Università degli Studi di Bologna
Dipartimento di Chimicha “G. Ciamician”

Dottorato di Ricerca in Scienze Chimiche (XIX Ciclo)
Settore Disciplinare: CHIM/02
Coordinatore: Chiar.mo Prof. Vincenzo Balzani

# SVILUPPO E APPLICAZIONE DI METODI QM/MM NELLO STUDIO DELLA FOTOCHIMICA DI SISTEMI BIOLOGICI E SUPRAMOLECOLARI

**Piero Altoè**

Relatore: Dr. Marco Garavelli

Bologna 2006

# Contents

# Prefazione

Circa tre anni fa durante una della tante discussioni tra il dott. Marco Garavelli, il dott. Marco Stenta e il sottoscritto ci ritrovammo in un vicolo cieco; non potevamo studiare alcuni dei sistemi molecolari a cui eravamo interessati perché non avevamo gli adeguati strumenti computazionali. L'unico modo per uscire da quest'*empasse* fu sviluppare un nostro codice capace di effetture i calcoli ibridi QM/MM di cui avevamo bisogno. Da questo seme iniziale sono nate molte linee di ricerca e nuove idee per migliorare ulteriormente il nostro programma. Questa tesi consta di due parti: nella prima viene descritta in dettaglio la metodologia QM/MM da noi implementata, nella seconda sono presentati i risultati ottenuti applicando sia metodi *ab initio* che QM/MM allo studio fotochimico di sistemi biologici e supermolecolari.

Desidero ringraziare tutti coloro che hanno contribuito alla realizzazione di tutto questo: *in primis* il Dott. Marco Stenta autore di più del cinquanta per cento del codice sorgente di COBRAM, il Dott. Marco Garavelli per il sostegno e i preziosi consigli, e tutti i miei colleghi del laboratorio di chimica computazionale.
Vorrei ricordare il mio primo Maestro, il Prof. Fernando Bernardi che con umanità e chiarezza mi ha iniziato allo studio della chimica computazionale.

# Introduction

The aim of this introduction is to give the basic informations about the chemical and biochemical systems treated in the rest of the thesis. It starts with an overview of what are the QM/MM methods and than continues with a short description of each chemical system faced on the book. A more focused introduction on the precise chemical aspects investigated can be found at the beginning of each chapter.

## 1.1. Quantum mechanics/molecular mechanics methods

The first example of combined quantum mechanics and molecular mechanic method (QM/MM) has been reported by Warshel in 1976[1], but we have to wait 20 years to find important applications. The recent increase on the number of publications in which hybrid methods are utilized indicate that now they are a permanent part of the theoretical toolbox. In a very recent review D. Truhlar[2] give a short but clear description of why this methods are employed:

*"Despite the increasing computational capability available now, molecular modeling and simulation of large, complex systems at the atomic level remain a challenge to computational chemists. At the same time, there is an increasing interest in nanostructured materials, condensed-phase reactions, and catalytic systems, including designer zeolites and enzymes, and in modeling systems over longer time scales that reveal new mechanistic details. The central problem is: can we efficiently accomplish accurate calculations for large reactive systems over long time scales? As usual, we require advances in modeling potential energy surfaces, in statistical mechanical sampling, and in dynamics. [...] . Models based on classical mechanical constructs such as molecular mechanical (MM) force fields that are based on empirical potentials describing small-amplitude vibrations, torsions, van der Waals interactions, and electrostatic interactions have been widely used in molecular dynamics (MD) simulations of large and complex organic and biological systems as well as inorganic and solid-state systems. However, the MM force fields are unable to describe the changes in the electronic structure of a system undergoing*

*a chemical reaction. Such changes in electronic structure in processes that involve bond-breaking and bond-forming, charge transfer, and/or electronic excitation, require quantum mechanics (QM) for a proper treatment. However, due to the very demanding computational cost, the application of QM is still limited to relatively small systems consisting of up to tens or several hundreds of atoms, or even smaller systems when the highest levels of theory are employed. Algorithms that combine quantum mechanics and molecular mechanics provide a solution to this problem. These algorithms in principle combine the accuracy of a quantum mechanical description with the low computational cost of molecular mechanics, and they have become popular in the past decades. The incorporation of quantum mechanics into molecular mechanics can be accomplished in various ways, and one of them is the so-called combined quantum mechanical and molecular mechanical (QM/MM) methodology."*

During the last three years in our research group we have developed a QM/MM code, which combining different commercial softwares permits an accurate description of systems up to 20000 atoms[3]. Since we are interested on excited state description, the algorithm used to combine the QM to MM parts has been designed also to include the effect of the electronic excitations, on the QM/MM force field. In particular the QM wave function feels the presence of the MM point charges (i.e. electrostatic embedding), and the MM part reacts to the variation of the charge distribution of the QM part, passing from one electronic state to a new one. A complete description of our code can be found in part 1 chapter 4.

We have applied our QM/MM program to a variety of interesting chemical problems: photoinduced supermolecule activation, solvation effects and biological photoisomerization in protein. In the next sections can be found some small introductions of the molecular system investigated on this thesis.

## 1.2. Rotaxanes

Rotaxanes are examples of interlocked molecules. Unlike classical molecular structures, they consist of two or more separate components which are not connected by chemical (i.e. covalent) bonds. These structures are, however, true molecules, as each component is intrinsically linked to the other – resulting in a mechanical bond which prevents dissociation without cleavage of one or more covalent bonds.

Rotaxanes consist of macrocyclic rings trapped onto a linear unit (the "thread") by two bulky substituents ("stoppers"). Figure 1 shows a schematic representation of a [2]rotaxane.

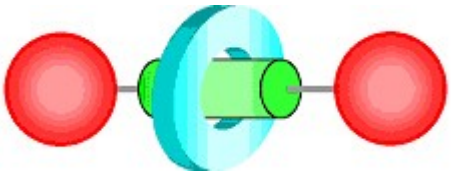


*Figure **1**: Cartoon representation of a [2]rotaxane.*

### 1.2.1. How are they made?

Up until the early 1980s, rotaxanes and other interlocked architectures were considered by most chemists to be academic curiosities with little potential for practical applications. The existing synthetic routes were low-yielding – relying simply on the chance interlocking of components during their formation or else on complex and protracted synthetic strategies. With increased interest in the developing field of supramolecular chemistry, however, chemists were able to apply their new-found knowledge of non-covalent interactions to chemical synthesis, resulting in the various templated synthesis methods which are used to efficiently produce rotaxanes and many other interesting compounds today[4,5,6].

Templated synthesis employs specific non-covalent interactions between components to hold the molecular precursors in the correct orientation for production of the interlocked product following a final covalent bond-forming reaction.

### 1.2.2. Why are they made?

Rotaxanes can now be constructed in high yields, using a variety of templating interactions: hydrogen bonds, metal-ligand interactions, π-π charge transfer interactions or hydrophobic forces. The acquired understanding of non-covalent interactions and their utility in molecular construction has been applied to the synthesis of ever-more complex architectures such as [n]rotaxanes, (i.e. rotaxanes involving several interlocked components) and knots. Synthetic challenges still remain however. The necessity for interactions between components naturally places restrictions on the structural and functional make-up of the final product and the ability to create rotaxanes composed of virtually any chemical units is an extremely attractive prospect and one which is being actively pursued in our group. There are also many other interesting architectures still to be reproduced at the molecular level.

The main driving-force for research on interlocked molecules today, however, is perhaps not the synthetic challenge but the interesting properties and potential applications of the molecules themselves. Interlocked molecules often exhibit markedly different properties to their non-interlocked analogues. This can include, for example, differences in spectroscopic responses, chemical reactivity or mechanical properties. These differences are a direct result of the interlocked architecture, yet perhaps the most exciting consequence of the mechanical bond, is the unique way in which different parts of the molecule can move with respect to the rest of the system (see Figure 2). Controlling these submolecular motions is a very active area of research in many groups across the world.

Controlled, submolecular motion of components in mechanically interlocked molecules is one approach to the creation of novel functional molecules – molecules which change their properties in response to some external stimulus (e.g. light, electricity or a chemical reagent). Such molecules will form the basis of the molecular machines and molecular devices which are predicted to be the key protagonists in the development of a "bottom-up" nanotechnology for the 21st century.

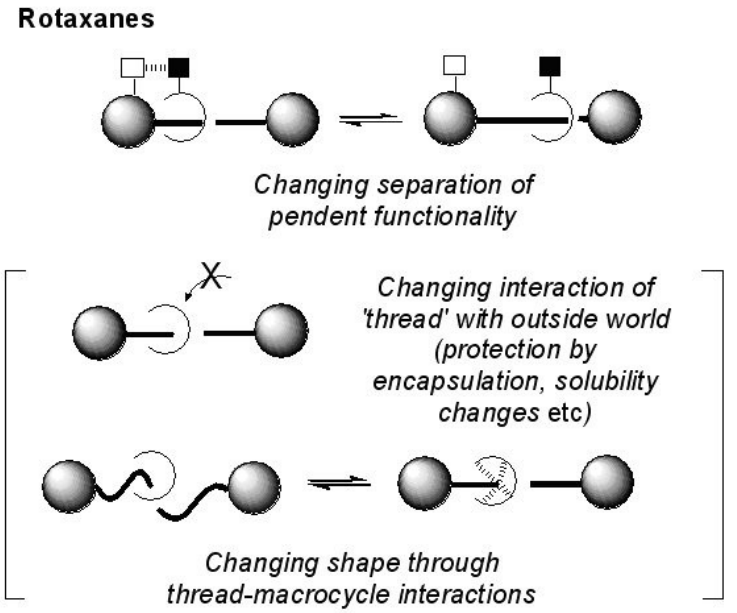



*Figure **2**: Mechanically interlocked architectures as components for molecular machines.*

### *1.2.3. Hydrogen Bonded Rotaxanes*

Figure 3b shows the X-ray crystal structure of a [2]rotaxane made by the Leigh group using hydrogen bonding interactions[7]. The macrocyclic ring is a tetra-amide ring while the thread also contains two amide groups. Two sets of bifurcated hydrogen bonds between the amide protons on the macrocycle and carbonyl oxygens on the thread are shown as dotted lines in the X-ray crystal structure.

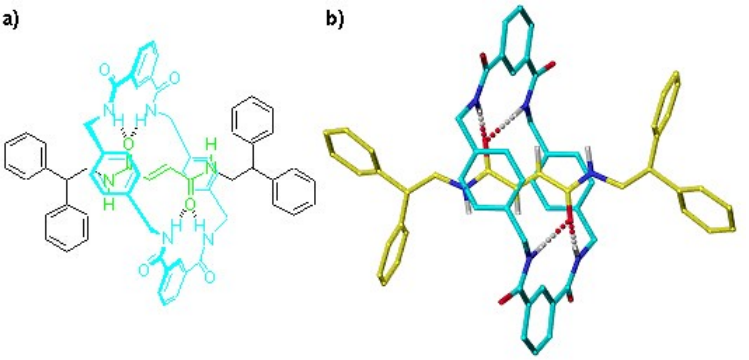



*Figure **3**: a) Chemical structure of and b) X-ray crystal structure of a hydrogen-bonded fumaramide [2]rotaxane. Crystal structure atoms: carbon (macrocycle), light blue; carbon (thread), yellow; oxygen, red; nitrogen, dark blue; amide and alkenyl hydrogen, white.*

This type of rotaxane is formed by a clipping strategy whereby the macrocycle component assembles around a preformed thread molecule[8]. When the acid chloride and diamine are combined, amide bonds begin to form between them. When the thread is present, the newly formed amide-containing fragments wrap themselves around it forming the maximum number of hydrogen bonds to stabilise the structure. This encourages an intramolecular reaction between two ends of the same fragment to form a macrocycle which is, perforce, locked around the thread. In the [2]rotaxane shown above, the carbon-carbon double bond in the thread preorganises the amide carbonyls in a close-to-ideal orientation for hydrogen bonding to the macrocycle so that the rotaxane is produced in 97% yield – an astonishingly efficient process for a five-component reaction.
On this class of rotaxane is observed a structural deformation caused by light. Exciting the system with a 250 nm light the central double bond of the thread isomerizes and the interaction between the two subunit decrease.
We are interested to study the photochemistry of this rotaxane to understand the isomerization mechanism. Starting from a small model system (i.e. fumaramide) and than passing to the entire rotaxane we have rationalize the complex photochemistry of this molecule.

## 1.3. Phytochrome

Phytochrome was first discovered in plants in 1959 as the photoreceptor that mediates plant growth and development in response to long-wavelength visible light[9]. Phytochrome measures the ratio of red light (R) to far-red light (FR), thereby allowing the plant to assess the quantity of photosynthetically active light and trigger shade avoidance responses. Phytochromes are found in all flowering plants and cryptophytes, and this important family of developmental regulators constitutes one of the three major classes of photoreceptors in higher plants, with the others being cryptochromes and phototropins.
More recently, phytochrome-related proteins have been isolated from other taxa. The first such protein to be discovered was the cyanobacterial chromatic adaptation sensor RcaE. Since this initial discovery, R/FR-sensing phytochromes have been discovered in cyanobacteria (Cph1/CphA, Cph2, and CphB/BphP), nonoxygenic bacteria (bacteriophytochromes or BphPs), and even fungi (Fphs), demonstrating that this class of photosensors is not limited to photosynthetic organisms. The true extent of the phytochrome family is only now becoming apparent with the advent of genome sequencing.
Although microbial phytochromes have proven amenable systems for biochemical and spectroscopic analyses, much remains to be determined about the function of many of these molecules in vivo. The BphPs from Rhodopseudomonas palustris regulate the biosynthesis of the photosynthetic apparatus in this organism [10,11], and BphPs regulate pigment biosynthesis in Deinococcus radiodurans and Rhodospirillum centenum [12,13]. The phytochrome from the filamentous fungus Aspergillus nidulans was recently implicated in sexual development[14]. These functions are thus conceptually analogous to functions of plant Phys: Phytochromes regulate the metabolic response of the organism to its light environment.

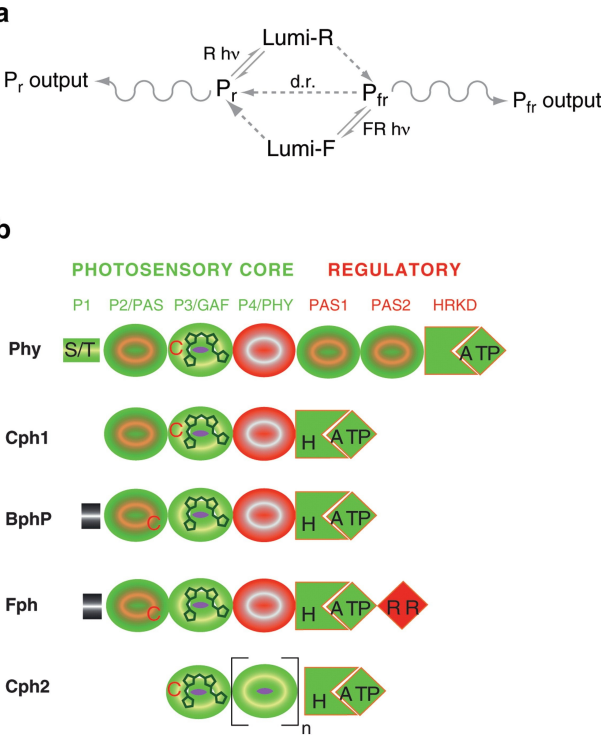



*Figure 4: Domain structure and chromophores of phytochromes. (a) The phytochrome photocycle. Illumination of $P_r$ phytochrome with red light (R) produces lumi-R as the primary photoproduct. This is subsequently converted to $P_{fr}$ via multiple light-independent steps. $P_{fr}$ can be converted into Pr either by illumination with far-red light (FR), producing lumi-F and then $P_r$ via subsequent thermal steps, or by an entirely thermal process known as dark reversion (d.r., center). The ratio between Pr and $P_{fr}$ (and hence between the two physiological outputs) is thus determined by the light environment and by the rate of dark reversion. (b) Domain architecture of the extended phytochrome family. The five classes of phytochromes possess an N-terminal photosensory core region and typically share regulatory output domains related to those found on two-component histidine kinases (HKRD). The P3/GAF domain is associated with the bilin chromophore and is highly conserved. All phytochromes except those found in the Cph2 subfamily share P2/PAS domains, whereas P4/PHY photosensory domains are specific to phytochromes and are thought to have folds similar to GAF domains (69). Plant phytochromes (Phys) possess two additional PAS domains within the regulatory region. Fungal phytochromes (Fphs) have a domain structure similar to those of the cyanobacterial phytochrome 1 (Cph1) and bacteriophytochrome (BphP) families, except for an additional C-terminal response regulator receiver domain (RR) extension and variable N-terminal extensions.*

### 1.3.1. Photoconversion: The "Central Dogma" of Phytochromes

An early key in defining the action of phytochrome in plant biology was the observation that both the spectrum of phytochrome preparations and the action spectrum of many plant responses were reversibly altered by illumination. Illuminating dark-grown tissues with R converts phytochrome from the R-absorbing $P_r$ form to the FR-absorbing $P_{fr}$ form, which triggers photomorphogenesis (Figure 4a). This change is reversible, with FR illumination restoring $P_r$, and involves both primary photochemistry and subsequent thermal steps.

R/FR photoreversibility is rarely observed in other organisms, because higher fluences or continuous light are typically required to maintain a threshold $P_r/P_{fr}$ ratio

for light responsiveness. This reflects a thermal process known as dark reversion, in which $P_{fr}$ phytochrome is slowly converted to $P_r$ phytochrome in the absence of light. While dark reversion is an intrinsic property of all phytochromes, plant phytochromes have apparently evolved to exhibit slower dark reversion. Interestingly, BphPs that possess far-red absorbance *maxima* in the thermal ground state with photoconversion to $P_r$-like species were recently described[11]. Whether this spectral inversion proceeds via a reverse of the normal dark reversion pathway or via some other mechanism is not yet clear for any of these bathyBphPs. However, despite these spectral variations, it is clear that the biological outputs from phytochromes reflect the ratio of the $P_r$ and $P_{fr}$ forms, and that this ratio is determined by the light environment, by the forward and reverse rates of photoconversion, and by the rates for thermal interconversion between these forms.

### 1.3.2. The Modular Domain Architecture of Phytochromes

The large and steadily growing number of phytochrome sequences now known permit classification into subfamilies [15] and delineation of conserved sequences and domains that are either ubiquitous among phytochromes or conserved in different subfamilies (Figure 4b). Plant Phys, Cph1s, and most BphPs share a common architecture consisting of an N-terminal photosensory region with three conserved domains (termed P2 or PAS domain, P3 or GAF domain, and P4 or PHY domain) and a C-terminal regulatory histidine kinase or histidine kinase-related domain (HKRD). Plant Phys have an additional N-terminal extension termed P1 known to inhibit dark reversion and two additional regulatory PAS domains important for nuclear localization[16]. Fungal Fphs have distinct N-terminal extensions and additional C-terminal response regulator domains (RR/REC) [14]. The other class of cyanobacterial phytochromes, the Cph2s, lack the N-terminal P2/PAS domain, but have other GAF domains duplicated C terminal to the P4/PHY domain.

### 1.3.3. Phytochrome Chromophore Structure

The characteristic absorbance spectra and photoconversion of phytochromes reflect their association with a linear tetrapyrrole bilin chromophore, which is normally covalently attached via a thioether linkage (Figure 5). Photoconversion involves a Z–E isomerization about the C15–C16 double bond of the bilin, as apophytochrome neither photoconverts nor exhibits a typical phytochrome absorbance spectrum [17]. The exact nature of this chromophore varies for different subfamilies of phytochromes: Plant Phys use phytochromobilin (PΦB) (Figure 5a), and Cph1s and Cph2s use phycocyanobilin (PCB). Both bilins are covalently attached at C31 to a conserved Cys in the P3/GAF domain of the photosensory core [18]. In contrast, BphPs and Fphs incorporate biliverdin IXα (BV) chromophores (Figure 5a). In these proteins, the more oxidized BV (see Sidebar) is attached to a conserved Cys upstream of the P2/PAS domain, apparently via a C32 linkage [19]. This linkage appears less stable in BphPs than the C31 linkage to PCB, based on recent evidence for its reversibility. Covalent attachment does not appear to be a prerequisite for photoconversion; indeed, a mutant BphP lacking the nucleophilic Cys residue functions as an enzyme for producing C15–C16 E bilins. Covalent attachment likely provides a more stable holoprotein that is better suited to reversible photoswitching.

Phytochromes are thus photoswitchable photosensors that utilize bilin chromophores.

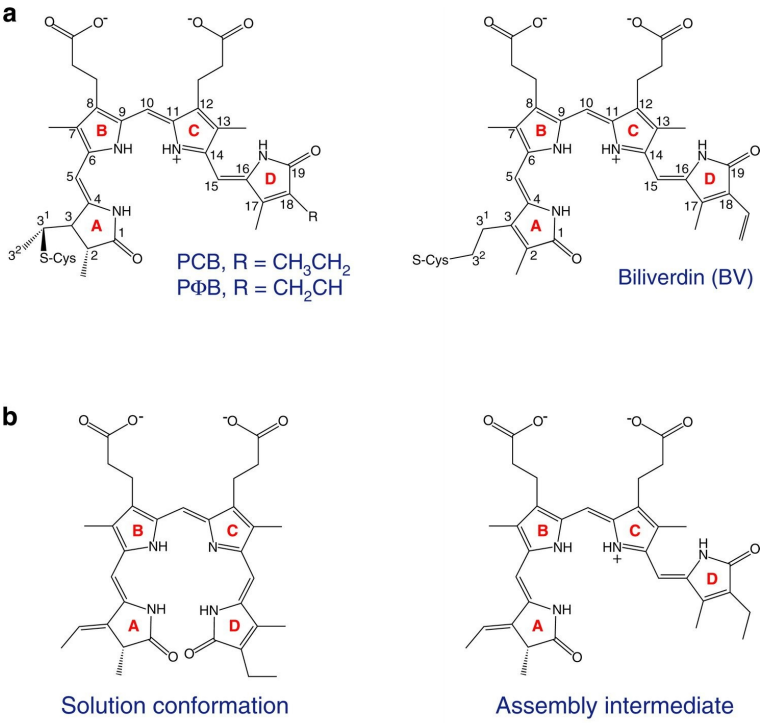



*Figure **5**: Chromophore structure and assembly. (a) The structures of the bilin chromophores utilized by known phytochromes are shown. Left: Phycocyanobilin (PCB) and phytochromobilin (PΦB) chromophores share a reduced A-ring and differ only at the C18 side chain. These chromophores are utilized by plant and algal Phys and cyanobacterial Cph1s and Cph2s. Right: The BphPs and Fphs instead utilize biliverdin (BV) as chromophore. All chromophores are shown in the C5–Z,syn C10–Z,syn C15–Z,anti configuration adopted in the Pr state (103). (b) Conformations of PCB thought to be present during the assembly reaction with Cph1 are shown (6). The cyclic, porphyrin-like C15–Z,syn species (left) is the most stable in solution at neutral pH and initially binds to apoCph1. After binding, the B-/C-ring system becomes protonated, driving adoption of a C15–Z,anti conformation (right), which is characterized by enhanced, red-shifted long-wavelength absorbance. This species then becomes covalently attached to Cys259 of Cph1 to give the Pr structure shown in (a). BV is bound to a different Cys upstream of the P2/PAS domain of BphPs (58, 103).*

We have foucus our attention only on the formation of the first photoproducts, studying the isolated chromophore in vacuo and than the absorption and emission spectra of the protein. The photochemistry of the phytochromobilin has been characterized using multiconfigurational method, deriving a clear picture of the possible deactivation mechanisms, while in protein the investigation has been limited on the initial excited state stretching relaxation.

## 1.4. Bacteriorhodopsin

Bacteriorhodopsin is a small (24 kDa) integral membrane protein composed of seven-transmembrane helical. The retinal was seen to extend transversely into the interhelical space, and the positions of bulkier residues defined the orientation of most of the helices and therefore the pathway of the transported proton (see Figure 6). Since its discovery in the early seventies[20], the conceptual simplicity and experimental advantages of bacteriorhodopsin have made this light-driven proton pump a testing ground for hypotheses of transport mechanisms and new experimental technologies.

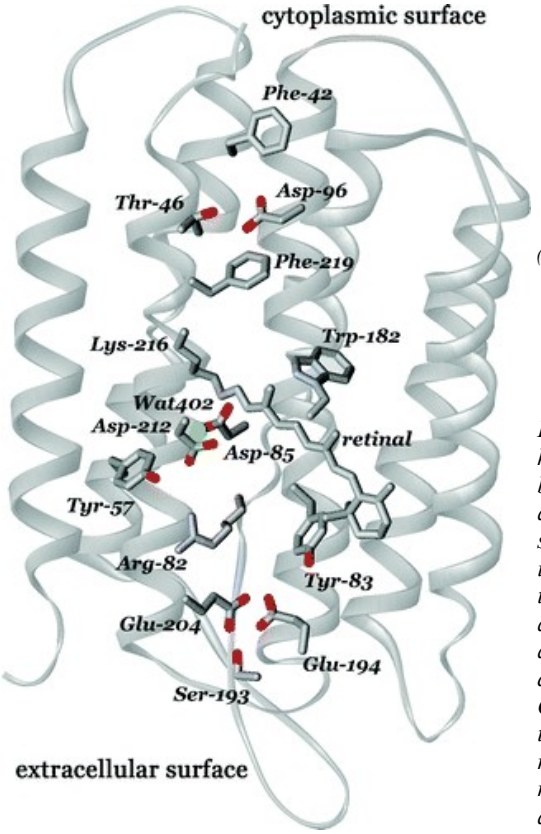



*Figure 6: Structure of bacteriorhodopsin, with functional side chains, the retinal, and Wat402 shown. Drawn from pdb coordinate file 1C3W.*

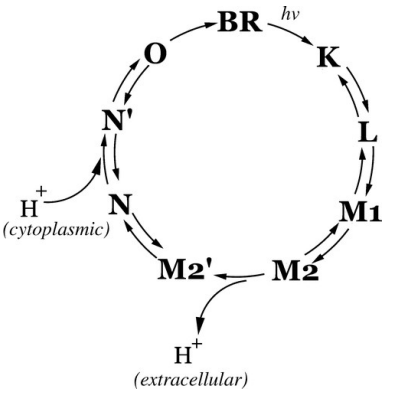



*Figure 7: Photocycle scheme from kinetics in the wild-type and mutant bacteriorhodopsins. The K, L, M, N, and O states are distinguished by their spectral properties in the visible, infrared, Raman, and NMR, and by their crystal structures when available. The substates of M and N are identified from kinetic or spectral data or by their structural differences. Only the BR and O states contain all-trans retinal; in all other states the retinal is 13-cis,15-anti. Proton release to the extracellular surface and uptake from the bulk on the cytoplasmic side are shown as they occur under physiological conditions (pH > 7).*

The wealth of spectroscopic data, together with the information from low-resolution electron diffraction maps, pointed to a well-defined transport mechanism. There was general agreement over the main steps [21,22]. After photoisomerization of the retinal from all-trans to 13-cis,15-anti, the Schiff base proton is transferred to Asp-85 located on the extracellular side, and a proton is then released to the bulk from a site near the surface. Figure 6 shows the overall structure of the protein with the most

important residues labeled. Reprotonation of the Schiff base is from Asp-96 located on the cytoplasmic side, aided by tilts of the cytoplasmic ends of helices F and G that were thought to result in increased hydration of this region. Reprotonation of the Schiff base through a proposed chain of water molecules is followed by reprotonation of Asp-96 from the cytoplasmic surface and reisomerization of the retinal to all-trans. Finally, transfer of a proton from Asp-85 to the vacant proton release site completes the cycle.

It has been known from the earliest works that the photocycle contains the spectrally distinguishable quasi-stable states K, L, M, N, and O. Once K is produced from J, which arises from the excited state, the rest of the reaction cycle is a sequence of thermal reactions, as in other transport systems( see Figure 7).

Our group studies since years the intrinsic behavior of the protonated schiff bases, now we have extended our attention on the photochemistry in protein. Thanks to the modern QM/MM techniques we could able to determinate the complete minimium energy path of the bacteriorhodopsin photoisomerization from the initial reagent to the first metastable intermediate K. We have also identify a possible structure for the J, a short life excited state intermediate. Our analysis include the estimation of the energy storage between BR and K and the variation of the dipole moment along the isomerization path.

# Part 1. Theoretical background

# 1. Quantum Mechanical methods

In this chapter is treated briefly the QM methods used to obtain ground and excited state properties of biological and supermolecular compounds described in the introduction.
It is assumed that the reader knows the basic information about quantum mechanic and their mathematical aspect. Furthermore here are only summarize the general approximations and tools common to all QM methods used to solve many body problems in chemistry (.i.e. Born-Oppenheimer approximation, LCAO approach, Slater Determinants, ...).
The multiconfigurational approach to solve the Schrödinger equation is the widely used to describe accurately the electronic excited state of molecules, in particular the configuration interaction (CI) and the multiconfigurational self consistent field (RASSCF and CASSCF) method are here described.

## 1.1. The general Hamiltonian

The total, non-relativistic Hamiltonian for a system of charged particles (electron and nuclei) can be written as[i]:

$$\hat{H} = -\sum_a \frac{1}{2\mathrm{m}_a} \nabla_a^2 + \sum_{a<b} \frac{q_a q_b}{|\vec{r}_a - \vec{r}_b|} \tag{1.1}$$

Where the $\nabla_a^2$ is the laplacian operator for the particle *a*:

$$\nabla_a^2 = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2} \tag{1.2}$$

i To simplify the notation, atomic unit are used here and in all chapter.

## 1.2. The Born-Oppenheimer approximation

Since the nuclear mass is much greater than those of the electrons, the nuclei move much more slowly and their motion can be separated from that of the electrons. In this picture, electrons react instantaneously to changes in nuclear positions and the electron distribution depends only on the position of the nuclei and not their momenta. This approximation is known as Born-Oppenheimer approximation (BO) as well as adiabatic approximation[23].
Applying the BO to the (1.1) the total Hamiltonian is split in two terms.
The nuclear:

$$\hat{T} = -\sum_{\mu} \frac{1}{2m_{\mu}} \nabla_{\mu}^2 \tag{1.3}$$

And the electronic:

$$\hat{H}_{el} = -\sum_{i} \frac{1}{2} \nabla_i^2 - \sum_{\mu,i} \frac{Z_{\mu}}{|\vec{r}_{\mu} - \vec{r}_i|} + \sum_{i<j} \frac{1}{|\vec{r}_i - \vec{r}_j|} + \sum_{\mu<\nu} \frac{Z_{\mu} Z_{\nu}}{|\vec{r}_{\mu} - \vec{r}_{\nu}|} = \sum_{i} \hat{h}_i + \sum_{i<j} \hat{g}_{ij} + \hat{h}_0 \tag{1.4}$$

Where the indices $\mu, \nu$ indicate the nuclei coordinates and $i,j$ the electrons coordinates.
In the right side of (1.4) the electronic Hamiltonian is condensed respectively in term of one-electron, two-electron and zero-electron operators.
Considering the nuclei coordinates constant (BO) the last term $\hat{h}_0$ is trivial and it will be omitted in he rest of the discussion.

## 1.3. Slater Determinant

The Molecular Orbital (MO) theory is an approach to Quantum Mechanics which uses one-electron functions to approximate the electronic wave function. These functions are formed by the product of a spatial part $\psi_i$(j) times a spin function $\alpha$(j) or $\beta$(j), and are therefore called spinorbitals (the index i refers to the i$^{th}$ spatial part, while j refers to the j$^{th}$ electron).
The antisymmetry property of the wave function is fulfilled using a Slater Determinant[24] (i.e. a antisymmetric product) composed of spinorbitals:

$$\Psi = (n!)^{-1/2} \begin{vmatrix} \psi_1(1)\alpha(1) & \psi_1(1)\beta(1) & \psi_2(1)\alpha(1) & \dots & \psi_{n/2}(1)\beta(1) \\ \psi_1(2)\alpha(2) & \psi_1(2)\beta(2) & \psi_2(2)\alpha(2) & \dots & \psi_{n/2}(2)\beta(2) \\ \dots & \dots & \dots & \dots & \dots \\ \psi_1(n)\alpha(n) & \psi_1(n)\beta(n) & \psi_2(n)\alpha(n) & \dots & \psi_{n/2}(n)\beta(n) \end{vmatrix} \tag{1.5}$$

In the Hartree-Fock (HF) method the total electronic wave function is formed by a single Slater Determinant of the occupied spinorbitals.

## 1.4. LCAO Expansion

Another approximation is to express the molecular orbitals $\Psi_i$ in terms of a set of N predefine one-electron functions $\chi_{\text{pi}}$, called basis functions (also basis set):

$$\psi_i = \sum_{p=1}^{N} c_{pi}\chi_p \tag{1.6}$$

where the coefficients $c_{pi}$ are the molecular orbital expansion coefficients.
When atomic orbitals are used as basis functions, the previous approach is defined as the Linear Combination of Atomic Orbitals (LCAO) approach.

## 1.5. The Hartree-Fock equations

The electronic energy expression deriving from the (1.4), after some mathematical work, can be written as:

$$E = \left\langle \Psi \left| \sum_i \hat{h}_i \right| \Psi \right\rangle + \left\langle \Psi \left| \sum_{i<j} \hat{g}_{ij} \right| \Psi \right\rangle = \sum_i \left\langle \psi_i \left| \hat{h} \right| \psi_i \right\rangle + \frac{1}{2} \sum_i \left\langle \psi_i \left| \hat{J} - \hat{K} \right| \psi_i \right\rangle \tag{1.7}$$

$\hat{J}$ and $\hat{K}$ are the total coulomb and exchange operators.
The (1.7) in the starting point to derive the HF equations, where the Fock operator is defined as:

$$\hat{F} = \hat{h} + \hat{J} - \hat{K} \tag{1.8}$$

Applying the variational principle to the (1.7) and substituting the Fock operator in the energy expression is possible to write it as:

$$\left\langle \delta\psi_i \left| \hat{F} \right| \psi_i \right\rangle = 0 \tag{1.9}$$

The variation of the wave function is:

$$\delta\psi_i = \sum_b \alpha_{ib}\psi_b + \sum_a \alpha_{ia}\psi_a \tag{1.10}$$

Where $\alpha$ an infinitesimal variation, and considering the wave function constituted by a Slater determinant $\Psi_b$ ,$\Psi_a$ are the occupied and virtual molecular orbitals (MO).
From the second term of (1.10) we get:

$$\left\langle \psi_a \left| \hat{F} \right| \psi_i \right\rangle = 0 \tag{1.11}$$

In can be concluded that the virtual orbitals do not participate during the variational procedure and then the (1.11) can be written as:

$$\hat{F}\,\psi_i = \sum_b \lambda_{bi}\,\psi_b \tag{1.12}$$

Let $\boldsymbol{V}$ denote the matrix which diagonalize $\boldsymbol{\lambda}$ and form new MOs:

$$\tilde{\psi}_b = \sum_i \psi_i V_{ib} \tag{1.13}$$

which leaves $\Psi$ unchanged. Taking the (1.12) multiplied by $V_{ib}$ and sum over $i$, leads to the most popular form of the Hartree-Fock equations[25] :

$$\hat{F}\,\tilde{\psi}_b = \epsilon_b\,\tilde{\psi}_b \tag{1.14}$$

To solve the HF equation with a wave function constructed using the Slater determinant and LCAO scheme, in a self consistent way the Roothaan-Hall equations[26] must be introduced. Starting from the (1.11):

$$\left\langle \psi_a \middle| \hat{F} \middle| \psi_i \right\rangle = \sum_{p,q} C^*_{pa} F_{pq} C_{qi} = \left( \boldsymbol{C}^\dagger \boldsymbol{F} \boldsymbol{C} \right)_{ai} = 0 \tag{1.15}$$

with:

$$F_{pq} = \left\langle \chi_p \middle| \hat{F} \middle| \chi_q \right\rangle \tag{1.16}$$

where $\chi_p, \chi_q$ are atomic orbitals.
One condition has to be satisfied, the orthonormality of the MOs:

$$\boldsymbol{C}^\dagger \boldsymbol{S} \boldsymbol{C} = \boldsymbol{1} \tag{1.17}$$

with the definition of overlap:

$$S_{pq} = \left\langle \chi_p \middle| \chi_q \right\rangle \tag{1.18}$$

The simultaneous solution of (1.15) and (1.17) is achieved if one solves the Roothaan-Hall equations:

$$\boldsymbol{F} \boldsymbol{C} = \boldsymbol{S} \boldsymbol{C} \boldsymbol{\epsilon} \tag{1.19}$$

The Roothaan-Hall equations need to be solved iteratively comparing the energies of the iterations $n^{th}$ i to the $n^{th}$-1 if they differs less than the threshold the procedure stops and the energy is converged.
The final energy, because the HF method consider only one of the possible Slater determinants (i.e. the ground state determinant).
The difference between the HF energy end the exact energy is called correlation

energy:

$$E^{corr} = E^{ecact} - E^{HF} \tag{1.20}$$

In conclusion the principals limitations of HF method are two:

1. It can describe only the ground state (single determinant approach).
2. It omits, by definition, all the correlation effects.

## 1.6. Configuration interaction

Several theories has been developed to solve the problem of correlation energy, some of them use the HF wave function has reference and then add a correction calculated using the variational principle (configuration interaction[25] = CI) or the perturbation theory (Møller-Plesset[25] = MP2,MP3,...).
The CI method use as wave function a linear combination of Slater determinant where the orbitals coefficients are the HF coefficients, for example the ground state wave function is:

$$\Psi_1 = \sum_L c_{L1}\,\phi_L \tag{1.21}$$

where $\phi_{L1}$ are the determinants of the ground state and all possible electronic excitations, and the Schrödinger equation become:

$$\sum_L c_{L1}\langle \phi_K|\hat{H}|\phi_L\rangle = E\sum_L c_{Li}\langle \phi_K|\phi_L\rangle \tag{1.22}$$

where $\langle \phi_K|\hat{H}|\phi_L\rangle$ is a generic element of the Hamiltonian matrix and $\langle \phi_K|\phi_L\rangle$ is the overlap matrix element between Slater Determinats.
The basis set constituted by the Slater Determinants is an orthonormal set $\langle \phi_K|\phi_L\rangle = \delta_{KL}$ and the (1.22) can be written in a compact way:

$$\boldsymbol{H}\boldsymbol{C}_1 = E\,\boldsymbol{C}_1 \tag{1.23}$$

The eigenvector $\mathbf{C_1}$ collect all the coefficients of the expansion, in other words the coefficients are the weight of each Slater Determinant.
The element $\mathbf{C_{11}}$ is associated to the HF wave function. The remaining coefficients are referred to the excited determinants and they represent the correction to the HF model.
The ground state energy obtained using a complete basis set is the the exact non relativistic energy of the system within the BO approximation, it consider all the correlation energy. Unfortunately, the above procedure cannot be implemented in practice because is impossible handle infinite basis set.
Working with a finite set of spinorbitals $\{\chi_i = 1,2,\ldots,2\mathrm{K}\}$ and N electrons, the $\binom{2\mathrm{K}}{N}$ determinants formed do not correspond to a complete basis. Nevertheless the

diagonalization of the Hamiltonian formed from this set leads to the exact solution in basis set approximation. This procedure is called full CI. Even for small molecule and basis set the number of determinants that must be included in a full CI calculation is extremely large. Thus in practice the CI expansion has to be truncated. The most common way is to included only the determinants referred to a certain type of excitations, like single excitations (CIS), double (CID) or single and double (CISD) and so on.
The above mathematical description as been done to obtain the ground state energy, but. in a similar way is possible also to calculate excited state properties solving the (1.23) for eigenvalue higher than the first where the $N^{th}$-eigenvalue is the energy of the $N^{th}$-state.
The CIS method if quite fast but not very accurate to describe electronic excitations because it can calculate only single-excitation states, while the CISD is more accurate, but the computational performance dramatically lowers down as the number of basis functions increases. A failure of the truncated CI methods is the lack of size-consistency in few words, the energy of two molecules placed at non-interacting distance is not the sum of the energies computed for each single molecule[ii].

## 1.7. Multiconfigurational self consistent field

In the previous method the determinants are formed from a set of canonical HF orbitals, but the resulting CI expansion unfortunately turns out to be slowly convergent (high number of configurations to get correct energy). A very good improvement of convergence to the correct energy comes from the multiconfigurational self consistent field[27] (MCSCF) methods, where together with the CI coefficients are optimized also the orbitals coefficients. Using this approach is possible to get accurate results with a limited number of determinants.
The mathematical machinery to solve the MCSCF equations is very complicate and lots of different way as been developed, here is described the easiest which optimize separately the two type of coefficients, albeit in the modern software is not used because it gives convergence problems.
The optimization of the CI coefficients can be done solving the equation (1.23) without any difference from a normal CI procedure. The troublesome part is the orbitals optimization, the approach is based on a unitary transformation of the MO coefficients that leads to equations Newton-Raphson like.
Assuming the orbitals orthonormal the unitary transformation can be written as:

$$\boldsymbol{\phi}' = \boldsymbol{U}\boldsymbol{\phi} \tag{1.24}$$

ii Without enter in mathematical details the non size-consistency can be explained with a simple example taken from the Szabo & Ostlund (page 261)[25].
*"Considering a supramolecule composed of two non interacting $H_2$ molecules. By definition the CID wave function of each of the monomers contains double excitations within the monomer. If we restrict the supramolecule trial function to double excitations, we exclude the possibility that both monomers are simultaneously doubly excited, since this represents a quadruple excitation. Thus the supramolecule wave function truncated at the CID level does not have sufficient flexibility to yield twice the monomer energy."*

where $\boldsymbol{\phi}$, $\boldsymbol{\phi}'$ are the orbitals before and after the transformation. The transformation has to be defined to minimize the energy at fix CI coefficients. The matrix **U** can be opportunely defined as:

$$\boldsymbol{U}=e^{\boldsymbol{A}}=\begin{pmatrix} e^{A_{11}} & \dots & e^{A_{1n}} \\ \dots & \dots & \dots \\ e^{A_{n1}} & \dots & e^{A_{nn}} \end{pmatrix} \tag{1.25}$$

where **A** is skew-symmetric matrix $\boldsymbol{A}^{\wedge}=-\boldsymbol{A}$, this constrain ensure the unitary condition for **U**[iii].
The matrix **A** can be represented as linear combination of elementary matrices $\mathbf{E}_{ij}$:

$$\boldsymbol{U}=e^{\sum_{ij} A_{ij}\boldsymbol{E}_{ij}} \tag{1.26}$$

where $\mathbf{E}_{ij}$ has element equal to 1 only for the $i^{th}$ row and $j^{th}$ column and all the rest are zero. Therefore the optimization of the orbitals coefficients correspond to determinate the $A_{ij}$ parameters.
The transformation (1.24) is through also for the wave function:

$$\left|\Psi'\right\rangle=e^{\boldsymbol{A}}\left|\Psi\right\rangle \tag{1.27}$$

and the energy of $\Psi'$ is:

$$E'=\left\langle\Psi'\left|\boldsymbol{H}\right|\Psi'\right\rangle=\left\langle e^{\boldsymbol{A}}\Psi\left|\boldsymbol{H}\right|e^{\boldsymbol{A}}\Psi\right\rangle \tag{1.28}$$

It has be clear which the (1.28) can be written, from the (1.26), as function of the parameters $A_{ij}$,and expanding in Taylor series and truncating to the second order:

$$E'=E(A)=E(0)+\sum_{ij}A_{ij}\left(\frac{\delta E}{\delta A_{ij}}\right)_0+\frac{1}{2}\sum_{ijkl}A_{ij}A_{kl}\left(\frac{\delta^2 E}{\delta A_{ij}\delta A_{kl}}\right)_0 \tag{1.29}$$

or in matrix form:

$$E'=E(0)+\boldsymbol{F}\boldsymbol{A}+\frac{1}{2}\boldsymbol{A}^{\boldsymbol{T}}\boldsymbol{G}\boldsymbol{A} \tag{1.30}$$

where **A** is the column vector of the parameters, **F** is a row vector of the first derivative of energy respect to the $A_{ij}$ and **G** is the matrix of the second derivatives respect to the same parameters.
The *minimum* of (1.30) is give satisfying the following condition:

iii In fact: $\boldsymbol{U}^{\dagger}\boldsymbol{U}=(e^{\boldsymbol{A}})^{\dagger}e^{\boldsymbol{A}}=e^{\boldsymbol{A}^{\dagger}}e^{\boldsymbol{A}}=e^{-\boldsymbol{A}}e^{\boldsymbol{A}}=\mathbf{1}$

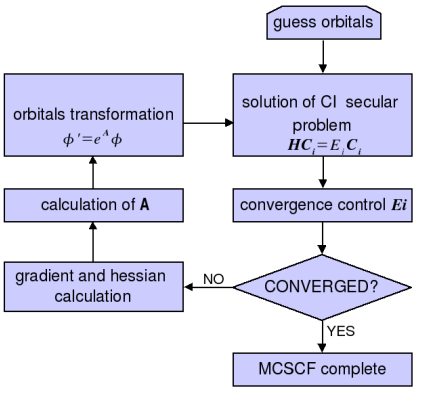
$$\boldsymbol{F}+\boldsymbol{G}\boldsymbol{A}=0 \tag{1.31}$$

from which we get **A**:

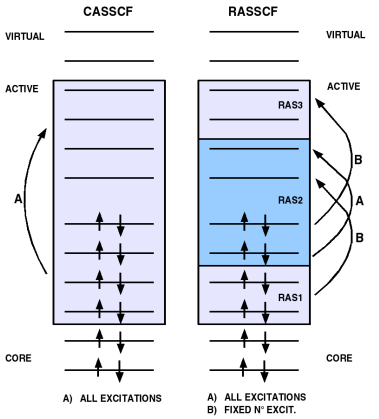
$$\boldsymbol{A}=-\boldsymbol{G}^{-1}\boldsymbol{F} \tag{1.32}$$

Once **A** is known is possible to expand in series $e^{\mathbf{A}}$ ( truncated to the second order) to get **U**:

$$\boldsymbol{U}=e^{\boldsymbol{A}}=\mathbf{1}+\boldsymbol{A}+\frac{1}{2}\boldsymbol{A}^2 \tag{1.33}$$




*Scheme 1.1:* Flowchart of a general procedure of MCSCF calculation.




*Scheme 1.2:* Excitations in CASSCF & RASSCF methods.

from the (1.33) it can demonstrated which **U** is a rotation matrix[iv]. In conclusion the unitary transformation which optimize the orbital coefficients are rotations, but like

iv To understand why **U** is a rotation matrix can be useful a simple example. If we consider a system of two orbitals **U** is a2x2 matrix, and we have also to remember which $A_{ij}=-A_{ji}$:

$$\boldsymbol{U}=e^{\boldsymbol{A}}=\mathbf{1}+\boldsymbol{A}+\frac{1}{2}\boldsymbol{A}^2=\begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix}+\begin{pmatrix}0 & A_{12}\\ -A_{12} & 0\end{pmatrix}+\frac{1}{2}\begin{pmatrix}0 & A_{12}\\ -A_{12} & 0\end{pmatrix}^2=\begin{pmatrix}\frac{1}{2}A_{12}^2 & A_{12}\\ -A_{12} & \frac{1}{2}A_{12}^2\end{pmatrix}=\begin{pmatrix}\cos A_{12} & sen\,A_{12}\\ -sen\,A_{12} & \cos A_{12}\end{pmatrix}$$

The matrix elements sum of the three initial matrices can be interpreted on the base of the $\mathrm{senA_{12}}$ and $\mathrm{cosA_{12}}$:

$$sen\,A_{12}=A_{12}-\frac{A_{12}^3}{3!}+\frac{A_{12}^5}{5!}-\ldots\simeq A_{12}$$

$$\cos A_{12}=1-\frac{A_{12}^2}{2!}+\frac{A_{12}^4}{4!}\ldots\simeq 1-\frac{1}{2}A_{12}^2$$

From this expansions is clear that for small value of $A_{12}$ the diagonal elements are close to the cos and the extra diagonal to the sen.

HF case, is not enough do it once, it has to be repeated iteratively since convergence, and the same for the CI coefficients. For a general MCSCF procedure see scheme 1.1. The modern software to solve the MCSCF equations use very complicate mathematical tricks where the CI coefficients calculation and the orbitals optimization are coupled.
The two most used approach to select the configurations to be introduced in the computation are the *complete active space self consistent field* (CASSCF) and the *restricted active space self consistent field* (RASSCF), in both cases the excitations are considered only between a defined number of orbitals.
In this two methods the orbitals era divided in three regions:

1. The core orbitals, which are treated at HF level
2. The active orbitals (occupied and empty), where the excitations are considered explicitly.
3. The virtual orbitals, which are empty HF orbitals.

The difference between CASSCF and RASSCF is the type of excited determinants used, while in CASSCF all the possible excitations in the active space are calculated, which correspond to full CI for the active orbitals region, the RASCF is possible to select the excitations introduced. In fact the active space of RASSCF is further on divided in three part (scheme 1.2):

1. RAS1, occupied orbitals where only a certain number of electrons are allowed to be excited, like in truncated CI method.
2. RAS2, occupied and empty orbitals, this part in treated at CASSCF level.
3. RAS3, empty orbitals, similar to RAS1, but in this case only a defined number of electrons can occupy these orbitals.

The orbitals to be introduced in the active space is the critical point of these methods. It is problem dependent and it is not possible to define which are the best ones only on the base of the HF orbitals energies. The chemical intuition has to guide the scientist to the selection, for example if we are interested in the description of the lowest excited state of benzene we have to consider the six $\pi$ orbitals of the molecule in our active space. Unfortunately not every time is possible to use the best active space, because the computational cost increase rapidly with the number of active orbitals, and to get a balanced proportion between accuracy and CPU time some orbitals, which can be important, have to be discarded. In that cases is necessary to be very careful in the selection in order to be able to describe all the chemical properties of interest.

## 1.8. Multiconfigurational second order perturbation theory

The MCSCF can describe the correlation energy only of the active part, the rest of the system is treated as in HF, to get rest of the correlation is essential the introduction of a post MCSCF method. The most used and robust way to consider this energy is correct the wave function with the perturbation theory, and in particular the CASPT2[27] method is one of the powerful method.
The starting point in CASPT2 is the CASSCF wave function from which it is possible to calculate the second order perturbation in like manner to the MP2 theory.

$$\hat{H} = \hat{H}_0 + \lambda \hat{H}_1 \qquad (1.34)$$

where $\hat{H}_0$ and $\hat{H}_1$ are the zeroth order and the perturbation Hamiltonians, and λ the weight of the perturbation. In the same way are defined the wave functions and the energies:

$$E_0 = \langle \Psi_0 | \hat{H}_0 | \Psi_0 \rangle \quad , \quad E_1 = \langle \Psi_0 | \hat{H}_1 | \Psi_0 \rangle \quad , \quad E_2 = \langle \Psi_0 | \hat{H}_1 | \Psi_1 \rangle \quad (1.35 \text{ a, b, c})$$

where the condition of orthogonality between the zeroth order wave function and perturbation is implied.
The $\hat{H}_0$ is define as:

$$\hat{H}_0 = \hat{F} = \sum_i f_i \qquad (1.36)$$

usually the Fock-operator is assumed diagonal in a HF space orbitals, but this is not the case for CASSCF reference wave function.
The CASSCF wave function is invariant against rotations among the core orbitals, the active orbitals, and the virtual orbitals. This freedom can be used to simplify the operator, dividing it in diagonal subsets:

$$\hat{F} = f_{cc} + f_{aa} + f_{vv} + f_{ac} + f_{vc} + f_{va} \qquad (1.37)$$

where the labels *c*, *a*, *v* are referred respectively to core, active, and virtual orbitals spaces. From this point the solution of the equations follow the normal perturbation theory, where the Hamiltonian of the perturbation is defined as the difference between the full Hamiltonian and $\hat{H}_0$, and the correction to the wave function is:

$$\Psi_1 = \sum_\mu C_\mu \phi_\mu \qquad (1.38)$$

where $\phi_\mu$ are all the configurations in the CI space, except those belonging to the active space. The coefficients are defined as:

$$C_\mu = -\frac{\langle \phi_\mu | \hat{H}_1 | \Psi_0 \rangle}{E_\mu - E_0} \qquad (1.39)$$

The CASPT2 gives very accurate results, but it has an enormous computational cost. In fact it is usually applied to correct the energy of a previous CASSCF calculation.

# 2. Molecular Mechanics

The Molecular Mechanics (MM) are based on classical physics, where the atoms are considered spheres connected each other by springs. From this simple idea have been developed several force fields, with different degrees of complexity, but with one common task: all force field are some order of magnitude faster than QM methods.
The weak point in MM theory are the parameters, in fact each atom, bond angle, ..., has to be parametrized, and usually more general are the parameters less accurate is the result; it is thought also the vice versa, but with the restriction to applicable only to a small amount of different atoms without re-parametrization.

## 2.1. The Amber Force Field

The Amber force field (AFF) functional form is described as minimalist by the same authors[28], with the bonds and angles represented by a simple harmonic expression, the van der Waals (VdW) interaction represented by a Lennard-Jones potential, electrostatic interactions described by a Coulombic interaction of atom-centered point charges, and dihedral angle modeled with a simple set of parameters, often only specified by the two central atoms.

$$E_{total} = \sum_{bonds} K_r (r - r_{eq})^2 + \sum_{angles} K_\theta (\theta - \theta_{eq})^2 + \sum_{dihedrals} \frac{V_n}{2}[1 + \cos(n\phi - \gamma)] + \sum_{i<j} \frac{A_{ij}}{R_{ij}^{12}} - \frac{B_{ij}}{R^6} + \frac{q_i q_j}{\epsilon R_{ij}} \tag{2.1}$$

Electrostatic and VdW interactions are calculated only between atoms in different molecules or for atoms in the same molecule separated by at least three bonds. Those non-bonded interactions separated by exactly three bonds ("1-4 interactions") are reduced by the application of a scale factor.
The parameters in the AFF are fitted in three different manners:

1. $r_{eq}$, $K_r$, $\theta_{eq}$, $H_\theta$, are parametrized to reproduce the experimental normal modes frequencies of a set of molecule.
2. $V_n$ is fitted to reproduce *ab-initio* energies.

3. $A_{ij}$, $B_{ij}$ are parametrized using *monte carlo* method.
4. The atomic partial charges derive from *ab-initio* calculation and subsequent RESP fitting.

The Amber force field has been designed to describe biological compounds in particular proteins and DNA, but the recent introduction of a new set of atom type and parameters, called general Amber force field (GAFF)[29], has improved the capability of this MM method toward all the organic molecules.

# 3. Exploring potential energy surfaces

The identification of stationary points on the potential energy surface (PES), which have a chemical meaning, remain a troublesome issue[30,31]. Two are the most important reasons of this difficulty, first a complete exploration of the PES is impossible due to the big number of internal coordinate of molecules (6N-6 where N is the number of atoms), therefore is the chemist who must identify, by the so called chemical intuition, the most relevant starting geometry to reach the desired point. Second the algorithms often fail locating *minima* and most often fail locating *transition states* (TS). In fact the localization of TS resembles more art than technique.
Not only stationary points are important to understand the reactivity of a molecular system. I f 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 and the second part of the chapter summarize their the most important properties.

## 3.1. Characterization of Stationary Points

The sufficient conditions to identify a *minimum* of a function $f(\boldsymbol{x})$ in *n* variable is the vanishing of the first derivative and a positive second derivative. This are the particular conditions for *minima,* but it can be defined a general role to identify stationary points[31]:

$$\left\{\begin{array}{c} \nabla f(\boldsymbol{x}')=0 \\ In\nabla^2 f(\boldsymbol{x}')=[n-k,k,0] \end{array}\right\} \Rightarrow stationary\ point \qquad (3.1)$$

where the gradient and the Hessian are given by

$$\nabla f(\boldsymbol{x}') = \begin{bmatrix} \frac{\partial f}{\partial x_1} \\ \vdots \\ \frac{\partial f}{\partial x_n} \end{bmatrix}, \qquad \nabla^2 f(\boldsymbol{x}') = \begin{bmatrix} \frac{\partial f}{\partial x_1 \partial x_1} & \cdots & \frac{\partial f}{\partial x_1 \partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial f}{\partial x_n \partial x_1} & \cdots & \frac{\partial f}{\partial x_n \partial x_n} \end{bmatrix} \quad (3.2)$$

and *In* **M** is the inertia of real symmetric matrix **M**.
A stationary point with *k* different from zero is called *k'th order saddle point,* where *k* indicate the number of direction with negative curvature. On the potential surfaces *minima* and first order saddle points correspond to equilibrium geometries and transition states (see Figure 3.1 and 3.2). Higher order saddle point on PES are of no interest.

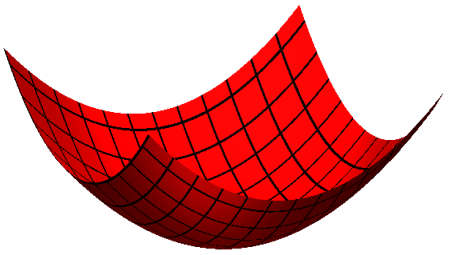


*Figure **3.1**: Example of two dimensional minimum.*

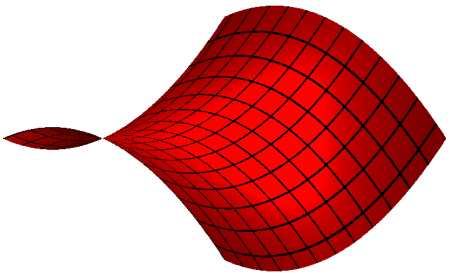


*Figure **3.2**: Example of two dimensional transition state*

## 3.2. The linear model

The simplest model applied only to locate *minima*, is the *local linear model*[31]:

$$m_A(\boldsymbol{x}) = f(\boldsymbol{x}_c) + \boldsymbol{g}_c^t \boldsymbol{s} \quad (3.3)$$

where

$$\boldsymbol{s} = \boldsymbol{x} - \boldsymbol{x}_c \quad \text{and} \quad \boldsymbol{g}_c = \nabla f(\boldsymbol{x}_c) \quad (3.4\ a,b)$$

The linear model require only the energy and the first derivative, but It does not contain the information about the curvature, and cannot give any information regarding the stationary point found (*minimum*, TS, ...).
Nevertheless thanks to its simplicity this approach is used by a large number of MM packages, in particular combined with the *line search* (see section 3.5) constitute the *steepest descent method.*

## 3.3. The quadratic model

The problem of the linear model is that it gives no information about the curvature of $f(\boldsymbol{x})$. This is provided by the more useful *local quadratic model* or *second order model(SO)*[31]:

$$m_q(\boldsymbol{s}) = f(\boldsymbol{x}_c) + \boldsymbol{g}_c^t \boldsymbol{s} + \frac{1}{2}\boldsymbol{s}^t \boldsymbol{B}_c \boldsymbol{s} \tag{3.5}$$

the symmetric matrix $\boldsymbol{B}_c$ is the exact Hessian at the geometry $\boldsymbol{x}_c$:

$$\boldsymbol{B}_c = \nabla^2 f(\boldsymbol{x}_c) \tag{3.6}$$

In some cases the Hessian matrix is computed though approximated method (see below).
To determinate the stationary points of the quadratic model we differentiate the model and set the result equal to zero. We obtain a set of linear equation:

$$\boldsymbol{B}_c \boldsymbol{s} = \boldsymbol{g}_c \tag{3.7}$$

which has a unique solution,

$$\boldsymbol{s} = -\boldsymbol{B}_c^{-1} \boldsymbol{g}_c \tag{3.8}$$

The quadric model therefore has one and only one stationary point. For example this is a *minimum* if $\boldsymbol{B}_c$ is positive define. $\boldsymbol{s}$ is called *Newton steep* (or *quasi Newton steep* if the Hessian is approximated), it form the basis of the *Newton's method*. Unfortunately the use of the *Newton steep* can lead to unexpected results (several time is to large), but it ensure that inside the step $\boldsymbol{s}$ it must be a stationary point.
The *second order model* it is computationally expensive since it require energy first and second derivative, but it give complete information about slope and curvature at point $\boldsymbol{x}_c$. For this reason the most part of the modern software use SO models with approximate Hessian matrix, several method has been developed to provide the best estimation of $\boldsymbol{B}_c$, the widely used is th BFGS discussed in the next section.

## 3.4. Hessian update

If the exact Hessian is unenviable or computationally expansive, it is possible to use an approximation. The update techniques are designed to determinate an approximate Hessian $\boldsymbol{B}_u$ at point

$$\boldsymbol{x}_u = \boldsymbol{x}_c + \boldsymbol{s}_c \tag{3.9}$$

we define the gradient difference:

$$\boldsymbol{y}_c = \boldsymbol{g}_u - \boldsymbol{g}_c \tag{3.10}$$

expanding $\boldsymbol{s}_c$ around $\boldsymbol{x}_u$ gives

$$\boldsymbol{y}_c = \boldsymbol{B}_u \boldsymbol{s}_c + O(\boldsymbol{s}_c^2) \tag{3.11}$$

which shows that the gradient difference $\boldsymbol{y}_c$ contains a component of finite-difference approximation to the exact Hessian along the direction $\boldsymbol{s}_c$.
Based on the (3.11) all the Hessian update are required to fulfill the *quasi-newton condition*:

$$\boldsymbol{y}_c = \boldsymbol{B}_u \boldsymbol{s}_c \tag{3.12}$$

To locate *minima* the most used Hessian update algorithm is the *Broyden-Fletcher-Goldfarb-Shanno* (BFGS) *update* given by:

$$\boldsymbol{B}_u = \boldsymbol{B}_c + \frac{\boldsymbol{y}_c \boldsymbol{y}_c^t}{\boldsymbol{y}_c^t \boldsymbol{s}_c} - \frac{\boldsymbol{B}_c \boldsymbol{s}_c \boldsymbol{s}_c^t \boldsymbol{B}_c}{\boldsymbol{s}_c^t \boldsymbol{B}_c \boldsymbol{s}_c} \tag{3.13}$$

While for transition states it can be appropriate the *Powell-symmetric-Broyden* (PSB) *update* given by:

$$\boldsymbol{B}_u = \boldsymbol{B}_c + \frac{(\boldsymbol{s}_c^t \boldsymbol{s}_c)\boldsymbol{T}_c \boldsymbol{s}_c^t + (\boldsymbol{s}_c^t \boldsymbol{s}_c)\boldsymbol{s}_c \boldsymbol{T}_c^t - (\boldsymbol{T}_c^t \boldsymbol{s}_c)\boldsymbol{s}_c \boldsymbol{s}_c^t}{(\boldsymbol{s}_c^t \boldsymbol{s}_c)^2} \tag{3.14}$$

where

$$\boldsymbol{T}_c = \boldsymbol{y}_c - \boldsymbol{B}_c \boldsymbol{s}_c \tag{3.15}$$

A condition to get reasonable geometry optimization using approximate Hessians is good starting matrix $\boldsymbol{B}_c$.
The Hessian update techniques are usually combined to *line search* and the resulting minimization algorithms are called *quasi-newton methods*.

## 3.5. Steep: direction and size

Both linear and quadratic model have a weak point, they can be applied only in the vicinity of the stationary point of interest, and the iterative application of them need a procedure to calculate the steep size and the direction for each iteration. In fact several steep, using linear or quadratic model, has to be performed to reach the stationary point
The difference between a good and a bad optimization algorithm is the strategy for calculating steep size and direction. Several methods have been developed for this purpose, in this section are summarized some of them.

The first and easiest way to define the direction is to choose the s*teepest descent (z)*:

$$\boldsymbol{z}^t \boldsymbol{g} < 0 \tag{3.16}$$

This direction is useful when linear model are applied. The size of the move in this case is calculated using an l*ine search* approximated algorithm.
An alternative to the steepest descent direction is the conjugate direction, which produces generally faster convergence than *steepest descent* directions. In practice after the first steep the new direction is not the *steepest descent* but a perpendicular direction respect to the previous move.
Usually the previous method are not adopted when a quadratic model is applied except the *line search* which is applied to determined the size and the direction is given by the direction of *Newton steep* (see section 3.3).
More sophisticated techniques can be used to determinate direction and size like the *trust region minimization* or the *rotational function minimization*. Without enter in the detail of these methods they are generally more accurate than previous one and they can also be used combined with *line search.*

## 3.6. Stopping criteria

The iterative procedure which consist in several single optimization steep, using one of the method above described, stops when certain condition are satisfied. The most straightforward require the norm of the gradient to be smaller than some threshold:

$$|\boldsymbol{g}| \leqslant \epsilon \tag{3.17}$$

or similarly a the displacement from one geometry to the subsequent has to be smaller than a threshold:

$$|\boldsymbol{s}| \leqslant \epsilon' \tag{3.18}$$

This two condition usually have to be satisfied at the same time.

## 3.7. Conical intersection and non crossing role

In diatomic molecules the PES of two states (*eg*. the ground state and the first excited state) will only intersect if the states have a different (spatial or spin) symmetry. However, an analogous statement is *not true* of polyatomic systems[32]: *two PES of a polyatomic molecule can in principle intersect even if they belong to states of the same symmetry and spin multiplicity*.
This sentence leaves open the question whether such intersections *ever* occur in polyatomic systems. We can try to give a quantitative analysis of this situation[33]. We imagine that all but two of the solutions of the Schrödinger equation for the electronic wave function have been found, and that $\Phi_1$ and $\Phi_2$ are any two functions which, together with the found solutions, constitute a complete orthonormal set. Of course the two missing solutions correspond to the two states (whose energy is $E_1$ and $E_2$) whose crossings we are interested in. Then it must be possible to express

each of the two remaining electronic eigenfunctions (which describes the states we want to examine) in the form:

$$\Psi = c_1 \phi_1 + c_2 \phi_2 \tag{3.19}$$

The well known secular eigenvalue equation is obviously expressed as:

$$\begin{bmatrix} H_{11}-E & H_{12} \\ H_{21} & H_{22}-E \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \end{bmatrix} = 0 \tag{3.20}$$

and after very simple passages we can write down the expressions for the energies $E_1$ and $E_2$ of the two states as:

$$E_1 = \frac{H_{11}+H_{22}-\sqrt{(H_{11}-H_{22})^2+4H_{12}^2}}{2} \tag{3.21}$$

$$E_2 = \frac{H_{11}+H_{22}+\sqrt{(H_{11}-H_{22})^2+4H_{12}^2}}{2} \tag{3.22}$$

where, for the matrix elements:

$$H_{11} = \langle \phi_1 | H | \phi_1 \rangle, \quad H_{22} = \langle \phi_2 | H | \phi_2 \rangle, \quad H_{12} = H_{21} = \langle \phi_1 | H | \phi_2 \rangle. \tag{3.23 a,b,c}$$

Thus, in order to have degenerate solutions (the radicand must vanish), it is necessary to satisfy two independent conditions:

$$H_{11} = H_{22}, \quad H_{12} = H_{21} = 0 \tag{3.24 a,b}$$

and this *requires the existence of at least two independently variable nuclear coordinates*. In a diatomic molecule there is only one variable coordinate - the interatomic distance - so the non-crossing rule follows: for states of different (spatial or spin) symmetry, $H_{12}$ is always zero and we have to verify the only condition $H_{11} = H_{22}$; this is possible for a suitable value of the single variable coordinate. Otherwise, if the two states have the same symmetry, they will not intersect; but in a system of three or more atoms there are enough degrees of freedom for the rule to break down: the two conditions can be simultaneously satisfied by choosing suitable values for two independent variables, while the other $n$-2 degrees of freedom ($n$=3$N$-6) are free to be varied without exiting from the crossing region.
If we denote the two independent coordinates by $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$, and take the origin at the point where $H_{11} = H_{22}$ and $H_{12} = H_{21} = 0$, the secular equations may be cast in the form:

$$\begin{bmatrix} W+h_1\boldsymbol{x_1}-E & l\boldsymbol{x_2} \\ l\boldsymbol{x_2} & W+h_2\boldsymbol{x_1}-E \end{bmatrix}\begin{bmatrix} \boldsymbol{c_1} \\ \boldsymbol{c_2} \end{bmatrix}=0 \qquad (3.25)$$

or

$$\begin{bmatrix} W+(m+k)\boldsymbol{x_1}-E & l\boldsymbol{x_2} \\ l\boldsymbol{x_2} & W+(m-k)\boldsymbol{x_1}-E \end{bmatrix}\begin{bmatrix} \boldsymbol{c_1} \\ \boldsymbol{c_2} \end{bmatrix}=0 \qquad (3.26)$$

where

$$m=\frac{1}{2}(h_1+h_2), \quad k=\frac{1}{4}(h_1-h_2) \qquad (3.27)$$

The eigenvalues are:

$$E=W+m\boldsymbol{x_1}\pm\sqrt{k^2\boldsymbol{x_1^2}l^2\boldsymbol{x_2^2}} \qquad (3.28)$$

and this is the equation of a *double cone* with vertex at the origin. For this reason, such crossing points are called *conical intersections*. Indeed, if we plot the energies of the two intersecting states against the two internal coordinates $\boldsymbol{x_1}$ and $\boldsymbol{x_2}$ (whose values at the origin satisfy the two conditions $H_{11}=H_{22}$ and $H_{12}=H_{21}=0$) we obtain a typical double cone shape (see Scheme 3.1)

Let's try to have a deeper insight into the physical meaning of the two conditions $H_{11}=H_{22}$ and $H_{12}=H_{21}=0$. If we consider $\phi_1$ and $\phi_2$ as the *diabatic* components of the *adiabatic* electronic eigenfunction (a *diabatic* function describes the energy of a particular spin-coupling, while the adiabatic function represents the surface of the real state), the crossing condition (real or avoided) is fulfilled when the two diabatic components $\phi_1$ and $\phi_2$ cross each other, and this happens when $H_{11}=H_{22}$, *i.e.* the energy of the two diabatic potentials ($H_{11}$ is the energy for the diabatic function $\phi_1$ and $H_{22}$ is the energy for the diabatic function $\phi_2$) is the same.

At the crossing of the diabatic functions ($H_{11}=H_{22}$, the expressions for the energies of the two real states (see eq. 3.21 and 3.22) become:

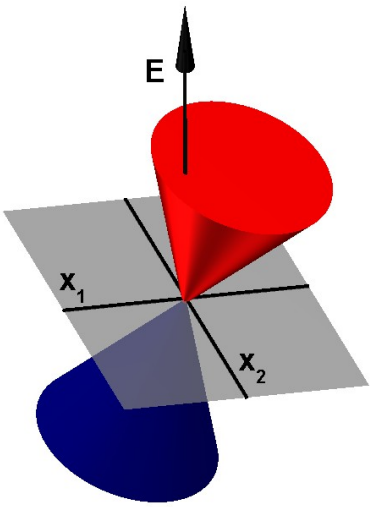
$$E_1=H_{11}-H_{12}, \quad E_2=H_{11}+H_{12} \qquad (3.29\ a,b)$$

and we see that the energy gap between the two real states is:

$$E_2-E_1=2H_{12} \qquad (3.30)$$

Thus, if the exchange term $H_{12}$ is *not zero*, the crossing will be avoided and the potential surfaces of the two real states will “diverge”, being one of the two energies slower and the other higher then the diabatic energy$H_{11}$. Moreover, the value of the exchange term $H_{12}$ determines how deep the avoided crossing *minimum* is (small

values will generate deep *minima*, big values shallow *minima*). In general, $H_{12}$ is zero (and the crossings will be real) when the two electronic states have a different (spatial or spin) symmetry, while it is usually assumed *not zero* for states of the same symmetry (which will generate avoided crossings). Anyway, we have shown that this rule is true only for diatomic molecules: in a polyatomic system we can have real crossings for suitable values of a pair of independent coordinates ($\boldsymbol{x}_1$ and $\boldsymbol{x}_2$), which will simultaneously satisfy the (3.24).
In conclusion the following statement is true:
*for a polyatomic system two states (even with the same symmetry) will intersect along a n-2 dimensional hyperline as the energy is plotted against the n internal nuclear coordinates.*




*Scheme **3.1**: Double cone representation of a conical intersection. Blue cone is the ground state while red cone is the excited state.*

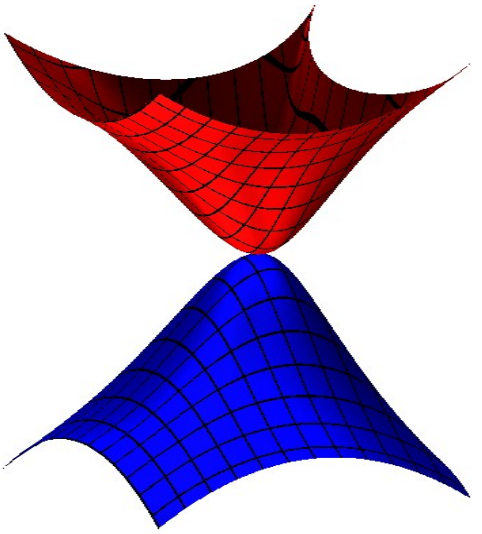


*Figure **3.3**: Example of two dimensional conical intersection.*

## 3.8. The "Physical Chemistry" of Conical Intersections

To understand the relationship between the surface crossing and photochemical reactivity, it is useful to draw a parallel between the role of a transition state in thermal reactivity and that of a conical intersection in photochemical reactivity[34].
In a thermal reaction, the transition state (TS) forms a bottleneck through which the reaction must pass on its way from reactants (R) to products (P) (Scheme 3.2). A transition state separates the reactant and product energy wells along the reaction path. An accessible conical intersection (CI) (Scheme 3.3) also forms a bottleneck that separates the excited state branch of the reaction path from the ground state branch. The crucial difference between conical intersections and transition states is that, while the transition state must connect the reactant energy well to a *single*

product well via a single reaction path, an intersection is a "*spike*" on the ground state energy surface (see Scheme 3.3) and thus connects the excited state reactant to *two or more* products (P and P') on the ground state via a branching of the excited reaction path into several ground state relaxation channels. The nature of the products generated following decay at a surface crossing will depend on the ground state valleys (relaxation paths) that can be accessed from that particular structure.
Theoretical investigations of surface crossings have required new theoretical techniques based upon the "mathematical" description of conical intersections and we now briefly review the central theoretical aspects. The double cone shape of the two intersecting potential energy surfaces can only be seen if the energies are plotted against two special internal geometric co-ordinates of the molecule ($\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ in Scheme 3.1). The co-ordinate $\boldsymbol{x}_1$ is the gradient difference vector:

$$\boldsymbol{x}_1 = \frac{\partial (E_1 - E_2)}{\partial \boldsymbol{q}} \tag{3.31}$$

while $\mathbf{x}_2$ is the gradient of the interstate coupling vector:

$$\boldsymbol{x}_2 = \left\langle \boldsymbol{C}_1 \frac{\partial \hat{H}}{\partial \boldsymbol{q}} \boldsymbol{C}_2 \right\rangle \tag{3.32}$$

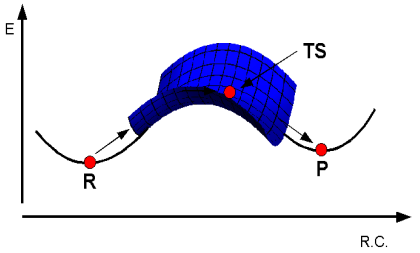
where $\boldsymbol{C}_1$ and $\boldsymbol{C}_2$ are the *configuration interaction* (CI) eigenvectors in a CI problem and $\hat{H}$ is the *CI* Hamiltonian. The vector $\boldsymbol{x}_2$ is parallel to the non-adiabatic coupling vector:

$$g(\boldsymbol{q}) = \left\langle \psi_1 \left| \frac{\partial \psi_2}{\partial \boldsymbol{q}} \right. \right\rangle \tag{3.33}$$

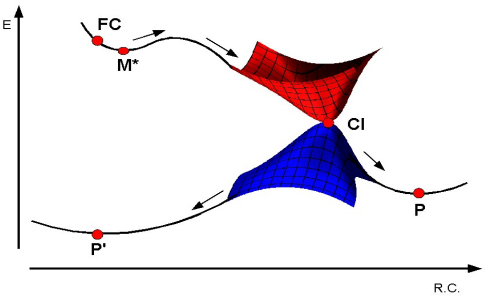
These geometric co-ordinates form the so called "branching space". As we move in this plane, away from the apex of the cone, the degeneracy is lifted (see Scheme 3.1), and ground state valleys must develop on the lower cone. In contrast, if we move from the apex of the cone along any of the remaining *n*-2 internal co-ordinates (where *n* is the number of degrees of freedom of the molecule), the degeneracy is not lifted. This *n*-2 dimensional space, called "intersection space", is a hyperline consisting of an infinite number of conical intersection points.
Non adiabatic events (transition from the excited state to the ground state at the CI) pose a serious challenge because the non-adiabatic transition is rigorously quantum mechanical without a well-defined classical analogue. At a crude level of theory[35] the probability of a surface hop is given as:

$$P = e^{-\frac{\pi}{4}\zeta} \tag{3.34}$$

where the Massey parameter is:




*Scheme **3.2**: typical behavior of a thermal reaction.*




*Scheme **3.3**: Typical behavior of a photochemical reaction.*

$$\zeta = \frac{\Delta E(\boldsymbol{q})}{\hbar |\dot{\boldsymbol{q}}| g(\boldsymbol{q})} \tag{3.35}$$

Thus this simple theory predicts that radiationless transitions will occur when the energy gap $\Delta E(\boldsymbol{q})$ is small and the scalar product between the velocity vector and the non-adiabatic coupling $g(\boldsymbol{q})$ is large. Practically it can be demonstrated that unless the $\Delta E(\boldsymbol{q})$ is less than 2 kcal/mol the decay probability is vanishingly small.
Often the chemically relevant conical intersection point is located along a valley on the excited state potential energy surface. The system fall down from the *Franck-Condon* (FC) region and arrive at the excited state *minimum* M*, which is connected to the CI trough a TS. However, there are situations where there is no transition state connecting M* to the intersection point or where an excited state intermediate on the upper energy surface does not exist. In such situations, mechanistic information must be obtained by locating the lowest lying intersection point along the *n*-2 intersection space of the molecule.

## 3.9. Conical intersection optimization

Often the chemically relevant conical intersection point is located along a valley on the excited state potential energy surface. The system fall down from the Franck-Condon region and arrive at the excited state *minimum* M*, which is connected to the CI trough a TS. However, there are situations where there is no transition state connecting M* to the intersection point or where an excited state intermediate on the upper energy surface does not exist. In such situations, mechanistic information must be obtained by locating the lowest lying intersection point along the *n*-2 intersection space of the molecule.
In practice, in order to properly locate these low-energy stationary points where two potential energy surfaces have the same energy, one must carry out constrained geometry optimizations where the geometry is optimized in directions orthogonal to the two directions $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$. It is important to appreciate that the gradient on the excited state PES will not be zero at an optimized conical intersection point, since it "looks like" the vertex of an inverted cone (seeScheme 3.3). Rather, it is the projection of the gradient of the excited state PES onto the orthogonal complement of

$\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ (*i.e.* the n-2 dimensional hyperline) that goes to zero when the geometry of the conical intersection is optimized. This situation is distinguished from an "avoided crossing *minimum*" of two surfaces which is a real *minimum* where the gradient on the excited state PES would go to zero.
Thus, in an optimized conical intersection point, two conditions have to be fulfilled:

$$E_2 - E_1 = 0\,, \quad \frac{\partial E}{\partial \boldsymbol{x}_3} = \frac{\partial E}{\partial \boldsymbol{x}_4} = ... = \frac{\partial E}{\partial \boldsymbol{x}_n} = 0 \qquad (3.36\ \text{a,b})$$

an alternative and related method which has been implemented in the *Gaussian03*[36] program package is the following. For the minimization $E_1$ and $E_2$ in the $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ plane we have:

$$\frac{\partial (E_a - E_b)^2}{\partial \boldsymbol{q}} = 2(E_a - E_b)\boldsymbol{x}_1 \qquad (3.37)$$

where $\boldsymbol{x}_1$ is the gradient difference vector. The length of $\boldsymbol{x}_1$ has no significance - only its direction. ($|\boldsymbol{x}_1|$ will be large if the potential energy surfaces have opposite slope but very small if they have nearly the same slope). This means that the size of the step should only depend upon $E_2 - E_1$, and suggests that we should take the gradient along the step to the *minimum* of $E_2 - E_1$ to be:

$$\boldsymbol{f} = \frac{(E_1 - E_2)\boldsymbol{x}_1}{\sqrt{\boldsymbol{x}_1 \boldsymbol{x}_1}} \qquad (3.38)$$

Clearly $\boldsymbol{f}$ will go to zero when $E_2 = E_1$, independently of the magnitude of $\boldsymbol{x}_1$. If we now define the projection $\boldsymbol{P}$ of the gradient of $E_2$ onto the (n-2) orthogonal complement to the plane $\boldsymbol{x}_1$, $\boldsymbol{x}_2$ as:

$$\boldsymbol{g} = \boldsymbol{P}\frac{E_1}{\partial \boldsymbol{q}} \qquad (3.39)$$

then the gradient to be used in the optimization becomes:

$$\boldsymbol{g} = \boldsymbol{g} + \boldsymbol{f} \qquad (3.40)$$

Notice that, starting from an arbitrary initial structure (initial point) the method locates, loosely, the closest intersection point. After that it moves downhill along the intersection space.

# 4. QM/MM method

The study of large molecular systems takes advantage of the use of computational methods based on Molecular Mechanics (MM)[37,38]. These methods simulate atoms almost as rigid charged spheres; the interactions between atoms are modeled on the basis of chemical connectivity, using simple harmonic potentials (or sometimes more complex functions) to describe bonds, angle bending and torsions. Non-bonded electrostatic and Van der Waals interactions are accounted for on the basis of the charge (or dipole) assigned to the atoms using a simple Culombic potential and by means of a Lennard-Jones potential (or similar), respectively. The analytic functional form of the equations used to compute energies and forces make the MM calculation fast even for large molecules. Anyway, the drawback of these methods is their inability in describing processes involving a change of the "nature" of atoms. Thus, chemical reactions (where bonds are broken and formed, and where connectivity and atom-types are not preserved) cannot be described by MM methods. On the other hand, computational methods based on Quantum Mechanics (QM)[2] have proved to be successful in dealing with such problems because they explicitly treat the electrons (and their couplings) by means of the calculation of the associated wave function.

Unfortunately, the QM methods are much more expensive than MM ones in terms of computational cost and cannot be used to study very large systems. Thus, a problem arises when studying the chemical reactivity of large molecular systems; in response to this need the last three decades were spent on theoretical studies for the development of new computational methods. A promising technique is the partitioning of the whole system (called *real* in the following discussion) in two regions (see Figure 4.1): a small part, containing the atoms involved in the chemical process, is described at QM level, while the remaining atoms are treated at MM level, in order to speed-up the calculation and simulate (although at a lower level) the influence of the environment on the reactive core. This hybrid approach is usually called "QM/MM"[39,40,2,41,1]. Anyway, this method has to overcome a difficult technical problem, which is often a source of significant errors and approximations: the correct description of the boundary region. Thus, the approximations adopted to deal with the interface between the two regions have a fundamental importance for a correct

simulation of their reciprocal influence. Consequently, great care is addressed to this problem when developing QM/MM methods[40,2,42], because it strongly influences the capability of reproducing the effects of the surrounding regions on the investigated process.
The wide literature[39,40,2,41,1,42,43,44,45] on hybrid methods gave us a complete landscape on past attempts to develop a solid and general QM/MM approach. We took advantage of this experience for developing a new versatile QM/MM algorithm and in this paper we describe the details of our implementation, focusing into its advantages as well as its limitations. More generally, our algorithm is a hybrid approach that acts as an interface between different programs; these programs execute the single phases of the whole calculation and are allowed to communicate each other by the interface, whilst the partial data are gathered and manipulated to give the final result. Modularity is the main feature of this approach, allowing the user to tailor the computational level by selecting (and combining together) specific programs according to the specific requirements of the investigation. The final purpose being to perform the computational task with the more appropriate (and efficient) computational tools available. A similar approach was used by P. Sherwood and co-workers to construct their own software, called ChemShell[41,46].

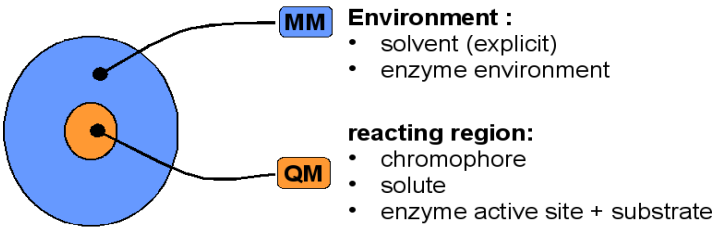



*Figure **4.1**: Schematic representation of the hybrid QM/MM approach.*

## 4.1. The QM/MM partitioning scheme

In order to construct a versatile QM/MM scheme we adopted a modular approach, combining different available codes to exploit their recognized ability in dealing with a specific aspect of the whole problem. Such an approach allows a great flexibility in the partitioning of the system, because the computation of energy and forces on one hand, and the projection of the new geometry on the other, are performed separately at different points of the whole QM/MM calculation (see Figure 4.2). More specifically, energy and forces are calculated according to a two-region partitioning of the system, where a QM and an MM calculation are independently performed on the two aforementioned sub-regions. This information is then processed (see below for the details) to obtain the total QM/MM energy and the complete set of forces. These data are then used to generate the new geometry and during this step the system is also partitioned into two independent regions, which do not necessarily coincide with the first ones: the external/larger one (**Opt1**) is handled by means of a fast and rough algorithm, like the Steepest Descents, while the smaller region (**Opt2/MD**) (including the **QM** sub-region) is treated with a more sophisticated

optimization algorithm, like BFGS[46], or with a molecular dynamics code based on the Beeman or the Velocity-Verlet algorithms[38].
In our approach the **QM** region is a subset of the **Opt2/MD** region. When **QM** and **Opt2/MD** do coincide, then the system is divided in two layers called respectively *high*, treated at QM level and optimized with the accurate algorithm, and *low*, treated at the MM level and optimized independently using the fast algorithm. When **QM** and **Opt2/MD** do not coincide (i.e. **Opt2/MD** is larger than **QM**), then some MM atoms are involved, together with the QM atoms, in the accurate geometry optimization or MD (i.e. these MM and QM atoms are coupled together). This buffer (intermediate) region of MM atoms, called *medium*, represents a significant improvement and makes the difference in respect to the three layers ONIOM(MO1:MO2:MM) approach from Morokuma[43,44], where the system is divided in three layers: the inner one (MO1, which is treated at high QM level), the outer (MM, which is treated at MM level) and the intermediate layer (MO2, which is treated at a low QM level, often semiempirical). This scheme is intended to improve the simulation of the electrostatic interaction between QM and the MM regions by interposing a buffer layer treated at a low QM level. In our experience we found that a simpler scheme, like our one, is efficient if an appropriate MM force field is used and if an Electrostatic Embedding[2,41] scheme (see below) is adopted. In our approach the intermediate region has a different function in respect to Morokuma's approach and it is used to improve the reliability and efficiency in the optimization (or molecular dynamic) phase. The capability of handling coupled together the *high-medium* region (**Opt2/MD**) allows the explicit treatment of large molecular motions around the reactive region without increasing the computational cost, because energy and forces of non-reacting (MM) atoms included in the **Opt2/MD** region are computed at the MM level.

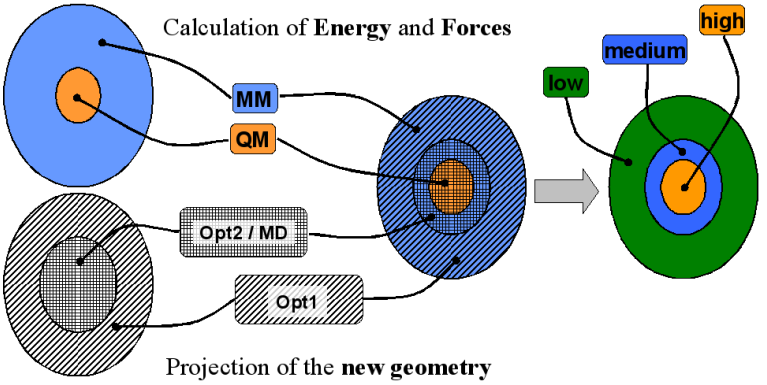



*Figure **4.2**: Graphical representation of the tree-layer scheme adopted in our code.*

As mentioned above, our scheme uses the Electrostatic Embedding[40,2,41,1] approach to account for the electrostatic influence of the surrounding MM region on the QM layer. Analogously, the effect of charge changes occurring during the chemical process in the *high* (QM) layer is accounted for by using in the MM calculations the

QM atomic point charges coming from QM computations. Furthermore, our approach also allows a full independent optimization of the *low* layer (using the fast algorithm) per each optimization step of the *high-medium* layer; this feature nearly resembles the so called "micro-iteration"[47] technique of ONIOM that, for example, can be useful in the study of a QM solute in a MM solvent. Some authors[48] pointed out the need of taking into account, when studying complex systems like enzymes, the contribution from the protein reorganization energy. We developed the aforementioned optimization scheme to deal with "the linear response of the protein to the movement along the reaction coordinate", in the aim of improving the description of this phenomenon.
The algorithm is intended to deal with all the possible combinations of layers in order to customize calculations according to the specific problem under investigation. The full set of possible combinations actually supported is reported in Table 4.1; the calculation level is indicated using up to three figures, which are the first figures of the name of the layers used. This notation will be used all through the paper.

*Table **4.1**: Calculations type*

| **Calculation type** | **Layers used** | | |
|---|---|---|---|
| | *high* | *medium* | *low* |
| HML | X | X | X |
| HM | X | X | - |
| ML | X | - | X |
| H | X | - | - |
| M | - | X | - |

For example, HML refers to a three layers system: this can be useful, for example, to study the reaction mechanism of an enzymatic process when including a small part of the active site and the substrate in the *high* layer, assigning the remaining part of the active site to the *medium* layer and leaving all the other residues of the enzyme in the *low* layer. An HM calculation can be employed for the accurate optimization, or for the molecular dynamics, of the whole molecular system and allows a great saving in computational resources by treating at the QM level (*high* layer) only the reactive sub-region, while the remaining part of the system is treated at the MM level (*medium* layer). As mentioned above, the HL scheme can give good results for the study of a solute (*high* layer) in a solvent (*low* layer) composed by a large number of explicit molecules. The ML calculation is fully performed at the MM level; it can be used, for example, to obtain a preliminary guess for a critical point on the Potential Energy Surface (PES) associated to the reaction under study. In the H (full QM) calculation the energy and force evaluation from a code is combined with the optimization (or MD) algorithm of a different code to study, for example, photoinduced ultra-fast processes with QM-MD simulations. The M (fully MM) calculation can be used to get a guess geometry for an H calculation. ML and M

calculations can be used only for *minima* or conformational transition states (with no change in the chemical connectivity); they also give the opportunity to use the force field from an MM code and the optimization or MD algorithm from a different one.

## 1.1. The boundary region

As stated in the introduction, handling the boundary between the QM and MM regions need extreme caution, because wrong assumptions can easily lead to unphysical results. In some cases the boundary does not go through a covalent bond: this is the case of a solute (QM level) immersed in a large number of explicit solvent molecules (MM level); this case is very easy to handle and do not need special assumptions However, in many cases, one cannot avoid passing the QM-MM boundary crosswise one (or more) covalent bond(s), like it is for enzymatic systems. Two strategies have been developed to overcome this problem: a) the *atom-link* approach[1,49,50,51] and b) the modified orbital methods[1,52,53,54]. We will focus on the former approach because we adopted it to handle boundary regions in our QM/MM method.

When the QM-MM boundary cut a covalent bond, then the valence of the frontier QM atom (Q1 in Figure 4.3) remains unfilled; we call *model* this dangling QM structure and we can observe that it coincides with the *high* layer, as mentioned above. To perform the QM calculation we need to saturate the dangling bond with a "frontier atom". This link atom is usually a hydrogen atom, in which case from *model* we obtain *model-H* (Figure 4.4). The *atom-link* approach has, like each boundary treatment, strong advantages as well as weaknesses[2]; it is widely used and proved to give good results, it is straightforward to understand and it is easy to implement. However, the introduction of artificial atoms that are not present in the original system, can give problems in obtaining the total QM/MM energy value and in geometry optimizations.

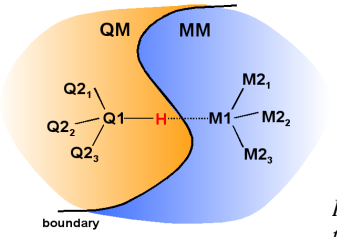



*Figure **4.3**: Boundary region crossing a covalent bond.*

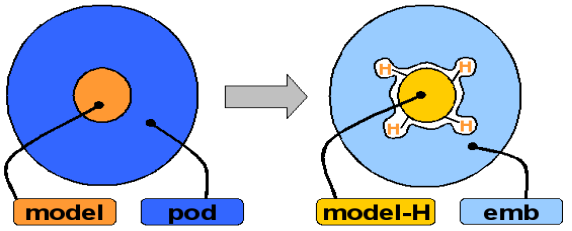



*Figure **4.4**: The atom-link approach implies a change in the original charges (pod) of the MM region to be introduced in the QM calculation, according to the Electrostatic Embedding scheme. The procedure to obtain the new set of charges (emb) is described in the text.*

### *4.1.1. The atom-link position*

Obtaining the correct QM/MM energy can be difficult because *model-H*, used for QM calculation, do not coincide with the *high* layer, due to the presence of additional hydrogen link atoms and because the MM atomic point charge at the boundary have

to be modified[2] (see further sections). We adopted a subtractive approach, similar to the ONIOM scheme [43,44], but not identical to it, because of a different partitioning scheme (see subsection 3).
To improve the geometry optimization procedure and avoid the artifacts arising from the presence of foreign H atoms, we tried to eliminate their extra degrees of freedom by making their position depending by the position of the Q1 (QM) and M1 (MM) atoms. Thus, during a geometry optimization or a molecular dynamics the position of the hydrogen link atom is reassigned at each step and is chosen to be on the line connecting Q1 an M1 atoms, removing two degrees of freedom (see Figure 4.3). The last degree of freedom can be eliminated in several ways, but the simplest way is to keep fixed the Q1-*H* distance. Herein, we impose a fixed distance of 1.09 Å, which is a standard value for a C($sp^3$)-H bond; in principle, to reduce the error associated with this technique, this choice implies the boundary bond type to be only C($sp^3$)-C($sp^3$). We are now developing new recipes for selecting different boundary bonds[2]. It is worth noting that the bond itself between Q1 and M1 is calculated molecular mechanically.

### 4.1.2. Handling QM/MM non-bonding cross-terms

In the presented algorithm, all the energy components of the *high* layer (QM region) are given by a QM calculation on *model-H*; the bonded (stretching, bending, torsions) and non bonded (electrostatic and Van der Waals) interactions within the *medium* and *low* layers (MM region) are handled at the MM level. Managing the cross terms is more difficult and can be done with different approaches. In our scheme, all the bonding and the Van der Waals terms are treated at the MM level, while electrostatic interactions between the QM and the MM region are computed at the high QM level. For this purpose, an Electrostatic Embedding[2] scheme was adopted: QM computation are performed on *model-H* surrounded by the atomic point charges of the MM layers (i.e. the *emb* charges). In our notation *pod* and *emb* are two subset of atomic point charges of the MM regions. The *pod* set contains the unmodified atomic point charges from the MM force field, while the *emb* set of charges do differ form the *pod* one (see Figure 4.4) only at the boundary region. Indeed, it has been shown that *pod* charges, when coupled to the *atom-link* method, lead to an unrealistic hyperpolarisation of the wave function at the Q1-*H* bond. This unphysical effect is mainly due to the presence of the atomic point charge on the M1 atom near the hydrogen atom link (see Figure 4.3). To avoid this problem, the charge on M1 is redistributed (i.e. it is spread) on the neighbouring MM atoms (M2) while it is set to zero on M1, so that the total charge of the system is preserved. It is worth to specify that the charge is not equally redistributed on the M2 atoms, but the added fractional charge is proportional to the module of the original MM charge hosted on that atom. We realized that this charge redistribution scheme (called *emb*) gives better results if compared to other strategies, like the zeroing of all the charges on the M2 atoms. Anyway, basing on recent publications[40,2,1,44], we are now trying to improve this redistribution scheme in order to preserve the polarity of the original Q1-M1 bond as well as the one of the M1-M2 bonds; this will be particularly important for the study of systems in which the frontier does cut bonds other than the standard (substantially non-polar) C($sp^3$)-C($sp^3$) ones.

## 4.2. QM/MM formalism

The QM/MM total energy follows a subtractive scheme[4] and can be written as:

$$E^{tot} = E_{QM}^{model-H} + E_{MM}^{real} + E_{QM}^{el._{model-H|emb}} - E_{MM}^{el._{model|pod}} - E_{MM}^{model-H} \qquad (4.1)$$

Where the first and third term are calculated at the QM level. $E_{QM}^{model-H}$ is the QM energy of *model-H in vacuo*, while $E_{QM}^{el._{model-H|emb}}$ is the electrostatic interaction between *model-H* and the charge embedding (*emb*). In this way the wave function is perturbed by the surrounding charges, ensuring the polarization of the QM part by the MM environment. This method is usually referred to as Electrostatic Embedding[7] and, practically speaking, the two $E_{QM}^{model-H}$ and $E_{QM}^{el._{model-H|emb}}$ terms are computed together via a single QM computation on *model-H* in the bath of *emb* charges.
The second, forth and fifth terms of eq. (4.1)( $E_{MM}^{real}$, $E_{MM}^{el._{model|pod}}$, $E_{QM}^{model-H}$) are MM energies. In particular, they can be divided into the following single contributions:

$$E_{MM} = E_{MM}^{real} - E_{MM}^{el._{model|pod}} - E_{MM}^{model-H} = E_{MM}^{pod} + E_{MM}^{VdW_{model|pod}} + E_{MM}^{bond,bend.,tors._{model|pod}} \qquad (4.2)$$

Where the MM contribution to the total energy has been split into three terms: the first ( $E_{MM}^{pod}$ ) involves only MM atoms, the second ( $E_{MM}^{VdW_{model|pod}}$ ) and the third ( $E_{MM}^{bond,bend.,tors._{model|pod}}$ ) are cross contributions between the QM and MM atoms (Van der Waals and frontier covalent bond terms, respectively).
Two approximations are implied in this description:

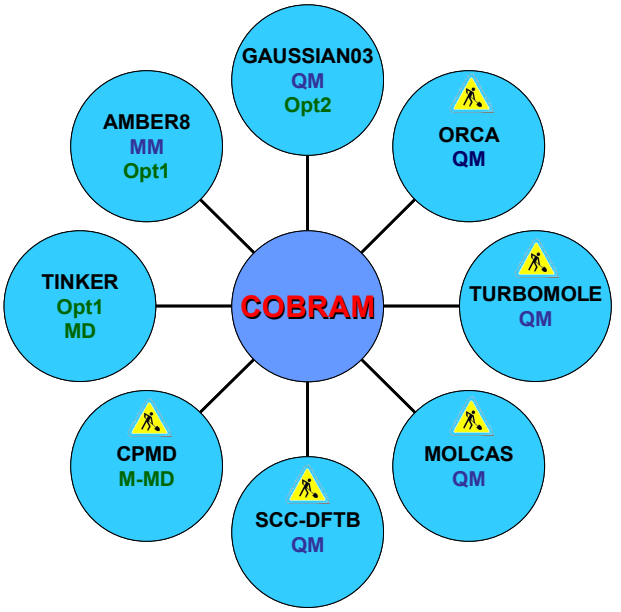
- The *atom-link* scheme (i.e. the atom added to saturate the QM part) is a hydrogen link atom (see section 4.2.1).
- The charges on M1 atoms (i.e. the frontier MM atoms covalently bond to the QM region) are redistributed to its nearest-neighbors (M2, see section 4.2.1).

Analytical derivatives of $E^{tot}$ lead to the forces used for the QM part, while the MM region is optimized on the base of a pure MM calculation. During this later (classical MM) step, the electrostatic interaction between the QM and MM part is computed using point charges derived from the QM wave function, according to the CHELPG or ESP schemes. The *atom-link* charge is redistributed over all the QM atoms in order to preserve the total charge (see Appendix 1 for technical details).

## 4.3. Implementation

A series of standard commercial packages such as MOLCAS55, Gaussian03[36], Turbomole[56], and ORCA[57,58](for QM calculations) and Amber[59] and Tinker[60] (for the MM part) may be plugged to the developed software, which we have called COBRAM (COmputational BRidge between Ab-initio and Molecular-mechanic) (see Scheme 4.1). In particular, an efficient strategy resulted when using Gaussian and Tinker algorithms to perform, respectively, geometry optimizations or molecular dynamics on the *high-medium* region, while Amber for MM calculations (i.e. optimization of the *low* layer and evaluation of MM energy and forces), since it

allows the treatment of systems whose parameters are not included in many force fields: indeed, the standard Amber force field (*ff03*)[61] is fully compatible with the Generalized Amber Force Field (GAFF)[29] developed to describe almost every organic molecule.
The flowchart of the QM/MM interface is presented in Scheme 4.2: in particular, points 2, 3, 4, 5 do need external MM programs to calculate energies and forces, while point 6 uses a QM software to get the electrostatically embedded wave function and its energy and first derivatives. Point 7 merges all these results to give the total energy and gradient. Finally, in the 9th section, an external program for geometry optimization or molecular dynamics is used to generate the new geometry to restart the cycle.



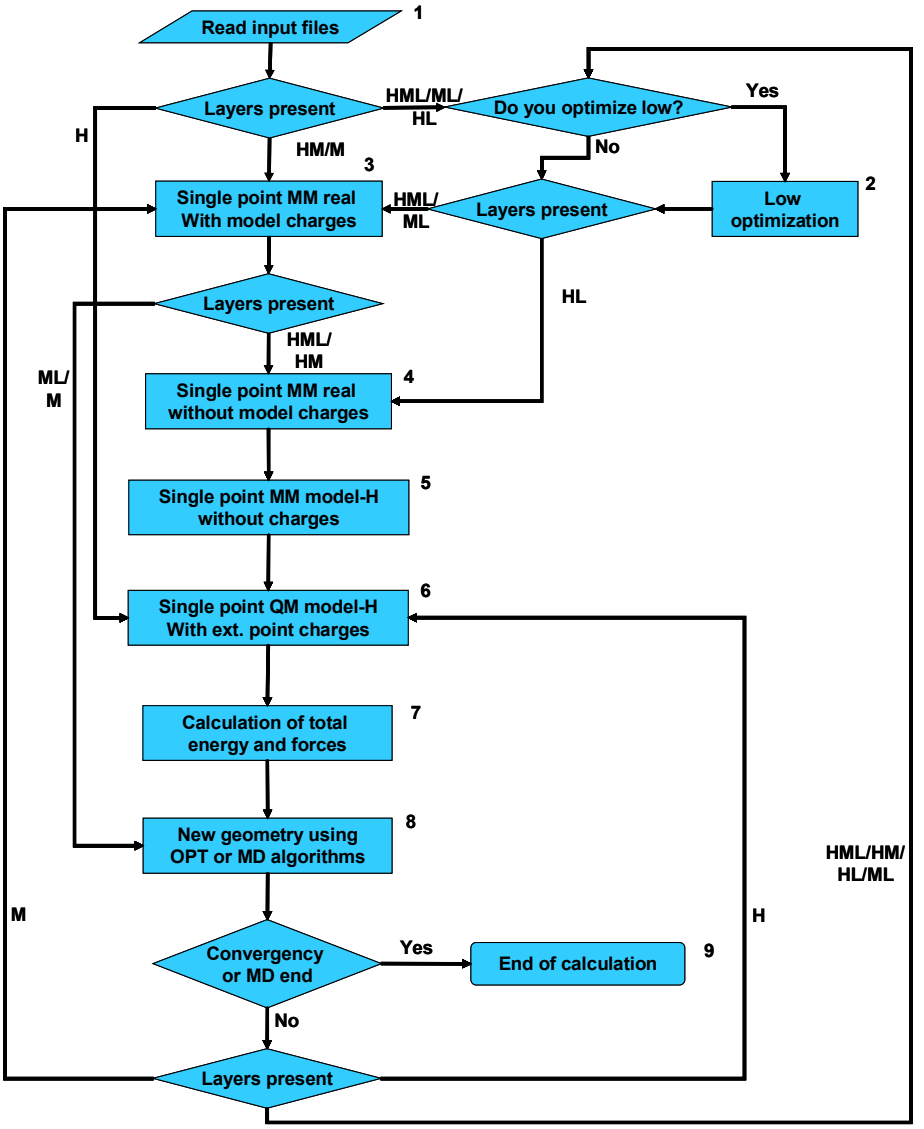

*Scheme **4.1**: Software Linked to COBRAM*

The calculation terminates when the convergence criteria are satisfied or when the time for a MD run is over.

## 4.4. Second derivatives

The calculation of the second derivative, of the wave function respect to nuclear positions, is mandatory to evaluate the curvature of the PES and to investigate molecular properties which depend directly from curvature (examples are IR and Resonance Raman frequency, zero point energy, ...).
The calculation of the Hessian matrix in our QM/MM program is carried out using a numerical procedure. It consist in a simple evaluation of the energy and first derivative in the region around the chosen point of the PES (i.e. *minimum* or saddle

point). Practically 6N single point calculation has to be performed, where N is the number of atroms of the *high* and *medium* layers. This approach is computationally expensive. In fact for each atom movement we have to reevaluate the wave function.




*Scheme **4.2**: QM/MM implementation flowchart.*

To move around this bottleneck we have introduced a simple approximation, which consists in keep unattached the QM wave function (and the its gradient) when the move involve only a MM atom. In this way we perform 6N single point where N is

the number of the atoms only of the *high* layer.
This approach leads to a quite modest approximation, because it involve only a small variation in the energy calculation of the term $E_{QM}^{el.,model-Hlemb}$ in the equation 4.1, and it safe a huge amount of computational time.

# Part 2. QM/MM Photochemistry

# 1. Hydrogen bonded rotaxane

Establishing methods for controlling aspects of large amplitude submolecular movements is a prerequisite for the development of artificial devices that function through rotary motion at the molecular level.
Large amplitude internal rotations that resemble to some extent processes found in authentic machinery have recently inspired analogic molecular versions of gears[62], turnstiles[63], brakes[64], ratchets[65], rotors[66], and unidirectional spinning motors[67,68,69,70] and are an inherent characteristic of many catenanes and rotaxanes[71]. Establishing methods for controlling aspects of such movements is a prerequisite for the development of artificial devices that function through rotary motion at the molecular level.

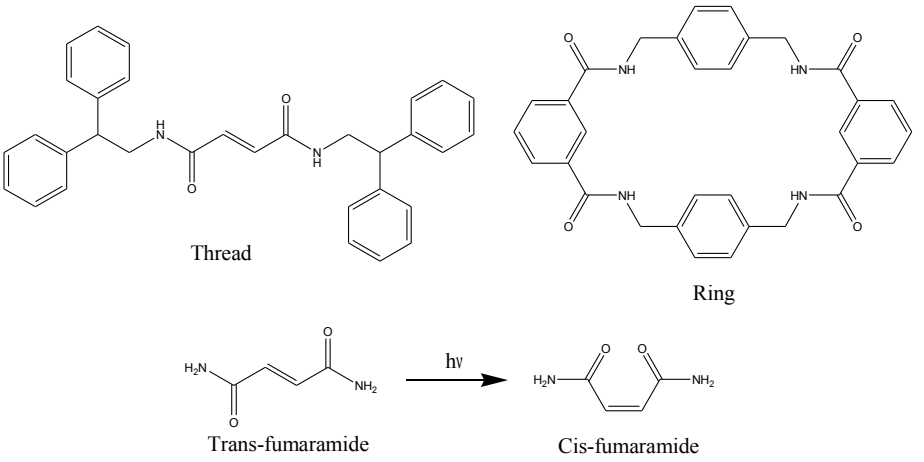



*Figure **1.1**: Structural formula of the two subunit of the rotaxane, ring and thread (above). Photo induced isomerization of the fumarammide (below).*

Fumaramide threads template the assembly of benzylic amide macrocycles around them to form rotaxanes in high yields[7]. This cheap and simple preparative procedure (suitable threads are prepared in a single step from fumaryl chloride and a bulky

primary or secondary amine) is particularly efficient because the trans-olefin fixes the two hydrogen bond-accepting groups of the thread in an arrangement that is complementary to the geometry of the hydrogen bond-donating sites of the forming macrocycle. However, the feature of the fumaramide unit that makes it such an effective template also provides an opportunity to enforce a geometrical change in the thread after rotaxane formation, thus altering the nature and strength of the interactions between the interlocked components. Isomerization of the olefin from E- to Z- must necessarily disrupt the near-ideal hydrogen bonding motif between macrocycle and thread and therefore also change any internal dynamics governed by those interactions.To test this idea, the photochemical isomerization of fumaramide in vacuo and fumaramide-based thread and rotaxane was investigated (see Figure 1.1).
Stand alone fumaramide has been be considered the model system to investigate the photoisomerization process which occurs in the isolated thread.
This chapter is divided in two part: the first analyze the peculiar photochemistry of the fumaramide as a model of the rotaxane thread. The second part describe the behavior on the excited state of the entire rotaxane. The investigation of the rotaxane properties could not possible without the help of a hybrid QM/MM approach.

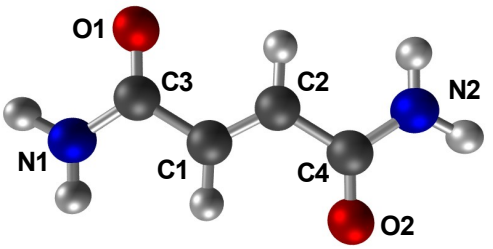



*Figure **1.2**: $C_2h$ fumaramide atoms numbering.*

## 1.1. Computational details

The fumaramide is a small molecule of 14 atoms and can be studied at high ab initio level. We have adopted a complete active space-self consistent field (CASSCF)/6-31G* level of theory, the active space is composed of 14 electrons in 10 orbitals, all the $\pi$ orbitals of the molecule and two lone pair of the oxygens atoms (see Figure 1.3). Multiconfigurational second-order perturbation theory reevaluation of the energy carried out with the complete active space second-order perturbation theory (CASPT2) method is then used to increase the accuracy of the path energy profile. To minimize the influence of weakly interacting intruder states the so-called imaginary level-shift technique was employed. The software packages used are Gaussian03 and MOLCAS6[55].
To compare the energies deriving from calculation using a different number of roots ,

we have performed state average single point at CASPT2//CASSCF/ano-s level on 5 roots.

## 1.2. Fumaramide: a thread model system

Since the interplay between the three low-lying excited-state surfaces in fumaramide is rather complicated, we begin our discussion with a simplified overview of our results before discussing the details. We have collected all the structural and energetic data Table 1.1 and 1.2 and in a schematic representation of the potential surface shown in Figure 1.4 and in Figure 1.5. Acronyms are used in the table and the figures to denote the energetic ordering of the states at the tabulated geometries.

### *1.2.1. Franck-Condon region and Characterization of the spectroscopic state.*

The first four electronic excited states have different nature, the first two (almost degenerate) are characterized by a single electronic transition $n \rightarrow \pi^*$ between orbitals 3-8 and 4-8, while the higher two by single $\pi \rightarrow \pi^*$ excitations. In particular $S_3$ is a bright (oscillator strength ~0.1) state which absorbs at 239nm (see Table 1.1). This value is close to the experimental maximum of the UV/VIS spectrum of the tread in solvent ($CH_2Cl_2$) of 254nm[72]. The red shift of the thread respect to the fumaramide can be attributed to the hyperconjugation effect of the substituents.
The spectroscopic state in mainly described by a *Homo-Lumo* excitation (orbitals 7 and 8 in Figure 1.3), they are located respectively on the C=C central bond (*C1-C2*) and on *C1-C3, C2-C4*.

### *1.2.2. Minimum energy path "a static point of view"*

The Figure 1.4 summarize the photochamistry of the fumaramide based only on the *minimum* energy points found on the potential energy surfaces. After photoexcitation to $S_3$ (in absence of triplet sensitizer), s-trans-fumaramide relaxes, only by stretching modifications, to **CI $S_3/S_2$**, a conical intersection which connect the $\pi \rightarrow \pi^*$ $S_3$ to the $n \rightarrow \pi^*$ state $S_2$. From this point the system can relax along $n \rightarrow \pi^*$ state or it can remain on $\pi \rightarrow \pi^*$ state and rotate the central C=C up to 90 degrees reaching a second CI. The geometry of **CI $S_3/S_2$** is characterized, respect to the ground state *minimum*, by an increasing of the *C1-C2* bond distance and a general breaking of the original $C_2h$ symmetry. Following the $n \rightarrow \pi^*$ state, the change in the force field (from $S_3$ to $S_2$) rearrange again the bond distances leading to a new conical intersection between $S_2$ and $S_1$ called **CI $S_2/S_1$ planar**. Then the last part of the excited state relaxation take place on the first excited state with $n \rightarrow \pi^*$ nature arriving on a *minimum* with planar structure (**Min-$S_1$**). All this part of the path, from **FC** to **Min-$S_1$**, is characterized by bond distances deformation and no torsion are involved. While, if we analyze the second way from the **CI $S_3/S_2$**, as mentioned above, it occur a rotation of the central C=C, this rotation does not require energy and lead to a 90 degrees twisted structure which correspond to a CI between $S_2$ and $S_1$ (**CI $S_2/S_1$ 90°**). Through a new relaxation the molecule fall into a new degeneracy region **CI $S_1/S_0$ 90°**, which is one of the possible path for the trans-cis fumaramide isomerization.

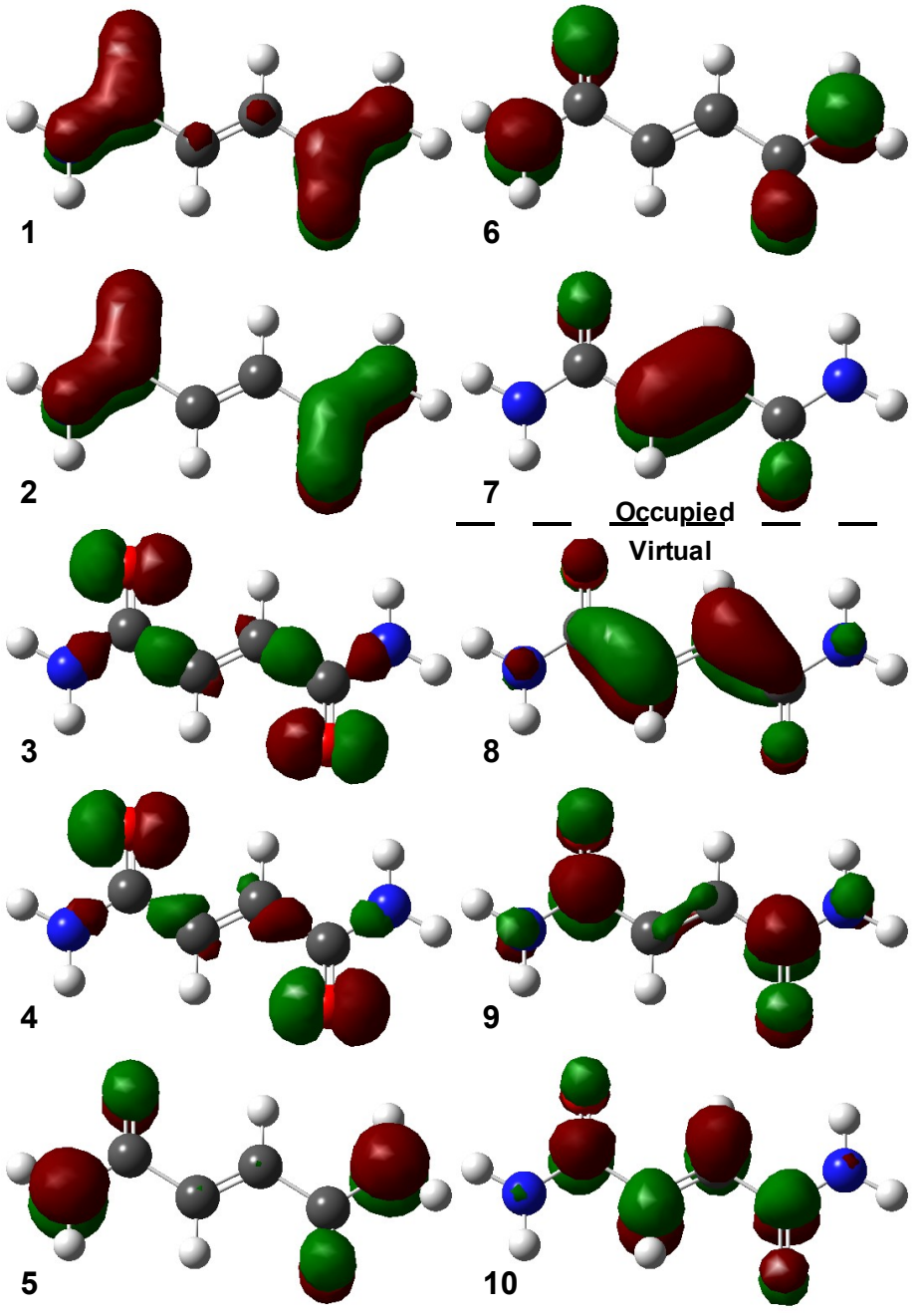



*Figure **1.3**: Active space MOs of the fumaramide*

*Table **1.1**: CASSCF and CASPT2 energies of fumaramide.*

| **Geometry** | | *State* | *CASSCF (ANO-S 3s2p1d)* | *CASPT2 (ANO-S 3s2p1d)* | *ΔE (kcal/mol) CASSCF* | *ΔE (kcal/mol) CASPT2* |
|---|---|---|---|---|---|---|
| **FC** | $A_g$ | S0 | -413.7670356036 | -414.9076535320 | 0.00 | 0.00 |
| | $B_g$ | S1 | -413.5839146556 | -414.7550634121 | 114.91 | 95.75 (299 nm) |
| | $A_u$ | S2 | -413.5789407935 | -414.7482691840 | 118.03 | 100.01 (286 nm) |
| | $B_u$ | S3 | -413.5150201586 | -414.7171293294 | 158.14 | 119.55 (239 nm) |
| | $A_g$ | S4 | -413.5126419040 | -414.7148590877 | 159.63 | 120.98 |
| | | T1 | -413.6349738427 | -414.7798028590 | 82.87 | 80.23 |
| | | T2 | -413.5920893713 | -414.7581152752 | 109.78 | 93.84 |
| | | T3 | -413.5879027053 | -414.7519170289 | 112.41 | 97.72 |
| | | T4 | -413.5637293311 | -414.7317599870 | 128.57 | 110.37 |
| | | T5 | -413.5573038138 | -414.7313386542 | 131.61 | 110.64 |
| **Min-$S_1$** | | S0 | -413.7114857237 | -414.8642809089 | 34.86 | 27.22 |
| | | S1 | -413.6398038284 | -414.7809352010 | 79.84 | 79.52 |
| | | S2 | -413.5621850356 | -414.7267525681 | 128.54 | 113.52 |
| | | S3 | -413.5526109618 | -414.7295974336 | 134.55 | 111.73 |
| | | S4 | -413.5269732329 | -414.6619648899 | 150.64 | 154.17 |
| | | T1 | -413.6466191308 | -414.7852708445 | 75.56 | 76.8 |
| | | T2 | -413.6393324995 | -414.7842537512 | 80.13 | 77.43 |
| | | T3 | -413.5648903224 | -414.7339874618 | 126.85 | 108.98 |
| | | T4 | -413.5619065706 | -414.7326357506 | 128.72 | 109.82 |
| | | T5 | -413.5201755508 | -414.7038818188 | 154.90 | 127.87 |
| **TS to CI $S_0/S_1$** | | S0 | -413.6431866203 | -414.7912824260 | 77.72 | 73.02 |
| | | S1 | -413.6257809034 | -414.7635583537 | 88.64 | 90.42 |
| | | S2 | -413.5336864335 | -414.6865756847 | 146.43 | 138.73 |
| | | S3 | -413.5298056754 | -414.7117934990 | 148.86 | 122.90 |
| | | S4 | -413.5181685604 | -414.6626253921 | 156.16 | 153.76 |
| **CI $S_0/S_1$ 90°** | | S0 | -413.6358739024 | -414.7897637363 | 82.30 | 73.98 |
| | | S1 | -413.6311881220 | -414.7782842658 | 85.24 | 81.18 |
| | | S2 | -413.5242014190 | -414.7018044565 | 152.38 | 129.17 |
| | | S3 | -413.5205589938 | -414.6806761893 | 154.66 | 142.43 |
| | | S4 | -413.5039748331 | -414.6680033289 | 165.07 | 150.38 |
| **CI $S_0/S_1$ 30°** | | S0 | -413.5965636873 | -414.7516539741 | 106.97 | 97.89 |
| | | S1 | -413.5906616143 | -414.7484781123 | 110.67 | 99.88 |
| | | S2 | -413.4907481209 | -414.6477179235 | 173.37 | 163.11 |
| | | S3 | -413.4360496166 | -414.6511272191 | 207.69 | 160.97 |
| | | S4 | -413.4116389043 | -414.6184258514 | 223.01 | 181.49 |
| **CI $S_2/S_1$** | | S0 | -413.7294942789 | -414.8812978851 | 23.56 | 16.54 |

| **Geometry** | *State* | *CASSCF (ANO-S 3s2p1d)* | *CASPT2 (ANO-S 3s2p1d)* | *ΔE (kcal/mol) CASSCF* | *ΔE (kcal/mol) CASPT2* |
|---|---|---|---|---|---|
| **planar** | S1 | -413.6107894143 | -414.7749988565 | 98.04 | 83.24 |
| | S2 | -413.6050596130 | -414.7657083858 | 101.64 | 89.07 |
| | S3 | -413.5543205145 | -414.7348162505 | 133.48 | 108.46 |
| | S4 | -413.5452185964 | -414.7360113411 | 139.19 | 107.71 |
| **CI $S_2/S_1$ 90°** | S0 | -413.6568348567 | -414.8170929486 | 69.15 | 56.383 |
| | S1 | -413.5879934127 | -414.7407636643 | 112.35 | 104.72 |
| | S2 | -413.5693104716 | -414.7389646913 | 124.07 | 105.85 |
| | S3 | -413.5155625643 | -414.7082705353 | 157.80 | 125.11 |
| | S4 | -413.4770017812 | -414.6933199063 | 182.00 | 134.49 |
| **CI $S_3/S_2$** | S0 | -413.7163695769 | -414.8710668583 | 31.79 | 22.96 |
| | S1 | -413.6204884764 | -414.7737600184 | 91.96 | 84.02 |
| | S2 | -413.5697286298 | -414.7507983380 | 123.81 | 98.43 |
| | S3 | -413.5610709272 | -414.7368704184 | 129.24 | 107.17 |
| | S4 | -413.5060404541 | -414.7015650924 | 163.77 | 129.32 |
| **$T_1$ min** | S0 | -413.6781984749 | -414.8196172021 | 55.75 | 55.24 |
| | S1 | -413.5723820019 | -414.7326167362 | 122.15 | 109.84 |
| | S2 | -413.5707183642 | -414.7277271876 | 123.19 | 112.90 |
| | S3 | -413.5229059856 | -414.6946816618 | 153.19 | 133.64 |
| | S4 | -413.5203908206 | -414.6921518040 | 154.77 | 135.23 |
| | T1 | -413.6792918045 | -414.8183826457 | 55.06 | 56.02 |
| | T2 | -413.5733559963 | -414.7310005048 | 121.53 | 110.85 |
| | T3 | -413.5724817391 | -414.7291904897 | 122.08 | 111.99 |
| | T4 | -413.5235493980 | -414.6944625591 | 152.79 | 133.78 |
| | T5 | -413.5211403648 | -414.6904781969 | 154.30 | 136.28 |

*Table **1.2**: Fumaramide bond distances.*

| *Bond* | *Geometries* | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | **FC** | **Min $S_1$** | **TS to CI $S_0/S_1$** | **CI $S_0/S_1$ 90°** | **CI $S_0/S_1$ 30°** | **CI $S_2/S_1$ planar** | **CI $S_2/S_1$ 90°** | **CI $S_3/S_2$** | **$T_1$ min** |
| *C1-C2* | 1.340 | 1.409 | 1.482 | 1.474 | 1.460 | 1.427 | 1.455 | 1.458 | 1.466 |
| *C1-C3* | 1.492 | 1.359 | 1.330 | 1.349 | 1.376 | 1.403 | 1.399 | 1.370 | 1.471 |
| *C2-C4* | 1.492 | 1.471 | 1.473 | 1.478 | 1.507 | 1.397 | 1.481 | 1.458 | 1.470 |
| *O1-C3* | 1.207 | 1.368 | 1.365 | 1.324 | 1.352 | 1.288 | 1.342 | 1.346 | 1.210 |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| *O2-C4* | 1.207 | 1.209 | 1.209 | 1.209 | 1.198 | 1.284 | 1.209 | 1.215 | 1.209 |
| *N1-C3* | 1.359 | 1.390 | 1.388 | 1.384 | 1.378 | 1.364 | 1.343 | 1.339 | 1.372 |
| *N2-C4* | 1.359 | 1.377 | 1.379 | 1.377 | 1.349 | 1.372 | 1.380 | 1.368 | 1.378 |

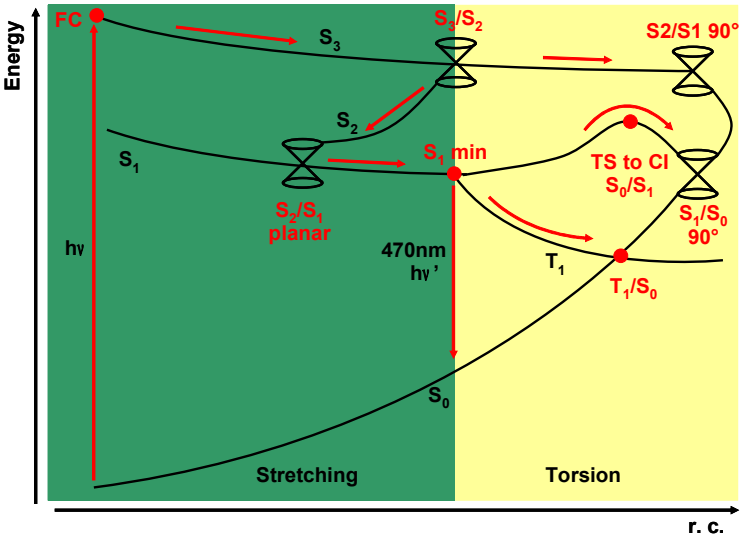



*Figure **1.4**: "Static interpretation" of the fumaramide photochemistry.*

The same funnel can be reached from the **Min-$S_1$** overcoming a barrier (**TS to CI $S_0/S_1$**) of 11 kcal/mol. Exists other two possible mechanism to recover the ground state from the **Min-$S_1$** , the emission at 470nm and the intersystem crossing between $S_1$ and $T_1$ with subsequent $T_1$ relaxation and back intersystem crossing $T_1$ to $S_0$. We can divide the photochemical paths in two families. The first involve only stretching motions is composed by: **CI $S_3/S_2$**, **CI $S_2/S_1$ planar** and **Min-$S_1$**, trhough these points the system start form an higher state and in few femtoseconds (barrierless process) it arrive on $S_1$ *minimum*, following the Kasha rule. The second family groups the geometries where the central C=C is rotated of 90 degrees (**CI $S_2/S_1$ 90°** and **CI $S_1/S_0$ 90°**), this points are unfavored respect to the previous for three reasons:

- ✔ The **CI $S_2/S_1$ 90°** is located at higher energy respect to **CI $S_2/S_1$ planar**.
- ✔ The system has to pass 11 kcal/mol TS to moves from **Min-$S_1$** to **CI $S_1/S_0$ 90°**.
- ✔ Rotation of the C=C needs time to occur because the kinetic energy has to be redistributed from the stretchings, which are initially populated, to the torsional mode. Besides the kinetic energy is not redistributed only on the C=C torsion, but it is dissipate on several normal mode.

Experiments support our hypothesis of unfavored C=C rotation, in fact the photoisomerization quantum yield of the thread from trans to cis fumaramide is only 0,044[72].

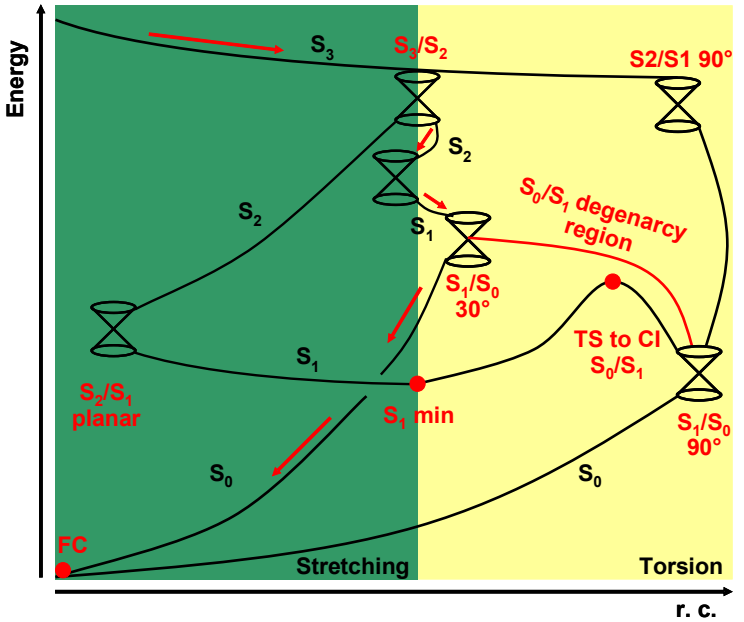



*Figure **1.5**: "Dynamic interpretation" of the fumaramide photochemistry.*

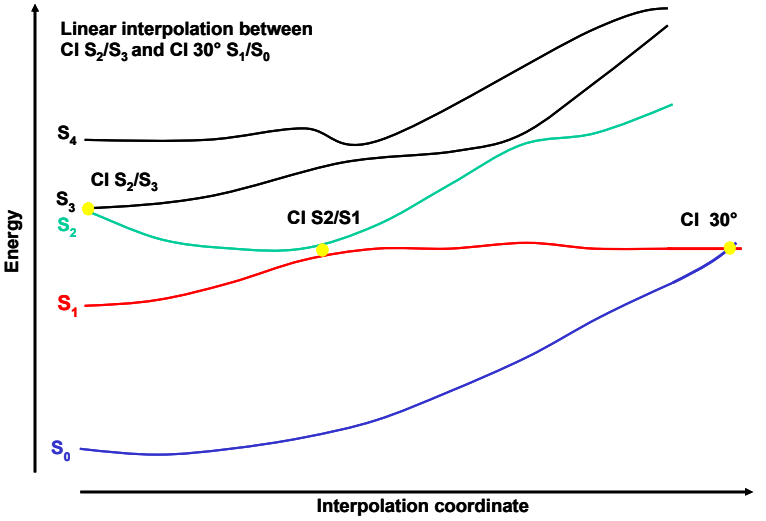



*Figure **1.6**: liner interpolation between **CI $S_3/S_2$** and **CI $S_1/S_0$ 30°**.*

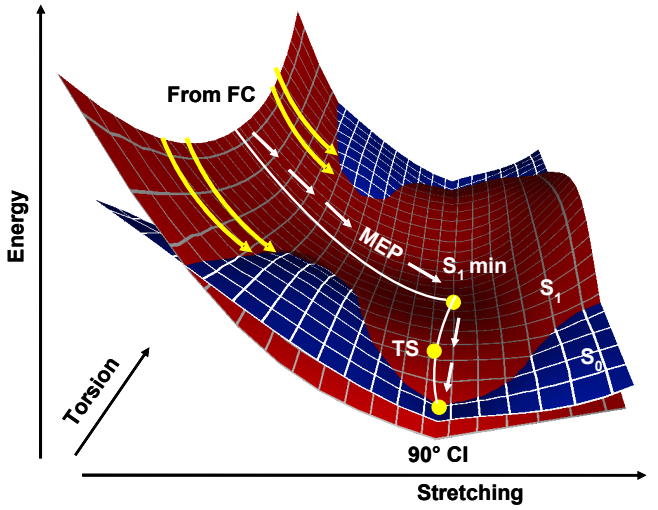



*Figure 1.7: Cartoon comparison of the "static and dynamic interpretation" of the fumaramide deactivation channels.*

It can be concluded that the **Min-S$_1$** collects the most part of the population and the more probable deactivation mechanisms involve or emission of intersystem crossing. Unfortunately both these chance are not back by experiments because no emission is observed and the photoisomerization quantum yield on singlet and sensitized triplet is very different (0,044 for singlet and 0,2 for triplet)[72].
The mechanism picture described above can be considered a static view the trans-fumaramide photochemistry, because it consider as populated points only the lower in energy. To interpret correctly the experimental results it is mandatory to explore also regions which are not on the *minimum* energy path.

### *1.2.3. Minimum energy path "a dynamic point of view"*

The inconsistency between experiment and theory, reported in the previous section, suggests us to explore different regions of the PES in order to identify new deactivation funnel. The low isomerization quantum yield indicates the presence of a conical intersection where the C=C central bond is not strongly twisted. Besides the absence of fluorescence and the different quantum yield between singlet and triplet point out that the system does not remain a long on the **Min-S$_1$**. Starting from these premises we have investigated the intersection space between $S_0$ and $S_1$, performing series of constrained CI optimization at different value of the C=C. Thanks to this procedure we have identified a intersection space which span from 90 (**CI S$_0$/S$_1$ 90°**) to 150 (**CI S$_0$/S$_1$ 30°**) degrees. In our hypothesis the system after photoexcitation relax to the **CI S$_3$/S$_2$** and then does not follow the the *minimum* energy valley on $S_2$ but it cross immediately $S_1$ and the fall in the wide $S_1/S_0$ CI region. This could happen because the molecule has a huge amount of kinetic energy, and it can populate easily zone of PES far from the MEP. A schematic representation of this behavior is reported in Figure 1.5. To support our reactivity scheme we have done a linear interpolation between **CI S$_3$/S$_2$** and **CI S$_0$/S$_1$ 30°** (see Figure 1.6) and we have

found that the system can pass from the first to the last point without overcame any barrier. We have also identified along the rigid scan CI between $S_2$ and $S_1$ suggesting that this deactivation path is very probable.
This "dynamic interpretation" of the disexcitation of fumaramide can be approximatively described by two, three dimensional, surfaces ($S_1$ and $S_0$) (see Figure 1.7), where on the x-axis is reported the stretching coordinate on the y-axis the rotation of the C=C and on the z-axis the energy. On this surfaces exist a region where for small torsional variation they cross and it is possible to pass from one to the other without energetic leap. So the system before reaching the *minimum* of $S_1$ can drop down to $S_0$. Under this hypothesis it come clear why no emission is observed. Moreover the low photoisomerization quantum yield is a consequence of the scarce rotation of the C=C at the CI geometry.

## 1.3. Fumaramide based rotaxane

### *1.3.1. Rotaxane: system setup*

To investigate the photo-switchable rotaxane based on a fumaramidic group we have employed our QM/MM code[3]. The system has been divided in two parts: the dimethyl-fumaramide treated at QM level (for detail see the section computational details on this chapter), and the rest of the thread and the ring are treated using the generalized amber force field (GAFF) at MM level (see Figure 1.8). To minimize the atom-link error the connection between the two part has been placed on a C-C bond. A two layer (high-medium) approach has been adopted (for a detailed description of the method see chapter 4 part I), where the MM part is optimized couplet to the QM one.

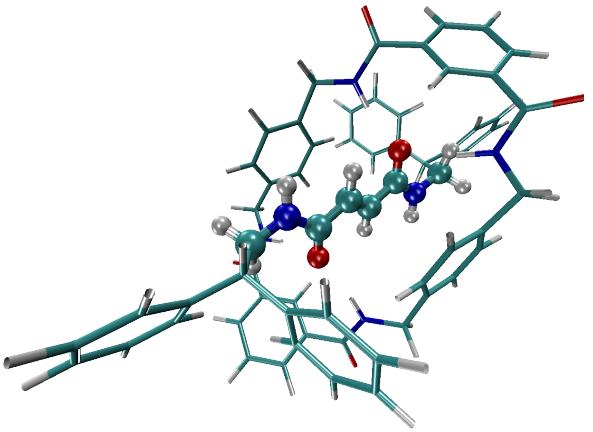


*Figure **1.8**: Fumaramide based rotaxane structure. The QM part is represented in ball & stick, while the MM part in tube.*

### *1.3.2. Ground state minimum*

We have tested the accuracy of our QM/MM approach comparing the ground state

*minimum* calculated at CASSCF/GAFF level to the previous published mm3 and crystallographic structures[73]. To measure the quality of our results we have compared the hydrogen bond distances, which connect the ring to the thread, obtained computationally and experimentally (see Table 1.3). the outcome is satisfying, the errors are less than 0,3 Å.

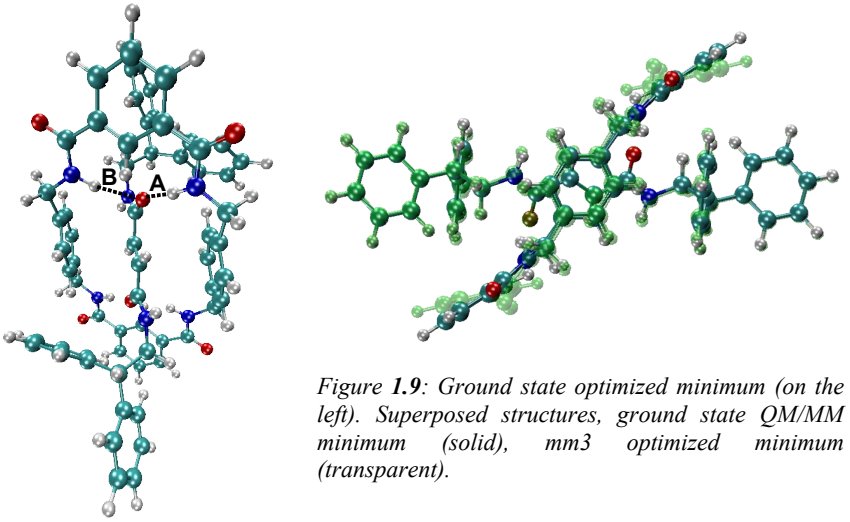



*Figure **1.9**: Ground state optimized minimum (on the left). Superposed structures, ground state QM/MM minimum (solid), mm3 optimized minimum (transparent).*

*Table **1.3**: Hydrogen bond distances obtained from QM/MM CASSCF/GAFF calculation and previous published mm3 and crystallographyc data. (numbering in Figure 1.9)*

| Method | H-bond A | H-bond B |
|---|---|---|
| QM/MM | 1,90 | 2,12 |
| X-ray | 2,22 | 1,94 |
| mm3 | 1,99 | 2,03 |

Starting from this result we have calculated the absorption spectra, the bright state (251 nm) is red shifted respect to the fumaramide in vacuo (239 nm) and in perfect agreement to the experimental value for the thread in solution (254 nm) (the rotaxane does not present *maxima* on its absorption spectra but only a rising band up to 200nm)[72].

### *1.3.3. First excited state*

Optimizing the geometry of the first excited state starting from the Franck-Condon point, leads to a *minimum* very similar to the **Min-$S_1$** of the fumaramide (see Figure 1.10). The excited state does not preserve the ground state symmetry and the

asymmetry is evident also in the wave function, where the main excitation is between a *n* orbital located only on one oxygen, to the $\pi^*$ orbital of the central C=C bond.
The differences between rotaxane and fumaramide become evident when starts the rotation of the C=C, in fact while in the fumaramide the rotation at 90 degrees leads to a CI, in the rotaxane we have not individuate any degeneracy (see Figure 1.11). Moreover in the supermolecule the energy grows continuously from 0 to 90 degrees, while in the isolated system after a TS the energy decrease until the CI. This trend can explained by steric interactions between the ring and the thread which play against the C=C rotation. A second reason for the unfavored twisting is the presence of hydrogen bonds and stacking interaction between the fumaramide and the ring, they are not as strong as in the ground state, nevertheless bonding interaction subsist also on $S_1$. Since the hydrogen bonds are mainly coulombic interactions, the reduction of the strength of the hydrogen bonds occurs only on the side of the molecule, where the oxygen is deprived of one electron and its charge decrease from -0,65 a.u. on $S_0$ to -0,27 a.u. On $S_1$.

### 1.3.4. $T_1$ state

The triplet state as $\pi \rightarrow \pi^*$ nature and the orbitals involved are the C=C bonding and antibonding, the result is a lowering of the C=C bond strength and the *minimum* structure on the fumaramide is rotated of 90 degrees on the C=C. There is a wide region in which the $T_1$ surface is degenerate to $S_0$ between 80 to 110 degrees. The behavior described above is typical of several molecule on the triplet state (stilbene, azobenzene, ecc.), but investigating the rotaxane $T_1$ state we have found that the impediment to the C=C rotation lead to a *minimum* rotated of only 50 degrees. For the $T_1$ state the strength of the non bonding interaction between the to subunit is similar to the ground state, because the excitation is located only on the central double bond leaving almost untouched the region of the molecule where the hydrogen bonds are are formed. Therefore the partial rotation of the *minimum* structure derive from equilibrium of two different factor: the force directly connected to the wave function nature which push the toward 90 degrees twisting, and the non bonding forces which try to maintain planar the fumaramidic group.

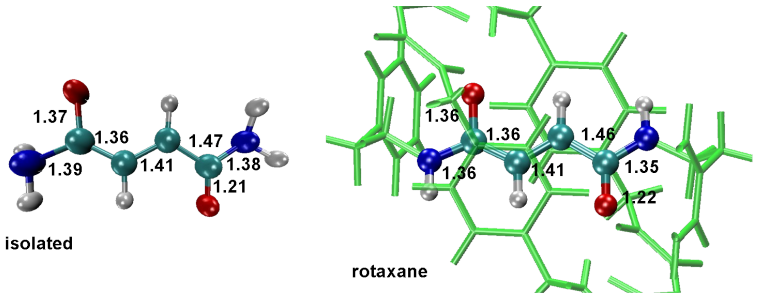



*Figure **1.10**: Bond distances of the $S_1$ minimum of fumaramide (on the left) and rotaxane (on the right)*

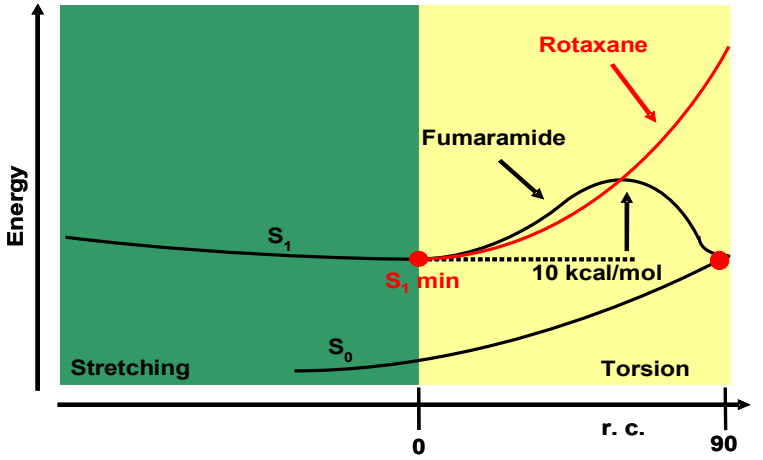



*Figure **1.11**: $S_1$ PES comparison between rotaxane and isolated fumaramide*

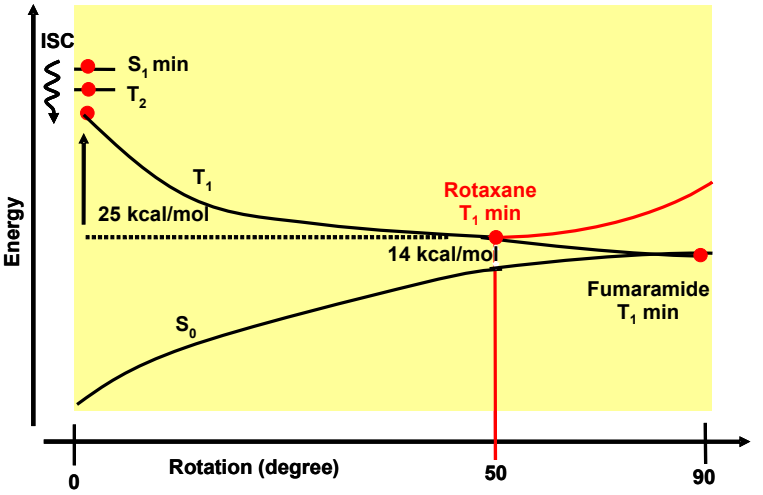



*Figure **1.12**: $T_1$ PES comparison between rotaxane and isolated fumaramide*

## 1.4. Conclusions

The investigation of the fumaramide photochemistry reveals several intriguing aspects, where different excited state are involved. We have first identified a pattern of possible deactivation mechanisms which involves emission, internal conversion and intersystem crossing. They are based on the analysis of the *minimum* energy path, where the main part of the population is collected on the $S_1$ *minimum*. From this region three are the way to return on the ground state: IC (trough C=C rotation), emission and ISC. All these possibilities could be conceivable, but experimental evidences are in contrast with them. So we suggest a different mechanism where the Kasha role breaks down. After the photoexcitation on the bright state ($S_3$) the system

does not reach the **Min-$S_1$**, but it cross earlier the $S_0$ PES returning on ground state *minimum*.
This is not the first example of Kasha role failure, nevertheless it is a very interesting molecule because it is small and it can be studied using very accurate method giving us a precise picture of its photochemistry.
From our study we can also suggest some experiments to verify our theoretical investigation. In particular up to now the experiments provide informations only exciting the bright state, what's happen if only $S_1$ is excited? In our opinion two different deactivation mechanisms could active the emission and the ISC, because the CI populated descending from $S_3$ is too high if we excite at $S_1$ wavelength. And what about the life time? Also in this case something different would be observed probably a longer lifetime, because the ISC and emission are usually slower than IC where no barriers are present.
From the theoretical point of view we are planning to execute a molecular dynamic simulation of the fumaramide in vacuo on the excited state at ab initio level. The inclusion of kinetic energy will be fundamental to understand if the non equilibrium points calculated since here are populated or not.
The complex photochemistry of fumaramide is present also on the rotaxane, but with an important difference: the environment effects play a fundamental role in the geometrical deformation. In fact while for the stretching relaxation we have not observed no differences between the isolated fumaramide and the rotaxane, when the chromophore start to rotate the PES of the supermolecule is completely different from the isolated one. The investigation of the rotaxane photochemistry is now limited to $S_1$, but in the next future we will study the higher states to reach a complete comprehension of the problem.

# 2. Phytochrome

The phytochromobilin (PΦB) and the biliverdin (BV) belong to the family of chromophores responsible for the phytochrome biological activity. Phytochrome exists in two thermostable forms: the physiologically inactive red-light absorbing from ($P_f$, $\lambda_{max}$ = 660 nm) and the physiologically active far-red-light absorbing form ($P_{fr}$, $\lambda_{max}$ = 730 nm)[74]. These two forms are photointerconvertible, and $P_f/P_{fr}$ thus constitutes a photochromic system. The phototransformation produce variation in proteic structure triggering a cascade of reactions. This photoinduced change is at the base of several photomorphogenic[75,76] functions of bacteria and plants. For instance, it stimulate the organism to search light sources (this mechanism is known as phototropism).

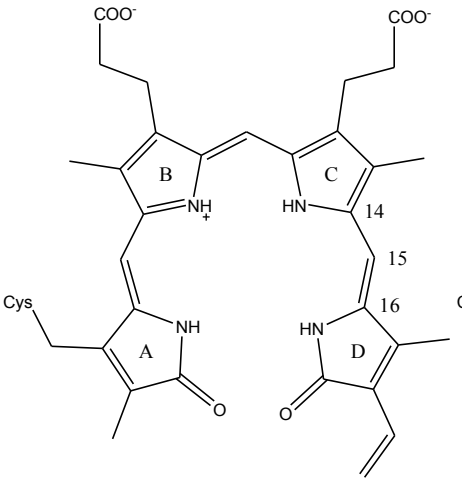



*Figure **2.1**: Structural formula of the biliverdin.*

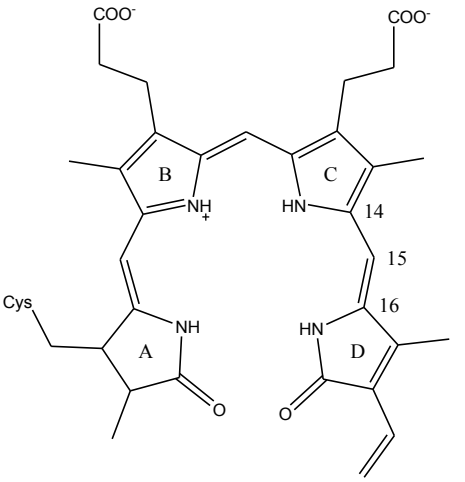



*Figure **2.2**: Structural formula of the phytochromobilin.*

The interconversion between the two different thermically stable forms ($P_f \leftrightarrow P_{fr}$) involves a cis-trans photoisomerization of the bilinic chromophore, which is linked to

the cysteine residue of the apoprotein[18], leading to a large conformation rearrangement of the phytochrome.
The PΦB (Figure 2.2) and the BV (Figure 2.1) chomophores have a open chain tetrapyrrolic structure, the only difference being a double bound located on the ring A. They are characterized by a remarkable conformational freedom, and a high polarizability of the π system. In the last three decades the conformation of the chromophores in the $P_f$ form has been studied through several spectroscopic techniques[77,78,79,80,81,82,83,84], but only the recent X-ray structure[85] has finally identified the correct conformation in protein (ZZZssa). Whereas the position of the isomerization is clear since years[86], the chromphore undergo two photoinduced isomerizations, the Z → E between C15 and C16, and the *anti* → *syn* between C14 and C15. What remains uncertain is the order of the events, namely if they are step wise of concerted. The conversion process involve several metastable intermediates, but only the first one (called lumi-R or $I_{700}$)[79,83,87,88] need light to be formed. Spectroscopic[81,80] and computational[89] studies support the hypothesis that formation of the lumi-R correspond to the Z → E isomerization, while the second steep, *syn* → *anti,* takes place subsequently on the ground state[90].
Studies to determinate the lifetime of the excited state (e.g. transient absorption and emission spectroscopy)[91,90,92,93] shows a multiexponential decay both in solvent and protein. In protein the fastest component is ~3 ps, but a second deactivation channel seems to be responsible of the formation of the primary photoproduct (lumi-R), which occurs in ~30 ps. Several interpretation has been suggested for this complex decay. For example, two different ground state species or excitation to two different excited states with different deactivation paths, have been suggested as possible hypothesis but the issue remains open. As it is in protein, also in solvent a multiexponential decay of similar lifetimes has been observed[93,94].
This chapter presents the results of the computational studies for the photochemical activity of phytochromobilin in vacuo and the biliverdin in protein with particular attention to the isomerization of the exocyclic double bonds. The spectroscopic state of both chromophores is characterized and full optimizations on $S_1$ has been performed to investigate the possible deactivation paths. In fact, while in protein the mechanism involve the C15-C16 bond, in solvent or vacuo we have discovered several possible paths which could recover the system to the ground state.
From our results we also suggest a possible interpretation of the picoseconds biexponential decay in solvent. In fact, the rotation of the two central bonds (between the rings B-C) are not equivalent. While one leads directly to a conical intersection with the ground state, the other drives the systems into a *minimum* close to a sloped deactivation funnel. Both these points can be reached by a common excited state intermediate (i.e. an S1 *minimum*) and without overcame high energy barriers.

## 2.1. Computational details

### *2.1.1. The model system*

The chromophore (PΦB) has been simplified to make feasible investigations using high level *ab initio* methods. For this purpose we have followed a decreasing approach: first we have optimized the unreduced PΦB and calculate the absorption

maximum (i.e. the vertical energy gap between the ground state *minimum* and the bright state $S_1$), then we have iterated the same procedure on several reduced systems. Firstly, we have cut the isopropilic groups, then also the A ring methyls have been dropped (and substituted by hydrogen atoms); finally, also methyls of the B and C rings have been removed. Comparing computed geometries and $S_0$-$S_1$ absorption energies for the four systems, we conclude that the molecule shown in Figure 2.3, where the two isopropilic groups and the two methyl groups of ring A have been replaced by hydrogen atoms, represents the best compromise between accuracy and computational cost. Here, all the other substituents of the original molecule have been retained to maintain the correct steric repulsion (and therefore the correct conformation) between the rings (see Figure 2.3).
Since in the present work we are not interested in the conformational analysis of the single exocyclic bonds, we have chosen the ZZZasa conformation as a representative starting point (see Figure 2.3).

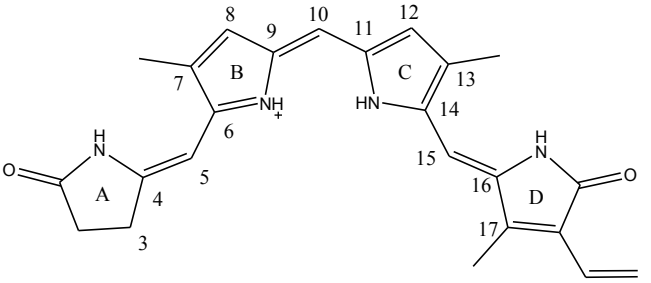



*Figure **2.3**: Structural formula of the phytochromobilin model ZZZasa.*

## *2.1.2. Methods*

We have performed DFT and TD-DFT calculations to define the model system (see section 2.1.1) by using the BP86 functional and the resolution of identity approximation with a 6-31G(d) basis set. These computations have been carried out using *TURBOMOLE 5.6*.[56].
Model system full unconstrained optimizations have been carried out using different *ab initio* and DFT-B3LYP methods. A 6-31G(d) basis set has been always adopted. We have performed optimizations using different methods to get a more complete set of data and compare the results. In particular, for ground state we have used HF, MP2 DFT and CASSCF, while for excited state CIS and CASSCF. These computations have been performed using the *Gaussian 03*[36] suite of programs.
To account for correlation effects, the energies of all the optimized geometries (whatever is the method used to generate these structures) have been reevaluated by using the CASPT2 approach available in the *MOLCAS6*[55] programs package. The successful performance of CASPT2 in computing experimentally accurate energies (with errors usually below 4-5 kcal/mol) and spectroscopic properties is well-established and is the reason for using this method in evaluating the energies of the investigated system. The choice of the active space is critical for CASSCF and CASPT2 computations. Unfortunately, in a system as large as РФВ, it is not possible

to include the full π system, which is composed by 28 electrons in 25 orbitals. Thus, we have drastically reduced it to a 8 orbitals in 8 electrons active π space. Unfortunately, this choice may create problems for the correct geometrical description of the system during CASCF geometry optimizations, because it seems not to be perfectly balanced. For this reason, we preferred also to use other than CASSCF optimized geometries, such as MP2 and CIS ones. In any case, single point CASPT2 calculations have been always performed on the optimized geometries to account for correlation energy and have a uniform energy scale (the 8 electrons / 8 orbitals active space has been reevaluate for each point including on it the 8 π orbitals with higher weight resulting from the CIS wave function analysis). To minimize the influence of weakly interacting intruder states the so-called imaginary level-shift technique was employed.

## 2.2. Results and discussion

### *2.2.1. Ground state geometry and vertical excitation*

The most relevant geometrical parameters of the ground state *minimum* are reported in Table 2.1 The MP2 and DFT results indicate as stable structure the resonance hybrid, where the positive charge is delocalized between the B and C rings. Both methods show a quasi symmetrical geometry respect to the central carbon atom C10. This behavior is justified by the structure of the A and D rings which differs each other only by a double bond and the allyl group. Oppositely, the CASSCF optimization in completely unbalanced in favor of one Lewis structure. We have attributed this comportment to a failure in the active space selection. In fact the 8 electrons 8 orbitals active space used include only the π orbitals of the C15-C16 bond, and not the specular C4-C5. We have also inverted the selection and after the optimization we have obtained the other Lewis structure. A further validation of the MP2 structure comes from its CASPT2 energy (which we consider as the most accurate method applied in our model) as compared to the CASPT2 values computed for the CASSCF *minima*: these calculations show that the Møller-Plesset geometry is about 15 kcal/mol lower than the multiconfigurational one (see Table 2.2) So it is clear that in this case we have to be careful in trusting CASSCF geometries. Unfortunately, it is not possible to further extend the active space because of the computational cost.

The first excited state is characterized by a Homo-Lumo (see Figure 2.4-2.5) single excitation; both orbitals are delocalized in the two central rings and the exocyclic bonds C4-C5, C15-C16, while the external rings are not involved. The transition has no effect in the dipole moment because of the two orbitals are located in the same region of the molecule and no intramolecular charge transfer take place. Nevertheless the state can be considered ionic.

Interestingly, the vertical absorption calculated for the MP2 geometry (599 nm) see Table 2.2) is in good agreement with the experimental value of 590 nm observed for the absorption maximum in non polar solvent ($CH_3Cl$)[94]. This result confirm the quality of our approach (i.e. ground state MP2 geometry).

*Table **2.1**: exocyclic bond distances (angstrom) and dihedral angle (degrees) of the ground and first excited states optimized geometries. Numbering in Figure 2.3.*

| **Geometrical parameters** | **Method** | | | |
|---|---|---|---|---|
| | *DFT-B3LYP* | *MP2* | *CASSCF(8,8)* | *CIS $S_1$* |
| C4-C5 | 1,366 | 1.362 | 1.355 | 1.355 |
| C5-C6 | 1.435 | 1.438 | 1.433 | 1.431 |
| C9-C10 | 1.393 | 1.394 | 1.342 | 1.394 |
| C10-C11 | 1.396 | 1.397 | 1.441 | 1.394 |
| C14-C15 | 1.432 | 1.435 | 1.474 | 1.417 |
| C15-C16 | 1.370 | 1.368 | 1.345 | 1.366 |
| C3-C4-C5-C6 | 178 | 177 | 179 | 177 |
| C4-C5-C6-C7 | -32 | -45 | -33 | -35 |
| C8-C9-C10-C11 | 169 | 165 | 178 | 168 |
| C9-C10-C11-C12 | 167 | 163 | 150 | 167 |
| C13-C14-C15-C16 | -31 | -44 | -66 | -28 |
| C14-C15-C16-C17 | 176 | 174 | 179 | 174 |

*Table **2.2**: CASPT2 energies of $S_0$ and $S_1$, dipole variation and oscillator strength of the ground state geometries optimized at DFT, CASSCF and MP2 level.*

| **Method** | *CASPT2 $S_0$ energy* | *CASPT2 $S_1$ energy* | *Dipole variation* | *$\Delta E$ $S_0/S_1$ kcal/mol (nm)* | *Oscillator strength* |
|---|---|---|---|---|---|
| CASSCF(8,8) | -1294.69833 | -1294.61155 | 8,9 | 54.45(525) | 0,60 |
| MP2 | -1294.72205 | -1294.64602 | 2,0 | 47.71(599) | - |

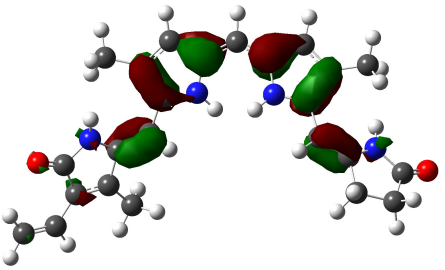


*Figure **2.4**: HF Homo orbital of the (PΦB) model.*

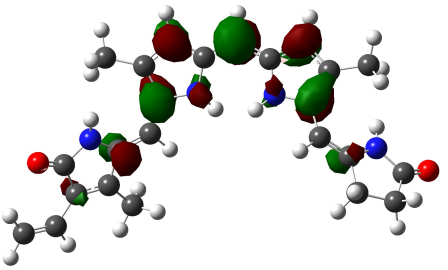


*Figure **2.5**: HF Lomo orbital of the (PΦB) model.*

### *2.2.2. $S_1$ minimum and deactivation paths*

A *minimum* has been located on $S_1$ by CIS optimizations from the FC point (**$S_1$-min**), which is geometrically very similar to FC and only 4,7 kcal/mol (PT2 energy values, see Table 2.3) below it. In this structure there is a small scratching rearrangement, the major differences being attributed to two torsional angles C4-C5-C6-C7 and C13-C14-C15-C16, which pass respectively from -45 to -35 degrees and from -44 to -28 degrees. This behavior suggests a slight improvement in the double bond character of C5-C6 and C14-C15.

At the **$S_1$-min,** the $S_0$-$S_1$ energy gap ΔE (708 nm) very well matches with the emission wavelength experimentally observed($\lambda_{max}$ 680-690 nm)[94]. Indeed, the **$S_1$-min** may be seen as the emitting region where the system stands for a while (few picoseconds) before it leaves this zone to fall in the deactivation funnel and ground state recovery occurs. In particular, four possible paths have been located, which lead to $S_0$/$S_1$ conical intersection funnels, each one involving the twisting of one extra-annular bond (dihedral angles C3-C4-C5-C6, C8-C9-C10-C11, C9-C10-C11-C12, C14-C15-C16-C17, respectively).

In particular, four twisted *minima* have been located at the CIS level about these bonds (which are called **CI-A**, **CI-B**, **CI-C** and **CI-D,** see Figure 2.11) After single point CASPT2 corrections, **CI-A**, **CI-B**, **CI-C** substantially appear as conical intersections where the S1 and S0 energies are degenerate (or almost degenerate). While **CI-D** does not present energy degeneracy, a conical intersection point has anyway been located very closed to this point and lower in energy.

To connect each CI point to the common excited state intermediate (**$S_1$-min**) we have performed four relaxed scans at the CIS level. These are only preliminary results because we have calculate only the CIS energies and no PT2 corrections have been applied yet; nevertheless, they can deliver qualitative informative about the presence of barriers which the system must overcame to reach the deactivation funnels. The rotation around the bonds linking the B-C rings are more favorite

To connect each CI point to the common excited state intermediate (**$S_1$-min**) we have performed four relaxed scans at the CIS level. These are only preliminary results because we have calculate only the CIS energies and no PT2 corrections have been applied yet; nevertheless, they can deliver qualitative informative about the presence of barriers which the system must overcame to reach the deactivation funnels. The

rotation around the bonds linking the B-C rings are more favorite(Figure 2.7 and 2.8) (the TS are ~7 kcal/mol) respect to the paths to **CI-A** and **CI-D** which present higher barriers (Figure 2.6 and 2.9). We have to be very careful in the interpretation of the CIS results, because the relative positions of the **CI-B** and **CI-C** respect to the **$S_1$-min** are inverted respect to the CASPT2 values (compare Table 2.3 to Figure 2.7 and 2.8) this different behavior can lead to a different position of the TS along the path and different values of the barriers (usually much lower at the PT2 level, and they can even disappear).
On the base of the CASPT2 results we can hypothesize a possible explanation of the two fast deactivation paths experimentally observed (i.e. biexponential decay). After the initial relaxation (following vertical excitation) to the **$S_1$-min,** the system can proceed in two directions. The first to the **CI-C** deactivation funnel (i.e. a conical intersection: here in fact the the two roots have already swap and a crossing region is intercepted on the way to **CI-C**) and, through this region, directly to the ground state. This route would correspond to the experimental shortest component of ~3ps. The second path leads to the **CI-B** *minimum* (here in fact the two root have not yet swap, but it is very close to a sloped CI region) where the system remains for few picoseconds and may partially equilibrate before hitting the degeneracy region and return onto the ground state (see Figure 2.10).

*Table **2.3**: CASPT2 energies of the optimized (CIS) excited state structures and ground state optimized (MP2) photoproducts.*

| **Geometry** | *CASPT2 $S_0$ energy* | *CASPT2 $S_1$ energy* | *$\Delta E$ of $S_0$ respect to $S_0$-min kcal/mol* | *$\Delta E$ of $S_1$ respect to $S_0$-min kcal/mol* | *$\Delta E$ $S_0/S_1$ kcal/mol (nm)* |
|---|---|---|---|---|---|
| **FC** | -1294,722054 | -1294,646026 | 0,00 | 47,71 | 47,71 (599) |
| **Min-$S_1$** | -1294,714118 | -1294,653512 | 4,98 | 43,01 | 38,03 (708) |
| **CI-A** | -1294,638999 | -1294,629346 | 52,12 | 58,17 | 6,06 |
| **CI-B** | -1294,664311 | -1294,656962 | 36,23 | 40,85 | 4,16 |
| **CI-C** | -1294,661319 | -1294,669608 | 38,11 | 32,91 | -5,20 |
| **CI-D** | -1294,669240 | -1294,635192 | 33,14 | 54,51 | 21,37 |
| **Prod-syn** | -1294,728894 | -1294,673365 | -4,29 | 30,55 | 34,84 (821) |
| **Prod-anti** | -1294,726250 | -1294,661872 | -2,63 | 37,76 | 40,40 (708) |

### *2.2.3. Photoproducts*

We have limited the investigation of the possible photoproducts to the two biological important isomers: ZZEasa (i.e. isomerization about the C15-C16 bond only) and ZZEass (i.e. isomerization about both the 14-15 and 15-16 bonds). Calculations show that the both products are more stable than the reactant (see Table 2.3) The

absorption *maxima* of the ZZEasa and ZZEass isomers are red shifted respect to the reagent, which is the same trend experimentally observed in protein.

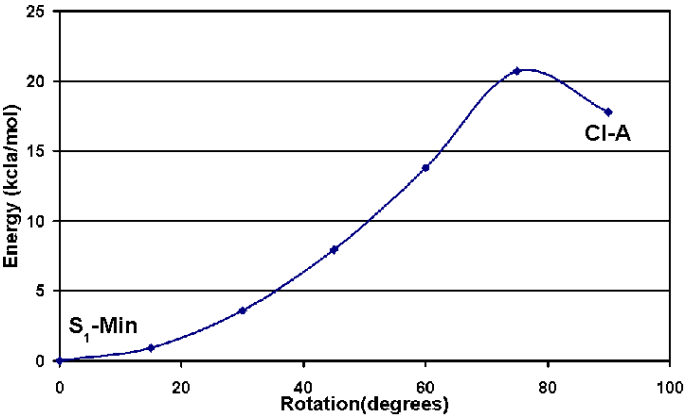



*Figure **2.6**: Optimized scan between **$S_1$-Min** and **CI-A**.*

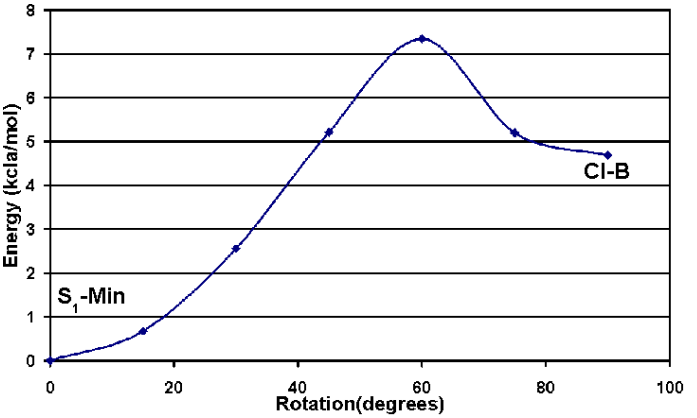



*Figure **2.7**: Optimized scan between **$S_1$-Min** and **CI-B**.*

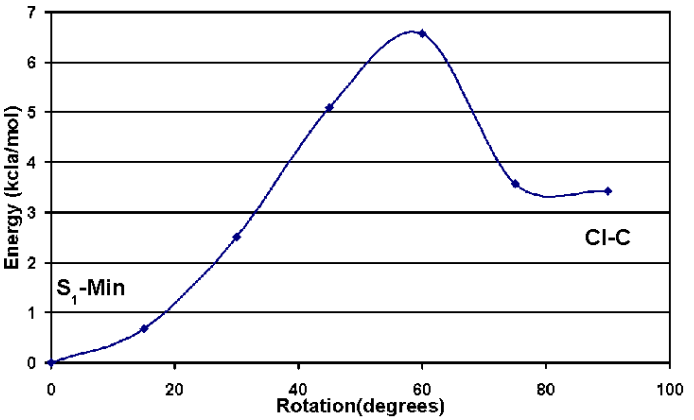



*Figure **2.8**: Optimized scan between **$S_1$-Min** and **CI-C**.*

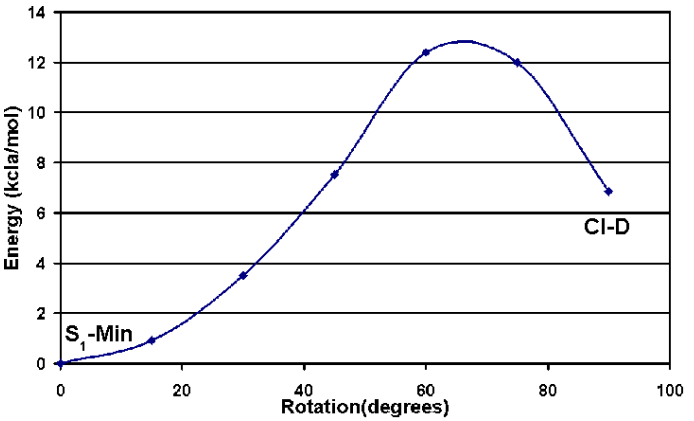



*Figure **2.9**: Optimized scan between **$S_1$-Min** and **CI-D**.*

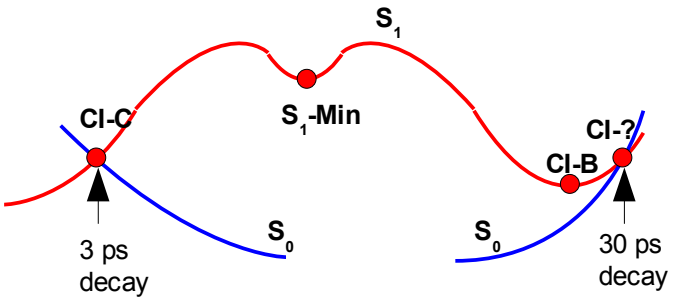



*Figure **2.10**: Hypothesis of possible deactivation paths of the PΦB model in solvent. From the **$S_1$-min** the system can move directly to CI (**CI-C**) or to a minimum (**CI-B**) close to an hypotized CI( **CI-?**). (The TS in figure are only qualitative).*

## 2.3. QM/MM setup

Only one crystallographic structure[v] (with a resolution of 2.5 Å) has been documented up to date for the active subunit of a phytochrome. Unfortunately, in this protein the chromophore is not the phytochromobilin, but the biliverdin (see Figure 2.1). This mean that, strictly speaking, we cannot compare directly the above results computed in vacuo (section 2.2) to the ones computed for this protein because the two chromophores are different (the choice of a different chromophore for the computations in vacuo is simply due to the crystal structure has become available only after our initial study on the isolated system). Nevertheless, the two molecules are very similar (both structurally and electronically) and their general photochemical behavior is predicted to be pretty much the same. Therefore, we think we can freely compare the results computed on these two systems. Obviously, in order to study a realistic system, we have used the biliverdin chromophore for the calculations in the protein.

To simulate the absorption and emission spectra of the available phytochrome structure we have employed our QM/MM code. Using the standard Amber package tool the protein has been solvated with a water cup centered on the biliverdin and with a 20 Å radius. The hydrogens atoms has been added using the H++[vi,95] program. The protonation state of the residues close to the chromophore (HIS260, ASP207, HIS290) and the orientation of HOH12 and HOH18 has been corrected in order to preserve the hydrogen bond network (see Figure 2.12).

In order to minimize spurious steric interactions between nearby protein residues and the biliverdin chromophore, 10 Å of residues around the biliverdin have been optimized at MM (amber ff99)[59] level. We have divided the resulting system in three parts: the QM layer is the biliverdin except the two isopropilic groups, the medium layer treated at MM level is represented by the two isopropilic groups the water molecule HOH12 and HOH18 and the side chain of the LYS24, which is covalently linked to the QM part. The third and last part is obviously the rest of the system which is kept frozen during all calculations.

The experience collected in our previous study in vacuo drove us to use the MP2 method for ground state optimizations, while on the excited state both a CASSCF (with an 8 electrons in 8 orbitals active space) and a CIS approach have been employed. In all cases a 6-31G* basis set was used. The active space include 6 $\pi$ orbitals of the rings B and C, and the two $\pi$ orbitals of the exocyclic double bond between C and D.

v The pdb code is 1ZTU and can be downloaded free of charge at the protein data bank website: http://www.rcsb.org/pdb/home/home.do.

vi H++ is an automated system that computes pK values of ionizable groups in macromolecules and adds missing hydrogen atoms according to the specified pH of the environment. http://biophysics.cs.vt.edu/H++/index.php

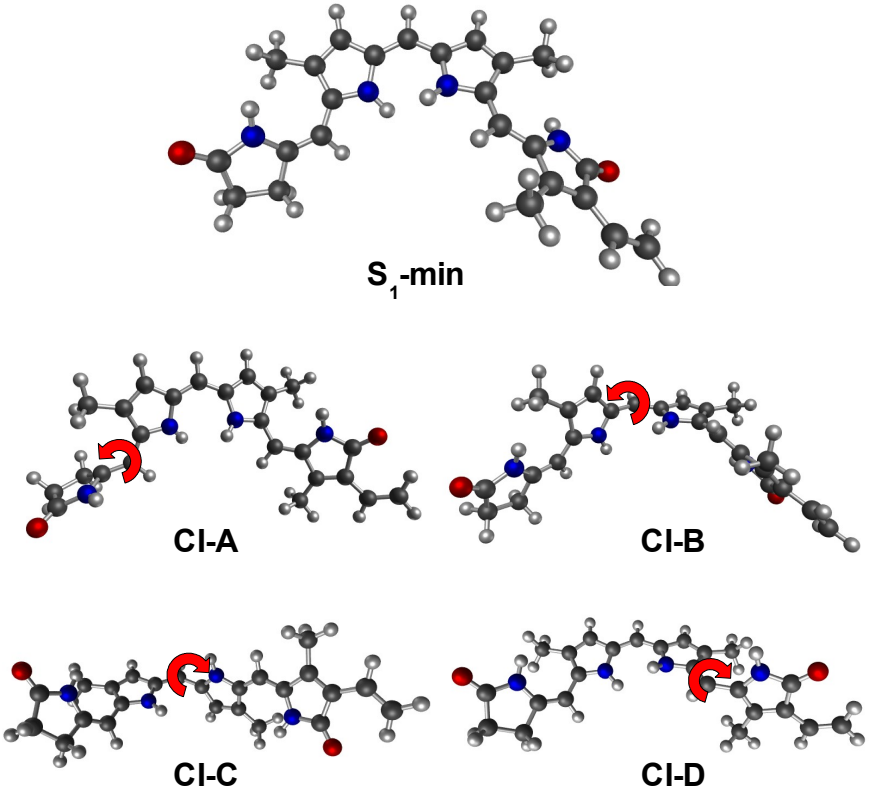



*Figure **2.11:** Ball & stick representation of the excited state minimum and of the four conical intersections.*

## 2.4. Results and discussion

### *2.4.1. Ground state and vertical excitation*

The optimization of the ground state leads to a geometry which is similar to the PΦB model geometry previously optimized in vacuo, where the resonance formula is predominant.

Analyzing the structure of the binding pocket (the complex hydrogen bond network is totally preserved during the optimization) is clear that the BV is constrained to be almost planar except for the ring D which is partially rotated. This rotation can represent a structural key factor in assisting and triggering the observed phoisomerization (see Figure 2.12).

Remarkably, the calculated absorption maximum (709 nm) is in surprise agreement with the experimental value (700 nm)[96], so we consider this structure as a good starting point for excited state investigations.

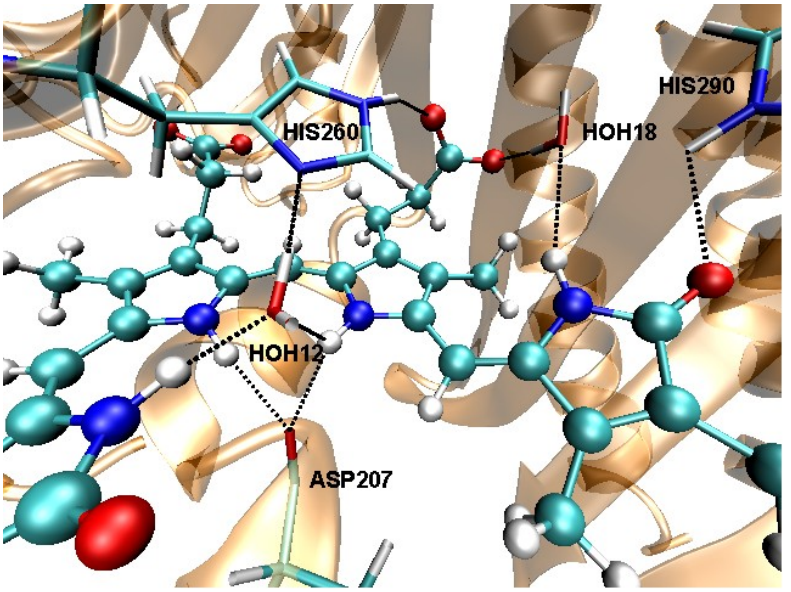



*Figure **2.12**: hydrogen bond network of the biliverdin (ball &stick) binding pocket( tube)*

### *2.4.2. Excited state relaxation end emission*

Starting from the Franck-Condon region, very limited out-of-plane structural geometrical modifications take place when the system relax into an excited *minimum* which remains substantially planar. In fact, we have observed that the forces in the FC region are mainly located on skeletal stretching coordinates. The scarce structural modification can be assigned to the constrained provided by protein structure which surrounds the chromophore and lock the BV, but also to the presence of a real planar *minimum* on S1, as already revealed for the isolated model. We have calculated the emission from the *minimum* structure (742nm).

Differently than seen in isolated conditions, it is apparent that the rigidity of the protein cavity prevents rotation about the extra-annular bonds between the A-B and B-C rings, thus locking the two most favored deactivation routes seen in vacuo. This suggests that the deactivation path involved in the protein must be addressed to a different channel: this is the rotation about the C15-C16 bond, as it is in fact observed experimentally, since this is the unique twisting motion (of the four ones found in the isolated system) which my be accommodated into the protein banding pocket. On the other hand, it has been shown in the previous section that this path is involving an energy barrier in isolated conditions, i.e. it is a thermally activated

process which cannot account for the shortest (picoseconds) component of the decay observed in the protein. This also means that the protein environment must accelerate this path, by catalyzing and triggering the photoisomerization about the C15-C16 bond, thus reducing (if not completely removing) the barrier observed in vacuo along this path. This catalytic effect of the binding pocket, which would account for the ultrafast photoisomerization and internal conversion observed, may be due to both steric and electrostatic effects. Our study, which is still in progress, is aimed to unveil this catalytic effects, thus providing a rationale for the BV bio-activity observed in the protein.

## 2.5. Conclusion

Our study is the first full comprehensive investigation of the intrinsic (i.e. isolated conditions) excited state behavior of a realistic model of the phytochromobilin PΦB chromophore. We demonstrated that the computation of CASPT2 energies in combination with MP2 and CIS methods for the investigation of ground and excited state geometries, respectively, is a good compromise between accuracy and computational cost in the exploration of the photochemically relevant PES driving the photo-induced motion of the system and its decay. Remarkably, all the computational results are in good agreement with the available experimental data. Furthermore, they do provide a rationale for the photochemical behavior observed. In particular, several important features emerge from our study:

a) The ground state structure of the PΦB model is an almost perfect resonance hybrid of two limiting structures, where the charge is perfectly delocalized on the two central (B and C) rings.

b) The first excited state (i.e. the bright state) is a single homo-lumo excitation involving orbitals located on the central B and C rings. It results that this state does not involve any intramolecular charge transfer.

c) We have found four possible deactivation routes which may all trigger the deactivation of the system to the ground state by a twisted conical intersection. In fact, each conical intersection structure involves a 90 degrees rotation of one of the two external exocyclic double bonds (**CI-A**, **CI-D**) or of one of the two central bonds of intermediate single/double bond character (**CI-B**, **CI-C**).

d) The four rotations are not equivalent and we have shown that **CI-B**, **CI-C** are far exceedingly the most favorite ones, involving the lowest energy points: the ultrafast biexponential (radiationless) decay observed has been assigned to these two energetically favorite routes.

e) On the other hand, investigation in protein shows that the residues surrounding the chromophore constrain the A, B and C rings on the same plane. This fact prevents a deactivation through the same funnels seen for the system in vacuo. Therefore, we suggest that the protein has a catalytic effect promoting the C15-C16 photoisomerization which is otherwise unfavored in isolated conditions (and, we guess, in solvent).

# 3. Chromophores in solvents

The optical properties of an organic chromophore are among the ones that can be highly influenced by the environment (such as the solvent or the protein). This effect is particularly pronounced for charged molecules where, usually, the bright excited state has (at least partially) a charge transfer character. Thus, for example, while the solvent may play a role in stabilizing (by solvation) the ground state of the charged chromophore, it may play in the opposite direction for the excited state. This behavior leads generally to a blue shift of the absorption maximum going from *vacuo* to solvent. For the protein a similar effect may be envisaged.

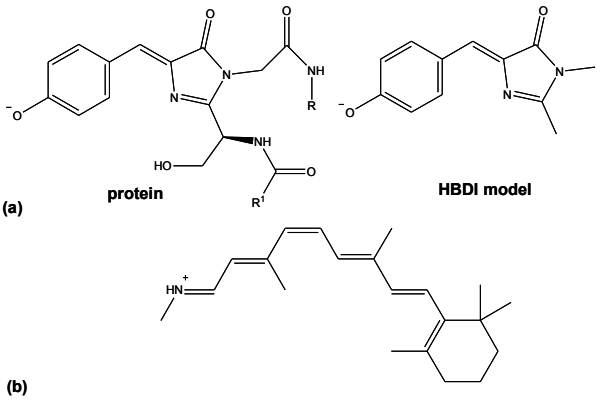



*Figure **3.1**: Chemical structure of the a) GFP chromophore and its isolated model (HBDI) and b) the protonated Schiff base of 11-cis retinal.*

Two biologically important chromophores belong to this class of compounds: the anionic 4'-hydroxybenzyldiene-2,3-dimethyl-imidazolinone (HBDI, see Figure 3.1a) and the protonated Schiff base of retinal (RPSB see Figure 3.1b), which are the chromophores of the green fluorescent protein (GFP) and rhodopsin proteins, respectively. This section focuses on the description of the optical properties of these systems in different environments by means of QM/MM computations. An

advantage is that both experimental and computed spectra are available in *vacuo*, solvents, as well as protein[97,98,99,100,101]. Thus, it is possible to compare (and validate) our QM/MM results against the ones already appeared in the literature

## 3.1. Structure and optical properties for solvated HBDI and GFP.

Thanks to its highly fluorescent chromophore and the fact that this is directly formed from a tri-peptide unit (Ser65-Thr66-Gly67)[99] during the protein folding without cofactors, GFP is widely used in biochemistry as a fluorescent marker[99].The anion is considered to be the active form of the chromophore and, while in protein it strongly fluoresces, in solution the fluorescence is quenched. Therefore, the study of this chromophore in different environments is crucial to understand its photophysical/photochemical properties.

### *3.1.1. Solvation by HL and HML QM/MM computations*

HBDI (Figure 3.1a) has been chosen as a model for the GFP chromophore because it preserves all the conjugate system as well as the structure of the original protein-embedded chromophore and it has been widely investigated in water solution. QM/MM computations have been used to simulate solvation with the chromophore (anionic HBDI) embedded in a 16 Å TIP3P water box, using an MM $Na^+$ as the counter ion (as it is experimentally)[97]. The CASSCF(6-31G*) level has been extensively used for the chromophore during QM/MM optimizations and then the energy has been refined to account for correlation effects using single point CASPT2 computations. This CASPT2//CASSCF approach, which defines the *high* (QM) layer, has been widely used for the photochemistry of organic chromophores (proving to give reliable results of experimental accuracy)[102] and represents the state of the art to calculate excited state properties of organic systems. In order to obtain a good compromise between speed and accuracy, the full active π-space (16 electrons in 14 orbitals) has been reduced (according to our previous paper[103]) to a 12 electrons in 11 orbitals, where the orbital localized on the amidic group has been removed together with the highest and lowest energy π-orbitals of the benzene ring.

To generate a good starting point we have performed a classical molecular dynamics with periodic boundary conditions. While the chromophore was kept frozen at its original QM geometry *in vacuo* with point charges obtained using the AM1-BCC procedure[104,105], the solvent and counter ion were kept free to move, so that during the 1 ns MD the counter ion is able to diffuse in the water bulk. The lowest potential energy geometry was selected along the trajectory as a representative point of the sampling and as starting guess for QM/MM computations.

Three different QM/MM optimization types were performed on the system according to different and increasing levels of mobility (i.e. the parts comprising the atoms that are free to move and are optimized can be changed according to the partitioning scheme used): (i) HL optimizations, where only the *high* (QM) layer (i.e. the chromophore) is optimized, while all the other (MM) waters (i.e. the *low* layer) are kept frozen; (ii) HML optimizations, where the *high* (QM) layer (i.e. the chromophore) and a *medium* (MM) layer (i.e. the water molecules surrounding the chromophore up to a 3.5 Å radius) are optimized together, with all the other (MM)

waters (i.e. the *low* layer) kept frozen; (iii) HML optimizations, where the *high* (QM) layer (i.e. the chromophore) and a *medium* (MM) layer (i.e. the water molecules surrounding the chromophore up to a 3.5 Å radius) are optimized together, while all the other (MM) waters (i.e. the *low* layer) are free to move (independently with respect to the first two ones) up to a 10 Å radius around the chromophore (and the others are kept frozen to their initial positions). These three different computational levels (which are identified in Table 3.1 as HL, HML1 and HML2, respectively) highlight the high flexibility of the optimization procedure implemented in the current QM/MM code and, very remarkably, are shown to produce results of increasing accuracy (see Table 3.1).

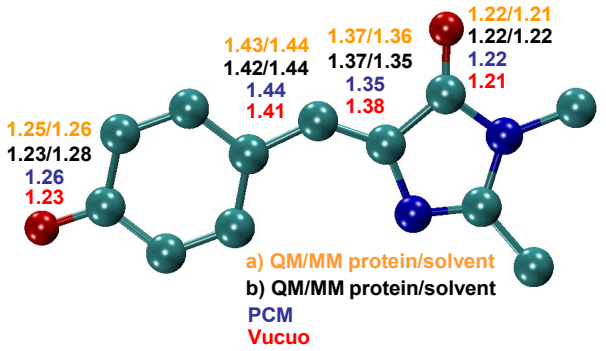



*Figure **3.2**: Bond distances values calculated for the GFP chromophore in vacuo, solvent (PCM and QM/MM) and protein: a) QM/MM values from ref. [51] and b) our QM/MM results.*

*Table **3.1**: Absorption maxima for the anionic GFP chromophore.*

| Structure | Absorption maximum (nm) | |
|---|---|---|
| | QM or *QM/MM* | Experimental |
| vacuo | 465[a], 493[b], | 479[d] |
| Solvent (bound ionic pair) | *434*[c] | 425[e]-432[f] |
| Solvent(loose ionic pair) | 460 (HM), 499 (HML1), 441 (HML2) | |
| Protein | *504*, *468*[c] | 495[g] |

*a) ref. [106]; b) ref. [103]; c) ref. [102]; d) ref. [98]; e) ref. [107];f) ref. [97] g) ref.[108]*

It is worth noting that the counter ion ($Na^+$) is 7 Å away from the chromophore in all the QM/MM optimized structures, revealing a weakly bound ionic pair. Although this is only one of the many possible stable solute/solvent configurations, it is remarkable that the computed absorption energy (499 nm, 460 and 441 nm for (i), (ii) and (iii), respectively) is progressively getting very close to the experimental value

recorded for the chromophore in water (425 nm) and does perfectly account for the blue shift observed on going from *vacuo* (where, again, the computed 465 nm absorption maximum [102,103] does nicely reproduce the experimental value of 479 nm [97]) to water (see Table 3.1). Figure 3.2 reports the relevant bond distances for HBDI *in vacuo*, explicit solvent and PCM[102,103,106]. It is apparent that the solvent has also a key role in tuning the chromophore molecular structure, since it shifts the geometry to the individual resonance structure carrying the negative charge on the phenolic oxygen, while a resonance hybrid is more likely to represent the chromophore *in vacuo*. Furthermore, it results that implicit (PCM) as well as explicit (QM/MM) techniques do yield similar results[102,103].

### *3.1.2. GFP by HML QM/MM computations*

Finally, QM/MM computations have been used to investigate the geometry and absorption maximum of the chromophore in the protein (GFP). The crystallographic structure (code 1GLF)[109] available in the protein data bank archive has been selected. After protonation, using the **H++**[95] procedure, hydrogen atoms have been relaxed in order to get a reliable starting structure. A three-layer (HML) partitioning scheme has been adopted: the *high* (QM) layer (i.e. the electrostatic embedded chromophore, see Figure 3.3) and the *medium* layer (see tube representation in Figure 3.3) constitute the mobile part (following the receipt of Sinicropi et al.[102]), while all the remaining (MM) protein is kept frozen and represents the *low* layer. The same QM level as for HBDI in solution has been adopted for the chromophore, while protein and water molecules are computed with the Amber force field. Very remarkably, the computed 504 nm absorption maximum for the I anionic state of wild type GFP does nicely reproduce the experimental value of 495 nm[108] (see Table 3.1). The same applies also for the chromophore structure (see Figure 3.2).

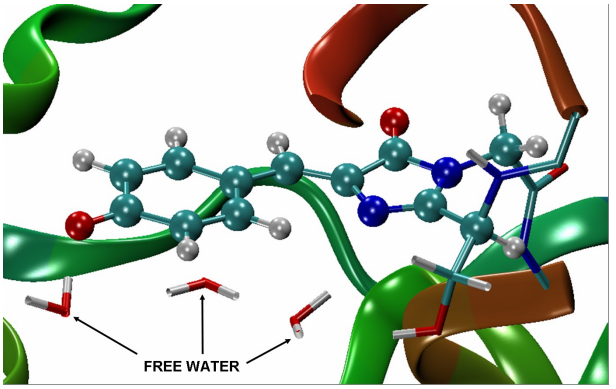



*Figure **3.3**: Model system used fot the QM/MM study of the anionic GFP. Ball&stick is the QM part, the tube is the movable medium layer, while the rest is kept frozen.*

### *3.1.3. cis retinal PSB in solution*

A similar QM/MM investigation as before has been performed for solvated RPSB. Specifically, in the present study our effort is focused to simulate the effect of methanol in the spectral tuning of the 11-cis-retinal chromophore. Thus, for this purpose, QM/MM computations have been used with the chromophore (RPSB) embedded in a cubic 45 Å methanol box, using an MM $Cl^-$ as the counter ion (as it is experimentally)[101]. Similarly to the examples above and previous computational works[110,111], a CASPT2//CASSCF(6-31G*) approach (with electrostatic embedding) has been adopted for the chromophore (using the full π-space of 12 π-electrons in 12 π-orbitals), which defines the *high* (QM) layer, during QM/MM optimizations.
To generate a good starting point for QM/MM optimizations we have performed a classical molecular dynamics with periodic boundary conditions. While the chromophore was kept frozen at its original QM geometry *in vacuo* [112] with point charges obtained using AM1-BCC procedure[104,105], the solvent and counter ion were kept free to move, so that the counter ion is able to diffuse in the methanol solvent. After a MD run of 1 ns, we see (at least for this specific trajectory) that most of the time the chloride counter ion resides near the retinal nitrogen (i.e. close to the cationic head, in a tight ionic pair arrangement). Thus, a low energy snapshot with the counter ion close to nitrogen has been selected as a representative point of the sampling and as starting guess for QM/MM computations.
A two layers (HL) QM/MM optimization has been performed: the *high* (QM) layer is given by the chromophore (as stated above), while all solvent molecules and the counter ion constitute the *low* (MM) layer (thus excluding a *medium* buffer), which has been left free to move independently up to a distance of 10 Å from the chromophore (while the rest is kept frozen at its initial geometry).

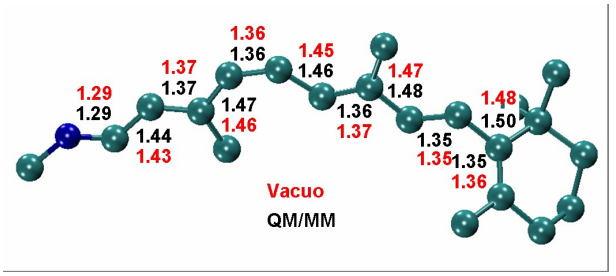



*Figure **3.4**: Bond distances computed for RPSB in vacuo (red) and in protein (black).*

Remarkably, the resulting absorption maximum (453 nm) is in good agreement with the experimental value of 440 nm[101] and with previous QM/MM computations[113] (see Table 3.2), but in this case the optimized geometry does not undergo significant variations on going from *vacuo* to solution (see Figure 3.4): while in HBDI geometrical modifications come from a different weight of the two resonance structures, in the retinal chromophore the structure does not change a lot, because the geometry is described only by one resonance structure(i.e. the one with the positive charge on nitrogen). This means that the change of the absorption spectrum on going

from *vacuo* to solvent (see Table 3.2) is addressed only to the perturbation of the wave function by the electrostatic effects of the external point charges and that electrostatic embedding is enough to get a correct description of the chemico-physical properties of the chromophore in solution.

*Table **3.2**: Absorption maxima for RPSB.*

| Structure | Absorption maximum (nm) | |
|---|---|---|
| | QM or *QM/MM* | Experimental |
| Vacuo | 545[a] | 610[c] |
| Solvent (bound ionic pair) | *453*, *429*[b] | 440[d] |
| Solvent(loose ionic pair) | 442 | |

*a) ref. [112]; b) ref. [113]; c) ref. [100]; d) ref.[101]*

To explore the effects of the counter ion ($Cl^-$) position on the absorption energy of the chromophore, we have also investigated the optical properties of a loose ionic pair: thus, a low energy snapshot with this arrangement has been selected along the previous MD run as a starting point for new QM/MM optimizations (which have been performed at the same HL level as before). Very interestingly, the absorption maximum computed for this QM/MM optimized loose ionic pair is substantially unchanged (442 nm) (see Table 3.2) and in agreement with the experimental value.

## 3.2. Solvated ion pairs: structure and properties

Interestingly, our results for the optimized geometry and the absorption maximum of solvated HBDI (section 3.1) match perfectly (see Figure 3.2 and Table 3.1) with the ones reported in previous QM/MM computations[102] where an analogous CASPT2//CASSCF approach was employed (see Table 3.1). Anyway, it is worth noting that there the counter ion was forced to stay at the vicinity of HBDI (i.e. a tight bound ionic pair arrangement was selected), instead of being far away as it is in our case. The same results appear for solvated RPSB, where the close and the loose ion pairs do lead to very similar optical and structural properties (seen section 3.2 above). This leads to the remarkable conclusion that the solvent in the loose ionic pair interacts with the solute as if it were the real counter ion, i.e. solvation shells form a *virtual counter ion* that has the same (electrostatic) effect as the real one in a tight bound ionic pair. This effect is due to the reorientation (i.e. polarization) of the polar solvent close to the chromophore (see Figure 3.5): the polarized permanent dipoles of the solvent act similarly to the counter ion, stabilizing the ground state respect to the charge transfer excited state.
This discussion implies that tight bound ionic pairs may be seen as qualitatively good models for solvated charged chromophores in general, as anticipated in our previous works [110,114]. This study represents a further validation of the aforementioned statement and is in line with previous suggestions by Sakuray and coworkers[115,116].

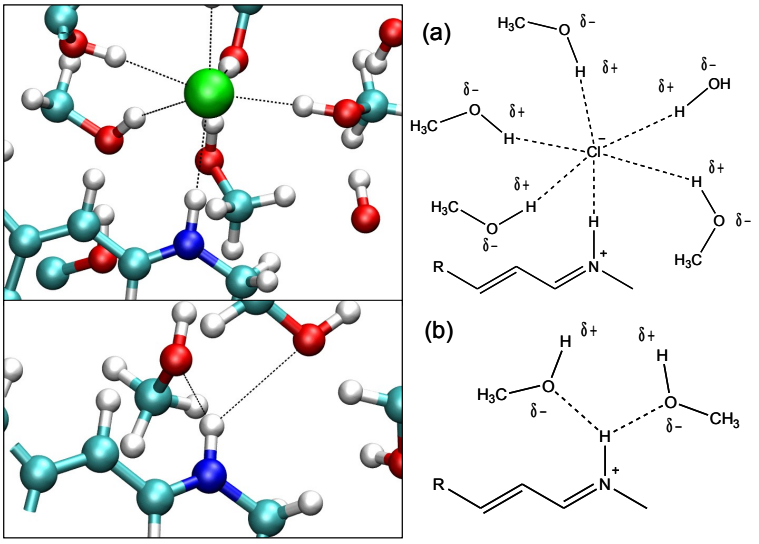



*Figure **3.5**: Schematic representation of interation between solvent, PSB and counter ion: (a) tight bound ionic pair, (b) dissociated couple.*

# 4. Bacterio-rhodopsin

The outwardly directed light-activated proton pump Bacteriorhodopsin (bR) is a small G-protein-coupled receptor contained in the purple membrane of *Halobacterium Salinarium*. Upon photoexcitation, the all-*trans* retinal chromophore (PSBT) bounded to Lys216 via a protonated Schiff Base linkage, converts in its 13-*cis* isomer (PSB13). The photoisomerization triggers a series of conformational changes that result in a proton translocation from the cytoplasmic to the extracellular domain[117].

The effect of the protein cavity on the retinal reactivity and its spectroscopic properties is an intriguing topic, which has been investigated by numerous theoretical works[118,119,120,121,122,123] with various QM/MM approaches. In fact, with respect to solution, the protein catalyzes the reaction by inducing stereoselectivity and by increasing both the reaction rate and the efficiency. While the photoisomerization of PSBT in solution leads to a mixture of photoproducts[124] (the quantum yields for 9-*cis*, 11-*cis* and 13-*cis* being 0.02, 0.14 and 0.01, respectively) in about 10 ps, only the 13-*cis* isomer is formed in bR with a higher quantum yield (~0.60[125]). The first "hot" photoproduct $J_{625}$ is formed within only 500 fs[126,127] and relaxes to the $K_{590}$ intermediate with a time constant of 3 ps[127]. The red-shift of the absorption maximum in going from solution (~450 nm, both in methanol[101] and hexane[128]) to the protein (568 nm) is another yet not fully understood effect. While explaining the details of the fine-tuning performed by the protein environment would represent a biologically relevant achievement *per se*, the interest has been dramatically boosted by the recent applications of this protein in nano-technologies[129,130].

In this work, the early three states $bR_{568}$, $I_{460}$ and $K_{590}$ along the bR photocycle are investigated through high level QM/MM computations. Recent studies on retinal models *in vacuo*[131,132] show that the CASPT2//CASSCF/6-31G* strategy is a reliable approach that yields good values for geometric parameters, energy differences, spectroscopic properties, change in dipole moment and, more generally, it gives an accurate description of the Potential Energy Surface (PES) that allows a rationalization of experimental data. Here the same strategy is coupled with the AMBER molecular mechanics force field, allowing for CASPT2//CASSCF/AMBER geometry optimization and excited state property evaluation *in protein*. We will show

that this method is capable to reproduce the *selective* isomerization of the $C_{13}$=$C_{14}$ double bond, the transient absorption and emission *maxima* and the energy storage in the $K_{590}$ state.
Briefly, the method is based on a hydrogen link-atom scheme with the frontier atom placed at the C-$C_\varepsilon$ bond of the Lys216 side chain (see Figure 4.1). The retinal, Schiff base Nitrogen and attached ε-carbon atom constitute the QM part of the system, described at the CASSCF/6-31G* level (with an active space of 12 π-electrons in 12 π-orbitals). The MM and QM segments interact in the following way: (i) the QM electrons and the full set of MM point charges interact via the one-electron operator, (ii) stretching, bending and torsional potentials involving at least one MM atom are described by the MM potential (iii) QM and MM atom pairs separated by more then two bonds interact via either standard or re-parametrized van der Waals potentials. Single State CASSCF/6-31G*/AMBER geometry optimization is carried out with the GAUSSIAN98[133] and TINKER[60] programs.

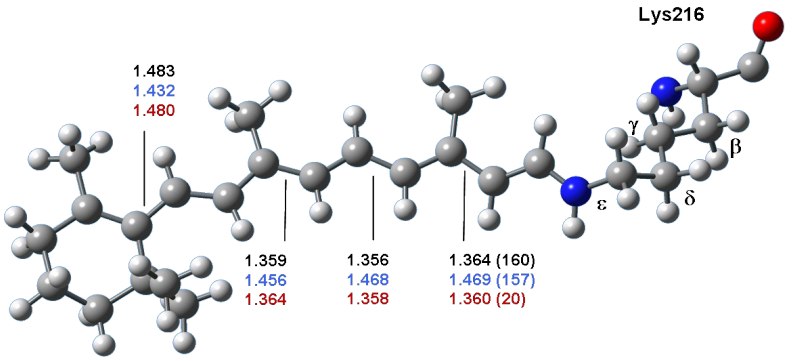



*Figure **4.1**: CASSCF/6-31G*/AMBER optimized structures for $bR_{568}$ (black), $I_{460}$ (blue) and $K_{610}$ (red). The dihedral angle $C_{15}$-$C_{14}$=$C_{13}$-$C_{12}$ is reported in parenthesis. The sidechain of Lys296 from $C_\beta$ to $C_\delta$ is free to relax and modeled with the AMBER force field, while the retinal, Nitrogen atom and $C_\varepsilon$ are treated at the QM level.*

## 4.1. QM/MM Setup and Computation Details.

The protein framework used in the bR computations is derived from the PDB file 1C3W[134], with the addition of the Thr157-Glu161 missing loop taken from the 1QM816[135] file[vii]. According to experimental[136] and previous theoretical works[120,121,122], the protonation state of Asp96, Asp115 and Glu204 is set neutral, while for all the other residues the standard protonation state is used. The resulting (+3) overall

vii The "All-Atoms Fit" option of the Swiss-PdbViewer tool (Guex, N.; Peitsch, M. C. Electrophoresis 1997, 18, 2714, http://www.expasy.org/spdbv/) has been used to superimpose the missing loop to the 156,162 and 163 residues, common to both sequences. The two "frontier" residues Thr-157 and Glu-161 and the side chain of Lys-159 are then MM optimized with the Tinker3.9 package using the Amber force field to better fit the protein surface.

charge is neutralized placing three chloride ions on the protein surface (one near Arg171, another in proximity of the Arg221, Arg223 and Lys26 and the other near Lys125). The position of the three water molecules W401, W402 and W406 is optimized at the HF/6-31G* level and kept frozen during all the following geometry optimizations (the Lys296 side chain and the chromophore are free to relax). Two sets of State-averaged CASPT2 computations (equal weights have been used for $S_0$, $S_1$, $S_2$, $S_3$ and $S_4$) have been performed on the CASSCF/Amber optimized structures (Single State calculations have been used for geometry optimizations) using the MOLCAS5[55] package. In the first set, the protein electrostatic environment provided by the RESP-charges distribution contributes to the description of the CASPT2 Hamiltonians via the one-electron operator. In the second set, the electrostatic bath is not considered (i.e. the QM part of the system is taken alone without the RESP charges): this accounts for the *in vacuo* description of the system optimized in protein. Oscillator Strengths are computed within the RASSI approach[137].

## 4.2. bR Induced Stereoselectivity and Spectral Shift.

The 160° twisting of the $C_{13}$=$C_{14}$ double bond in the $bR_{568}$ optimized structure reported in Figure 4.1 shows that the protein cavity tilts this bond from planarity of about 20°, which prompts the isomerization of that *specific* double bond after photoexcitation. This suggests that the stereospecificity observed in bR is due to the protein cavity steric effects leading to a *structural* deformation of the chromophore backbone already in the ground state and not to electrostatic effects taking place on the excited state and affecting the retinal wavefunction.

The 53.6 kcal/mol (see Figure 4.2) required for the vertical transition to the spectroscopic state $S_1$ (note that the 0.411 Oscillator Strength value reported in Figure 4.1 clearly labels this as an allowed transition) compare very well with the 568 nm (50.3 kcal/mol) experimentally determined. It is interesting to note that the transition energy computed *for the same structure placed in vacuo* (that is, the protein electrostatic bath is removed as described above for "set 2") yields an excitation energy of 43.6 kcal/mol (see dotted lines in Figure 4.1). This indicates that the *electrostatic* effect of bR *blue shifts* the absorption maximum of about 7 kcal/mol, which is in agreement with previous computational results[42,119,138]. Therefore, the observed red(opsin)-shift with respect to solution data should mainly be attributed to the protein *structural* effect which forces the β-ionone ring in a planar s-*trans* conformation. This makes PSBT a six conjugated double bonds chromophore, whereas in solution a skewed s-*cis* conformation for the ring is expected[139] that reduces the conjugation of the π-system to only five double bonds. The blue-shift (compared to the vacuo) induced by the electrostatic environment of the protein on the $S_1$ state is due to the complex counterion system (Asp85, Asp212, Arg82) located close to the Schiff Base (see Ref. 27 for a detailed description of the effects of a counterion on retinal models PSBs). Since in the ground state (covalent nature) the positive charge is located close to the counterion, while in the $S_1$ state (ionic nature) it is partially transferred towards the β-ionone part of the π-system (as previously shown[131] for retinal models *in vacuo*), the overall effect of the counterion is to increase the $S_0$-$S_1$ energy gap. Therefore, the bR counterion effect is not counterbalanced by the *overall* protein environment, as instead it appears to occur in

Rhodopsin[140].

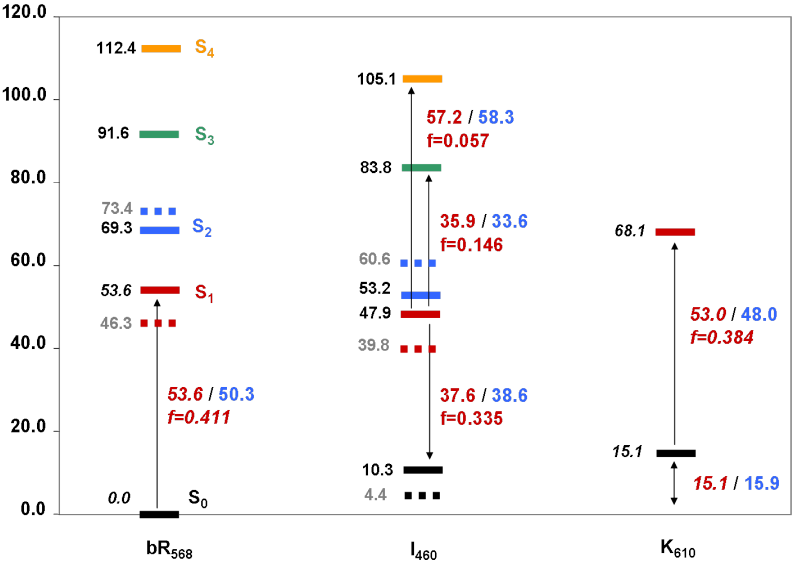



Figure 2

*Figure **4.2**: CASPT2//CASSCF/6-31G*/AMBER energy profile for the $bR_{568}$, $I_{460}$ and $K_{610}$ states relative to the $bR_{568}$ ground state. Energies reported in Italic are determined with Single State computations, while all the other values come from State Averaged computations with equal weights for the first five states. The energies for the system in vacuo (see the text) are reported with dotted lines. The computed values for the indicated transitions are reported in red (together with the computed Oscillator Strength)[137], while experimental data (see references in the text) are shown in blue for comparison.*

## 4.3. Transient Absorption and Fluorescence.

The relaxation process from the Frank Condon region leads to a shallow excited state *minimum* (see blue values in Figure 4.1) in which the backbone is relaxed along the stretching coordinate, but the twisting of the $C_{13}$=$C_{14}$ bond has not yet proceeded. The barrier that leads out of this *minimum* is less than 1 kcal/mol, in striking agreement with low-temperature experiments which account for ca. 1 kcal $mol^{-1}$ barrier[141,142]. This structure can be tentatively identified as the $I_{460}$ state responsible for the fluorescence and transient absorption experimentally observed. In fact, the ~740 nm (38.6 kcal $mol^{-1}$) fluorescence maximum experimentally determined by Shenkl *et al.*[143] is in excellent agreement with the 37.6 kcal $mol^{-1}$ energy gap computed for the $S_1$->$S_0$ transition. Moreover, a transient absorption peaking at 850 nm (i.e. 33.6 kcal $mol^{-1}$) has been described by Hasson *et al.*[141]. This band can be attributed to the $S_1$->$S_3$ transition, which is computed to be 35.9 kcal $mol^{-1}$ and shows a significant Oscillator Strength. The transient near-IR absorption seen by Logunov *et al.*[144] and peaking at 490 nm (i.e. 58.3 kcal $mol^{-1}$) can be attributed here to the $S_1$->$S_4$ transition,

which shows an energy gap of 57.2 kcal $mol^{-1}$. The excellent agreement (1-2 kcal $mol^{-1}$) between the computed and experimental data, both for the emission and transient absorption, call for this structure as the $I_{460}$ intermediate, i.e. a structure which has been relaxed along the stretching coordinates, but that has not yet started its motion along the torsional coordinate.

## 4.4. Decay to the Ground State and $K_{610}$ Energy Storage

By overcoming the small barrier the retinal can reach a $S_1/S_0$ Conical Intersection (CI), which allows for a radiationless decay to the ground state. The $C_{13}$=$C_{14}$ dihedral angle in this structure is twisted by 120 degrees, that is, the decay point is still on the *reactant* side of the reaction coordinate. Interestingly, after the CI, the $S_1$ and $S_0$ states remain parallel and very close each other (less than 10 kcal $mol^{-1}$ separation) up to a 90 degrees twisting. This allows the suggestive hypothesis that an extended seam of $S_1/S_0$ crossings exist, spanning different twistings for the $C_{13}$=$C_{14}$ double bond (i.e. from 120 to 90 degrees), which is responsible for the ultrafast decay to the ground state. Here, only the more twisted crossings would lead to photoproduct (PSB13) formation. This may explain why Birge *at al.*[145] found at 77K a lower Quantum Yield (~0.3) than experiments[125] performed at room temperature (~0.6).
Ground state optimization of the chromophore structure towards the *product* side leads to a *minimum* on $S_0$ (red numbers in Figure 4.1) in which the chromophore has completed its isomerization process (i.e., the $C_{13}$=$C_{14}$ dihedral is now in the *cis* conformation, with a 20 degrees twistings). The energy difference between this structure and the $bR_{568}$ is 15.1 kcal $mol^{-1}$, in striking agreement with the 15.9 kcal $mol^{-1}$ energy storage determined for $K_{590}$ by Birge *et al.*[145] with spectroscopic and photocalorimetric measures. This agreement allows us to tentatively label this structure as the $K_{590}$ intermediate. The computed $S_0$-$S_1$ energy gap for this structure is 53.0 kcal $mol^{-1}$, while the 595 nm experimentally determined[125] correspond to a smaller 48.0 kcal $mol^{-1}$ transition.

## 4.5. Conclusions

In conclusion, using a CASPT2//CASSCF/AMABER approach, we have been able to determine the structures of the $bR_{568}$, $I_{460}$ and $K_{590}$ first photocycle states and to accurately reproduce their spectroscopic properties and energy storage. Furthermore, we have shown that the pure electrostatic effect of the bR cavity is to blue-shift (with respect to vacuo) the absorption maximum, due to the complex counterion system close to the PSB region. The observed red-shift in bR with respect to solution absorption is addressed to the extended conjugation of the chromophore due to the β-ionone ring planarization with respect to its skewed s-*cis* conformation in solvent.

# Conclusions

In this work is presented a new implementation of a QM/MM code that permits the investigation of large system. Moreover it can be considered more than a standard QM/MM code, in fact our philosophy is to create a general tool linking each other different commercial software in order to capitalize the benefits of each program.
Up to the present the implemented features are:

- QM/MM calculation up to three layers, (QM/MM/MM = high, medium, low)
- Electrostatic embedding, (i.e. perturbation of the wave function through the external MM point charges).
- QM/MM Geometry optimization and transition state search, using analytical first derivative, through Gaussian03 optimization tool kit.
- Numerical second derivative for the layers high and medium. We have also developed an approximated algorithm to calculate the numerical Hessian matrix, which is faster and it introduces a limited error in the results.
- Excited states calculation, where the MM part feels the electrostatic rearrangement of the QM wave function.
- QM and QM/MM classical molecular dynamic.

Thanks to this program I have investigated different interesting biochemical and supermolecular problems.
Comparing the results obtained for the chromphores in vacuo to the calculations using the QM/MM method, where the environment effects are taken in count explicitly, we have noted important differences on theirs photochemical behavior. Firstly we have observed steric effects, like in rotaxane, that modify the topology potential energy surface, up to cancel a deactivation path presents in the fumaramide (90 degrees twisted conical intersection). A second fundamental contribution of the MM part has been reveled in the study of the retinal and GFP chromophores in solution, where the effect of the point charges of the solvent modify the absorption wavelength, causing a general blue shift, attributable to a stabilization of the ground

state and concomitant destabilization of the excited state.
The investigation conducted on the bacteriorhodopsin has highlighted the presence of both the effects above described (i.e. steric and electrostatic). The electrostatic effect generated by the protein cavity leads to a blue shift in the absorption spectrum, not as strong as in polar solvent, while the geometrical constrains facilitate the photoisomerization of the retinal.
The application of the QM/MM allow us to comprehend, with high accuracy level, the photochemistry of complex systems, overcoming the limit of the traditional *ab initio* methods.

# Appendix

## Appendix 1. The construction of the *real* system

A particular care is needed in redistributing charges if the boundary region cuts one or more covalent bonds. Specifically, MM computations need a topology file for the whole system (*real*) that contains the connectivity data and the atomic point charges assigned on the basis of the adopted force field. The sum of all the atomic charges gives the total charge of the *real* system ($\mathbf{Q}^{real}$), which is necessarily a net charge and must be a constant value all through QM/MM calculations. Thus, before starting calculations, atomic point charges for the whole system atoms have to be assigned and their sum must be a net constant value; but while the charges for the atoms in the MM region will remain constant throughout the course of all QM/MM calculations (at least this is our case, since we don't use a polarizable force field), charges for the QM atoms are generated at each step from the wave function and may change accordingly, which may lead to a variable fractional total charge for the *model* system (note that when the *real* system is split in *model* and *pod* (see Figure 4.4), the corresponding total charges, $\mathbf{Q}^{model'}$ and $\mathbf{Q}^{pod'}$, are not necessarily net constant values although it is their sum $\mathbf{Q}^{real} = \mathbf{Q}^{model'} + \mathbf{Q}^{pod'}$). QM calculations are performed on *model-H*, whose net charge $\mathbf{Q}^{model\text{-}H}$ and multiplicity are imposed by the user. To prepare a suitable topology file and grant invariability of the system total net charge, in the current implementation we force the total charge of *model* ($\mathbf{Q}^{model}$) to be equal to that of *model-H* by redistributing the charge difference $\mathbf{\Delta Q} = \mathbf{Q}^{model\text{-}H} - \mathbf{Q}^{model'}$ over the QM atoms (we do that proportionally to the magnitude of each QM atom charge, but alternative strategies are also possible). Thus:

$$\mathbf{Q}^{model'} + \mathbf{\Delta Q} = \mathbf{Q}^{model\text{-}H} = \mathbf{Q}^{model}$$

Obviously, to conserve the total charge of the system ($\mathbf{Q}^{real}$), also the total charge of *pod* ($\mathbf{C}^{pod}$) has to be modified by redistributing an opposite charge value ($-\mathbf{\Delta Q}$) over

the MM atoms:

$$\mathbf{Q}^{pod} = \mathbf{Q}^{real} - \mathbf{Q}^{model} = \mathbf{Q}^{real} - (\mathbf{Q}^{model'} + \Delta\mathbf{Q}) = (\mathbf{Q}^{real} - \mathbf{Q}^{model'}) - \Delta\mathbf{Q} = \mathbf{Q}^{pod'} - \Delta\mathbf{Q}$$
$$- \Delta\mathbf{Q} = \mathbf{Q}^{pod} - \mathbf{Q}^{pod'}$$

For *pod* (i.e. the MM atoms) this is done only once, i.e. for the starting geometry during the initial set up stage and, as stated above, these charges do not change any more. This charge redistribution process may be accomplished according to different strategies since the choice is not unique (mostly depending by the system under study). Here, we can draw some general guidelines on the basis of our experience. When studying an enzymatic system the procedure is quite complex, because all the residues have a net charge. In this case, $-\Delta\mathbf{Q}$ is equally redistributed only on the MM atoms of *pod* belonging to the residues crossed by the boundary line (thus leaving a net charge on ach one) and these new atomic point charges of *pod* will not change anymore during QM/MM calculations (i.e. we don't use a polarizable force field); then, *emb* is generated from *pod* at each step (see Figure 4.4), to be included in the QM calculation according to the Electrostaic Embedding scheme[2].
Another crucial point is how to handle charges in the QM region. In many cases the charges of this region are not correctly parameterized in the adopted force field, or they are completely unavailable. In these cases it is necessary to get a guess of the charges to perform the first calculation step and different approaches are possible for accomplishing this task; for example, we found that computing the wave function of *model-H* at the AM1-BCC[105,104] (or a low *ab initio*) level gives a good starting point. Anyway, during the QM/MM calculation, the wave function of *model-H* is computed at the higher level (including the effect of the Electrostatic Embedding) and its charges reevaluated. Thus, according to the flowchart of our code (Scheme 1), at the end of each cycle the complete set of charges of *real* is reassembled using for *model* the charges computed at the high QM level. It is worth repeating that, QM calculations are performed on *model-H* and the obtained charge includes H link atoms. According to the procedure described above, at every cycle we redistribute the charges of the H link atoms ($\Delta\mathbf{Q}$) on the *model* (QM) atoms. Thus, by adopting all these precautions, the total charge of the system ($\mathbf{Q}^{real}$) is preserved, as well as the total charge of *model* ($\mathbf{Q}^{model} = \mathbf{Q}^{model\text{-}H}$) and *pod* ($\mathbf{Q}^{pod}$). The atomic point charges of *pod* are preserved all over the QM/MM calculations, while the charges of *model* do change at every step, according to the computed QM electronic distribution.

# References


1. Warshel, A. & Levitt ,M. *Journal of Molecular Biology* ,**1976**,*103*, 227-249.
2. Lin, H. & Truhlar ,D.G. *Theoretical Chemistry Accounts* ,**2007**,*2*, 117.
3. Altoe', P.; Stenta, M.; Bottoni, A.; Garavelli, M. *Theor. Chem Acc.* ,,*accepted*, .
4. Dietrich-Buchecker, C.O. & Sauvage ,J.P. *Chemical Reviews* ,**1987**,*87*, 795-810.
5. Anderson, S.; Anderson, H.L. & Sanders ,J.K.M. *Accounts of Chemical Research* ,**1993**,*26*, 469-475.
6. Fyfe, M. & Stoddart ,J. *Accounts of Chemical Research* ,**1997**,*30*, 393-401.
7. Gatti, F.; Leigh, D.; Nepogodiev, S.; Slawin, A.; Teat, S. & Wong ,J. *Journal of the American Chemical Society* ,**2001**,*123*, 5983-5989.
8. Brancato, G.; Coutrot, F.; Leigh, D.A.; Murphy, A.; Wong, J.K.Y. & Zerbetto ,F. *PNAS* ,**2002**,*99*, 4967-4971.
9. Butler, W.L.; Norris, K.H.; Siegelman, H.W. & Hendricks ,S. *PNAS* ,**1959**,*45*, 1703-1708.
10. Giraud, E.; Fardoux, J.; Fourrier, N.; Hannibal, L.; Genty, B.; Bouyer, P.; Dreyfus, B. & Vermeglio ,A. *Nature* ,**2002**,*417*, 202-205.
11. Giraud, E.; Zappa, S.; Vuillet, L.; Adriano, J.; Hannibal, L.; Fardoux, J.; Berthomieu, C.; Bouyer, P.; Pignol, D. & Vermeglio ,A. *J. Biol. Chem* ,**2005**,*280*, 32389-32397.
12. Davis, S.J.; Vener, A.V. & Vierstra ,R.D. *Science* ,**1999**,*286*, 2517-2520.
13. Jiang, Z.; Swem, L.R.; Rushing, B.G.; Devanathan, S.; Tollin, G. & Bauer ,C.E. *Science* ,**1999**,*285*, 406-409.
14. Blumenstein, A.; Vienken, K.; Tasler, R.; Purschwitz, J.; Veith, D.; Frankenberg-Dinkel, N. & Fischer ,R. *Current Biology* ,**2005**,*15*, 1833-1838.

15. Montgomery, B.L. & Lagarias ,J.C. *Trends in Plant Science* ,**2002**,*7*, 357-366.
16. Chen, M.; Tao, Y.; Lim, J.; Shaw, A. & Chory ,J. *Current Biology* ,**2005**,*15*, 637-642.
17. Sineshchekov, V.A. *Biochimica et Biophysica Acta (BBA) - Bioenergetics* ,**1995**,*1228*, 125-164.
18. Lagarias, J.C. & Rapoport ,H. *Journal of the American Chemical Society* ,**1980**,*102*, 4821-4828.
19. Lamparter, T.; Carrascal, M.; Michael, N.; Martinez, E.; Rottwinkel, G. & Abian ,J. *Biochemistry* ,**2004**,*43*, 3659-3669.
20. Oesterhelt, D. & Stoeckenius ,W. *PNAS* ,**1973**,*70*, 2853-2857.
21. Mathies, R.A.; Lin, S.W.; Ames, J.B. & Pollard ,W.T. *Annual Review of Biophysics and Biophysical Chemistry* ,**1991**,*20*, 491-518.
22. Oesterhelt, D. *Current Opinion in Structural Biology* ,**1998**,*8*, 489-500.
23. Born, M. & Oppenheimer ,R. *Annual Phys.* ,**1927**,*84*, 457-484.
24. Boys, S.F. *Proc. Roy.Soc. (London)* ,**1950**,*200*, 542.
25. Szabo, A. & Ostlund, Neil, S. *Modern quantum chemistry: an introduction to advanced electronic structure theory*, McGraw-Hill, **1989**.
26. Roothaan, C.C.J. *Reviews of Modern Physics* ,**1951**,*23*, 69.
27. Roos, B.O. *European Summer School in Qunatum Chemistry. Book 2*, Lund University, **2005**.
28. Cornell, W.D.; Cieplak, P.; Bayly, C.I.; Gould, I.R.; Merz, K.M.; Ferguson, D.M.; Spellmeyer, D.C.; Fox, T.; Caldwell, J.W. & Kollman ,P.A. *J. Am. Chem. Soc.* ,**1995**,*117*, 5179-5197.
29. Wang, J.; Wolf, R.M.; Caldwell, J.W.; Kollman, P.A. & Case ,D.A. *J Comput Chem* ,**2004**,*25*, 1157-1174.
30. Schlegel, H.B. *Adv. chem. Phys.* ,**1987**,*67*, 249.
31. Helgaker, T. *European Summer School in Qunatum Chemistry. Book 2* , Lund University, **2005**.
32. Teller, E. *J. Phys. Chem.* ,**1937**,*41*, 109.
33. Herzberg, G. & Longuet-Higgins ,H.C. *Trans. Faraday Soc.* ,**1936**,*35*, 77.
34. Bernardi, F.; Olivucci, M. & Robb ,M.A. *Chem. Soc. Rev.* ,**1996**,*25*, 321-328.
35. Desouter-Lecomte, M. & Lorquet ,J.C. *The Journal of Chemical Physics* ,**1979**,*71*, 4391-4403.
36. Frisch, M.J. **2004**
37. Jansen, F. *Introduction to Computational Chemistry*, John Wiley & Sons, UK, **1999**.
38. Leach, A.R. *Molecular Modelling: Principles and Applications*, Pearson Education EMA, UK, **2001**.
39. Bakowies, D. & Thiel ,W. *J. Phys. Chem.* ,**1996**,*100*, 10580-10594.
40. Gao, J.; Lipkowitz, K.B. & Boyd ,D.B. *Reviews in Computational Chemistry*

,**1995**,*7*, 119-185.

41. Sherwood, P. & J Grotendorst *NIC series* ,**2000**,*3*, 285-305.
42. Gao, J. & Truhlar ,D.G. *Annual Review of Physical Chemistry* ,**2002**,*53*, 467-505.
43. Maseras, F. & Morokuma ,K. *Journal of Computational Chemistry* ,**1995**,*16*, 1170-1179.
44. Svensson, M.; Humbel, S.; Froese, R.D.J.; Matsubara, T.; Sieber, S. & Morokuma ,K. *J. Phys. Chem.* ,**1996**,*100*, 19357-19363.
45. Vreven, T.; Mennucci, B.; da Silva, C.O.; Morokuma, K. & Tomasi ,J. *Journal of Chemical Physics* ,**2001**,*115*, 62-72.
46. Sherwood, P.; de Vries, A.H.; Guest, M.F.; Schreckenbach, G.; Catlow, C.R.A.; French, S.A.; Sokol, A.A.; Bromley, S.T.; Thiel, W. & Turner ,A.J. *Journal of Molecular Structure* ,**2003**,*632*, 1-28.
47. Peng, C.; Ayala, Philippe Y Schlegel, H Bernhard & Frisch ,M.J. *Journal of Computational Chemistry* ,**1996**,*17*, 49-56.
48. Vreven, T.; Morokuma, K.; Farkas; Schlegel, H.B. & Frisch ,M.J. *Journal of Computational Chemistry* ,**2003**,*24*, 760-769.
49. Klahn, M.; Braun-Sand, S.; Rosta, E. & Warshel ,A. *J. Phys. Chem. B* ,**2005**,*109*, 15645-15650.
50. Singh, U.C. & Kollman ,P.A. *Journal of Computational Chemistry* ,**1986**,*7*, 718-730.
51. Field, M.J.; Bash, P.A. & Karplus ,M. *Journal of Computational Chemistry* ,**1990**,*11*, 700-733.
52. Ferre, N. & Olivucci ,M. *Journal of Molecular Structure* ,**2003**,*632*, 71-82.
53. Gao, J.; Amara, P.; Alhambra, C. & Field ,M.J. *J. Phys. Chem. A* ,**1998**,*102*, 4714-4721.
54. Pu, J.; Gao, J. & Truhlar ,D.G. *J. Phys. Chem. A* ,**2004**,*108*, 5454-5463.
55. Karlstrom, G.; Lindh, R.; Malmqvist, P.; Roos, B.O.; Ryde, U.; Veryazov, V.; Widmark, P.; Cossi, M.; Schimmelpfennig, B.; Neogrady, P. et al. *Computational Materials Science* ,**2003**,*28*, 222-239.
56. Ahlrichs, R.; Bar, M.; Haser, M.; Horn, H. & Kolmel ,C. *Chemical Physics Letters* ,**1989**,*162*, 165-169.
57. Neese, F. **2006**
58. Sinnecker, S. & Neese ,F. *Journal of Computational Chemistry* ,**2006**,*27*, 1463-1475.
59. Case, D.A.; Cheatham, T.E.; Darden, T.; Gohlke, H.; Luo, R.; Merz, K.M.; Onufriev, A.; Simmerling, C.; Wang, B. & Woods ,R.J. *Journal of Computational Chemistry* ,**2005**,*26*, 1668-1688.
60. Dudek, M.J. & Ponder ,J.W. *Journal of Computational Chemistry* ,**1995**,*16*, 791-816.
61. Duan, Y.; Wu, C.; Chowdhury, S.; Lee, M.C.; Xiong, G.; Zhang, W.; Yang

RONG; Cieplak, P.; Luo, R.; Lee, T. et al. *Journal of Computational Chemistry* ,**2003**,*24*, 1999-2012.

62. Iwamura, H. & Mislow ,K. *Accounts of Chemical Research* ,**1988**,*21*, 175-182.
63. Bedard, T.C. & Moore ,J.S. *Journal of the American Chemical Society* ,**1995**,*117*, 10662-10671.
64. Kelly, T.R.; Bowyer, M.C.; Bhaskar, K.V.; Bebbington, D.; Garcia, A.; Lang, F.; Kim, M.H. & Jette ,M.P. *Journal of the American Chemical Society* ,**1994**,*116*, 3657-3658.
65. Kelly, T.; Sestelo, J. & Tellitu ,I. *The Journal of Organic Chemistry* ,**1998**,*63*, 3655-3665.
66. Schoevaars, A.; Kruizinga, W.; Zijlstra, R.; Veldman, N.; Spek, A. & Feringa ,B. *The Journal of Organic Chemistry* ,**1997**,*62*, 4943-4948.
67. Kelly, T.R.; De Silva, H. & Silva ,R.A. *Nature* ,**1999**,*401*, 150-152.
68. Koumura, N.; Zijlstra, R.W.J.; van Delden, R.A.; Harada, N. & Feringa ,B.L. *Nature* ,**1999**,*401*, 152-155.
69. Kelly, T.; Silva, R.; Silva, H.; Jasmin, S. & Zhao ,Y. *Journal of the American Chemical Society* ,**2000**,*122*, 6935-6949.
70. Koumura, N.; Geertsema, E.; van Gelder, M.; Meetsma, A. & Feringa ,B. *Journal of the American Chemical Society* ,**2002**,*124*, 5037-5051.
71. Vincenzo Balzani, Alberto Credi, Françisco M Raymo, J Fraser Stoddart, *Angewandte Chemie* ,**2000**,*39*, 3348-3391.
72. Buma, W. personal comunication.
73. Gatti, F.G.; Leon, S.; Wong, J.K.Y.; Bottari, G.; Altieri, A.; Morales, M.A.F.; Teat, S.J.; Frochot, C.; Leigh, D.A.; Brouwer, A.M. et al. *Proceedings of the National Academy of Sciences of the United States of America* ,**2003**,*100*, 10-14.
74. Wolfhart, R.; Fritz, T. *Angewandte Chemie International Edition in English* ,**1991**,*30*, 1216-1228.
75. Pratt, L.H. *Photochem. Photobiol.* ,**1995**,*61*, 10.
76. Quail, P.H.; Boylan, M.T.; Parks, B.M.; Short, T.W.; Xu, Y. & Wagner ,D. *Science* ,**1995**,*268*, 675.
77. Rüdiger, W.; Thümmler, F.; Cmiel, E. & Schneider ,S. *Proc. Natl. Acad. Sc. USA* ,**1983**,*80*, 6244.
78. Thümmler, F.; Rüdiger, W.; Cmiel, E. & Schneider ,S.Z. *Naturforsch* ,**1983**,*38c*, 359.
79. Kneip, C.; Hildebrandt, P.; Schlamann, W.; Braslavsky, S.E.; Mark, F. & Schaffner ,K. *Biochemistry* ,**1999**,*38*, 15185.
80. Matysik, J.; Hildebrandt, P.; Schlamann, W.; Braslavsky, S.E. & Schaffner ,K. *Biochemistry* ,**1995**,*34*, 10497.
81. Andel, F.; Lagarias, J.C. & Mathies ,R.A. *Biochemistry* ,**1996**,*35*, 15997.

82. Fodor, S.P.A.; Lagarias, J.C. & Mathies ,R.A. *Biochemistry* ,**1990**,*29*, 11141.
83. Mizutani, Y.; Tokutomi, S. & Kitagawa, T. 1994, 33, 153. *Biochemistry* ,**1994**,*33*, 153.
84. Siebert, F.; Grimm, R.; Rüdiger, W.; Schmidt, G. & Scheer, H. Eur. J. Biochem. 1990, 194, 921. *J. Biochem.* ,**1990**,*194*, 921.
85. Wagner, J.R.; Brunzelle, J.S.; Forest, K.T. & Vierstra ,R.D. *Nature* ,**2005**,*438*, 325-331.
86. Farrens, D.L.; Holt, R.E.; Rospendowski, B.N.; Song, P.S. & Cotton ,T.M. *Journal of the American Chemical Society* ,**1989**,*111*, 9162-9169.
87. Zhang, C.; Farrens, D.L.; Björling, S.C.; Song, P. & Kliger, D. S. J. Amer. Chem. Soc. 1992, 114, 4569. *J. Amer. Chem. Soc.* ,**1992**,*114*, 4569.
88. Zeidler, M.; Lamparter, T.; Hughes, J.; Hartmann, E.; Remberg, A.; Braslavsky, S.E.; Schaffner, K. & Gärtner ,W. *Photochem. Photobiol* ,**1998**,*68*, 857.
89. Durbeej, B.; Borg, A.O. & Eriksson *Phys. Chem. Chem. Phys.* ,**2004**,*6*, 5066.
90. Bischoff, M.; Hermann, G.; Rentsch, S. & Strehlow ,D. *Biochemistry* ,**2001**,*40*, 181.
91. Buchler, R.; Hermann, G.; Lap, D.V. & Rentsch ,S. *Chem. Phys. Lett.* ,**1995**,*233*, 514.
92. Bischoff, M.; Hermann, G.; Rentsch, S.; Strehlow, D.; Winter, S. & Chosrowjan ,H. *J. Phys. Chem. B* ,**2000**,*104*, 1810.
93. Rentsch, S.; Hermann, G.; Bischoff, M.; Strehlow, D. & Rentsch ,M. *Photochem. Photobiol* ,**1997**,*66*, 585.
94. Ditto, M.; Brunner, H. & Lippitsch ,M.E. *Chem. Phys. Lett.* ,**1991**,*185*, 61.
95. Gordon, J.C.; Myers, J.B.; Folta, T.; Shoja, V.; Heath, L.S. & Onufriev ,A. *Nucleic Acids Research* ,**2005**,*33*, 368 371.
96. Inomata, K.; Hammam, M.A.S.; Kinoshita, H.; Murata, Y.; Khawn, H.; Noack, S.; Michael, N. & Lamparter ,T. *J. Biol. Chem.* ,**2005**,*280*, 24491.
97. He, X.; Bell, A.F. & Tonge ,P.J. *J. Phys. Chem. B* ,**2002**,*106*, 6056-6066.
98. Nielsen, S.B.; Lapierre, A.; Andersen, J.U.; Pedersen, U.V.; Tomita, S. & Andersen ,L.H. *Phys Rev Lett* ,**2001**,*87*, 228102.
99. Tsien, R.Y. *Annual Review of Biochemistry* ,**1998**,*67*, 509-544.
100. Andersen, L.H.; Nielsen, I.B.; Kristensen, M.B.; ElGhazaly, M.O.A.; Haacke, S.; Nielsen, M.B. & Petersen ,M.A. *J. Am. Chem. Soc.* ,**2005**,*127*, 12347-12350.
101. Freedman, K.A. & Becker ,R.S. *Journal of The American Chemical Society* ,**1986**,*108*, 1245-1251.
102. Sinicropi, A.; Andruniow, T.; Ferre, N.; Basosi, R. & Olivucci ,M. *J. Am. Chem. Soc.* ,**2005**,*127*, 11534-11535.
103. Altoe, P.; Bernardi, F.; Garavelli, M.; Orlandi, G. & Negri ,F. *J. Am. Chem. Soc.* ,**2005**,*127*, 3952-3963.

104. Jakalian, A.; Jack, D.B. & Bayly ,C.I. *Journal of Computational Chemistry* ,**2002**,*23*, 1623-1641.
105. Jakalian, A.; Bush, B.L.; Jack, D.B. & Bayly ,C.I. *Journal of Computational Chemistry* ,**2000**,*21*, 132-146.
106. Martin, M.E.; Negri, F. & Olivucci ,M. *J. Am. Chem. Soc.* ,**2004**,*126*, 5452-5464.
107. Schellenberg, P.; Johnson, E.; Esposito, A.P.; Reid, P.J. & Parson ,W.W. *J. Phys. Chem. B* ,**2001**,*105*, 5316-5322.
108. Creemers, T.M.H.; Lock, A.J.; Subramaniam, V.; Jovin, T.M. & Volker ,S. *Nat Struct Mol Biol* ,**1999**,*6*, 557-560.
109. Feese, M.D.; Faber, H.R.; Bystrom, C.E.; Pettigrew, D.W. & Remington ,S.J. *Structure* ,**1998**,*6*, 1407-1418.
110. Cembran, A.; Bernardi, F.; Olivucci, M. & Garavelli ,M. *J. Am. Chem. Soc.* ,**2004**,*126*, 16018-16037.
111. Vreven, T.; Bernardi, F.; Garavelli, M.; Olivucci, M.; Robb, M.A. & Schlegel ,H.B. *J. Am. Chem. Soc.* ,**1997**,*119*, 12687-12688.
112. Cembran, A.; Gonzalez-Luque, R.; Altoe, P.; Merchan, M.; Bernardi, F.; Olivucci, M. & Garavelli ,M. *J Phys Chem A Mol Spectrosc Kinet Environ Gen Theory* ,**2005**,*109*, 6597-6605.
113. Andruniow, T.; Ferre, N. & Olivucci ,M. *Proceedings of the National Academy of Sciences of the United States of America* ,**2004**,*101*, 17908-17913.
114. Cembran, A.; Bernardi, F.; Olivucci, M. & Garavelli ,M. *Proc Natl Acad Sci U S A* ,**2005**,*102*, 6255-6260.
115. Houjou, H.; Inoue, Y. & Sakurai ,M. *J. Am. Chem. Soc.* ,**1998**,*120*, 4459-4470.
116. Houjou, H.; Sakurai, M. & Inoue ,Y. *Chemistry Letters* ,**1996**,, 1075-1076.
117. Oesterhelt, D. & Stoechenius ,W. *Nat. New Biol.* ,**1971**,*233*, 149-152.
118. Rajamani, R. & Gao ,J. *Journal of Computational Chemistry* ,**2002**,*23*, 96.
119. Houjou, H.; Inoue, Y. & Sakurai ,M. *J. Phys. Chem. B* ,**2001**,*105*, 867.
120. Vreven, T. & Morokuma ,K. *Theor. Chem Acc.* ,**2003**,*109*, 125.
121. Hayashi, S. & Ohmine ,I. *J. Phys. Chem. B* ,**2000**,*104*, 10678.
122. Hayashi, S.; Tajkhorshid, E. & Schulten ,K. *Biophys. J.* ,**2002**,*83*, 1281.
123. Hayashi, S.; Tajkhorshid, E. & Schulten ,K. *Biophys. J.* ,**2003**,*85*, 1440.
124. Koyama, Y.; Kubo, K.; Komori, M.; Yasuda H. & Mukai Y. Photochem. Photobiol. 1991, 54, 433-443. *Photochem. Photobiol.* ,**1991**,*54*, 433-443.
125. Logunov, S.L.; El-Sayed, M.A. & Song ,L. *J. Phys. Chem.* ,**1996**,*100*, 2391.
126. Mathies, R.A.; Brito Cruz, C.H.; Pollard, W.T. & Shank ,C.V. *Science* ,**1988**,*240*, 777.
127. Dobler, J.; Zinth, W.; Kaiser, K. & Oesterhelt ,D. *Chem. Phys. Lett.*

,**1988**,*144*, 215.

128. Kandori, H.; Katsuta, Y.; Ito, M. & Sasabe, H. 1995, 117, 2669. *J. Am. Chem. Soc.* ,**1995**,*117*, 2669.
129. Hampp, N. *Chem. Rev.* ,**2000**,*100*, 1755.
130. Dugave, C. & Demange ,L. *Chem. Rev.* ,**2003**,*103*, 2475.
131. Gonzalez-Luque, R.; Garavelli, M.; Bernardi, F.; Merchan, M.; Robb, M.A. & Olivucci ,M. *Proc. Natl. Acad. Sci. USA* ,**2000**,*97*, 9379.
132. Cembran, A.; Bernardi, F.; Olivucci, M. & Garavelli ,M.J.A.C.S. *J. Am. Chem. Soc.* ,**2003**,*125*, 12509.
133. Frisch, M.J. *Gaussian 98 , Gaussian, Inc., Pittsburgh PA*, , **1988**.
134. Luecke, H.; Schobert, B.; Richter, H.; Cartailler, J. & Lanyi, J. K. 1999, 291, 899. *J. Mol. Biol.* ,**1999**,*291*, 899.
135. Takeda, K.; Sato, H.; Hino, T.; Kono, M.; Fukuda, K.; Sakurai, I.; Okada, T. & Kouyama ,T. *J. Mol. Biol.* ,**1998**,*283*, 463.
136. Sasaki, J.; Lanyi, J.K.; Needleman, R.; Yoshizawa, T. & Maeda ,A. *Biochemistry* ,**1994**,*33*, 3178.
137. Malmqvist, P.Å.I. *J. Quantum Chem.* ,**1986**,*30*, 479.
138. Houjou, H.; Koyama, K.; Wada, M.; Sameshima, K.; Inoue, Y. & Sakurai ,M. *Chem. Phys. Lett.* ,**1998**,*294*, 162.
139. Albeck, A.; Livnah, N.; Gottlieb, H. & Sheves, M. 1992, 114, 2400. *J. Am. Chem. Soc.* ,**1992**,*114*, 2400.
140. Ferré, N. & Olivucci, M. , 125, 6868. *J. Am. Chem. Soc.* ,**2003**,*125*, 6868.
141. Hasson, K.C.; Gai, F. & Anfinrud ,P.A. *Proc. Natl. Acad. Sci. USA* ,**1996**,*93*, 15124.
142. Shapiro, S.L.; Campillo, A.J.; Lewis, A.; Perreault, G.J.; Spoonhower, J.P.; Clayton, R.K. & Stoeckenius ,W. *Biophys. J.* ,**1978**,*23*, 383.
143. Schenkl, S.; Portuondo, E.; Zgrablić, G.; Chergui, M.; Haacke, S.; Friedman, N. & Sheves ,M. *Phys. Chem. Chem. Phys.* ,**2002**,*4*, 5020.
144. Logunov, S.L.; Volkov, V.V.; Braun, M. & El-Sayed ,M.A. *Proc. Natl. Acad. Sci. USA* ,**2001**,*98*, 8475.
145. Birge, R.R.; Cooper, T.M.; Lawrence, A.F.; Masthay, M.B.; Vasilakis, C.; Zhang, C. & Zidowetzki ,R. *J. Am. Chem. Soc.* ,**1989**,*111*, 4063.

# Figure Index

**INTRODUCTION**

**PART 2.**

# Schemes Index

**PART 1.**

# Tables Index

**PART 1.**



**PART 2.**

# Abbreviations

| | |
|---|---|
| AFF | Amber Force Field |
| BFGS | Broyden-Fletcher-Goldfarb-Shanno |
| BO | Born-Oppenheimer Approximation |
| bR | Bacteriorhodopsin |
| CASPT2 | Complete Active Space Second order Perturbation |
| CASSCF | Complete Active Space Self Consistent Field |
| CI | Configuration Interaction or Conical Intersection |
| COBRAM | COmputational BRidge between Ab initio and Molecular mechanic |
| FC | Franck Condon |
| GAFF | Generalized Amber Force Field |
| H | High |
| HF | Hartree Fock |
| HL | High-Low |
| HM | High-Medium |
| HML | High-Medium-Low |
| LCAO | Linear Combination of Atomic Orbitals |
| MCSCF | Multi Configurational self consistent field |
| MD | Molecular Dynamic |
| MM | Molecular Mechanic |

| | |
|---|---|
| MO | Molecular Orbital |
| PES | Potential Energy Surfaces |
| PSB | Protonated Schiff base |
| QM | Quantum Mechanic |
| QM/MM | Quantum Mechanic/Molecular Mechanic |
| RASSCF | Restricted Active Space Self Consistent Field |
| SO | Second Order |
| TS | Transition State |

# Index