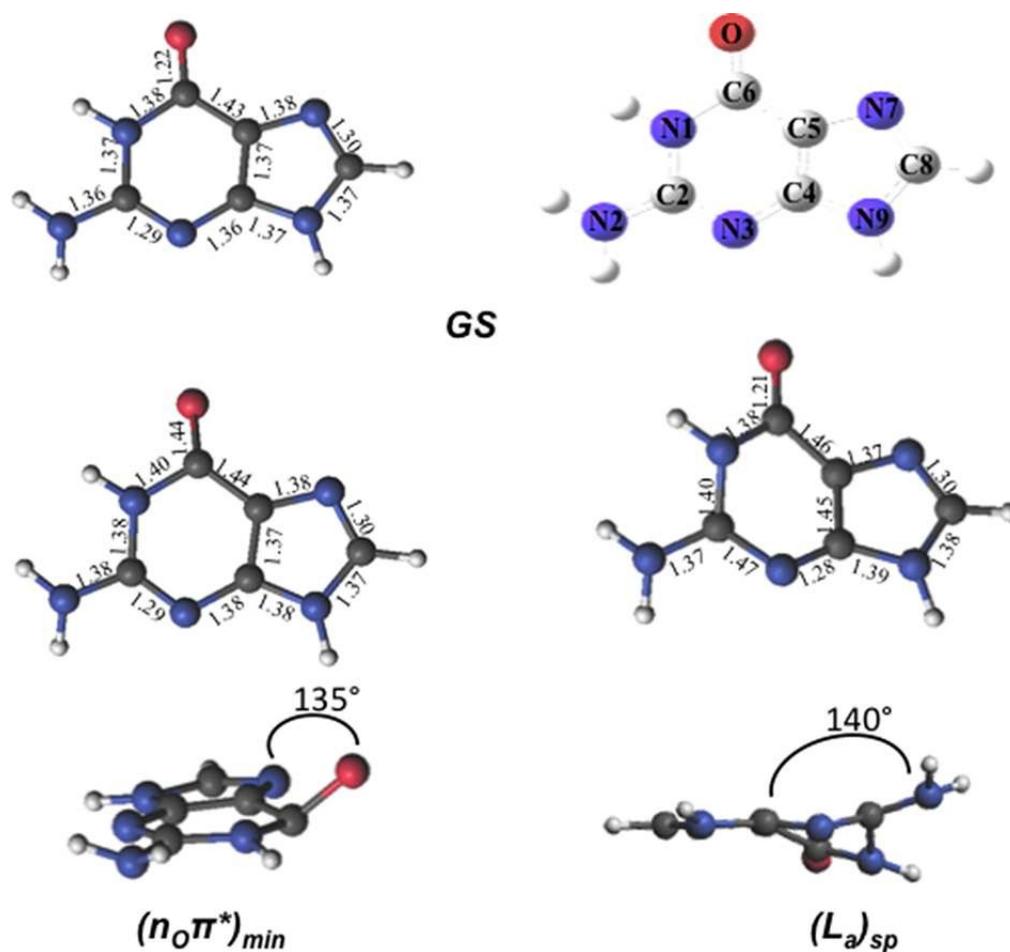
Figure 5.3: Geometries and main geometrical parameters of the Franck-Condon structure and excited-state stationary points characterized in GMP for its lowest-lying excited states computed at the CASSCF/6-31G* level of theory.

and could be shielding that molecular moiety from the surrounding water molecules thus mitigating its exposure as compared to the six-member ring and justifying why the latter experiences such pronounced changes upon solvation. The table reported in figure 5.5 contains the vertical excitation energies at the CASPT2 level computed in the present study, together with several other computations and experimental values reported in literature. We have used Plattâs nomenclature [177], where L_a represents the $\pi\pi^*$ excited state characterized by the largest contribution of the configuration HOMO (H) $\rightarrow$ LUMO (L) to the CASSCF wave function, whereas L_b identifies the $\pi\pi^*$ excited state with a predominance of H $\rightarrow$ L+1 and H-1 $\rightarrow$ L configurations. By inspection of the table in figure 5.5, it can be readily seen

that there is a qualitative difference in the relative order of the lowest-lying excited states *in vacuo* and in solution: both cases feature the $\pi\pi^*$ L_a state as the energetically lowest-lying excited state, at 4.93 and 4.77 eV respectively, but while *in vacuo* S2 has $n_O\pi^*$ nature (5.54 eV, involving the oxygen lone pair of the purinic base), in solution S2 is the $\pi\pi^*$ L_b state, placed at 5.44 eV. This effect has been widely reported in the literature being related to a bathochromic effect [178], with the $n_O\pi^*$ featuring a dipole moment of 5.69 D (3.51 D *in vacuo*) being less stabilized than the $\pi\pi^*$ L_b state (7.96 D, 4.92 D *in vacuo*) as it can be also noticed by looking at their change in dipole moment.

	N1-C2	C2-N2	C2-N3	N3-C4	C4-C5	C5-C6	C6-O	C6-N1	C5-N7	N7-C8	C8-N9	N9-C4
GS *in vacuo*[a]	1.37	1.38	1.31	1.36	1.39	1.44	1.21	1.41	1.38	1.30	1.38	1.37
GS in solution [b]	1.37	1.36	1.29	1.36	1.37	1.43	1.20	1.38	1.38	1.30	1.37	1.37
$(L_a)_{sp}$ in solution[b]	1.40	1.37	1.47	1.28	1.45	1.46	1.21	1.38	1.37	1.30	1.38	1.39
$(L_a/GS)_{CI\text{-}1}$ *in vacuo*[a]	1.41	1.40	1.45	1.29	1.45	1.47	1.20	1.41	1.39	1.29	1.39	1.37
$(L_a/GS)_{CI\text{-}1}$ in solution[b]	1.37	1.38	1.50	1.24	1.48	1.47	1.23	1.37	1.37	1.30	1.38	1.42
$(L_a/GS)_{CI\text{-}2}$ *in vacuo*[a]	1.41	1.40	1.38	1.28	1.45	1.35	1.34	1.36	1.39	1.29	1.38	1.38
$(L_a/GS)_{CI\text{-}2}$ in solution[b]	1.42	1.36	1.49	1.27	1.48	1.48	1.21	1.39	1.38	1.30	1.39	1.39
$(L_b/L_a)_{CI}$ *in vacuo*[a]	1.38	1.40	1.32	1.31	1.45	1.43	1.24	1.36	1.35	1.44	1.42	1.40
$(L_b/L_a)_{CI\text{-}1}$ in solution[b]	1.39	1.37	1.28	1.37	1.38	1.38	1.39	1.39	1.38	1.30	1.36	1.38
$(L_b/L_a)_{CI\text{-}2}$ in solution[b]	1.46	1.35	1.37	1.21	1.59	1.54	1.18	1.30	1.30	1.47	1.41	1.39
$(n_O)_{min}$ *in vacuo*[a]	1.38	1.38	1.28	1.37	1.36	1.44	1.38	1.41	1.36	1.30	1.37	1.36
$(n_O)_{min}$ in solution[b]	1.38	1.38	1.29	1.38	1.36	1.44	1.41	1.40	1.38	1.30	1.37	1.80
$(n_O/L_a)_{CI}$ *in vacuo*[a]	1.41	1.40	1.38	1.28	1.45	1.35	1.34	1.36	1.39	1.29	1.38	1.38
$(n_O/L_a)_{CI}$ in solution[b]	1.39	1.37	1.28	1.37	1.38	1.38	1.39	1.39	1.38	1.30	1.36	1.38
$(n_O/L_b)_{CI}$ in solution[b]	1.40	1.35	1.30	1.36	1.38	1.42	1.24	1.36	1.37	1.30	1.37	1.37

[a]Values obtained at the CASSCF/6-31G* level of theory by Serrano-Andrés et al. (2008).
[b]Values computed in the present study.

Figure 5.4: Bond distances (in Å) characterizing the key structures involved in the photoinduced events of GMP upon UV-light irradiation.

The solvent therefore affects the absorption spectra prominently at the FC region by blue-shifting the $n_O\pi^*$ state by 0.23 and red-shifting the $\pi\pi^*$ L_b by 0.33 eV with respect to the gas-phase, a change comparable to those previously reported in other QM/MM studies in solution [153, 178]. It is worth noting that the oscillator strengths associated to the transitions to the L_a and L_b states are also affected by the effect of the solvent. Whereas L_a features as the brightest lowest-lying excited state in the gas phase and the L_b state presents a slightly smaller oscillator strength (0.158 and 0.145 respectively), in solvent L_a and L_b display a reverse order in the values for the oscillator strength associated to their transitions (0.09 and 0.17 respectively), this being in agreement with the experimental data reported by femtosecond broad-band transient spectroscopy [152]. The effect of the basis set on the excitation energies has also been studied by employing both

6-31G^* and ANO-L 321/21 basis sets. The 6-31G^* CASPT2 computations report values of 4.77, 5.44, and 5.77 eV for the L_a , L_b , and $n_O\pi^*$ transitions respectively, whereas the ANO-L yields absorption energy values of 4.50, 5.10, and 5.71 eV for the L_a , L_b , and $n_O\pi^*$ states respectively, consistent with those reported experimentally [152], which place the absorption maximum of L_a at 4.50 eV. The table in fig. 5.5 also reports several theoretical approaches found in literature employing different methods to simulate solvated GMP. As can be seen, more sophisticated DFT/MRCI methods agree qualitatively with the values here reported at the CASPT2 level of theory, yielding values within 0.33 and 0.02 for the L_a and L_b states respectively with respect to the CASPT2/ANO-L values here obtained, and a slight energetic deviation is also found when comparing to the values computed at a TDDFT level. This small difference present in TDDFT that results employing a polarization continuum model (PCM) (see Table 5.5) could also be due to the fact that [152] used a methylated guanine to model the effect of the phosphate group whereas here the phosphate group is explicitly included even if just at the MM level. On overall we can conclude that the absorption values reported here at the CASPT2/ANO-L level are consistent with the experimental data and with the highest-level theoretical estimates present in the literature, thus highlighting the capabilities of the CASSCF/CASPT2 protocol to treat excited states in a balanced manner [179] and its usefulness for mapping photochemical reaction paths [163].

5.3.2 Excited-state evolution

The excited-state evolution is here tracked by means of static approaches through excited-state optimizations and the characterization of the conical intersections representing the crossings among the energetically lowest-lying excited states, that are the ultimate accountables for the deactivation photoprocess. Additionally, the MEPs connecting the initially accessed states and subsequent photoinduced events have been mapped making use of the IRC technique. The lowest-lying excited state is the $\pi\pi^*$ L_a state, which is expected to be the main spectroscopic state due to its lowest- lying position and relatively large oscillator strength. This state is generally assumed to present an easily accessible CI along its relaxation pathway toward the ground-state characterized by a ring-puckering motion (see 5.6) widely featured in the DNA/RNA nucleobases [180] *in vacuo*. This profile is slightly altered in the presence of polar environments as it has been previously reported for other purine nucleobases [158], where a shallow stationary point $(L_a)_{sp}$ arises along the MEP close to the CI with the ground state being placed at 1.15 eV

vertically and adiabatically at 3.30 eV from the ground state (see 5.7). Two different CIs have been characterized in the vicinity of this stationary point, one optimized directly in solvent corresponding to the minimum energy conical intersection (MECI) also reported by Serrano-Andres et al. (2008) [25], and another relating to the CI found in the same study along the MEP *in vacuo*, which we have tentatively named $(L_a/GS)_{CI-1}$ and $(L_a/GS)_{CI-2}$ in the present study, respectively.

Both CIs as well as the $(L_a)_{sp}$ are characterized by ring-puckering structures in the A6 cycle yielding slight bond-length alterations (C2-N3, N1-C2, N3-C4, and C4-C5) as compared to the FC structure and featuring prominent N1-N-C2-N3 dihedral angle distortions at 133° , 132° , and 140° for the $(L_a/GS)_{CI-1}$, $(L_a/GS)_{CI-2}$, and $(L_a)_{sp}$ structures, respectively. These distortions are quantitatively different to those reported *in vacuo*, stressing out the importance of the solvent where the $(L_a/GS)_{CI-1}$ presents a 143° dihedral angle as compared to its 133° solvated counterpart, together with slightly pronounced bond-length differences as can be seen in the Table reported in figure 5.4.

Both CIs here characterized display very similar structures (see figure 5.6) and charge distributions very close to those found in the gas phase, presenting an inversion in their energetic order in solution being $(L_a/GS)_{CI-1}$ the energetically lowest-lying adiabatically at 3.31 eV, $(L_a/GS)_{CI-2}$ being placed 0.3 eV higher in energy and 3.60 eV adiabatically from the FC region. Two different CIs connecting the $\pi\pi^*$ L_a and $\pi\pi^*$ L_b states have also been located. Direct CI optimization leads to $(L_b/L_a)_{CI-2}$, its geometry presenting a deformation near the N9- C8-N7 angle with respect to the ground state. The bond lengths suffer from large distortions, observing a shortening of the N1-C6, C5- N7, C4-N3, N3-C2, C6-O, and C2-N bonds, and an increase in the length of the remaining bonds with respect to the FC region (see figure 5.6). This CI is placed at 7.62 eV vertically from the FC region, which is higher in energy than both $\pi\pi^*$ L_a and $\pi\pi^*$ L_b vertical excitation energies, placed at 4.77 and 5.44 respectively, thus hinting toward a sloped CI profile preventing their non- adiabatic interaction. A second CI, $(L_b/L_a)_{CI-1}$, obtained along the Lb MEP computed within a 4-in-4 active space to avoid excessive wave function mixing, has been located being placed 5 eV adiabatically from the FC region. This CI presents geometrical similarities with its *in vacuo* counterpart and provide an accessible channel to funnel the initially populated L_b state down to L_a . Further attempts to optimize the L_b state towards a possible minimum have been fruitless due to the large wave function mixing and root flipping problems experienced along the optimization procedure, preventing us to obtain further information on this

States	ΔE	f	μ
9H-GUANINE *IN VACUO*[a]			
GS	0		5.81
$\pi\pi^* L_a$	4.93	0.158	5.23
$n_O\pi^*$	5.54	0.002	3.51
$\pi\pi^* L_b$	5.77	0.145	4.92
GMP IN SOLVENT (6-31G·)[b]			
GS	0		8.58
$\pi\pi^* L_a$	4.77	0.09	8.97
$n_O\pi^*$	5.77	0.00	5.69
$\pi\pi^* L_b$	5.44	0.17	7.96
GMP IN SOLVENT (ANO-L 321/21)[b]			
GS	0		8.88
$\pi\pi^* L_a$	4.50	0.17	9.62
$n_O\pi^*$	5.71	0.00	5.64
$\pi\pi^* L_b$	5.10	0.20	8.31
EXPERIMENT[c]			
$\pi\pi^* L_a$	4.50	0.094	
$n_O\pi^*$			
$\pi\pi^* L_b$	4.96	0.167	
9Me-GUANINE·5H2O (TDDFT/PCM)[c]			
$\pi\pi^* L_a$	4.68	0.16	
$n_O\pi^*$	5.77		
$\pi\pi^* L_b$	5.18	0.34	
9H-GUANINE (TDDFT/TZVP)[d]			
$\pi\pi^* L_a$	4.93	0.158	
$n_O\pi^*$	5.36	0.106	
$\pi\pi^* L_b$	5.28	0.117	
9H-GUANINE (DFT/MRCI/TZVP)[d]			
$\pi\pi^* L_a$	4.83	0.22	
$n_O\pi^*$	5.32	0.001	
$\pi\pi^* L_b$	5.08	0.368	

[a]Values obtained at the CASSCF/6-31G· level of theory by Serrano-Andrés et al. (2008).
[b]Results obtained in the present study.
[c]Results obtained by Karunakaran et al. (2009).
[d]Results obtained by Parac et al. (2010).

Figure 5.5: Vertical absorption energies (ΔE, in eV), oscillator strengths (f) and dipole moments (μ, in Debye) computed at the FC region, together with previous theoretical and experimental data.

specific state. As previously stated, $n_O\pi^*$ represents the third energetically excited state for the GMP in solvent. The optimized $n_O\pi^*$ minimum found in solvent is prominently characterized by an elongation of the C6-O bond distance with respect to the ground state. The table in figure 5.4 shows the main differences in bond lengths, where the C6-O bond is elongated in this minimum from 1.22 Åat the FC region to 1.41 Å. The minimum of this
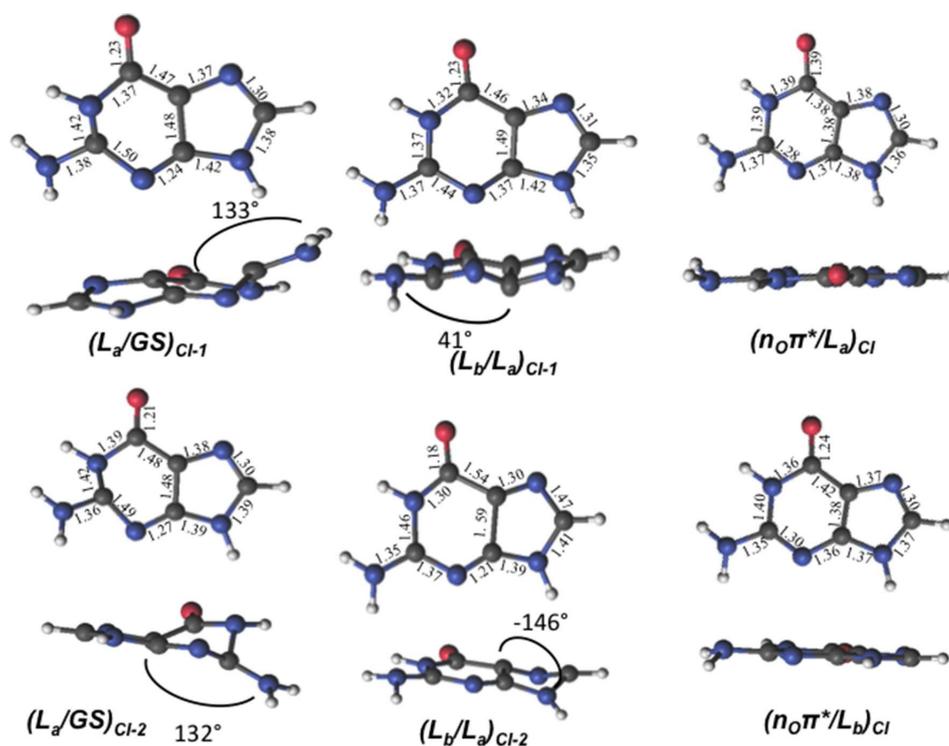
Figure 5.6: GMP geometries and main geometrical parameters of the characterized conical intersections between the lowest-lying excited states computed at the CASSCF/6-31G^* level of theory.

excited state in solvent is very similar to the one reported *in vacuo* [25], as would be expected given the small effects provided by the solvent on n_π^* excited states. Energetically it is placed 3.97 eV adiabatically and 2.67 eV vertically with respect to the ground state (see figure 5.8), which constitutes a stabilization of 1.8 eV from its initial value at the FC region. We have also located the CI between the states $n_O\pi^*$ and $\pi\pi^*$ L_a, $(n_O/L_a)_{CI}$, its geometry presenting a planar structure like the ground state, but showing an elongation of the C6-O bond and a shortening of the C6-C5 bond. We observe similarities between this geometry and that of the CI optimized *in vacuo*, as they both present the same kind of geometrical distortions given the small solvent effects in the $n_O\pi^*$ state, yielding a relative energy of 4.90 eV with respect to the FC region. The CI connecting the $n_O\pi^*$ and $\pi\pi^*$ L_b states, $(n_O/L_b)_{CI}$, has also been located, featuring a planar structure and differences in the bond lengths in the six-member ring, in particular in the differences in the N1-C2, N1-C6 bonds. The differences observed in the geometries of the excited state minima and of the different CIs characterized

both *in vacuo* and in solvent highlight the importance of the solvent that is considered explicit in our calculations, together with the explicit presence of the ribose and phosphate groups on GMP, providing more accurate estimates to relate to the photoinduced events in the cellular system.

5.3.3 Photophysics and photochemistry of GMP: elucidating its intrinsic photostability

Figure 5.9 features a scheme with the tentative photo-processes occurring upon UV irradiation in GMP based on our present computations and the theoretical and experimental data available in literature. Experimentally, time-resolved transient absorption in solution by Karunakaran et al. (2009) [152]suggest either bi- or triexponential decay depending on the spectral region measured, providing τ_1 =0.25ps, τ_2 = 1.0ps,and τ_3 = 2.5ps for the 270â400 nm region in the non-protonated GMP at neutral pH. [152] justify these ultrafast signals in terms of TDDFT calculations due to ring-puckering CIs between the $\pi\pi^*$ L_a and GS states, in accordance to what was previously proposed in vacuo by Serrano-Andres et al. (2008) [25] on the basis of CASSCF/CASPT2 computations. Molecular dynamics simulations by Serrano-Andres et al. (2008) [25] proposed that the ultra- fast decay from the $\pi\pi^*$ L_a state could occur already within the first 100 fs, an assumption that has been challenged by Barbatti et al. (2011) [26], pushing the timescale towards the 500 fs mark. (Lan et al., 2009) [141], on the other hand, support a biexponential decay based on their MD simulations where they obtain a faster 190 fs and a slower 400 fs decays through two different CIs between the $\pi\pi^*$ L_a and GS states, thus highlighting the complex process under study and the difficulty to simulate it coherently [181]. The most recent experimental data to our knowledge is based on the photoelectron spectroscopy of GMP both *in vacuo* and in solution [151], yielding a biexponential decay of τ_1 = 50fs and τ_2 = 600fs in vacuo and τ_1 = 120fs and τ_2 = 680 fs in solution, which reveals a striking similarity among the two different set of values as they are within experimental error and therefore provide evidence of the negligible role played by the solvent in the photoinduced decay paths present in this system. This biexponential decay is obtained through a 4.66 eV pump, which also suggests the possibility of only probing those kinetic processes undergone after direct $\pi\pi^*$ L_a population, whereas the time-resolved transient data by Karunakaran et al. (2009) yields a third exponent related to a slower timescale possibly arisen through an initial population of the $\pi\pi^*L_b$ state. Our present results point toward an ultrafast deactivation along the L_a MEP characterized by a ring-puckering motion

centered in the C2 atom, which brings the excited-state population to inter-
act non-adiabatically with the ground state, funneling down the population
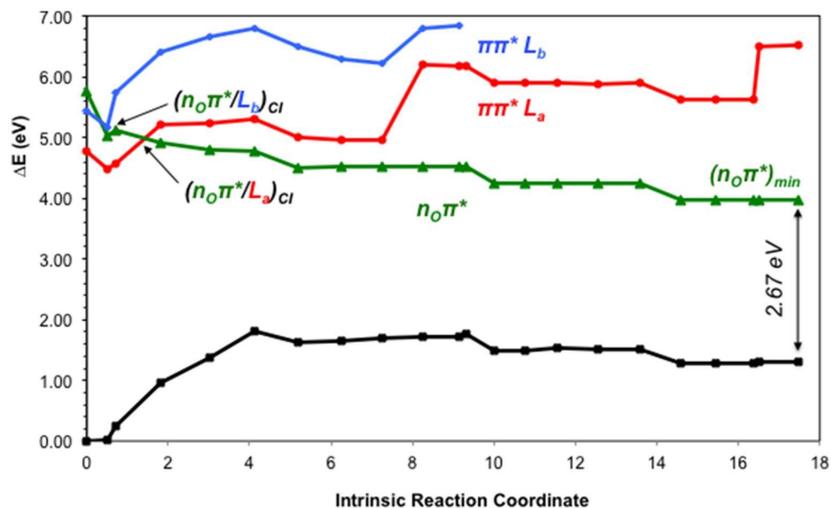that is further relaxed to the FC region.

Figure 5.7: CASPT2 energies of the ground (GS) and lowest-lying singlet excited states
($\pi\pi^*$ L_a, $\pi\pi^*$ L_b, and $n_O\pi^*$) of the GMP from the FC geometry and along the $\pi\pi^*$ L_a
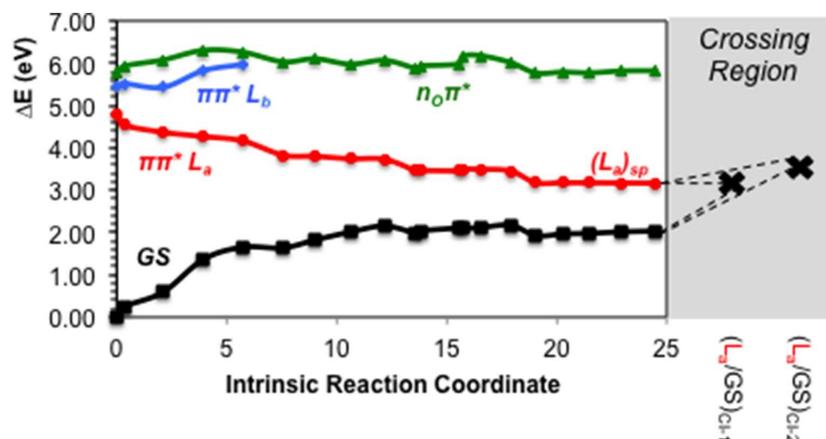CASSCF MEP.

Figure 5.8: CASPT2 energies of the ground (GS) and lowest-lying singlet excited states
($\pi\pi^*$ L_a, $\pi\pi^*$ L_b and $n_O\pi^*$) of the GMP from the FC geometry and along the $n_O\pi^*$
CASSCF MEP.

This kind of process has been widely attributed to the DNA/RNA nu-
cleobases and it is permanently linked with their intrinsic photostability[22]
and therefore extends to the photostability of the genomic material. Our
calculations in the presence of the solvent and the sugar and phosphate

groups reveal a very shallow stationary point at the end of the MEP right before reaching the intersection seam (see figure 5.7), which has also been previously documented in other nucleobases at a TDDFT level of theory to be related to a transition state with an imaginary reaction coordinate that drives the system towards the ring-puckering CI [182]. This stationary point is explained in terms of the planarity of the $\pi\pi^*$ potential energy hypersurfaces of all nucleobases [180, 156], and it is expected to be overcome in order to reach the CIs with the ground state in an ultrafast manner. In contrast, we have shown recently ([158, 166]) that the corresponding internal conversion path in adenine is controlled by a larger barrier, which may explain its longer lifetimes. Two different CIs have been located between the $\pi\pi^*$ L_a and GS states, namely $(L_a/GS)_{CI-1}$ and $(L_a/GS)_{CI-2}$.

These two CIs are analogous to those found *in vacuo* by [25] but they are characterized by an energetic inversion due to the embedding effect of the solvent. Both are expected to belong to the same intersection seam given their similar structures, and both are considered to be accessible from the L_a stationary point previously characterized. $(L_a/GS)_{CI-1}$, being the lowest-lying energy is expected to embody the fastest decay route, relating to the $\tau_1 = 0.25$ ps recorded through time- resolved transient and $\tau_1 = 120$ fs in time-resolved photoelectron spectroscopy experiments. This assignment is analogous to that previously reported for 9H-Guanine *in vacuo*, and supports recent experiments reporting analogous excited-state decays for GMP in gas-phase and in solution [151]. The second conical intersection with the ground state, $(L_a/GS)_{CI-2}$, might be responsible for the second exponential measured at $\tau_2 = 1.0$ ps and $\tau_2 = 680$ fs in time-resolved transient and photoelectron spectroscopy, respectively, given its higher-lying energetic position and a relatively small barrier of 0.3 eV to be overcome in order to access it once reaching the intersection seam. Nevertheless, the existence of another CI between the La and GS states not characterized in the present contribution could better explain that second lifetime, which will be related to the L_a surface as it has been recently reported experimentally [151]. These assignments agree with the MD studies by [141], reporting two different decay routes with their associated timescales to two distinct CIs between the $\pi\pi^*$ L_a and the ground state, and are in line with other theoretical CASPT2 and TDDFT studies that point toward the L_a state as the responsible for both ultrafast decay timescales. The $\pi\pi^*$ L_b state is predicted to be involved in the photoinduced events being related possibly only to the longest-lived τ_3. This is explained in terms of an initial population of the L_b state, which presents a sizable oscillator strength, and subsequent funneling of the population toward the L_a state mediated by the $(L_b/L_a)_{CI-1}$, placed at 5
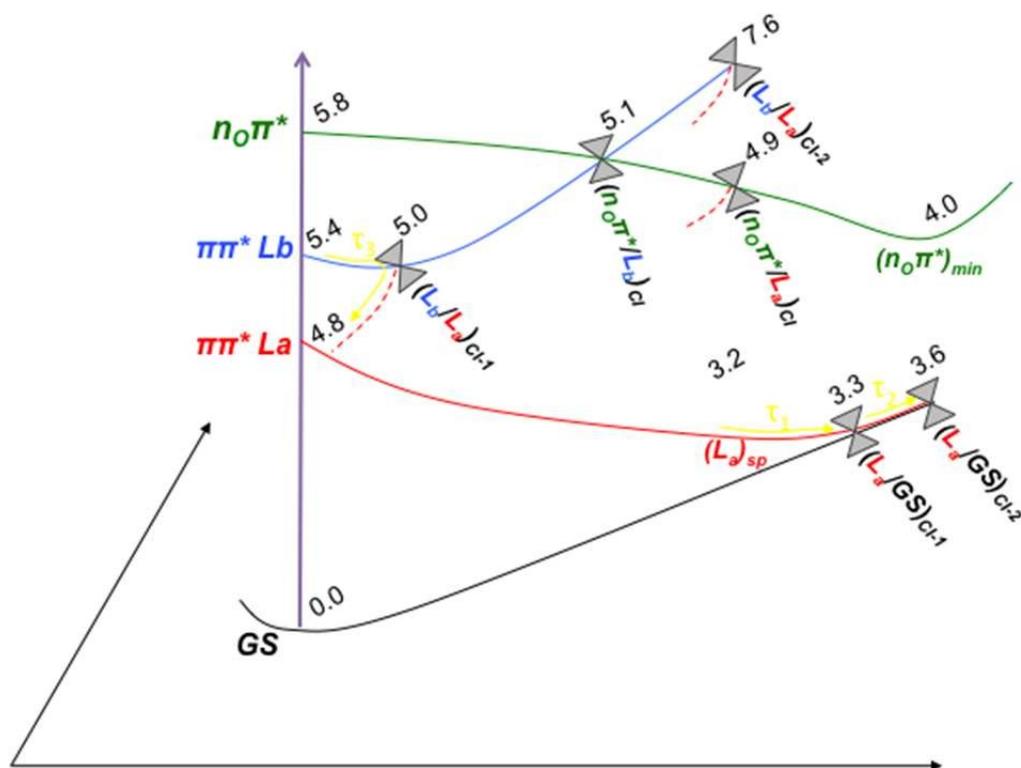
Figure 5.9: Scheme of the photoinduced processes occurring in solvated GMP upon UV-light irradiation. The energetic values (in eV) and the different crossing points and minima depicted refer to the CASPT2 computations carried out in the present study. The associated time constants marked in yellow have been taken tentatively from the ultrafast pump-probe transient absorption experiments reported by Karunakaran et al. (2009)[152]. Three different decay channels have been assigned to the experimental evidence: τ_1 refers to the ultrafast decay path from an initially accessed $\pi\pi^*$ L_a excitation to the ring-puckering CI with the GS, τ_2 to the initial population of the $\pi\pi^*$ L_b state and subsequent non-adiabatic population transfer to the L_a state finalizing in the CI between L_a and GS, and τ_3 that is tentatively assigned mainly to deactivation processes along the L_b state (See text for details).

eV along the L_b relaxation pathway. The second CI characterized between L_b and L_a states, $(L_b/L_a)_{CI-2}$, is expected to be irrelevant to the photoprocess due to its high-lying energetic position preventing its accessibility. The lone-pair excited state, $n_O\pi^*$, is blue-shifted in solution as has been already mentioned above. Its elevated vertical absorption energy prevents it to be one of the main spectroscopic states, yet its close-lying position to the absorbing $\pi\pi^*$ L_b state facilitates a possible population of this state and subsequent relaxation toward its minimum, $(n_O\pi^*)_{min}$. This minimum is placed at 3.97 eV adiabatically and 2.67 eV vertically from the GS, and

could be partly responsible of the longest-lived signal reported experimentally at $\tau_3 = 2.5$ps or at $\tau_3 = 167$ps at low pH,given that such $n\pi^*$-mediated processes have been already characterized experimentally to be close to the 100 ps timescale on pyrimidines both *in vacuo* and in solution [144]. The present study elucidates the photoinduced events in GMP in terms of an ultrafast decay along the main spectroscopic and initially accessed $\pi\pi^*$ L_a state characterized by a ring-puckering motion, which would cover the experimental timescales τ_1 and τ_2 through different CIs with the ground state, whereas the longest-lived component would be attributed to the decay routes mediated through the $\pi\pi^*$ L_b and, to a minor extent, to the $n_O\pi^*$ state.

5.4 Conclusions

The present study encompasses a photophysical and photochemical appraisal of water-solvated GMP by means of theoretical multireference perturbation theory QM/MM techniques. An initial MD simulation has been carried out in order to characterize the geometrical parameters of the FC region. The vertical excitation energies have been computed and compared to recent data found in literature and to the results obtained *in vacuo*, highlighting the importance of the environment yielding qualitative differences for the $\pi\pi^*$ L_a and $\pi\pi^*$ L_b states being red-shifted and for the $n_O\pi^*$ state being blue-shifted as compared to their gas-phase counterpart. The $\pi\pi^*$ L_a state is predicted to be the main spectroscopic state driving the ultrafast deactivation processes occurring in GMP upon UV-light irradiation based on a ring-puckering motion that enhances its non-adiabatic interaction with the ground state in a radiation- less manner. A shallow stationary point toward the end of the $\pi\pi^*$ L_a MEP has been characterized, together with two different CIs with the ground state that help rationalize the two fastest decay times measured experimentally. Upon initial L_b absorption, two CIs between the $\pi\pi^*$ L_b and L_a states have also been located, one of them along the L_b decay path pointing towards a population funneling down to the L_a state and another being too high in energy to contribute prominently in the photo-process. The CIs connecting the $n_O\pi^*$ state and the $\pi\pi^*$ L_b and L_a states have also been characterized along its relaxation route, yielding a minimum in the $n_O\pi^*$ state expected to emit vertically at 2.7 eV. Both $\pi\pi^*$ L_b and $n_O\pi^*$ are suggested to contribute to the longest-lived experimental timescale, the latter providing a lesser contribution given the relatively fast kinetic timescale and the long-lived timescales expected in $n\pi^*$ fluorescent states. Upcoming QM/MM dynamics simulations are expected to shed some more light on the photoinduced events occurring in water-solvated GMP and

its specific decay timescales in order to provide a more specific molecular counterpart to the experiment and better explain the photochemical and photophysical processes resulting in the intrinsic stability of the genomic material.

6 Photoactivity of retinal protonated Schiff bases in solvent

Retinal protonated Schiff bases (RPSBs) are polyene chromophores responsible for light reception in opsins, the G protein-coupled receptors (GPCR) of the retinylidene protein family. The radiationless ultrafast (sub-ps timescale) photoisomerization of the retinal chromophore in Rh is the primary event in the process of vision and a paradigm for fundamental studies of the molecular basis of vision in vertebrates. Actually, the understanding of the intrinsic photophysical and photochemical properties of retinal chromophores and of the environmental effects of the surroundings (proteins, solvent, organicâinorganic supports,etc.) is crucial for the design of tailored and high-performance photo-sensitive devices. The discrepancy between solution and protein behavior has been largely attributed to the complexity of the protein pocket with its three-dimensional arrangement of amino acids resulting in a unique steric and dielectric environment [6, 7]. The tunability of absorption spectra, isomerization yield, and reaction speed makes retinal the ideal candidate for investigations aiming at disclosing the molecular and structural origins of efficient photochemistry [6, 7, 8, 9, 10, 11, 12, 13, 14, 15]. As a consequence, retinal has become a paradigm for understanding the origins of activation and suppression of ultrafast relaxation processes, which is essential for the rational engineering of photoreactivity.

6.1 Introduction

The drastically different photo-reactivity of the retinal chromophore in solution compared to the protein is nowadays poorly understood. It has been recently demonstrated that the addition of a methyl group to the $C = C$ backbone of an all-trans retinal protonated Schiff base may accelerate the electronic decay in solution. In particular, methylation at position C10 of the polyene chain can speed up by an order of magnitude the excited state radiationless decay of the solvated all-trans retinal protonated Schiff-base (RPSB), the chromophore of the proton pump bacteriorhodopsin (bR)

that converts light into chemical energy. While in the protein environment the photoexcited âativeâ chromophore undergoes internal conversion via a trans-to-cis isomerization along the C13-C14 double bond in the subpicosecond time scale (approx. 0.5 ps) and with 0.64 quantum yield [142], in solution the photoisomerization produces the 11-cis isomer and it is significantly slower (ca. 4 ps) and less efficient (0.16 quantum yield) than in bR. However, solvated RPSB with a methyl group at position C10 (10Me-RPSB) shows recovering of the protein-embedded excited state with sub-ps dynamics and production of the 11-cis isomer with low quantum yield (0.09).

It is apparent that the addition of a methyl group to the retinal backbone (to the C10 position) results in protein-like photophysics and dynamics for the all-trans RPSB in solution. At the same time, this modification has the advantage of leaving the absorption spectrum unchanged, thus leading to an overall minimal perturbation of the native system.
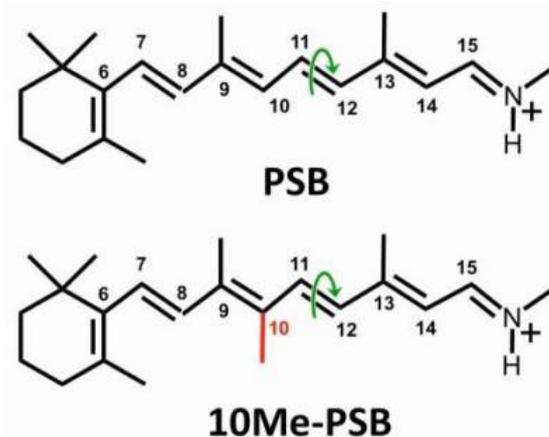

Figure 6.1: Native and 10-Me-RPSB molecular structures

The work below provides a rational for the experimental results and the effects of C10-methylation that are behind the observed protein-like speed up in the retinal photoisomerization and non-radiative decay. Moreover, the modelling of such systems in solvent, due to the great number of molecules and the complex architecture created by the solvent around the chromophore are extremely challenging. The following study is also aimed at formulating a general protocol and computational setup suitable for the study of complex systems in solvent by using hybrid QM/MM methods.

6.2 Computational details

6.2.1 MM dynamics and sampling

MM simulations were performed both for the native retinal and the 10-methylated retinal in methanol (the same solvent used in the experiments) to obtain a representative starting geometry to be employed for all subsequent computations and analyses. The MM dynamics calculations were carried out using the Amber-11 suite of programs making use of the Parm99 force field [46, 161]. Initially, a cubic solvent box consisting of molecules of explicit methanol was designed. A molecule of the two forms of retinal had previously been optimized in vacuum at a CASSCF level with an active space of 12 electrons in 12 orbitals (the full p orbitals of the retinal were involved in the a.s.). After the optimization of S0, also S1 was optimized in vacuum and the geometry of S1 (kept frozen) was used for the MM dynamics. The chromophore charges exploited for the MM dynamics are those computed at QM level for excited state S1. The entire system was then heated from 0 to 300 K for 1 ns at constant volume and constant pressure (1 atm), and then finally a 50-ns simulation run was performed recording the snapshots every 200 fs. After an energy analysis, the snapshot characterized by the lowest potential energies was selected as the initial structure for the following calculations. The MM dynamics were conducted using the geometry of the S1 state, instead of the ground state, to stabilize the ABL geometry (see the discussion below) of the retinal.

6.2.2 QM/MM calculations

QM/MM calculations were performed using the COBRAMM suite [48]. A three-layer approach (high, medium, and low) was used throughout [48]: almost all the RPSB was included in the high QM layer (figure 6.2), whereas the medium layer included the movable MM atoms that comprise all atoms within a 10 Åradius surrounding the retinal (resulting in 8 molecules of MeOH overall). The remainder of the MM system was kept frozen in the low layer during all optimizations, The boundary between the QM and MM regions has been set in order to include all the retinal atoms of the polyenic chain in the QM region, while the alkyl moiety of the B-ionon ring is placed in the MM medium layer: the link atoms employed for that are displayed

Equilibrium geometries and photoreaction paths [163] were determined by using fully unconstrained and constrained optimizations. CI optimizations were performed with the gradient projection algorithm by [165] as implemented in COBRAMM at the QM/MM level as described in the next
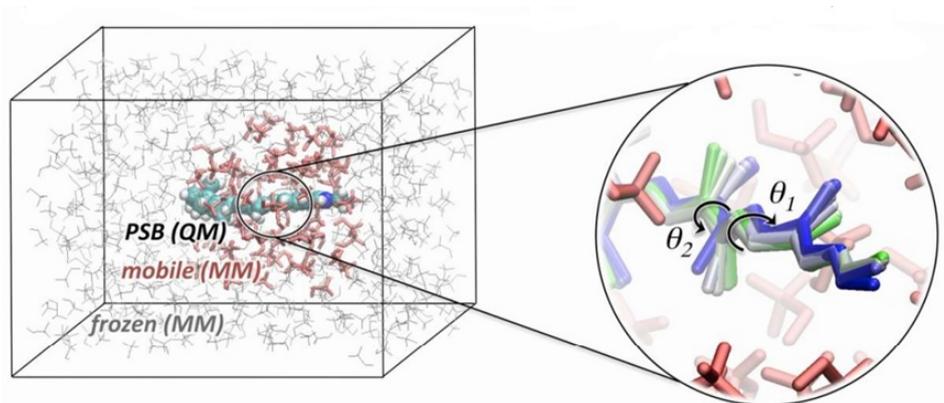
Figure 6.2: Retinal-containing box with the QM/MM partition scheme

chapter. The QM gradients employed in geometry optimizations and dynamics (including the non-adiabatic coupling terms) were computed making use of the complete active space self-consistent field (CASSCF) approach as implemented in the MOLPRO- 2010 [169, 170], while energies on top of the optimized geometries and MEP points were refined to account for dynamic correlation by single point complete active space second-order perturbation theory (CASPT2) calculations [167, 168] employing the MOLCAS-7 [171, 172] suite of programs. The ground and excited-state energies of the considered molecular systems were computed by using multiconfigurational second-order perturbation theory through the CASPT2 method.23 The 6-31G^* basis set was employed throughout the work. The active space comprises the full π systems, which represents a total of 12 π electrons distributed among 12 π orbitals. The geometry optimizations for the ground and the excited states were performed at the CASSCF(12,12) level.26 Both single-root(for S0) and state-average three-roots (weights 1-1-1 for S1) CASSCF(12,12) calculations were used. Excitation energies were computed at the CASPT2//CASSCF/6-31G^* level using state-average of three roots wave functions. To minimize the influence of weakly interacting intruder states in the state-average PT2 calculations, the so-called imaginary level-shift (of 0.2 a.u.)technique was employed.

Non-adiabatic dynamics on S1 have been carried out using the CO-BRAMM package with a $1-fs$ time step. An active space consisting of 12 electrons in 12 orbitals has still been used, with the same QM/MM setup used for previous calculations. The dynamics have been propagated dis-

abling surface hopping. More details on the dynamics will be discussed later. A velocity Verlet algorithm has been implemented in order to propagate trajectories on S_1, at constant energy, by equally weighing the first two roots in a state average CAS(12/12) procedure.

6.2.3 The choice of the medium layer

The choice of the number of the molecules of solvent to be included into the medium layer has been non-trivial since suttle changes in the hydrogen bonding network around the chromophore may induce an energy discontinuity in the MM contribution, thus preventing direct comparison between adjacent optimized points along a MEP. Indeed, if a large number of MM solvent molecules move around the retinal, they influence the total energy of the system that can thus display significantly different values even for similar structures. To select the minimum number of movable solvent molecules that should be eventually considered in the medium layer, a preliminary MEP mapping have been carried out around the C10-C11-C12-C13 dihedral angle employing a very large movable layer of methanols around the chromophore. Molecules of solvent around the retinal in the various points of the MEP have been analyzed and compared in order to spot those which move the most. Figure 6.3 shows the solvent and the chromophore geometries of two superimposed points of the MEP; only some solvent molecules (those stressed in yellow) are those which move the most. Such analysis has been carried out both for the native retinal and for the methylated one and has demonstrated that only 7-8 molecules of solvent are those which undergo the most significant spatial variations: such molecules have been selected and included in the Medium Layer for the subsequent calculations.

6.3 RPSB and 10-Me RPSB: ground and excited state optimizations and MEP calculations

The main geometrical parameters for the optimized ground state minimum of the RPSB and 10Me-RPSB systems in methanol are reported in Table 6.1. The ground-state equilibrium structure (S0-Min) of the RPSB and 10Me-RPSB show the same basic features: bond alternation is the one expected for a ground-state polyene and the main chain includes five conjugated double bonds. Note that these structures are characterized by geometrical parameter value very similar to that observed and computed in vacuum at the CASSCF (12,12) level [183, 184]: the effect of the solvent does not affect
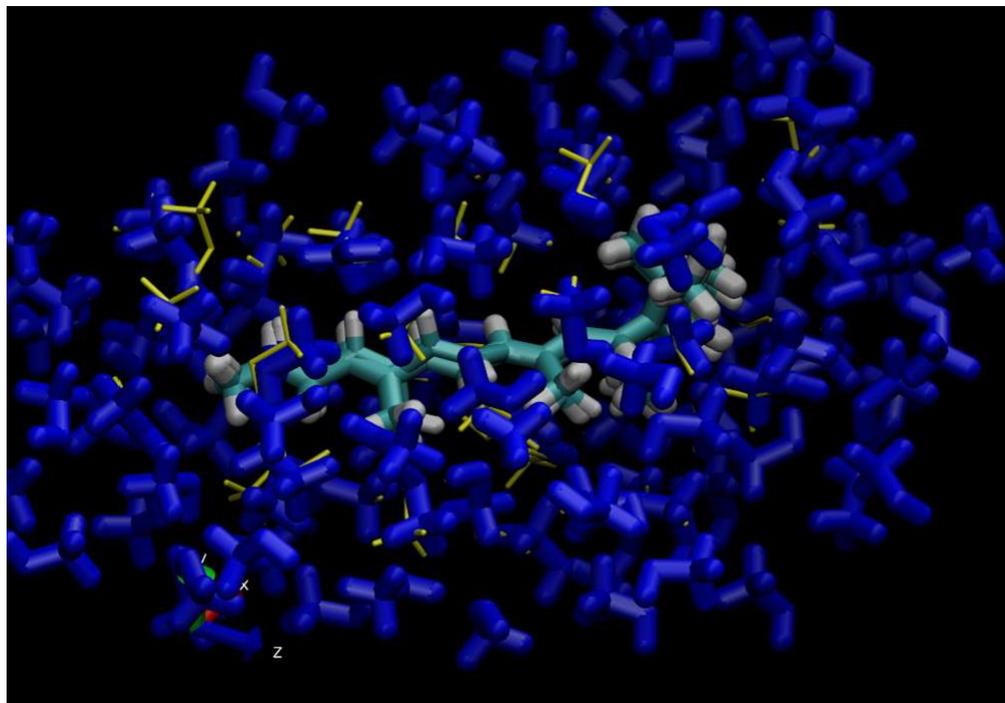
Figure 6.3: Solvent and chromophore geometries of two superimposed points of the MEP. The molecule stressed in yellow are those which move the most

Table 6.1: Bond lengths for retinal in vacuum and in solvent (MeOH)

	C6-C7	C7-C8	C8-C9	C9-C10	C10-C11	C11-C12	C12-C13	C13-C14	C14-C15	C15-N16
Vacuum	1.49	1.35	1.47	1.36	1.45	1.36	1.46	1.37	1.43	1.29
RPSB in MeOH	1.50	1.35	1.48	1.36	1.47	1.36	1.47	1.36	1.45	1.28
10-Me-RPSB in MeOH	1.50	1.35	1.48	1.37	1.48	1.36	1.47	1.36	1.44	1.28

the ground state structures. Moreover, the substitution of a âCH3 in stand of an hydrogen does not influence the alkylic chain structure.

The reported experimental values of lambda max absorption in methanol for RPSB and 10Me-RPSB are 2,78 eV and 2,69 eV respectively. The theoretical values calculated, after CASPT2 correction, are 2,65 eV and 2,58 eV, with a difference between the measured value and the calculated one equal to 0,13 eV for the native retinal and to 0,11eV for the methylated one. It is apparent that the QM/MM models and level of theory used could well reproduce the corresponding vertical excitation energy [185].

Upon photoexcitation, the all-trans retinal polyene chain with alternate single/double bonds in the ground state undergoes bond inversion, allowing twisting along C-C single bonds (i.e. the ground-state double bonds)

and thus facilitating the trans-cis isomerization. Therefore, the bond-length alternation (BLA) associated with the excited state minimum would affect the retinal photoisomerization rate and efficiency. The retinal BLA is numerically defined as the difference between the average distances of single, approx. 1.45 Å, and the double, approx. 1.35 Å, bonds in the chain, with values close to zero indicating the presence of bond delocalization and values close to 0.1 corresponding to fully alternated bonds.

This outcome (see figures 6.4) suggests that the photoinduced bond relaxation in the solvated 10Me-RPSB reaches an excited state minimum with alternate bond-lengths (ABL), from which bond-twisting can easily lead to a twisted CI. In contrast, BLA fluctuations of RPSB suggest that the native chromophore, after the rapid increase of the ground-state BLA and the initial bond-stretching fluctuations, remains trapped in an excited state minimum with even bond-lengths (EBL). The clear difference between the bond-length alternations in 10Me-RPSB and RPSB explains why only the 10-methylated chromophore can very rapidly (¡1ps) reach the CI seam.

The S1 minimum along the CASPT2//CASSCF potential energy surface (PES) of both native and 10-methylated all-trans retinals have been characterized. Geometry optimizations starting from the FC region confirmed the sizeable effect of the 10-methylation on the S1 PES: an ABL minimum is obtained for the 10Me-RPSB, in analogy with native RPSBs in proteins [13] and in gas-phase, while an EBL structure is found for the native RPSB. The resulting two different minima arise from distinctive linear interpolation between the bond-alternated FC geometry, EBL and the bond-inverted ABL structure.
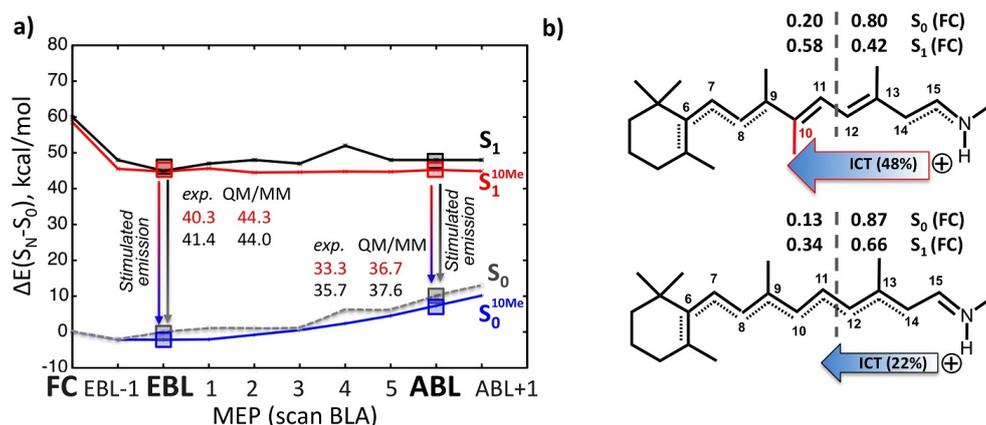

Figure 6.4: a) MEP calculated along the BLA, b) Mulliken charges distribution on RPSB and 10-Me RPSB on the FC geometries.

Figure 6.4 shows the CASPT2 corrected C-C bond relaxation MEP of both RPSB and 10Me-RPSB, calculated along the BLA coordinate connecting the FC and the ABL geometry on the S1 surface. In contrast with the RPSB energy profile, showing the presence of an activation barrier (of ca. 7 kcal/mol) connecting the EBL and the ABL minima, with the former being more stable than the latter by 3.2 kcal/mol, the 10Me-RPSB MEP is completely flat in this region. Thus, mapping the MEP of the C-C bond relaxation already provides a rationale for the different sub-ps excited state dynamics observed for the solvated retinals: the native RPSB stretching coordinate is characterized by a local minimum with even bond lengths (EBL, Figure 6.4a), from which the complete bond-inversion requires overcoming of a significant activation barrier, whereas the excited state potential energy surface of 10Me-RPSB involves a flat pathway for bond-inversion, directly leading to the photoactive bond-inverted ABL structure. The ABL minimum has a larger charge-transfer (ionic) character with respect to the EBL structure, since the C-C bond inversion favors the transfer of the positive charge from the protonated nitrogen atom towards the β-ionone ring. The partial (Mulliken) charge changes induced by the 10-methylation are reported in Figure 6.4b, indicating how the presence of the methyl group increases significantly (from 22 to 48%) the intramolecular charge transfer (ICT) that moves the positive charge from the SB side (C12âN) to the β-ionone ring (C5âC11) in the excited state. Thus, the presence of the electron-donating methyl group in position C10, i.e. in the molecule segment where the positive charge is transferred upon photoexcitation (from C5 to C11), is the source of the ABL structure stabilization along the 10Me-RPSB MEP. The occurrence of two possible excited state minima along the S1 surface has been previously detected for Rhodopsin mimics [186] and it is corroborated by the observation of two stimulated emissions (SE) in time-resolved experiments of RPSB and 10-Me-RPSB. As showed in Figure 6.4a, the QM/MM calculations provide theoretical SE values in good agreement with experimental data, with differences between only 1-4 kcal/mol. The overall appearance of the double emissions in the experimental time-resolved pump-probe maps is partially affected by the concurrent (and opposite sign) signals associated with S1-SX excited state absorptions [185]. Likewise, QM/MM calculations indicate the presence of excited state absorptions in both the native and the 10-methylated retinals from either the ABL or the EBL S1 minima, with highest oscillator strengths around 0.5-0.8, less than half of those of the S1-S0 emissions (around 2.0). In an effort to explore the consequences of the methylation at C10 on the retinal photoisomerization, other than a pure inductive electronic effect that stabilize

the ABL excited state minimum, the MEP of the all-trans to 11-cis photoi-
somerization, described by the clockwise rotation along the torsional angle
θ_1 have been computed. The photoisomerization MEP of solvated RPSB
and 10Me-RPSB is showed in Figure 6.5. The calculated energy profiles
clearly indicate that, once the ABL minimum is reached (second point of
the path), the clockwise twisting along the torsional angle θ_1 is barrierless
for both RPSBs, and a CI region is reached after a rotation of $\approx 70°$. How-
ever, the photoisomerization pathways exhibit a significant difference in the
topography of the CI region: the 10Me-RPSB MEP shows a steep downhill
energy profile while approaching the peaked CI seam, which is typical of
fast and effective photoisomerization processes, whereas the MEP of native
RPSB is associated with a sloped CI [187, 188], indicating an inefficient
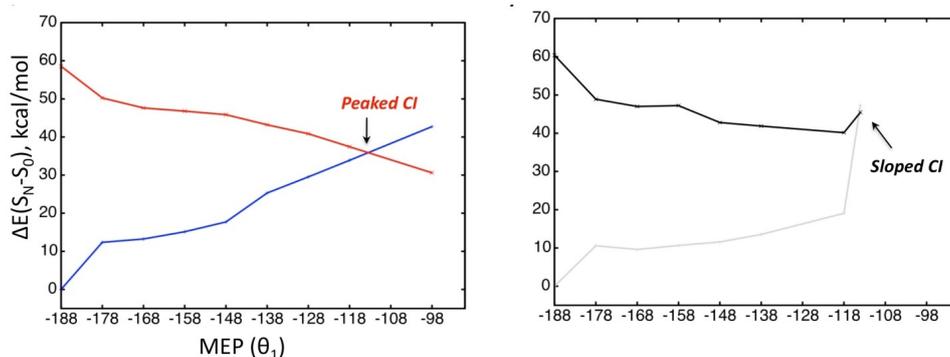S1âS0 population transfer and thus a slower excited state decay.

*Figure 6.5: MEP at CASPT2 level around the C9-C10-C11-C12 dihedral angle for the 10-Me
RPSB (left) and native RPSB (right).*

In fig. 6.6 we can observe the variation of the dihedrals of the alkyl chain
of the retinal along the MEP previously discussed. For the methylated, the
mechanisms that lead from the FC to CI entail that the rotation of the
C_{10}- C_{11}- C_{12}- C_{13} dihedral be combined with the rotation of the C_8- C_9-
C_{10}- C_{11} dihedral. Instead, for the native the mechanisms entail that the
rotation of the C_{10}- C_{11}- C_{12}- C_{13} dihedral imply also the rotation of the C_9-
C_{10}- C_{11}- C_{12} - C_{13} and C_{12}- C_{13}- C_{14}- C_{15} dihedrals. The mechanism here
observed will be discussed and compared below with the results obtained
from QM/MM dynamics.

MEPs around the C_8- C_9- C_{10}- C_{11} and C_{12}- C_{13}- C_{14}- C_{15} dihedrals,
both for the native and the methylated have also been done. It can be
observed that they are essentially flat in the case of methylated, and so is
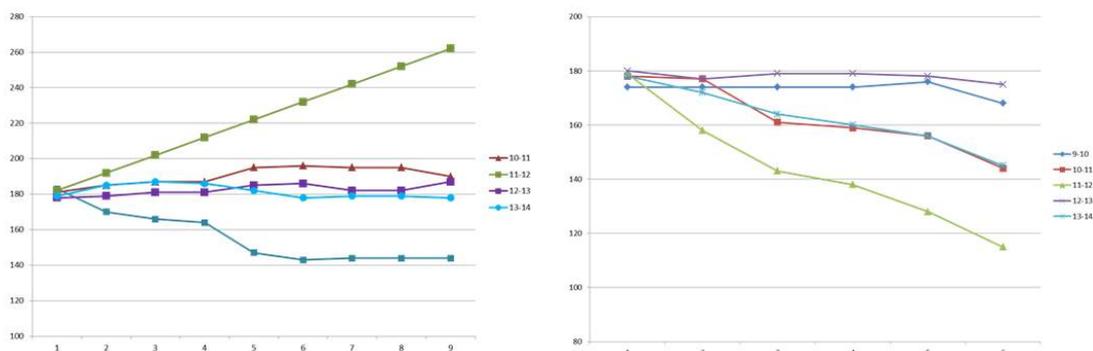the path to the conical intersection. On the contrary, for the native the

Figure 6.6: Variations in the dihedral along the MEPs shown in Figure 6.5

conical intersection is sloped in this case also.
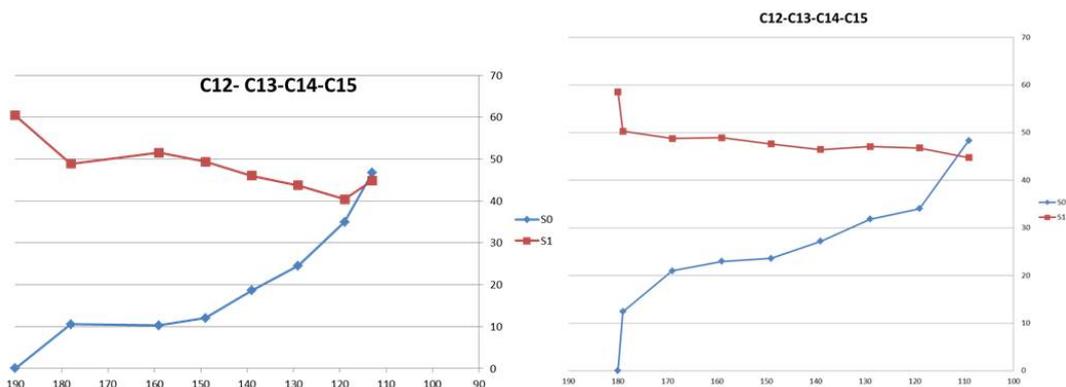
Figure 6.7: MEP at CASPT2 level around the C8-C9-C10-C11 and C12-C13-C14-C15 dihedrals for 10Me-RPSB (left), and for RPSB (right).

6.4   QM/MM dynamics

Non-adiabatic dynamics have been carried out at $0K$ both on the native and methylated retinal by using the COBRAMM packages on S1 with a time step of 1 fs. In such dynamics, unlike the MEP, the movable medium-layer solvent molecules have been assumed to be those located within a 10-Angstrom radius around the retinal. Such choice has been made to ensure that all the retinal atoms remain surrounded by the solvent during all the simulation despite the rotation of the alkyl-chain dihedral occurring during the simulation.

Observing the two trajectories there are differences between the photoisomerization process of PSB and 10Me-PSB in methanol solution. The
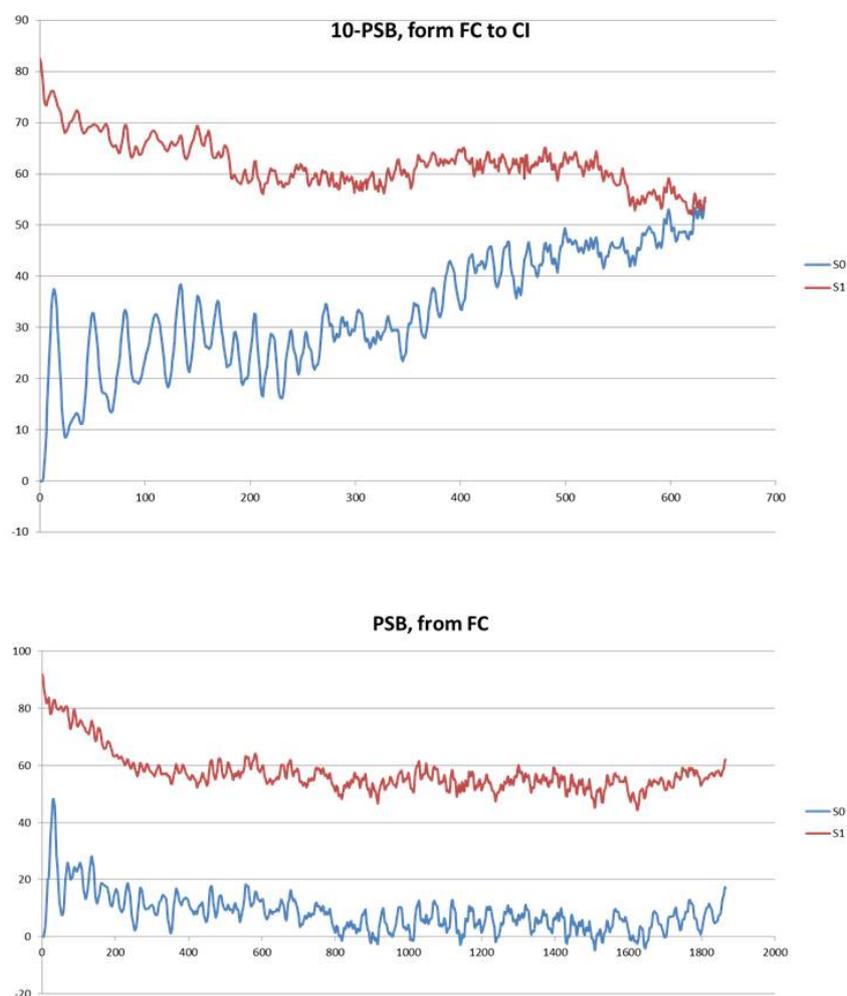
Figure 6.8: QM/MM MD for 10-Me PSB and PSB

evolution of the S1 energy relative to the ground state (S0) clearly indicates that a CI is reached within 0.6 ps for the 10Me-PSB. Instead, for the all-trans PSB, the S1/S0 energy gap keeps significantly large (≈ 50 kcal/mol) for more than 1.8 ps. These results are fully consistent with experimental observations showing an S1 lifetime of 0.7 and 4 ps for 10Me-PSB and PSB, respectively. As showed in Figure 6.9, the geometrical changes of the 10Me-PSB during the S1 evolution are associated with typical torsions along C-C single bonds (see mechanistic details below) yielding to a twisted CI geometry.

Figure 6.9 shows the fluctuations of the BLA observed during the QM/MM MD simulations of both the RPSB and the 10Me-RPSB solvated retinal

Figure 6.9: Analysis of BLA value during the dynamics for 10Me-RPSB and RPSB

chromophores. In the case of the 10Me-RPSB the initial BLA associated with the ground-state (approx. -0.1) rapidly increases upon photoexcitation and then oscillates significantly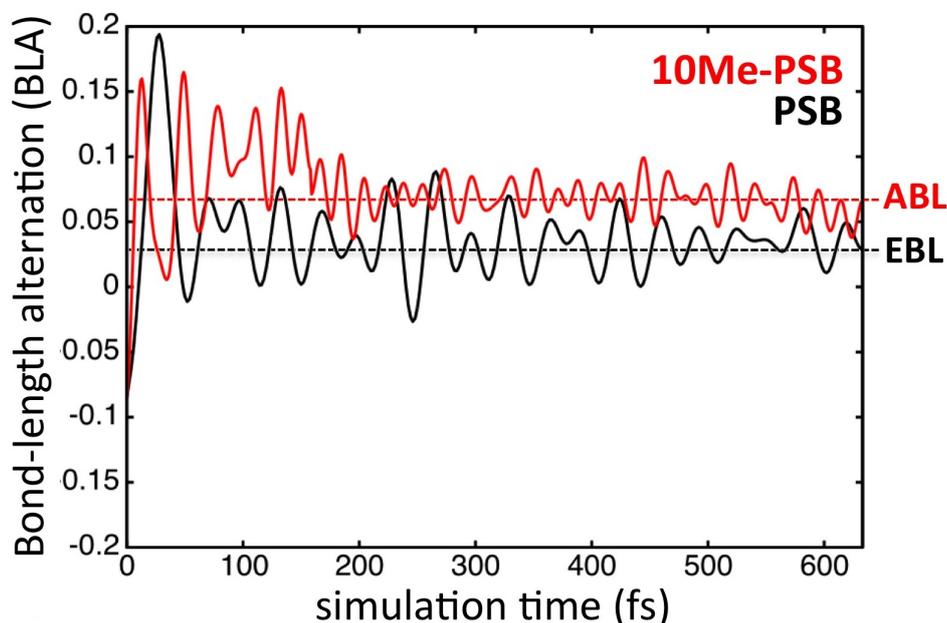 until converging to an approximate value of 0.07 within 0.23-0.25 ps. This BLA value is maintained almost constant all through the rest of the MD simulation until reaching the CI region. The 10Me-RPSB in solvent quickly arrives to an excited state ABL minimum; on the contrary BLA fluctuations of RPSB suggest that the native chromophore, after the steep increase of the ground-state BLA and the initial bond-stretching fluctuations, remains caught in an excited state EBL minimum with a BLA of approximately 0.03. The marked difference between the alternations in the bond lengths in the cases of 10Me-RPSB and RPSB accounts for the reason why only the 10-methylated chromophore is capable of quickly (¡1ps) reaching the CI seam.

Figure 6.10 shows the time-dependent evolution of the torsional angles during the photoisomerization process of 10Me-RPSB, as found in the QM/MM MD trajectory. The simulation indicates that, after an initial C-C bond relaxation, the 10-methylated retinal remains almost planar for the first 0.05-0.07 ps, with all three torsional angles close to 180°. The θ_2 angle time-evolution indicates that the torsion around the C9-C10 bond is undergoing an irreversible variation, with a counterclockwise rotation of 40°. While the torsion around the C13-C14 bond (θ_3) is also a counterclockwise
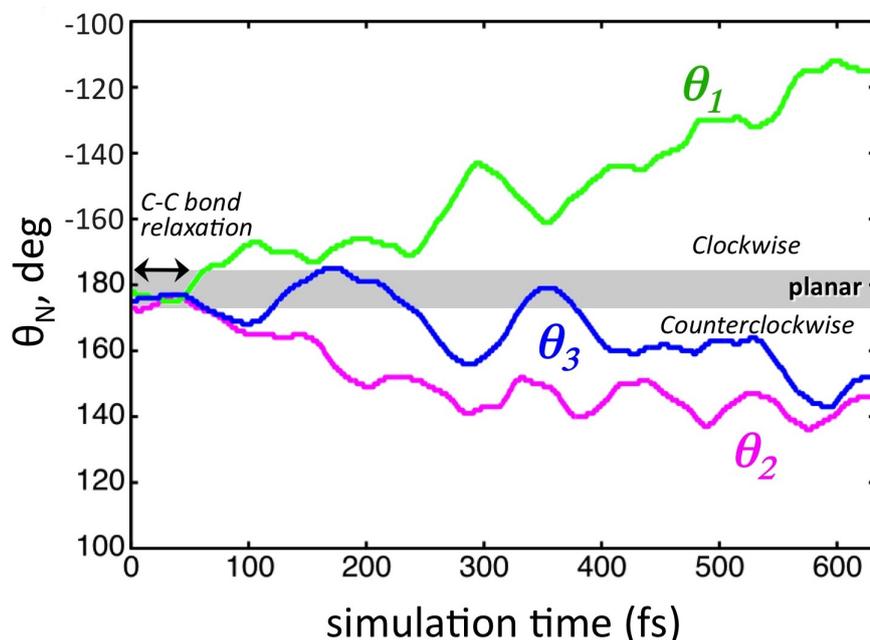
Figure 6.10: Evolution during the dynamics for θ_1 (torsional angle around the C11-C12 bond), θ_2 (torsional angle around the C9-C10 bond) and θ_3 (torsional angle around the C13-C14 bond).

rotation, the angle values oscillate more than those of (θ_2 and reach a 40° rotation only after 0.6 ps. The central θ_1 torsional angle shows, instead, a clockwise motion, with a minor 10° rotation in the 0.07-0.1 ps time range, a significant oscillation of 30° between 0.25 and 0.35 ps coupled with a similar, but reverse, oscillation of θ_3 and finally the main rotation that leads to a $\approx$ 70° twisted configuration along the C11-C12 bond. Thus, the asynchronous and opposite rotations of the C9-C10 and C11-C12 bonds, reminiscent of the asynchronous bicycle-pedal mechanism of Rh [189], are here accompanied by the rotation along the C13-C14 bond in the same direction of the C9-C10 one. The twisting along the central bond associated with a less-pronounced reverse-twisting of both adjacent double bonds defines an asynchronous double bicycle-pedal motion [13]. Notably, this mechanism here observed for the 10Me-RPSB in solution has been previously proposed for the photoisomerization of the all-trans RPSB retinal chromophore in the bR protein [13]. The QM/MM MD simulations indicate that the RPSB dynamics is trapped in the EBL minimum in the sub-ps regime, while the 10Me-RPSB excited state dynamics involve an ABL minimum, from which a CI region is reached within 0.6 ps following a double bicycle-pedal mechanism.

Such result agrees with that obtained from the MEP as previously described where, besides having done constrained optimizations, there was also a smaller number of movable solvent molecules (just 7 as compared to the movable 10-Angstrom radius layer around retinal employed in the dynamics). Investigations on the methylated have suggested to attempt to analyze, although in a qualitative manner, the quantum yield of the photoreaction: starting from a set of points characterized by a small energy difference between the ground-state and S1, where an efficient S1 → S0 hopping is assumed to occur, dynamics on the ground state have been started. However, out of ten dynamics carried out starting from ten hopping geometries, no one led to the photoproduct: all of them reverted back to the reagent. Such results is in good agreement with what observed, that is in the methylated retinal the modification caused a 50% decrease in the isomerization yield, thus suggesting an inverse correlation between reaction speed and yield [160]. Removal of backbone substituents increases the total isomerization yield, while on the contrary the addition of substituents reduces the yield. Dynamics starting from CT geometries where one of the dihedrals was already partially rotated in order to ease the photoisomerization around another dihedron have been effected. Such activity was aimed at studying the reaction mechanism in the case dihedrals other than C10-C11-C12-C13 rotated. As a results the dynamics reverted to the geometry characterizing the planar CT and it would subsequently photoisomerize around C11-C12. This is is line with very recent observations [185] showing that only the 11-cis photoproduct is formed in the case of the 10-methylated system.

For what the retinal in its native form is concerned, as already stated, the photoisomerization process is difficult since the solvent stabilizes the EBL form, that is the non-photoactive one: indeed the dynamics shows that the EBL form remains stable for 2000fs and the process does not goes on. However, since it is interesting to study the mechanism that eventually allows the native system to reach the CI (please note that such a mechanism has never been observed and studied for the retinal in solvent before), we switched off all the charges of the solvent molecules around in order to stabilize the ABL form (as it happens in gas phase, where no electrostatic effects exist) and therefore drive the dynamics to the conical intersection. The dynamics carried out by zeroing the charges that are associated to the solvent atoms let the cromophore be affected by steric interactions alone (mechanical embedding). It is apparent that these *biased* dynamics cannot provide information concerning the life-times of the excited states as well as the photoisomerization quantum yields. Still, they do provide reliable information concerning the photoreaction mechanism. Figure 6.11 shows

the difference between the ground state energy and that of the S1 state as a function of time and demonstrates that, with charges turned off, the CI is eventually reached in 800fs. While this timing cannot be compared with experiments, still the analysis of the variation of the dihedral angles allows to describe the photoreaction mechanism. Figure 6.11 shows that the C10-C11-C12-C13 dihedral angle twists along with the C8-C9-C10-C11 and C12-C13-C14-C15 dihedral (although in opposite direction). This is exactly the very same mechanism previously observed for the methylated retinal and is a further confirmation supporting the double bycicle-pedal path illustrated above.

6.5 Conclusions

In conclusion, it has been shown why 10-methylation in *all-trans* RPSB retinal triggers a dramatic change in the excited state subpicosecond dynamics. The methyl group in 10-position stabilizes an excited state minimum with a large charge-transfer character and with alternated C-C bonds, favoring an efficient photoisomerization. Remarkably, the photoisomerization space-saving double bycicle-pedal mechanism matches the one also observed in space constrained environments such as retinal proteins (e.g., Rh and bR). The proposed QM/MM models are fully consistent with time-resolved spectroscopy experiments, showing a broad double-peaked stimulated emission band matching the two excited state (EBL and ABL) minima. An detailed analysis of the photoisomerization mechanism of all-trans retinals in solution is provided, showing that production of the 11-cis isomer is achieved by a double bicycle-pedal mechanism, although with a very small quantum yield.
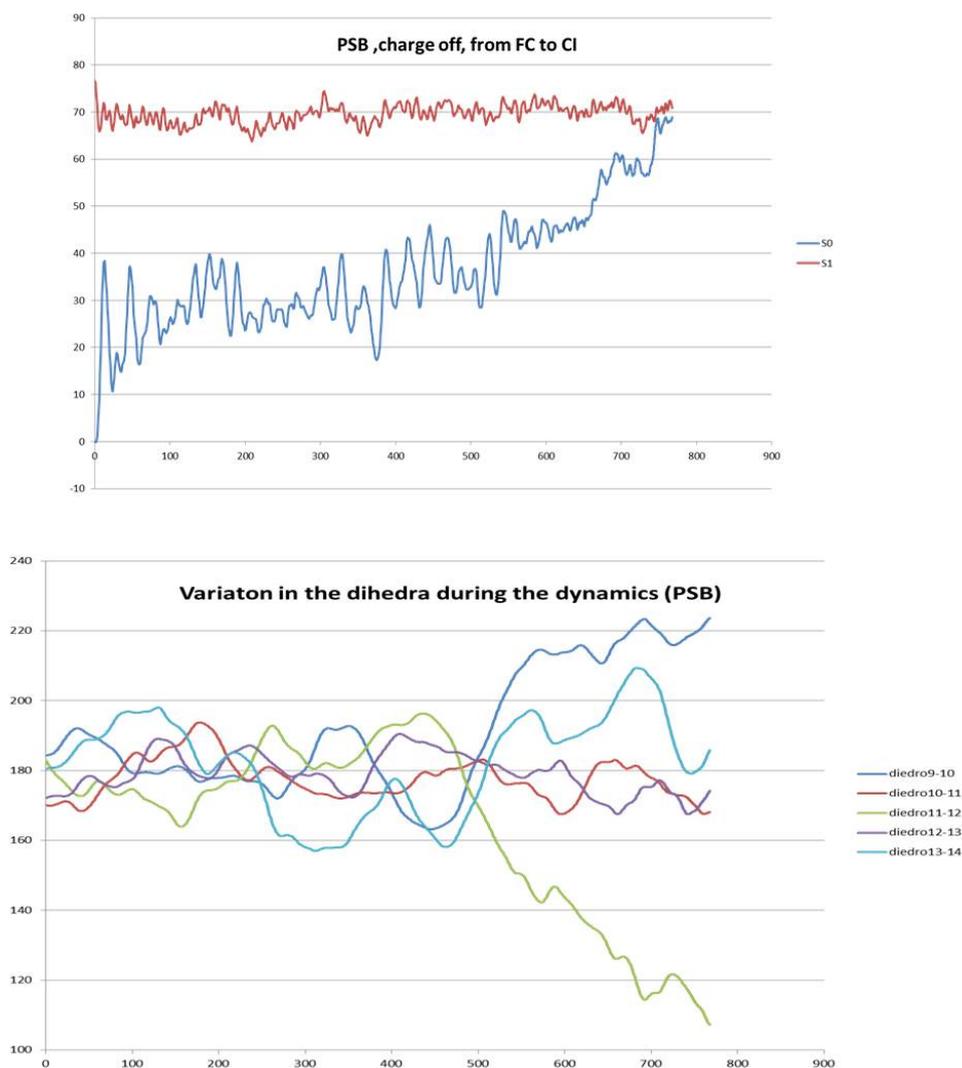
Figure 6.11: Difference between the ground state energy and that of the S1 state for the system RPSB in solvent without charges. The analysis of the variation of the dihedrals is depicted at the bottom: the C10-C11-C12-C13 dihedral rotation occurs along with the C8-C9-C10-C11 and C12-C13-C14-C15 dihedrals rotation.

7 Implementation of the developed tools for QM/MM calculations

As previously stated, COBRAMM (Computational Bridge between Ab-initio and Molecular mechanic) is a modular software aimed at performing hybrid QM/MM calculations. Modularity and flexibility are two of its leading characteristics. The program has been devised to be an interface between different programs that execute single phases of the whole calculation and exchange data by means of the interface itself and providing their seamless integration. This way the user can choose the most suitable programs to tailor the calculation steps to the requirements of the specific computational process required and chemical problem to study. COBRAMM currently supports recognized tools such as Amber and Molcas among the others, but support for other tools to come can be ensured by adding the supporting Python scripts to the COBRAMM system. The following section details the implementation for QM/MM conical-intersection optimizations in COBRAMM.

7.1 Conical intersection optimization in COBRAMM

The gradient projection method is a robust and fast method for locating minimum energy CI. It was originally proposed by Bearpark and co-workers [165] and is implemented in a number of quantum-chemical packages. The geometry optimization procedure follows an effective gradient g composed of two terms:

$$\mathbf{g} = 2(E_I - E_J)\mathbf{g}_{IJ}/|\mathbf{g}_{IJ}| + \mathbf{P}_{IS}(\nabla_R E_2) \qquad (7.1)$$

In the first term g_{IJ} is the difference vector of the gradients of states I and J. The multiplication of the normalized gradient difference vector times the energy difference $E_I - E_J$ assures that the first term vanishes when both states become degenerate. The second term represents the projection of the excited state gradient onto the orthogonal complement to the plane spanned by the gradient difference g_{IJ} and the scaled derivative coupling

vector $\mathbf{h}_{IJ} = (E_I - E_J)\langle\Psi_J|\nabla_R|\Psi_I\rangle$ .Both vectors define to first order approximation the branching plane in which a geometry deformation lifts the degeneracy between states I and J. Following the projection of the gradient onto the subspace orthogonal to the branching space ensures that the optimized structure is the lowest degenerate point. Technically the projection is performed via the projection matrix $\mathbf{P}_{IS} = 1 - \mathbf{A}(\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T$, where A is a 3Nx2 matrix with columns g_{IJ} and h_{IJ}. Implementing the procedure at QM/MM level is straightforward. Thereby, the gradient difference vector g_{IJ} is constructed out of the gradients of the QM and movable MM subsystems $\mathbf{g}_{IJ}^{QM}$ and $\mathbf{g}_{IJ}^{MM}$ ($\mathbf{g}_{IJ} = \left[\mathbf{g}_{IJ}^{QM}, \mathbf{g}_{IJ}^{MM}\right]$). The construction of the projection matrix P requires also the knowledge of h_{IJ} over the QM and MM sub-systems. As computing hIJ requires a wavefunction description of the system, $\mathbf{h}_{IJ}^{QM}$ is set to zero ($\mathbf{h}_{IJ} = \left[\mathbf{j}_{IJ}^{QM}, \mathbf{0}^{MM}\right]$). No weighting factors are used. [190]. The implemented algorithm includes the effect of the MM atoms on the structure and energy of the CI both implicitly via the one electron Hamiltonian (electrostatic embedding) and explicitly through the force on the movable MM atoms $\mathbf{g}_{IJ}^{MM}$. The only approximation is the neglect of structural deformations in the movable MM sub-system, which couple states I and J for the contruction of h_{IJ} and, thus, for the construction of the projection matrix $\mathbf{P}_{IS}$. The tool for the CI optimization uses the program Molpro for the calculation of all the QM gradients and a subroutine was implemented in Cobramm for the calculation of the of the QM/MM gradients and of g such as described in (1). In detail the steps are at first calculation of the QM gradients by Molpro, after that, the new script in Cobramm, using the QM gradients, calculates (1) for the QM/MM system and at the last the g (1) is passed to Gaussian for the geometrical parameters optimization.

7.1.1   Evaluation of the performance of the script

A test to verify the correct results and correct operating principle of the total script was done. To this aim, the CI of a very well-known system was calculated with Cobramm. The CI of 2,3-di-tert-butylbuta-1,3-diene was found both at a QM level, with Molcas and Gaussian09, and a QM/MM level with Cobramm using a 6-31g^* basis set and a 4-electron, 4-orbital active space (the butadiene moiety included in the QM layer, while the two tert-butyl substituents described at the MM level in the medium layer). The aim was to compare the results, in term of geometrical parameters, obtained at a full QM level with the one obtained using a QM/MM setup with Cobramm. The results shown in the table shown in figure 7.1 evidence

	TotQMMolpro	QM/MM	TotQM Gaussian	Other QM/MM
R 1-2	1,467	1,465	1,471	1,472
R 2-3	1,439	1,437	1,444	1,449
R 3-4	1,467	1,468	1,469	1,470
A 1,2,3	106,8°	107,3°	106,6 °	105,1°
A 3,4,5	115,07°	114,5°	115,3 °	114,8°
D 1,2,3,4	110,6°	109,5°	110,2°	Np

Figure 7.1: The table shows the geometrical parameters obtained at different levels of calculations with different QM programs. The geometrical parameters taken into account are bond length, angle and dihedral angle. The number 1, 2, 3 and 4 are referred to the C atom in the principle chain of the molecule, as shown in fig. 7.2

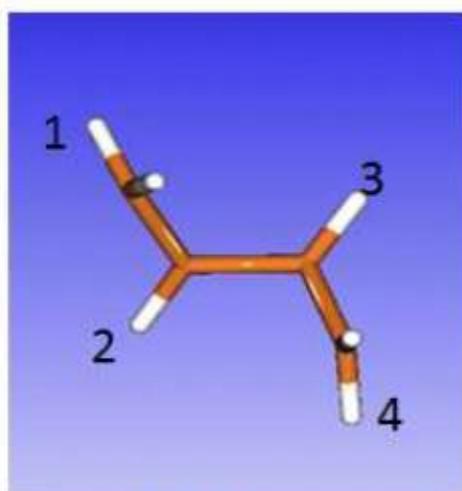that the there is good agreement between the results obtained with the tools utilized.

Figure 7.2: Reference molecule for QM/MM calculation performance evaluation

7.2   IRC in COBRAMM

In order to verify the nature of a transition state that has been optimized
with one of the local methods described before, the Hessian needs to display
the required number of negative eigenvalues. Aside from this local criterion,
it is also necessary to identify the minima connected through the transition
state. This latter part is usually performed through calculation of some kind
of reaction coordinate. Intrinsic Reaction Coordinate (IRC) is defined as the
minimum energy reaction pathway (MERP) in mass-weighted Cartesian co-
ordinates between the transition state of a reaction and its reactants and
products. It can be thought of as the path that the molecule takes mov-
ing down the product and reactant valleys with zero kinetic energy. The
Gonzalez-Schlegel method for following the IRC is used in Gaussian using
the *irc* keyword. A routine aimed at executing the IRC on QM/MM sys-
tems has been implemented within the COBRAMM package. In particular,
the CBF.py script has been modified. Such script carries the information
that are used to create the input files for the programs that are launched
by COBRAMM to perform the requested calculations. According to a spe-
cific keyword in the COBRAMM input file (cobram.command), the CBF.py
script creates an ad-hoc input file for Gaussian, which is the application
which actually effects the IRC on the QM/MM system by means of a sys-
tem QM/MM gradient computed at step 0 by COBRAMM. The generated
input file contains both the required and the optional keywords to run the
applications (that can be modified by the cobram.command script) and also
the properly-formatted QM/MM gradient, so that it can be correctly in-
terpreted by Gaussian to carry out the IRC. A supplementary option has
been added during step 0 in order to have the IRC follow a specified vector,
assigned as a gradient. Such gradient is written on the gradient.dat file, and
it can be computed by means of a script based on the *.xyz* files which con-
tain the description both of the desired starting and target geometries. The
IRC thus implemented, in the two mentioned options, have been verified
and validated on the GMP system in solvent and subsequently used by the
staff of the theoretical photochemistry and photophysics research group.

7.3   Simulated transient spectra

The developed program is capable of building transient spectra, using the
energies and the oscillator strengths for different state transitions at differ-
ent times, as a superposition of Gaussian functions. Basically, a multidi-
mensional Gaussian spectrum is a plot representing a function in the form

$Z = f(x, y)$. In this case, x represents time; y is the state transition as will be described later in greater detail. Z is the result of the calculation. Its value is color-coded according to a color bar put besides the plot. An example of such Gaussian spectrum is depicted in figure 7.3. In order to
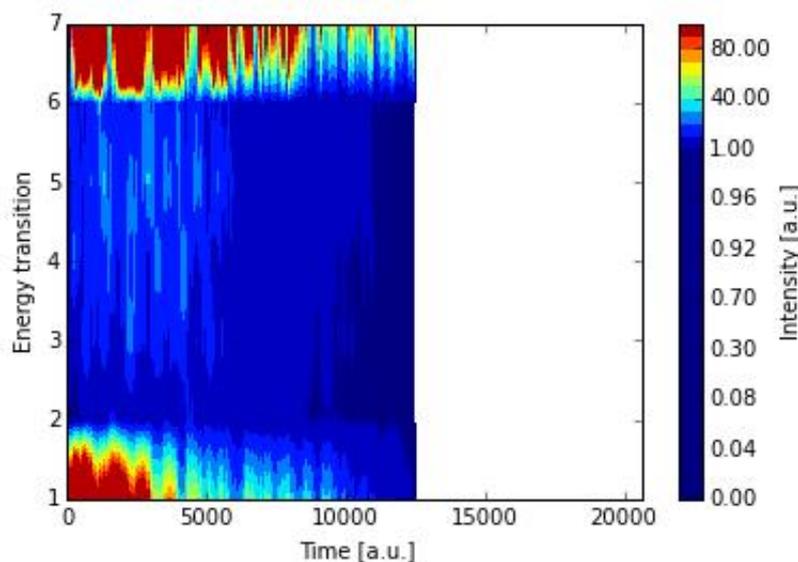
Figure 7.3: A typical spectrum obtained

ease visually spotting the higher- and lower-value areas, a contour plot has been used instead of a simpler scatter plot. The system developed consists of two Python scripts:

- Spectramaker.py, which reads the dynamics files, computes the multi-dimensional gaussians and eventually writes them to a file;

- Spectraplot.py, which reads the multidimensional gaussians file and displays the spectrum. For this purpose Spyder has been used. Spyder is part of Anaconda, a multi-platform Python interpreter that includes graphical output support. The MS Windows version has been used;

The workflow is depicted in figure 7.4.

7.3.1   Spectramaker.py

Spectramaker.py is the script which produces the multidimensional Gaussian spectrum output file based on the input data characterizing the dynamics. The script expects data to be organized as follows:

Figure 7.4: Workflow of the multidimensional Gaussian spectrum computation

- A base directory, thereon marked as {basedir}, exists where the script itself is located.

- The overall dynamics is composed of several dynamics (from different starting geometries), whose description is found in the {basedir}/geom_nnn/dinamica/out subdirectories.

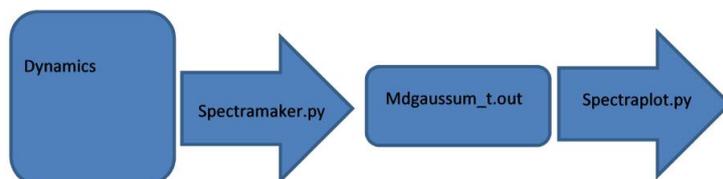- Several subdirectories of the {basedir}/geom_nnn/dinamica/out subdirectory exist, one for each sampling time of the dynamics, i.e. one for each time where state transition energies and oscillator strengths have been determined for the dynamics; each of these time subdirectories is named after the point in time (expressed in femtoseconds) when that particular sampling has occurred. Each subdir name is caspt2_nnn.n, where nnn.n is the sampling time expressed in femtoseconds. Moreover each time subdirectory shall contain the respective caspt2.out file as produced by MOLCAS.

A file named spectra_config.dat can be edited to override the default configuration parameters of spectramaker. The parameters that are customizable in the spectra_config.dat file are:

- ROOT: the base state number

- STOPTIME: the final sampling time (in femtoseconds) at which oscillator strength and state energy are available

- TIMEDELTA: the sampling step size, i.e. the time interval between two consecutive sampling times

- FWHM: the full-width half-maximum value of the LASER

- T_INIT: the simulation starting time for the computation of the multidimensional gaussians

- T_FINAL: the simulation end time

- T_STEP: the simulation step size.

Given a base state (the *ROOT* parameter in spectra_config.dat file) , the script computes the oscillator strengths and energy differences during transitions from the base state to the others: such computations are done for each geometry for each sampling time. Let $f_{osc}^n(\tau_i, x_{bj})$ be the oscillator strength of the $n-th$ geometry at sampling time τ_i for the state transition from base state b to state j. Moreover let $\Delta E^n(\tau_i, x_{bj})$ be the energy transition for the n-th geometry at sampling time τ_i for the state transition from base state b to state j. Based on each couple $\{f_{osc}, \Delta E\}$ as previously defined and given a simulation time interval $(t_{initial}, t_{final})$, multidimensional Gaussian functions are computed in the form:

$$g_{MD}(t, x_{bj}) = \sum_n \left(\sum_i f_{osc}^n(\tau_i, x_{bj}) exp\left[-\frac{(\Delta E^n(\tau_i, x_{bj}))^2 + (t-\tau_i)^2}{2FWHM_{LASER}^2} \right] \right),$$
$$t \in |t_{initial}, t_{final}| \tag{7.2}$$

7.3.2 Spectraplot.py

The Spectraplot script reads the multidimensional gaussians file (mdgaussum_t.out) constructed by Spectramaker.py and plots it. The script is based on PyPlot, which is part of MatPlotLib, a Matlab-like set of libraries for Python. Such libraries are freely available from the internet. Since a graphic output is required, a Python distribution that includes a graphic host has been used. A MS Windows-based distribution has been used for convenience sake. Such distribution is called Anaconda, and can be downloaded free of charge from https://www.continuum.io/downloads. Anaconda includes the Spyder development environment, which is the one that has been employed. Basically, Spyder is a Python environment including a script editor, a Python interpreter and an output console that can display both text and graphical outputs. A typical Spyder window is depicted in Figure 7.5: The Spectraplot script reads the mdgaussum_t.out file (produced by the Spectramaker script) and first puts the file data into a matrix. In order to correctly display the spectrum, the correct columns of the mdgaussum_t.out file have to be read. It must be made sure that the column describing the transition from the base state to itself is not read. This affects the meaning of the energy transition numbering in the output spectrum. In the example plot shown in Figure 7.3, the base state was state 2 and the spectrum is calculated using 8 roots. In detail the spectra is referred to GMP solvated. 35 non-adiabatic dynamics were realized with a PT2 correction of the energies each 2.5 fs . The program allow to work with several number of roots and using any base state.
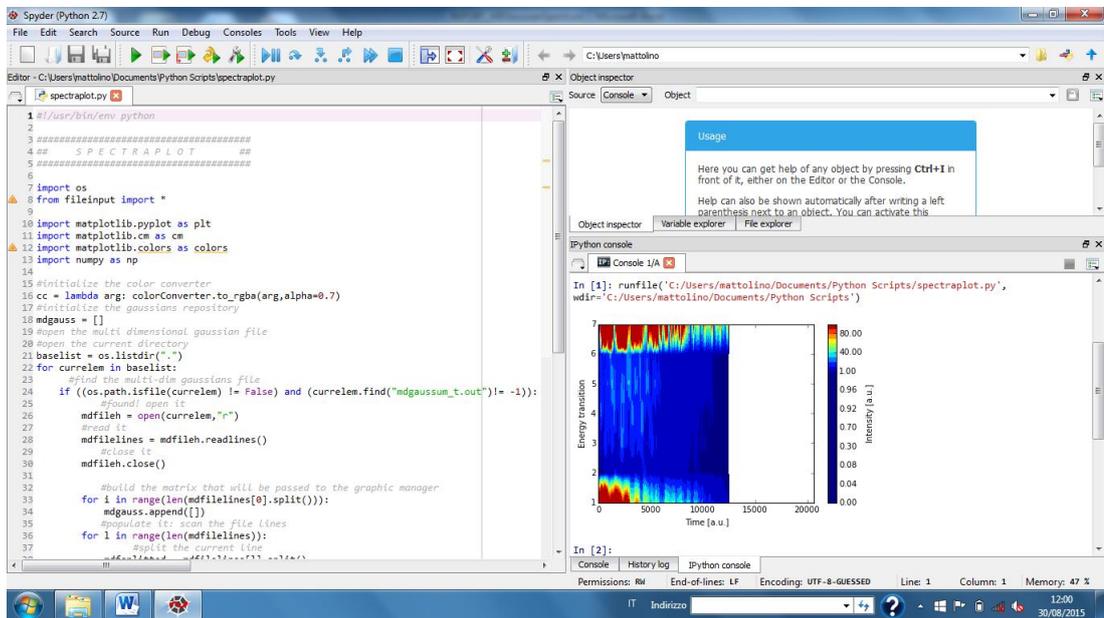
Figure 7.5: Spyder environment screenshot

8 Conclusions

Processes involved in solvated biological systems such as retinal PSB chromophores and nucleotides. Eventually, the purpose has been to provide the rationale behind their ultrafast decay and dynamics, highlighting the principles of the regulation, tuning and control of their photoactivity. This project has faced a two-fold challenge: on the one hand, a big effort has been devoted to overcome the difficulties connected to the size of such systems and to the related computational burden; on the other hand, the design of a QM/MM setup capable to properly (and realistically) model the experimental systems and reproduce the observed data has been developed, applied and generalized. Indeed, attempts have been made to account for experiments by developing ad-hoc tools and building âknow-howâ about the targeted topics. In particular, to our best knowledge, the photochemistry and photophysics of the real retinal chromophore in explicit solvent molecules has never been studied before at such a high computational level, nor the retinal photoisomerization mechanism in solvent has been yet disclosed. The lack of studies are mainly due to the extreme difficulties connected with the high number and mobility of solvent molecules. The computational protocol designed here for studying the native and 10-methylated PSB retinal chromophores, as well as GMP, in explicit solvent can be extended to other solvated chromophores (such as other retinals substituted in different positions and with different substituents) in order to predict their properties and draw guidelines for the design of molecular photoactive materials with tailored photochemical/photophysical properties for technological applications. Moreover, using the techniques, protocols and modelling tools developed here it becomes possible to simulate the transient spectra highlighting the photoinduced dynamics of those chromophores, thus assisting interpretation, understanding and rationalization of experimental data and spectra from transient spectroscopies. A tool to generate an experimental-like transient spectral map has been developed here. This tool is general and will be systematically applied to other studies involving spectral modeling and experimental comparison/interpretation. The study of nucleobases in solvent is extremely interesting and, also, necessary for the full understanding

of the photoinduced processes occurring in DNA; indeed, the environment in which DNA operates is water, and the intrinsic behavior of nucleotides in water is the cornerstone for any comparative multi-scale bottom-up study of genomic molecular materials. The study of other nucleobases, by means of the setup and the procedures that have been developed and tested for the GMP, could well be an attractive topic for future developments in order to analyze and compare the behavior of the different mononucleotides. A further future development is the study, in the same way and with QM/MM MD, of dinucleotides and DNA fragments, gradually broadening the size of the studied system in order to attain a better understanding of the underlying phenomena.

COBRAMM has been used for such simulations, and scripts have been developed ad-hoc in order to extend its capabilities for the QM/MM study of complex systems. In detail, COBRAMM scripts have been modified in order to allow QM/MM IRC (Intrinsic Reaction Coordinate) calculations and conical intersection optimizations. Such tools, which have been validated on top of the systems investigated here, have been and are nowadays used by other COBRAMM users for the study of several and diverse systems. In conclusion, it has been disclosed why the 10-methylation in all-trans RPSB retinal triggers a dramatic change in the excited state subpicosecond dynamics. The methyl group in 10-position stabilizes an excited state minimum characterized by a large charge-transfer and with alternated C-C bonds thus favoring an efficient photoisomerization. Noticeably, the photoisomerization space-saving double bicycle-pedal mechanism matches the one also observed in space constrained environments such as retinal proteins such as Rh and bR. The proposed QM/MM models are consistent with experimental results carried out by means of time-resolved spectroscopy techniques, showing a broad double-peaked stimulated emission band that matches the two excited state (EBL and ABL) minima. An detailed analysis of the photoisomerization mechanism of all-trans retinals in solution has been provided, showing that the production of the 11-cis isomer is achieved by a double bicycle-pedal mechanism, even though with a very small quantum yield. As to the GMP system, the discussed study gets ahead of the previous ones available from literature providing a photophysical and photochemical evaluation of water-solvated GMP using theoretical multireference perturbation theory QM/MM techniques. The study highlights the importance of the environment, displaying qualitative differences for the $\pi\pi^*$ L_a and $\pi\pi^*$ L_b states which are red-shifted and for the $n_O\pi^*$ state that is blue-shifted as compared to their gas-phase counterparts. The $\pi\pi^*$ L_a state is predicted to be the main spectroscopic state drives the ultra-fast deactivation processes to

which GMP undergoes upon UV-light irradiation based on a ring-puckering motion that enhances its non-adiabatic interaction with the ground state in a radiationless fashion. Moreover, a shallow stationary point towards the end of the $\pi\pi^*$ L_a MEP has been characterized, along with two different CI_s with the ground state that provide a rational basis for the two fastest decay times experimentally measured. Upon initial L_b absorption, two CIs between the $\pi\pi^*$ L_b and L_a states have also been located. The CIs between the $n_O\pi^*$ state and the $\pi\pi^*$ L_b and L_a states have also been characterized along its relaxation route, providing a minimum in the $n_O\pi^*$ state expected to vertically emit at $\approx$ 2.7 eV. Both $\pi\pi^*$ L_b and $n_O\pi^*$ are suggested to contribute to the longest-lived experimental timescale, the latter providing a lesser contribution given the relatively fast kinetic timescale and the long-lived timescales expected in $n\pi^*$ fluorescent states. The study provides a more specific molecular counterpart to the experiment and better explains the photochemical and photophysical processes promoting the intrinsic stability of the genomic material.

9 Appendix - Modifications to COBRAMM-5.2 for IRC calculations

Adding the IRC calculation involved modifications to several files of the CO-BRAMM suite. The modified files are CBF.py, cobram.py, gaussian.py, molpro.py and optxg.py.In the cobram.py file only comments have been added to ease code understanding during the preliminary code analysis. The same holds for gaussian.py where only debug writes to the standard output have been added for the understanding of the existing procedures. Notice: Line numbers in this file refer to COBRAM-beta_5.2_CI suite. A hashtag symbol in the python code snippet marks the beginning of a comment. Comments end implicitly at the end-of-line. In the mentioned files several changes have been effected in order to accomplish the conical intersection and the IRC calculations. Here they are reported and briefly discussed. Code line numbers where they are located are written in bracketed (**boldface**) characters.

9.1 Modifications to CBF.py

Modifications to CBF.py have been implemented so to add conical intersection and IRC computational capabilities to COBRAMM. A command option to instruct COBRAMM to perform CI has been added to calculation type *command[1]* array entry (**789**) and IRC step size option added to *command[65]* array entry (**237**).

```
command[65]=str(10)
```

Then energies for the two intersecting states have been computed and appended to the cobram.out file (**796-806**).

```
if (command[1] == 'ci'):
    tmp='%12f %12f\n' % (Ene_QM[0]-Uele_emb_emb,Ene_QM[1]-Uele_emb_emb)
    else:
tmp='%12f \n' % (Ene_QM-Uele_emb_emb)
```

```
if (command[1] == 'ci'):
    tmp='%12f %12f\n' % (E_tot[0],E_tot[1])#modified by Sal 03/11/2013
   else:
tmp = '%12f \n' %(E_tot)
```

```
if (command[1] == 'ci'):
tmp='%12f %12f \n' % (Ene_QM[0],Ene_QM[1])
    else:
tmp='%12f \n' % (Ene_QM)
```

The total energy has been computed and written to gradient.dat **(2449-2479)**. The gradients have been computed according to the type of calculation required: in the case of ML the CI is not implemented; else the gradient components for the QM and MM and the atomlinks are computed. Conical intersection gradient computation is also carried out based on the mentioned gradients and on the derivative coupling. This part of computation relies heavily on matrix mathematics therefore the *numpy* python package has been used **(2554-2823)** for such purpose, which allows a more straightforward treatment of matrix equations as if the matrix terms were scalar.

```
if lists[6][1]=='ci':
          DEn = totalenergy[0] - totalenergy[1]#Sal commented for debug
          #DEn = totalenergy[0] #Sal added for debug
          DE = float(DEn)
        ##open the file to be written
         f=open("gradient_ci.dat","w")
         #open the gradient source file
         S0 = open("gradient1.dat","r")
         #S1 = open("S1","r")
         S1 = open("gradient2.dat","r")
         #open dercoup
         dc = open("dercoup.dat","r")
         N = S0.readlines()
         Nl = len(N)
         #print "Nl"
         print Nl
         Nd=int(Nl)
         print 'Nd'
         print Nd
```

```
#chiudi tutti i file
#S0.close()
#S1.close()
#dc.close()
#f.close()

#now read the three matrices,notice that the first row is skipped
s0=np.loadtxt('gradient1.dat')
#s0=np.array(grad1)
#print 'grad1'
#print grad1
#s0 = np.fromfile(grad1,dtype=float,count=-1)#Sal
print 'grad1'
print s0
#s0=np.reshape(grad1,3*Nd)
#f.write("s0")
s2=np.loadtxt('gradient2.dat')
print 'grad1'
print s2
#s2=np.array(grad2)
#s2=np.fromfile(grad2,dtype=float,count=-1)#Sal
derco=np.loadtxt('dercoup.dat')
#computation of matrix difference
x1 = s0-s2
print 'scrivo x1:\n'#Sal debug
print x1#Sal debug
#scaling of the derco(x2 matrix)
x2 = DE * derco
#reopen the gradient file in append mode
#f=open("gradient_ci.dat","a")
#write matrices to such file
#f.write("x1\n")
#savetxt(f,x1,fmt='%12.10f')
#f.write("\n")
#f.write("x2\n")
#savetxt(f,x2,fmt='%12.10f')
#f.write("\n")

#compute the norm of the x1 vector
x1v=np.reshape(x1,3*Nl)
```

```python
 x1n = np.linalg.norm(x1v)
#savetxt(f,x1n,fmt='%12.10f')
#f.write("\n")
#let g = g1+g2
#compute g1
 g1=2*DE*(x1v/x1n)
#f.write("\n")
#f.write("g1\n")
 #savetxt(f,g1,fmt='%24.16f')
 #f.write("\n")
 #compute g2
 #compute P1sx2, the A matrix = x1v, P1s=I-A(A_T*A)^(-1)*A_T
 #step 1 - build the unity matrix
 I = np.eye(3*Nd)
 #step 2 - compute (A_T)
 x1v_T = np.transpose(x1v)
 #step 3 - compute (A_T * A)^(-1)
 TempVal = np.dot(x1v_T,x1v)
 TempVal = 1/TempVal
 #f.write("%24.16f"%TempVal)
 #f.write("\n")
 #f.write("\n")
 #step 4 - build up overall calculation
 x1TempMatrix = np.dot(x1v,TempVal)
 #savetxt(f,x1TempMatrix,fmt='%24.16f')
 #f.write("\n")
 P1 = I - np.dot(x1TempMatrix,x1v_T)
 #print'P1 is',P1
 x2v=np.reshape(x2,3*Nd)
 P1x2v=np.dot(P1,x2v)
 #f.write("Px2vector\n")
 #savetxt(f,P1x2v,fmt='%24.16f')
 #f.write("\n")
 #compute P(grad2)
 #step 1: compute P
 A=np.column_stack((x1v,P1x2v))
 #f.write("matrix A(x1,x2)\n")
 #savetxt(f,A,fmt='%24.16f')
 #f.write("\n")
 A_T= np.transpose(A)
```

```
        ATA=np.dot(A_T,A)
        ATAinv=np.linalg.inv(ATA)
        AATAinv=np.dot(A,ATAinv)
        P= I - np.dot(AATAinv,A_T)
        #savetxt(f,P,fmt='%24.16f')
        #f.write("\n")
        # P * grad2
        s2vec=np.reshape(s2,3*Nd)
        g2=np.dot(P,s2vec)
        #f.write("g2")
        #savetxt(f,g2,fmt='%24.16f')
        #f.write("\n")

        #G final
        G=g1+g2
        #f.write("final G\n")
        #savetxt(f,G,fmt='%24.16f')
        #f.write("\n")
        #write the matrix G
        G_reshaped=np.reshape(G,(-1,3))
    for i in range(len(lists[7])):
        x.append(G_reshaped[i][0])
            y.append(G_reshaped[i][1])
            z.append(G_reshaped[i][2])
        np.savetxt(f,G_reshaped,fmt='%24.16f')
        f.write("\n")
        f.close()
```

The model-H gradient has also been calculated taking into account the two states **(2842-2966)**. The new *CI* optimizer has been included among the iteration step options **(3576-3578)**.

```
elif lists[6][1] in ['optxg','freqxg','fullfreqxg','bomdxg', \\
                     'fullbomdxg','irc','ci'] :
    print 'step '+str(step)+' completed\n'
    optxg.checkconvergency(start,step,3)
```

Moreover, the charge computation for the CI case has been added, thus taking into account the two state energies **(3676-3684)**.

9.2   Modifications to molpro.py

Modifications to molpro.py include the different definition of the gradch array for accommodating the gradients for two states **(286-298)**.

```
#Different gradch management according to number of gradients
    if lists[6][1] == 'ci':
      gradch = [[[],[],[]],[[],[],[]]]
      current_linescount=0
      current_state_found=1
    else:
      gradch=[[],[],[]]
```

Derivative coupling data retrieving and parsing has been added **(417-445)**, of the charges to make gradient **(467-482)**:

```
if lists[6][1]=='ci':
        current_linescount=current_linescount+1
        print "current_linescount", current_linescount
        print "charge", chMED[a]
        if current_linescount == (len(chMED)/2)+1 :
          current_state_found=2
        print "current_state_found", current_state_found
        gradch[current_state_found-1][0].append(float(tmp2[2])*(-chMED[a]))
        gradch[current_state_found-1][1].append(float(tmp2[3])*(-chMED[a]))
        gradch[current_state_found-1][2].append(float(tmp2[4])*(-chMED[a]))
    else:
        gradch[0].append(float(tmp2[2])*(-chMED[a]))
        gradch[1].append(float(tmp2[3])*(-chMED[a]))
        gradch[2].append(float(tmp2[4])*(-chMED[a]))
    a=a+1
```

and of energy in the case of *CI* management **(1500-1663)** . Correct molpro command generation modified **(1732)**.

```
system('molpros9 -d$PWD/SINGLEPOINT -I$PWD/SINGLEPOINT \\
 -W$PWD/SINGLEPOINT < SINGLEPOINT/molpro.input > SINGLEPOINT/molpro.log 2>&1')
```

9.3   Modifications to optxg.py

Modifications to optxg.py have been carried out to include IRC calculations. First, *irc* command recognition has been added to the *command[1]* calcu-

lation option data register. This also creates the following input key for Gaussian **(280-283)**:

```
elif command[1]  in  ['irc'] and len(inputkey)==0:
        inputkey=['#p IRC=(FCCards,Readvector,MaxPoints='+maxoptstep+ \\
        ',StepSize='+command[65]+') Use=L115']
```

thus using the specified step size previously mentioned. Further Irc calculation data preparation is reported at lines **(327-440)**.

```
if command[1]=='irc':
        #geom.write('sono dentro\n')
         #script altra volta(cobram.irc)
         ##open the file to be written
         #f=open("gaussian.com","w")
         #open the gradient source file
         grad = open("gradient.dat","r")

         #write the header of gaussian.com
         #geom.write("First line of the header\n")
         #geom.write("Second line of the header\n")


         #read the gradient source file
         #the first line of the file is ENERGY
         #read it
         Energy = grad.readline()
         #the second line is the number of atoms
         NAT = grad.readline()
         NATd=int(NAT)

         #now write them to the matrix.dat file in the correct format
         float_var=float(Energy)
         geom.write("%24.16f\n"%float_var)

         #now read the gradient values
         k=1
         while 1:
          line_read=grad.readline()
          if not line_read:
           break;
```

```
      else:
       split_line=line_read.split()
       i=0
       #geom.write("Length of line %d"%len(split_line))
       while i<len(split_line):
         float_var=float(split_line[i])
         geom.write("%12.8f"%float_var)
         if k==6:
           geom.write("\n")
           k=1
         else:
           k=k+1
         i=i+1
      if NATd*3%6 != 0:
 geom.write("\n")
      #insert a blank row as a spacing
      #geom.write("\n")
      #NATd=2 per prova
      ##create the matrix
      #geom.write('inizio a creare matrice\n')
      m=[ [ 0 for h in range(3*NATd+1) ] for l in range(3*NATd+1) ]
      #print m
      i=0
      while i<=3*NATd:
       j=0
       while j<=i:
        if j==i:
          m[i][j]=1
        else:
          m[i][j]=0
        j=j+1
       i=i+1
      #print m
      ##write the matrix in the correct format
      k=0
      i=1
      while i<=3*NATd:
       j=1
       while j<=i:
        geom.write ("%12.8f"%(m[i][j]))
```

```
   k=k+1
   if k==6:
    geom.write("\n")
    k=0
   j=j+1
  i=i+1
 if (((NATd*3)*(NATd*3+1))/2)%6 != 0:
   geom.write("\n")
 #insert a blank row as a spacing
 grad.close()
 #reopen the file
 grad=open("gradient.dat","r")
 i=0
 while i<2:
  line_read=grad.readline()
  i=i+1
 #now we read the gradient
 #Energy=grad.readline()
 #NAT=grad.readline()
#print "riprendo il gradiente"
 geom.write("\n")
 k=0
 while 1:
  line_read=grad.readline()
  if not line_read:
    break
  else:
    split_line=line_read.split()
    i=0
    while i<len(split_line):
      float_var=float(split_line[i])
      geom.write("%10.6f"%float_var)
      k=k+1
      if k==8:
        geom.write("\n")
        k=0
      i=i+1
 grad.close()
```

Bibliography

[1]   Lasorne B., Worth G.A., and Robb M.A. "Excited-state dynamics". In: *WIREs Comput Mol Sci* 1 (2011), pp. 460–475.

[2]   Olivucci M. Bernardi F. and Robb M.A. In: *Chem. Soc. Rev.* 25 (1996), pp. 321–328.

[3]   Teller E. In: *J. Phys. Chem.* 41 (1937), p. 109.

[4]   Zewail A. H. "Femtochemistry: Ultrafast Dynamics of the Chemical Bond". In: *World Scientific: Singapore* 1-2 (1994).

[5]   Truhlar D.G. Lin H. In: *Theoretical Chemistry Accounts* 2 (2007).

[6]   Sakmar T.-Mathies R. Kochendoerfer G. Lin S. In: *Trends Biochem. Sci.* 24 (1999), p. 300.

[7]   de Groot H. Lugtenburg J. DeGrip W. Verhoeven M. Bovee-Geurts P. In: *J. Mol. Biol.* 98 (2006), p. 363.

[8]   Sheves M. Ottolenghi M. In: *J. Membr. Biol.* 112 (1989), p. 193.

[9]   Becker R. S. In: *Photochem. Photobiol.* 48 (1988), p. 369.

[10] Hoffmann M. Wanko M. et al. In: *J. Phys. Chem. B* 109 (2005), p. 3606.

[11] Ames J. Pollard W. Mathies R. Lin S. In: *Annu. Rev. Biophys. Biophys. Chem.* 20 (1991), p. 491.

[12] Lugtenburg J. Dawadi P. In: *Molecules* 15 (2010), p. 1825.

[13] Olivucci M. Garavelli M. Altoe P. Cembran A. In: *Proc. Natl. Acad. Sci. U.S.A.* 107 (2010), p. 20172.

[14] Warshel A. In: *Nature* 260 (1976), p. 679.

[15] Filippi C. Valsson O. In: *J. Chem. Theory Comput.* 6 (2010), p. 1275.

[16] Callis P. R. "Electronic states and luminiscence of nucleic-acid systems". In: *Annu. Rev. Phys. Chem.* 34 (1983), pp. 329–357.

[17] Kohler B. "Nonradiative decay mechanisms in DNA model systems". In: *J. Phys. Chem.* 1 (2010), pp. 2047–2053.

[18] Ravana J.-L. Cadet J. Mouret S. and Douki T. "Photoinduced damage to cellular DNA: direct and photosensitized reactions". In: *Photochem. Photobiol.* 88 (2012), pp. 1048–1065.

[19] Wolnicka-Glubisz A. Anver M. R.-Bahn J. Wielgus A. Noonan F. P. Zaidi M. R. et al. "Melanoma induction by ultraviolet A but not ultra-violet B radiation requires melanin pigment". In: *Nat. Commun.* 3 (2012), p. 884.

[20] Brash D. E. "UV signature mutations". In: *Photochem. Photobiol.* 91 (2015), pp. 15–26.

[21] Kamiya Y. and Asanuma H. "Light-driven DNA nanomachine with a photoresponsive molecular engine". In: *Acc. Chem. Res.* 47 (2014), pp. 1663–1672.

[22] Serrano-Andres L. and Merchan M. "Are the five natural DNA/RNA base monomers a good choice from natural selection? A photochemical perspective". In: *J. Photochem. Photobiol. C Photochem. Rev.* 10 (2009), pp. 21–32.

[23] McGown L. B. Gupta A. P. Taylor W. J. and Kempf J. G. "NMR studies of the chiral selectivity of self-assembled guanosinemonophosphate". In: *J. Phys. Chem. B.* 118 (2014), pp. 14243–14256.

[24] Tannahill D. Lam E. Y. Beraldi D. and Balasubramanian S. "G-quadruplex structures are stable and detectable in human genomic DNA". In: *Nat. Commun.* 4 (2013), p. 1796.

[25] Merchan M. Serrano-Andres L. and Borin A. C. "A three-state model for the photophysics of guanine". In: *J. Am. Chem. Soc.* 130 (2008), pp. 2473–2484.

[26] Szymczak J. J. Barbatti M. et al. "The decay mechanism of photoexcited guanine - a nonadiabatic dynamics study". In: *J. Chem. Phys.* 134 (2011).

[27] Stenta M. Garavelli M. Conti I. Altoe P. and Orlandi G. "Adenine deactivation in DNA resolved at the CASPT2/CASSCF/AMBER level". In: *Phys. Chem. Chem. Phys.* 12 (2009), pp. 5016–5023.

[28] Cramer C.J. *Computational Chemistry.* Wiley, 2004.

[29] Hartree D.R. In: *Proc. Cambridge Phil. Soc.* 24, 89, 111, 426 (1928).

[30] Slater J.C. In: *Phys. Rev.* 43, 830 (1930).

[31] Roothaan C.C.J. In: *Rev. Mod. Phys.* 23,69 (1951).

[32] Zerner M.C. In: *Theor. Chem. Acc.* 103, 219 (2000).

[33] Ostlund N.S. Szabo A. *Modern quantum chemistry: an introduction to advanced electronic structure theory*. McGraw Hill, 1989.

[34] Roos B.O. "European Summer School in Quantum Chemistry". In: *Lund University* 2 (2005).

[35] U. Burkert and V. Allinger. "Molecular Mechanics". In: *ACS Monograph* 177 (1982).

[36] AK Rappe and CJ. Casewit. *Molecular Mechanics Across Chemistry*. University Science Books, 1997.

[37] Thiel W. Bakowies D. In: *J. Phys. Chem* 100 (1996), pp. 10580–10594.

[38] Boyd D.B. Gao J. Lipkowitz K.B. In: *Reviews in Computational Chemistry* 7 (1995), pp. 119–185.

[39] Grotendorst J. Sherwood P. In: *NIC series* 3 (2000), pp. 285–305.

[40] Truhlar D.G. Lin H. In: *Theoretical Chemistry Accounts* 2 (2007), p. 117.

[41] Levitt M. Warshel A. In: *Journal of Molecular Biology* 103 (1976), pp. 227–249.

[42] Truhlar D.G. Gao J. In: *Annual Review of Physical Chemistry* 53 (2002), pp. 467–505.

[43] de Vries A.H. Sherwood P. et al. In: *Journal of Molecular Structure* 632 (2003), pp. 1–28.

[44] Oesterhelt D. and Stoeckenius W. In: *Proc Natl Acad Sci USA* 70 (1973), pp. 2853–2857.

[45] M. J. Frisch, G. W. Trucks, et al. *Gaussian 03, Revision C.02*. Gaussian, Inc., Wallingford, CT, 2004.

[46] Cheatam T.E. Case D.A. et al. In: *Journal of Computational Chemistry* 26 (2005), p. 26.

[47] Ponder J.W. Dudek M.J. In: *Journal of Computational Chemistry* 16 (1995), pp. 791–816.

[48] Bottoni A. Altoe P. Stenta M. and Garavelli M. "A tunable QM/MM approach to chemical reactivity, structure and psyco-chemical properties prediction". In: *Theor. Chem. Acc.* 118 (2007), pp. 219–240.

[49] Schlegel H.B. Hratchian H.P. "Finding minima, transition states, and
      following reaction pathways on ab initio potential energy surfaces".
      In: *Theory and Applications of Computational Chemistry: the First
      Forty Years*. Ed. by Frenking G. Dykstra C.E. et al. Elsevier, 2005,
      pp. 195–249.

[50] Wales D.J. *Energy Landscapes*. Cambridge University Press, 2003.

[51] Heidrich D. *The Reaction Path in Chemistry: Current Approaches
      and Perspectives*. Kluwer, 1995.

[52] Haile J.M. *Molecular Dynamics Simulation: Elementary Methods*.
      John Wiley and Sons, 1992.

[53] Hase W.L. (ed.) *Advances in Classical Trajectory Methods*. JAI Press,
      1992.

[54] Bunker D.L. "Classical trajectory methods". In: *Methods Comput.
      Phys.* 10 (1971), p. 287.

[55] Cross P.C. Wilson E.B. Decius J.C. *Molecular Vibrations: the Theory
      of Infrared and Raman Vibrational Spectra*. Dover Publications, 1980.

[56] Radom L. Hehre W.J. et al. *Ab Initio Molecular Orbital Theory*. John
      Wiley and Sons, 1980.

[57] Pulay P. Fogarasi G. "Geometry optimization in redundant internal
      coordinates". In: *J. Chem. Phys.* 96 (1992), pp. 2856–2860.

[58] Pulay P. Baker J. "Efficient geometry optimization of molecular clus-
      ters". In: *J. Comput. Chem.* 21 (2000), pp. 69–76.

[59] Thiel W. Billeter S.R. Turner A.J. "Linear scaling geometry opti-
      misation and transition state search in hybrid delocalised internal
      coordinates." In: *Phys. Chem. Chem. Phys.* 2 (2000), pp. 2177–2186.

[60] Baker J. Paizs B. et al. "Geometry optimization of large biomolecules
      in redundant internal coordinates." In: *J. Chem. Phys.* 113 (2000),
      pp. 6566–6572.

[61] Schlegel H.B. Kudin K.N. Scuseria G.E. "A redundant internal coor-
      dinate algorithm for optimization of periodic systems." In: *J. Chem.
      Phys.* 114 (2001), pp. 2919–2923.

[62] Angyan J.G. Bucko T. Hafner J. "Geometry optimization of periodic
      systems using internal coordinates." In: *J. Chem. Phys.* 122 (2005),
      p. 124508.

[63] Schnabel R.B. Dennis J.E. *Numerical Methods for Unconstrained Op-
      timization and Nonlinear Equations*. Prentice-Hall, 1983.

[64] Scales L.E. *Introduction to non-linear optimization*. Springer Verlag, 1985.

[65] Fletcher R. *Practical Methods of Optimization, 2nd ed.* John Wiley and Sons, 1987.

[66] Morokuma K. Maeda S. Ohno K. "An automated and systematic transition structure explorer in large flexible molecular systems based on combined global reaction route mapping and microiteration methods." In: *J. Chem. Teory Comput.* 5 (2009), pp. 2734–2743.

[67] Ayala P.Y. Peng C.Y. et al. "Using redundant internal coordinates to optimize equilibrium geometries and transition states." In: *J. Comput. Chem.* 17 (1996), pp. 49–56.

[68] Fukui K. In: *Acc. Chem. Res.* 14 (1981), p. 363.

[69] Schlegel H.B. Gonzales C. In: *J. Phys. Chem* 94 (1990), p. 5523.

[70] Longuet-Higgins H.C. Herzberg G. In: *Trans. Faraday Soc.* 35 (1936), p. 77.

[71] Amara P. Gao J. et al. In: *J. Phys. Chem. A* 102 (1998), pp. 4714–4721.

[72] Neese F. Sinnecker S. In: *Journal of Computational Chemistry* 27 (2006), pp. 1463–1475.

[73] Truhlar D.G. Pu J. Gao J. In: *J. Phys. Chem. A* 108 (2004), pp. 5454–5463.

[74] Lindh R. Karlstrom G. et al. In: *Computational Materials Science* 28 (2003), pp. 222–239.

[75] Bar M. Ahlrichs R. et al. In: *Chemical Physics Letters* 162 (1989), pp. 165–169.

[76] Neese F. In: *J. Am. Chem. Soc* 128 (2006), p. 10213.

[77] Kruizinga W. Schoevaars A. et al. In: *The Journal of Organic Chemistry* 62 (1997), pp. 4943–4948.

[78] Silva R.A. Kelly T.R. De Silva H. In: *Nature* 401 (1999), pp. 150–152.

[79] Zijlstra R.W.J. Koumura N. et al. In: *Nature* 401 (1999), pp. 152–155.

[80] Wu C. Duan Y. et al. In: *Journal of Computational Chemistry* 24 (2003), pp. 1999–2012.

[81] Mislow K. Iwamura H. In: *Accounts of Chemical Research* 21 (1988), pp. 175–182.

[82] Moore J.S. Bedard T.C. In: *Journal of the American Chemical Society* 117 (1995), pp. 10662–10671.

[83] Bowyer M.C. Kelly T.R. et al. In: *Journal of the American Chemical Society* 116 (1994), pp. 3657–3658.

[84] Tellitu I. Kelly T. Sestelo J. In: *Journal of Organic Chemistry* 63 (1998), pp. 3655–3665.

[85] Morokuma K. Vreven T. et al. In: *Journal of Computational Chemistry* 24 (2003), pp. 760–769.

[86] Braun-Sand S. Klahn M. et al. In: *J. Phys. Chem. B* 109 (2005), pp. 15645–15650.

[87] Kollman P.A. Singh U.C. In: *Journal of Computational Chemistry* 7 (1986), pp. 718–730.

[88] Karplus M. Field M.J. Bash P.A. In: *Journal of Computational Chemistry* 11 (1990), pp. 700–733.

[89] T.E. Alder B.J. Wainwright. In: *J. Chem. Phys.* 31 (1959), pp. 459–466.

[90] van Gunsteren W.F. Koehler J.E.H. Saenger W. In: *J. Mol. Biol.* 203 (1988), pp. 241–250.

[91] Egberts E. In: *Masters thesis, Univ. of Groeningen* (1988).

[92] Kollman P. Singh U.C. Weiner S.C. In: *Procl. Natl. Acad. Sci. USA* 82 (1985), pp. 755–759.

[93] Harvey S.C. et al. In: *Science* 223 (1984), pp. 1189–1191.

[94] Hermans J. (ed.) Polycristal Bookservice, 1985.

[95] Rao S.N. and others. In: *Nature* 328 (1987), pp. 551–554.

[96] van Gunsteren W.F. Escom Science Publishers, 1989, pp. 27–29.

[97] Jorgensen W.L. et al. In: *J. Chem. Phys.* 79 (1983), p. 926.

[98] Shanzer A. Lifson S. Felder C.E. In: *J. Am. Chem. Soc.* 105 (1983), p. 3866.

[99] Mazzarella L. Giglio E Liquori A.M. In: *Nuovo Cim.* 1 (1969), p. 135.

[100] Giglio E. Di Nola A. In: *Acta Crystallogr. A* 26 (1970), p. 144.

[101] Tildesly D.J. Allen M.P. Oxford University Press, 1989.

[102] Mandziuk M. Schlick T. Barth E. In: *Annu. Rev. Biomol.* 26 (1997), pp. 181–222.

[103] Berendsen H.J.C. Ryckaert J.P. Ciccotti G. In: *J. Comp. Phys.* 23 (1977), pp. 327–341.

[104] Verlet L. In: *Phys. Rev.* 159 (1967), p. 98.

[105] Verlet L. In: *Phys. Rev.* 165 (1968), pp. 201–223.

[106] Gao J. Garcia-Viloca M. et al. In: *Science* 303 (2004), pp. 186–195.

[107] K. E. Ranaghan and A. J. Mulholland. "Investigations of enzyme-catalysed reactions with combined quantum mechanics/ molecular mechanics (QM/MM) methods". In: *Int. Rev. Phys. Chem.* 29 (2010), 65â133.

[108] H. M. Senn and W. Thiel. In: *Quantum Chem. Mol. Simul.* 268 (2007), 173â290.

[109] Bonacic-Koutecky V. Michl J. *Electronic Aspects of Organic Photochemistry*. Wiley, 1990.

[110] Michl J. Klessinger M. *Excited States and Photochemistry of Organic Molecules*. Cambridge, 1995.

[111] Scaiano JC Turro NJ Ramamurthy V. *Excited States and Photochemistry of Organic Molecules*. Cambridge, 1995.

[112] Fukui K. "A formulation of the reaction coordinate". In: *J Phys Chem* 74 (1970), pp. 4161–4163.

[113] Hutter J. Marx D. *Ab Initio Molecular Dynamics: Basic Theory and Advanced Methods*. Kluwer, 1995.

[114] Tannor D.J. *Introduction to Quantum Mechanics: a Time-Dependent Perspective*. University Science Books, 2007.

[115] Worth G Meyer H-D Gatti F. *Multidimensional Quantum Dynamics: MCTDH Theory and Applications*. Wiley-VCH, 2009.

[116] Zimmerman H.E. "Molecular orbital correlation diagrams, Mobius systems, and factors controlling ground- and excited-state reactions". In: *II. J Am Chem Soc* 88 (1966), pp. 1566–1567.

[117] Michl J. "Photochemical reactions of large molecules. I. A simple physical model of photochemical reactivity". In: *Mol Photochem.* 4 (1972), pp. 243–256.

[118] Cederbaum LS Koppel H Domcke W. "Multi-mode molecular dynamics beyond the Born-Oppenheimer approximation". In: *Adv Chem Phys* 57 (1984), pp. 59–246.

[119] Yarkony DR. "Current issues in nonadiabatic chemistry". In: *J Phys Chem* 100 (1996), pp. 18612–18628.

[120] Truhlar DG Ben-Nun M Martinez TJ Hack MD Wensmann AM. "Comparison of full multiple spawning, trajectory surface hopping, and converged quantum mechanics for electronically nonadiabatic dynamics". In: *J Phys Chem* 115 (2001), pp. 1172–1186.

[121] Robb MA Worth GA Hunt P. "Nonadiabatic dynamics: a comparison of surface hopping direct dynamics with quantum wavepacket calculations". In: *J Phys Chem A* 107 (2003), pp. 621–631.

[122] Robb MA Worth GA. "Applying direct molecular dynamics to non-adiabatic systems." In: *Adv Phys Chem* 124 (2002), pp. 355–431.

[123] Cederbaum LS Worth GA. "Beyond Born- Oppenheimer: molecular dynamics through a conical intersection". In: *Annu Rev Phys Chem* 55 (2004), pp. 127–158.

[124] Zhu CY Truhlar DG Jasper AW Nangia S. "Non- Born-Oppenheimer molecular dynamics". In: *Acc Chem Res* 39 (2006), pp. 101–108.

[125] Lasorne B Worth GA Robb MA. "Solving the time dependent Schroedinger equation for nuclear motion in one step: direct dynamics of non-adiabatic systems". In: *Mol Phys* 106 (2008), pp. 2077–2091.

[126] Koppel H Cederbaum LS Burghardt I Worth GA Meyer HD. "Using the MCTDH wavepacket propagation method to describe multimode non-adiabatic dynamics". In: *Biophys J* 96 (2009), pp. 403–416.

[127] Preston RK. Tully JC. "Trajectory surface hopping approach to nonadiabatic molecular collisions: reaction of H+ with D2". In: *J Chem Phys* 55 (1971), pp. 562–572.

[128] Tully JC. "Mixed quantum-classical dynamics". In: *Faraday Discuss* 110 (1998), pp. 407–419.

[129] Schulten K Hayashi S Taikhorshid E. "Photochemical reaction dynamics of the primary event of vision studied by means of a hybrid molecular simulation". In: *Biophys J* 96 (2009), pp. 403–416.

[130] Thiel W. Fabiano E Groenhof G. "Approximate switching algorithms for trajectory surface hopping". In: *Chem Phys* 351 (2008), pp. 111–116.

[131] Miller WH Ali DP. "Effect of electronic transition dynamics on iodine atom recombination in liquids." In: *J Chem Phys* 78 (1983), pp. 6640–6645.

[132] George TF Miller WH. "Semiclassical theory of electronic transitions in low energy atomic and molecular collisions involving several nuclear degrees of freedom". In: *J Chem Phys* 56 (1972), pp. 5637–5652.

[133] Whitton WN Kuntz PJ Kendrick J. "Surface-hopping trajectory cal-
       culations of collision-induced dissociation processes with and without
       charge transfer". In: *Chem Phys* 38 (1979), pp. 147–160.

[134] Truhlar DG Zhu C Jasper AW. "Non-Born-Oppenheimer Liouville-
       von Neumann dynamics. Evolution of a subsystem controlled by lin-
       ear and population-driven decay of mixing with decoherent and co-
       herent switching". In: *J Chem Theory Comput* 1 (2005), pp. 527–
       540.

[135] Persico M. Granucci G. "Critical appraisal of the fewest switches
       algorithm for surface hopping". In: *J Chem Phys* 126 (2007), pp. 114–
       134.

[136] Persico M. Cattaneo P Granucci G. "Simulations of condensed phase
       photochemistry: cage effect and internal conversion in azoalkanes and
       nitrosamines". In: *J Phys Chem A* 103 (1999), pp. 3364–3371.

[137] Rothlisberger U. Tavernelli I Tapavicza E. "Nonadiabatic dynamics
       using time-dependent density functional theory: assessing the cou-
       pling strengths". In: *J Mol Struct Theochem* 914 (2009), pp. 22–29.

[138] Lischka H Pittner J Bonacic-Koutecky V. Zechmann G Barbatti M.
       "Multiple pathways in the photodynamics of a polar pi-bond: a case
       study of silaethylene." In: *Chem Phys Lett* 418 (2006), pp. 377–382.

[139] Lischka H. Barbatti M Ruckenbauer M. "The photodynamics of ethy-
       lene: a surface-hopping study on structural aspects." In: *J Chem Phys*
       122 (2005), pp. 174–307.

[140] Szymczak JJ Nachtigallova D Hobza P Lischka H. Barbatti M Aquino
       AJA. "Relaxation mechanisms of UV-photoexcited DNA and RNA
       nucleobases." In: *Proc Natl Acad Sci USA* 107 (2010), pp. 21453–
       21458.

[141] Thiel W. Lan Z Fabiano E. "Photoinduced nonadiabatic dynam-
       ics of pyrimidine nucleobases: on-the-fly surface-hopping study with
       semiempirical methods". In: *J Phys Chem B* 113 (2009), pp. 3548–
       3555.

[142] Weingart O Spillane KM Manzoni C Brida D-Tomasello G Orlandi
       G Kukura P Mathies RA Polli D Altoe P et al. "Conical intersec-
       tion dynamics of the primary photoisomerization event in vision." In:
       Nature 467 (2010), pp. 440–443.

[143] Schulten K. Hayashi S Taikhorshid E. "Photochemical reaction dynamics of the primary event of vision studied by means of a hybrid molecular simulation." In: *Biophys J.* 96 (2009), pp. 403–416.

[144] Crespo-Hernandez C. E. Hare P. M. and B. Kohler. "Internal conversion to the electronic ground state occurs via two distinct pathways for pyrimidine bases in aqueous solution". In: *Proc. Natl. Acad. Sci. U.S.A.* 104 (2007), pp. 435–440.

[145] Serrano-Andres-L. Robb M. A. Merchan M. and L. Blancafort. "Triplet-state formation along the ultrafast decay of excited singlet cytosine." In: *J. Am. Chem. Soc.* 125 (2005), pp. 8108–8109.

[146] Climent-T. Gonzalez-Ramirez I. Merchan M. Gonzalez-Luque R. and L Serrano-Andres. "Singlet-triplet states interaction regions in DNA/RNA nucleobase hypersurfaces." In: *J. Chem. Theory Comput.* 6 (2010), pp. 2103–2114.

[147] Gonzalez-L. Martinez-Fernandez L. and I. Corral. "An ab initio mechanism for efficient population of triplet states in cytotoxic sulfur substituted DNA bases: the case of 6-thioguanine." In: *Chem Commun* 48 (2012), pp. 2134–2136.

[148] Martinez-Fernandez-L. Pollum M. and C. Crespo-Hernandez. *Photochemistry of Nucleic Acid Bases and Their Thio- and Aza-Analogues in Solution.* Springer, 2014.

[149] P. Murat and S. Balasubramanian. "Existence and consequences of G- quadruplex structures in DNA." In: *Curr. Opin. Genet. Dev.* 25 (2014), pp. 22–29.

[150] Gustavsson-T. Banyasz-A. Miannay F.-A. and D. Markovitsi. "Excited-state dynamics of dGMP measured by steady-state and femtosecond fluorescence spectroscopy". In: *J Phys Chem A* 114 (2010), pp. 3256–3263.

[151] West-C. W.-Stavros V. G. Chatterley A. S. and J. R. R. Verlet. "Time-resolved photoelectron imaging of the isolated deprotonated nucleotides." In: *Chem Sci* 5 (2014), pp. 3963–3975.

[152] Kleinermanns K.-Improta R. Karunakaran V. and S. A. Kovalenko. "Photoinduced dynamics of guanosine monophosphate in water from broad-band transient absorption spectroscopy and quantum-chemical calculations." In: *J Am Chem Soc* 131 (2009), pp. 5839–5850.

[153] Doerr M.-Marian C. M. Parac M. and W. Thiel. "QM/MM calculation of solvent effects on absorption spectra of guanine." In: *J Comput Chem* 31 (2010), pp. 90–106.

[154] Barone V.-Lami A. Santoro F. and R. Improta. "The excited electronic states of adenine-guanine stacked dimers in aqueous solution: a PCM/TD-DFT study." In: *Phys. Chem. Chem. Phys.* 12 (2010), pp. 4934–4948.

[155] R. Improta. "Quantum mechanical calculations unveil the structure and properties of the absorbing and emitting excited electronic states of guanine quadruplex." In: *Chem. Eur. J.* 20 (2014), pp. 8106–8115.

[156] R. Improta and V. Barone. *Excited States Behavior of Nucleobases in Solution: Insights from Computational Studies*. Springer, 2014.

[157] Cohen B. Hare P. M. Crespo-Hernandez C. E. and B. Kohler. "Ultrafast excited-state dynamics in nucleic acids." In: *Chem Rev* 104 (2004), pp. 1977–2019.

[158] Altoe P. Stenta M. Garavelli M. Conti I. and G Orlandi. "Adenine deac- tivation in DNA resolved at the CASPT2//CASSCF/AMBER level." In: *Phys. Chem. Chem. Phys.* 12 (2010), pp. 5016–5023.

[159] Zelen T. Ruckenbauer M. Mueller T. Barbatti M. Hobza P. Nachtigallova D. et al. "Does stacking restrain the photodynamics of individual nucleobases?" In: *J Am Chem Soc* 132 (2010), pp. 8261–8263.

[160] M. Barbatti. "Photorelaxation induced by water-chromophore electron transfer." In: *J. Am Chem Soc* 136 (2014), pp. 10246–10249.

[161] Darden T. A. Cheatham T. E. Simmerling C. Wang J.-Duke R. Case D. et al. *Amber 11*. 2011.

[162] Stenta M. Bottoni A. Altoe P. and M. Garavelli. "COBRAMM: a tunable QM/MM approach to complex molecular architectures.Modelling the excited and ground state properties of sized molecular systems." In: *AIP Conf. Proc.* 963 (2007), pp. 491–505.

[163] M. Garavelli. "Computational organic photochemistry: strategy, achievements and perspectives." In: *Theor. Chem. Acc.* 116 (2006), pp. 87–105.

[164] Trucks G. W. Schlegel H. B. Scuseria G. E.-Robb M. A. Cheeseman J. R. Frisch M. J. et al. *Gaussian 09*. Gaussian, Inc. 2009.

[165] Robb M. A. Bearpark M. J. and H. Bernhard Schlegel. "A direct method for the location of the lowest energy point on a potential surface crossing." In: *Chem. Phys. Lett.* 223 (1994), pp. 269–274.

[166]   Nenov A. Hoefinger S. Flavio Altavilla S. Rivalta I. Dumont-E. et al
        Conti I. "Excited state evolution of DNA stacked adenines resolved at
        the CASPT2//CASSCF/Amber Level: from the bright to the excimer
        state and back". In: *Phys. Chem. Chem. Phys* 17 (2015), pp. 7291–
        7302.

[167]   Malmqvist P.-A. Andersson K. and B. O. Roos. "2nd-order perturbation-
        theory with a complete active space self-consistent field reference
        function." In: *J. Chem. Phys* 96 (1992), pp. 1218–1226.

[168]   Aquilante-F. Roca-Sanjuan D. and R Lindh. "Multiconfiguration second-
        order perturbation theory approach to strong electron correlation in
        chemistry and photochemistry." In: *Wiley Interdiscip. Rev. Comput.
        Mol. Sci* 2 (2012), pp. 585–603.

[169]   Knowles-P. J. Knizia G. Manby F. R. Werner H.-J. and M. Schuetz.
        "Molpro: a general-purpose quantum chemistry program package."
        In: *Wiley Interdiscip. Rev. Comput. Mol. Sci.* 2 (2012), pp. 242–253.

[170]   Knowles-P. J. Knizia G. Manby F. R. Schuetz-M. Celani P. Werner
        H.-J. et al. *MOLPRO, Version 2012.1. A Package of ab Initio Pro-
        gram.*

[171]   De Vico-L. Ferre N. Ghigo G. Malmqvist P.-A.-Neogrady P. Aquilante
        F. et al. "Software news and update MOLCAS 7: the next genera-
        tion." In: *J. Comput. Chem* 31 (2010), pp. 224–247.

[172]   Pedersen T. B. Veryazov V. Aquilante F. and R. Lindh. "MOLCASâ
        a software for multiconfigurational quantum chemistry calculations."
        In: *Wiley Interdiscip. Rev. Comput. Mol. Sci.* 3 (2013), pp. 143–149.

[173]   Segarra-Marti J. Serrano-Andres L. Merchan M.-Rubio-M. Gonzalez-
        Ramirez I. and D. Roca-Sanjuan. "On the N1âH and N3âH Bond
        dissociation in uracil by low energy electrons: a CASSCF/CASPT2
        study". In: *J. Chem. Theory Comput* 8 (2012), pp. 2769–2776.

[174]   Roca-Sanjuan D. Merchan-M. Segarra-Marti J. and R. Lindh. "On
        the photophysics and photochemistry of the water dimer." In: . *J.
        Chem. Phys.* 137 (2012), pp. 244–309.

[175]   J. Segarra-Marti and P. B Coto. "A theoretical study of the in-
        tramolecular charge transfer in 4-(dimethylamino)benzethyne". In:
        Phys. Chem. Chem. Phys 16 (2014), pp. 25642–25648.

[176] Dumez B.-Widmark P. O. Pierloot K. and B. O. Roos. "Density matrix averaged atomic natural orbital (ANO) basis sets for correlated molecular wave functions. IV. Medium size basis sets for the atoms H-Kr." In: *Theor. Chim. Acta* 90 (1995), pp. 87–114.

[177] J. R. Platt. "Classification of spectra of cataâcondensed hydrocarbons". In: *J. Chem. Phys.* 17 (1949), pp. 484–495.

[178] Arai G.-Yamazaki S. Nakayama A. and T. Taketsugu. "Solvent effects on the ultrafast nonradiative deactivation mechanisms of thymine in aqueous solution: excited-state QM/MM molecular dynamics simulations." In: *J. Chem. Phys.* 139 (2013), pp. 214–304.

[179] Andersson-K. Fulscher M. P. Malmqvist P. A.-Serrano-Andres L. Pierloot K. Roos B. O. et al. "Multiconfigurational perturbation theory: applications in electronic spectroscopy". In: *Adv. Chem. Phys* 93 (1996), pp. 219–331.

[180] Segarra-Marti-J. Roca-Sanjuan D. Giussani A. and M. Merchan. "Excitation of nucleobases from a computational perspective I: reaction paths." In: *Top.Curr. Chem.* 355 (2013), pp. 57–97.

[181] Richter-M.-Marquetand P. Mai S. and L. Gonzalez. *Excitation of Nucleobases from a Computational Perspective. II: Dynamics.* Springer, 2014.

[182] Barone-V.-Lami A.-Santoro F. Picconi D. and R. Improta. "The interplay between $\pi\pi^*/n\pi^*$ excited states in gas-phase thymine: a quantum dynamical study". In: *Chemphyschem* 12 (2013), pp. 1957–1968.

[183] Cembran A. et al. In: *J. Phys. Chem. A* 109 (2005), pp. 6597–6605.

[184] Andruniow T. et al. In: *Proc Natl Acad Sci USA* 101.52 (2004), pp. 17908–17913.

[185] Bassolino G et al. "Synthetic Control of Retinal Photochemistry and Photophysics in Solution". In: *JACS* 136.6 (2014), pp. 2650–2658.

[186] Huntress MM, Gozem S, et al. "Toward an Understanding of the Retinal Chromophore in Rhodopsin Mimics". In: *J Phys Chem B* 117.35 (2013), pp. 10053–10070.

[187] Ben-Nun M et al. "The role of intersection topography in bond selectivity of cis-trans photoisomerization". In: *Proc Natl Acad Sci USA* 99.4 (2002), pp. 1769–1773.

[188] Atchity GJ. "The role of intersection topography in bond selectivity of cis-trans photoisomerization". In: *J. Chem. Phys* 95 (1991), p. 1862.

[189]   Frutos LM, Andruniow T, et al. "Tracking the excited-state time evolution of the visual pigment with multiconfigurational quantum chemistry". In: *Proc Natl Acad Sci USA* 104.19 ().

[190]   A. Koslowski Keal T.W. and W. Thiel. "Comparison of algorithms for conical intersection optimisation using semiempirical methods". In: *Theoretical Chemistry Accounts* 118 (2007), pp. 837–844.