Universitat de Girona

Facultat de Ciències

Memòria del Treball Final de Grau

Títol del treball:

Optimization of Conical Intersections in Biological Systems with the ONIOM Scheme: Projected Gradient Implementation in the Microiteration Process

---

Estudiant: Sílvia Escayola Gordils

Grau en Química

Correu electrònic: silvieg7@gmail.com

Tutor: Dr. Lluís Blancafort San José

Cotutor*:

Empresa / institució: Universitat de Girona

Vistiplau tutor (i cotutor*):

Nom del tutor: Lluís Blancafort San José

Nom del cotutor*:

Empresa / institució: Universitat de Girona

Correu(s) electrònic(s): lluis.blancafort@udg.edu

*si hi ha un cotutor assignat

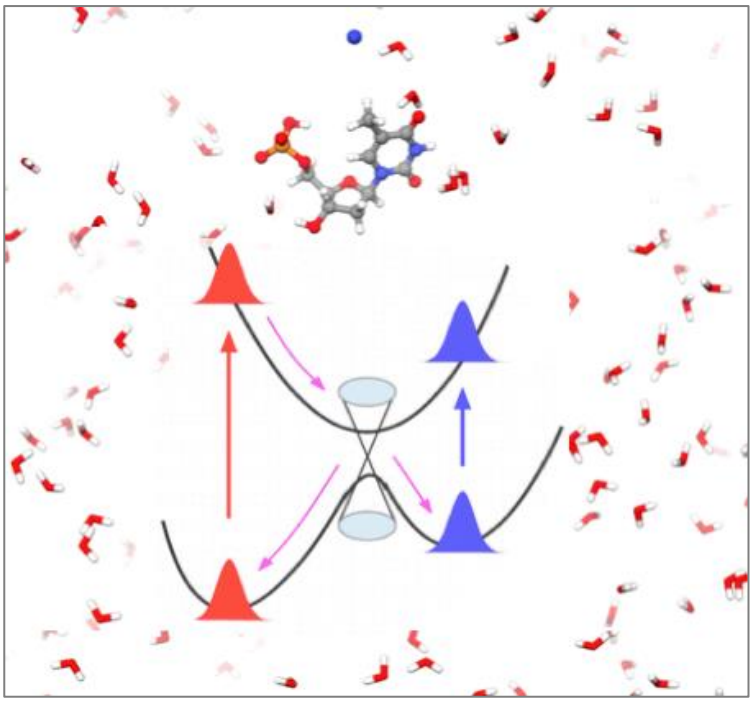
Data de dipòsit de la memòria a secretaria de coordinació: 22 de juliol de 2016

Universitat de Girona

Facultat de Ciències

Memòria del Treball Final de Grau

Optimization of Conical Intersections in Biological Systems with the ONIOM Scheme: Projected Gradient Implementation in the Microiteration Process



Sílvia Escayola Gordils

Tutor: Dr. Lluís Blancafort

SUMMARY

In recent years, the photochemical and photophysical properties of DNA have been a widely studied topic, as this gives a new insight at explaining DNA mutation by light. One fundamental theoretical element for the characterization of these processes are the conical intersections. The intersections between the ground and excited states which have been described up to now for the isolated nucleobases but not for the bases embedded in the large DNA system. For this reason, in this work a conical intersection optimization algorithm has been implemented in the ONIOM scheme, which will allow the study of biological systems such as DNA or proteins.

First, we have studied the photophysics of thymidine-phosphate in the gas phase and in solution, and we conclude that it is a good model system to test the application. The procedure used to improve the original conical intersection optimization method was to use a projected gradients in the microiteration steps of the ONIOM(QM:MM) scheme in order to avoid losing the energy degeneracy during the optimization. The gradient projection was first tested on an independent subroutine and then implemented in a development version of the Gaussian program. The CASSCF and AMBER methods are employed for the QM and MM calculations, respectively. This implementation was tested on the thymidine-phosphate model.

Finally, we have evaluated the results obtained for the implementation during the first optimization step, which includes the different microiteration cycles. The implementation performs properly. At the end of the optimization step, the energy difference is smaller when the projected gradient is used, in line with our goal. At the same time, the absolute energy obtained in the first optimization step including the projected gradient is somewhat higher than in the case where the gradient was not projected. The present results provide a first test of the new algorithm, which will be tested in the future with the more demanding case of thymidine-phosphate in solution.

RESUMEN

En los últimos años, el estudio de las propiedades fotoquímicas y fotofísicas del ADN ha sido un tema ampliamente tratado, ya que esto ofrece una nueva visión para explicar las mutaciones en el ADN producidas por la luz. Un elemento teórico fundamental para la caracterización de estos procesos son las intersecciones cónicas. Las intersecciones entre el estado fundamental y los estados excitados se han descrito para las bases nitrogenadas en fase gas pero no para las bases en sistemas de ADN. Por esta razón, en este trabajo se ha implementado un algoritmo para optimizar intersecciones cónicas dentro en método ONIOM, lo cual permitirá el estudio de sistemas biológicos como el ADN o las proteínas.

Primeramente, se ha estudiado la fotofísica de la timina monofosfato en fase gas i en solución y se ha concluido en que son buenos sistemas modelo para testear la implementación. El procedimiento que se ha realizado para mejorar el método original de optimización de intersecciones cónicas ha sido, incluir los gradientes proyectados en los pasos de microiteraciones de el esquema ONIOM(QM:MM) para evitar perder la degeneración de la energia. Para esto, se han empleado los métodos CASSCF y AMBER para los cálculos QM y MM, respectivamente. La implementación se ha probado para un sistema modelo de timidina monofosfato.

Finalmente, se han cualificado los resultados de la implementación durante el primer paso de optimización, el cual abarca diferentes ciclos de microiteraciones. Se ha observado que la implementación funciona correctamente. Pero a su vez, la proyección del gradiente durante las microiteraciones provoca que la energia absoluta del primer paso de optimización sea mayor que la obtenida sin usar la proyección del gradiente. Los resultados obtenidos sirven como test inicial del algoritmo implementado, el cual tiene que ser probado en un futuro con sistemas más completos como el caso de la timina monofosfato en solución.

RESUM

En els últims anys, les propietats fotoquímiques i fotofísiques de l'ADN han estat objecte de nombrosos estudis, ja que aquestes ofereixen una nova visió a l'hora d'explicar les mutacions de l'ADN produïdes per la llum. Un element teòric fonamental per a la caracterització d'aquests processos són les interseccions còniques. Per aquest motiu, en aquest treball s'ha dut a terme la implementació d'un algoritme d'optimització d'interseccions còniques dins del mètode ONIOM, això permetrà l'estudi de sistemes biològics com l'ADN o les proteïnes.

Primer de tot, s'ha estudiat la fotofísica de la timina monofosfat en fase gas i en solució i s'ha conclòs que són bons sistemes per provar la implementació. El procediment que s'ha seguit per dur a terme la millora del mètode d'optimització d'interseccions còniques original ha estat, incloure els gradients projectats dins dels passos de microiteracions en l'esquema ONIOM(QM:MM) per evitar perdre la degeneració de l'energia. Per això, s'han emprat els mètodes CASSCF i AMBER per als càlculs QM i MM, respectivament. La implementació s'ha provat per un sistema model de timidina monofosfat.

Finalment, s'han qualificat els resultats de la implementació durant el primer pas d'optimització, el qual engloba diferents cicles de microiteracions. S'ha vist que la implementació funciona correctament a l'hora de minimitzar la diferència d'energia existent entre els dos estats que formen la intersecció cònica. Però alhora, la projecció del gradient durant les microiteracions provoca que l'energia absoluta del primer pas d'optimització sigui major que l'obtinguda sense fer la projecció del gradient. Els resultats obtinguts serveixen com a test inicial del algoritme implementat, el qual s'ha de provar en un futur amb sistemes més complets com el cas de la timidina monofosfat en solució.

LIST OF ACRONYMS

BO	Born-Oppenheimer
CASSCF	Complete Active Space Self Consistent Field
CG	Composed Gradient
CoI	Conical Intersection
CS	Composed Step
FCI	Full Configuration Interaction
GS	Ground State
HF	Hartree Fock
MECI	Minimum Energy Conical Intersection
MM	Molecular mechanics
NR	Newton-Raphson
ONIOM	Our N-layer Integrated molecular Orbital molecular Mechanics
PES	Potential Energy Surface
QM	Quantum Mechanics
Thy	Thymine
TMP	Thyimidine-phosphate
UFF	Universal Force Field
UV	Ultraviolet

INDEX

1. INTRODUCTION

Solar radiation is one of the most powerful energy sources in our planet. A large number of processes and phenomena such as photosynthesis, human vision or Earth's global climate are directly dependent on the sunlight. This solar energy reaches the atmosphere in a wide range of wavelengths from the ultraviolet to infrared. Ultraviolet (UV) light, which is between 100 and 400 nm, plays a key role in many biological processes, however not all this radiation is helpful for life. One serious problem caused by exposure to a certain range of UV light (200-320 nm) is the damage at the molecular level of the DNA.[1]

Deoxyribonucleic acid (DNA) is a biopolymer that can be found in the cells of all known living organisms and stores the genetic instructions for their development and functioning. The DNA has a double-helix structure composed by nucleobase pairs and sugar-phosphate backbone. Unfortunately, when DNA is exposed to UV light the nucleobases can absorb this radiation and change their structure. As a result, DNA may lose its functionality and cells can die. Due to its importance, photoexcited DNA is a widely studied topic, which allows us to understand the nature of some DNA damage and repair mechanisms.

1.1. Photochemical Concepts

Photochemical processes result from the interaction between light and molecules or chemical systems. When light is absorbed by a molecule it is transformed into excitation energy and induces the transition from the ground state to an excited state. Once the molecule is excited there are different relaxation mechanisms that can occur, which can be illustrated with a Jablonski diagram (Figure 1).

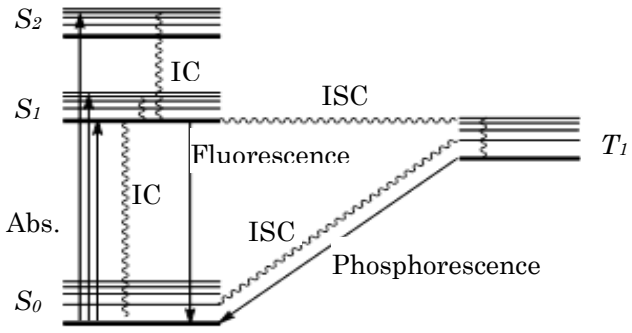


Figure 1. Jablonski diagram. IC-Internal conversion, ISC-Intersystem crossing (figure adapted from ref. 2).

The relaxation mechanisms can be classified as radiative or non-radiative. Principally, the way that the molecule follows to return to the ground state depends on the topology of the potential energy surface. In the case of nucleobases the most efficient processes are the non-radiative ones, where conical intersections (CoIs) play a central role. The existence of these CoIs provides a very efficient lane for the deactivation of the systems. One particular case is the relaxation of UV-photoexcited DNA nucleobases, since CoIs avoid the photochemical reactions that could lead to mutagenesis.[3] For this reason, these phenomena will be explained in more detail below.

1.1.1. Potential Energy Surfaces and Conical Intersections

The evolution of the molecular structure in the ground or excited states can be studied with the potential energy function depending on the molecular geometry, which representation is called potential energy surface (PES). In several cases, particularly in organic molecules, when a system is moving across the PES of an excited state there is a high probability of reaching a conical intersection, which is a region of surfaces crossing. Such crossings allow an extremely fast decay to the ground state and determine the reactivity of the molecule.[4]

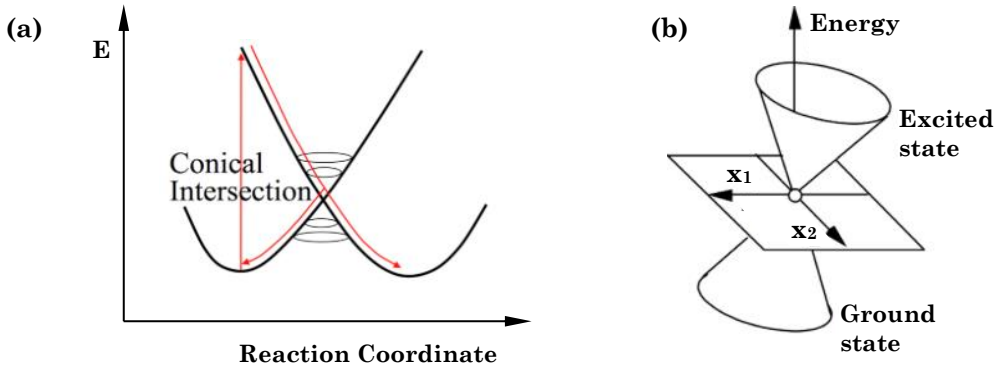


Figure 2. Schematic representation of CoIs. (a) Potential energy profile of a CoI. (b) Double cone shape of two intersecting PES, where the energy of the two states is plotted against x_1 and x_2 coordinates (figure adapted from ref. 4).

For a better understanding of the role of CoIs in photochemical reactions, these can be compared with the transition states of a thermal reaction. A transition state has the highest energy along the reaction path but it is a minimum in all the other directions of the PES. Similarly, CoIs have two directions analogous to the reaction path direction and so the optimization strategy of these points will be similar.

In a system of n internal coordinates there is an intersection space or seam, composed of n–2 internal coordinates, where the two electronic states are degenerate. Then, there are the two remaining internal coordinates that form the branching space, which has a broken degeneracy (Figure 2b). These two directions are the gradient difference (x_1) and the interstate coupling (x_2):

$$\boldsymbol{x_1} = \nabla(E_i - E_j) \qquad \boldsymbol{x_2} = \nabla\langle\Psi_i|\hat{H}|\Psi_j\rangle \tag{1}$$

Where E_i and Ψ_j are the energy and the wavefunction of the state i, respectively and $\hat{H}$ is the Hamiltonian. An important point, within the seam, to understand the photochemical behavior of the system is the minimum energy CoI (MECI). In order to find the MECI point, the energy difference between the two states and the energy of the seam space must be minimized simultaneously. However, one important point is that during the minimization of the energy, steps along x_1 and x_2 have to be avoided because otherwise the degeneracy between S_0 and S_1 would be lost. Therefore, it is necessary to project out the contributions to the gradient along x_1 and x_2 and calculating the so-called projected gradient. Therefore, the x_1 and x_2 vectors are important for the CoI optimization process, which will be described in Section 3.2.1.

1.2. Conical Intersections in Thymine

In the recent years a large amount of studies about the non-radiative decay of the DNA nucleobases adenine, cytosine, guanine and thymine (Figure 3) has been reported in the literature. Nucleobases have good photostability since they present a fast dissipation of the excitation energy to the surroundings in form of heat. As it was mentioned above, this rapid decay can be explained by the existence of conical intersections connecting the first excited states (*e.g.* $n\pi^*$, $\pi\pi^*$) with the ground state. This premise is also confirmed by several computational studies.[5]

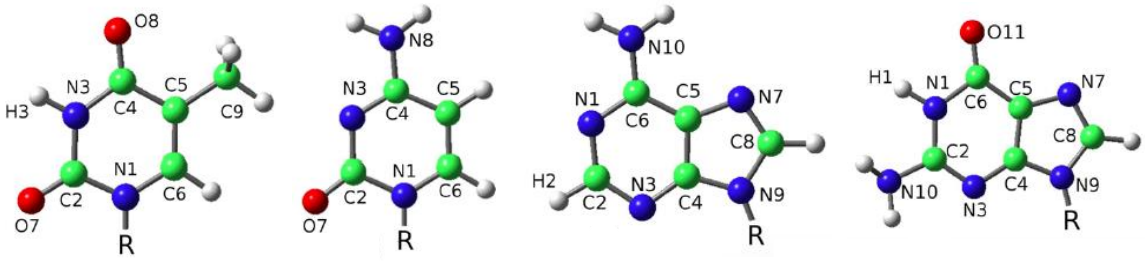


Figure 3. Representation of the four nucleobases of ADN. From left to right: thymine, cytosine, adenine and guanine when R is equal to hidrogen (Ref. 5).

For the particular case of thymine (Thy), the less energetic excited state is the S$n\pi^*$. This excited state comes from the transition of an electron from one of the lone pairs (n) of the carbonyl group, adjacent to the methyl group, to the π^* orbital. The ground state is connected to the S$n\pi^*$ and S$\pi\pi^*$ excited states by CoIs (Figure 4). The decay mechanism of the excited thymine is very complex and involves not only the S$n\pi^*$ state but also the S$\pi\pi^*$. In order to simplify the study, here we will center on the S$n\pi^*$ state and the S$n\pi^*$ /S_0-CoI.

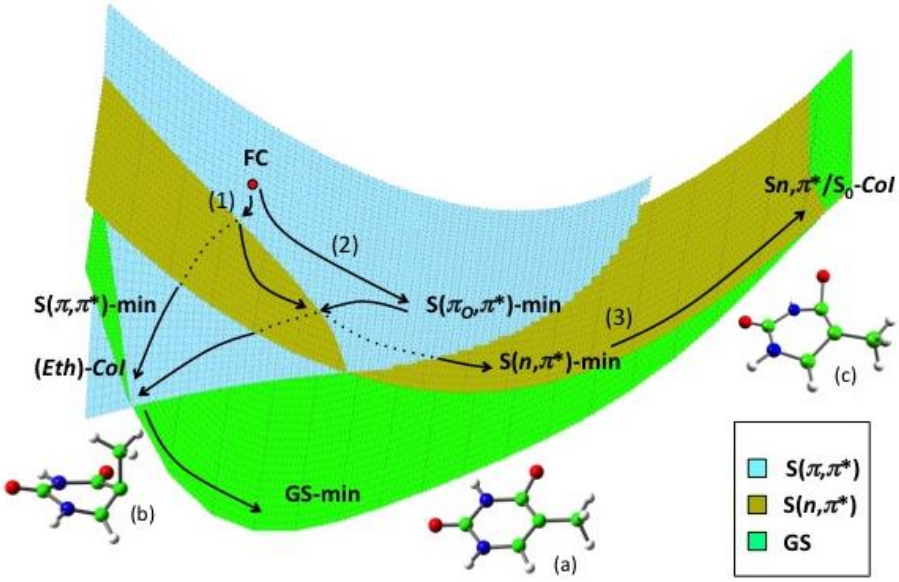


Figure 4. Schematic potential energy surfaces for the ground and excited states of Thy. Geometries of (a) GS minimum, (b) CoI between S$\pi\pi^*$ and S_0 (Eth-CoI) and (c) CoI between S$n\pi^*$ and S_0 (S$n\pi^*$/S_0-CoI). Representation of the decay pathways: (1) direct decay pathway, (2) indirect decay pathway and (3) S$n\pi^*$/S_0-CoI decay pathway (figure adapted from ref. 5).

1.2.1. Solvent and Environmental Effects

Biological processes are made of many chemical reactions and involve a large number of molecules. Consequently, even if we want to study a single reaction the environment of the reacting molecules can affect the reactivity of the system and lead to different reactions depending on whether it is taken into account or not.

The simplest approach to carry out the study of the UV-photoexcited DNA, at both experimental and theoretical level, is to consider the isolated nucleobases as a model system. Nevertheless, this approach is not good enough to understand the functional roles of the DNA and it is necessary to deal with more extensive systems such as nucleotides in solution or DNA strand fragments.

1.3. Computational Approaches

The computation of DNA-like systems is complex and requires of models that can reproduce as well as possible the real conditions. This computation is not trivial because the biomolecular systems are very large and comes with extremely high computational cost if we use methods based on Quantum Mechanics (QM) to carry out the calculations. Mainly, there are two ways to face this problem. On the one hand the system can be reduced to smaller models such as single nucleobases in gas phase, which can be easily calculated using QM. On the other hand, a cheaper computational method based on Molecular Mechanics (MM) can be used. However, both options still have limitations since the first one ignores the effects of the environment and the second one does not allow the description of CoIs. Then, we are losing a very important contribution that helps us to explain the environment effect to the molecular system. A possible solution is the use of a hybrid method based on the combination of both QM and MM. One example of this method is the so-called ONIOM (Our N-layer Integrated molecular Orbital molecular Mechanics) approach in which the molecular system must be divided in different layers.[6] This treatment allows us to work with very large systems and describe properly the bond cleavege of a smaller section.

In general, one of the most common approaches for the study of excited-state process is the characterization of the PES, in particular the location of excited state minima and CoIs. Over the last years, the group of Dr. Lluís Blancafort has been working on developing computational algorithms to locate and optimize CoIs on the PES. Recently, these algorithms have been implemented at the QM:MM level using UFF (Universal Force Field) to describe the MM part, which allowed the location of CoIs in crystal systems.[7] The present work is focused on extending the implementation to other force fields that contain charged atoms, like the so-called AMBER, allowing us to treat biological systems such as DNA and proteins. For this purpose, we will first evaluate different model systems based on the thymine chromophore to be used as tests of the implementation. Then, we will implement and test the CoI algorithm using CASSCF and AMBER levels of theory.

2. OBJECTIVES

The main goal of the present work is to extend and test the implementation of the CoI optimization algorithm in the ONIOM scheme using force fields that contain charged atoms. This implementation will be carried out within the Gaussian program.[8,9] Thus, the most important issues are the following:

- Choose a system to test the implementation. To this end, construct diverse model systems based on the thymine nucleobase and characterize the critical points of the PES, the ground state (S_0), lowest excited singlet state (S_1) and conical intersection between them (S_1/ S_0-CoI), at different levels of theory.

- Program a Fortran subroutine for the projected gradient calculation considering the memory requirements set by Gaussian.

- Implement the CoI optimization in the ONIOM(QM:MM) scheme including the gradient projection in the microiteration steps.

- Evaluate the effect of the microiterations on the energy degeneracy during the CoI optimization process in ONIOM with and without the gradient projection.

3. METHODOLOGY

3.1. Theoretical Methods

Computational chemistry is a branch of chemistry that uses different methods to determine the structure and predict the behavior of atoms and molecules. The main purpose of the present work is to extend the implementation of one of these methods, which combines both quantum and molecular mechanics calculations. For this reason, in the following section the basic concepts of quantum mechanics and molecular mechanics methods are briefly summarized.[10] Giving particular emphasis to the methods used in this work, the CASSCF, the AMBER force field and the ONIOM(QM:MM) scheme.

3.1.1. Electronic Structure Methods

The electronic structure methods are those based on quantum mechanics and basic physical constants. The goal of these methods is to find the solution of the time-independent Schrödinger equation, that gives a description of the studied quantum state of a system and its energy.

$$\hat{H}\Psi_i = \varepsilon_i \Psi_i \tag{2}$$

$$\hat{H} = \hat{T}_N + \hat{T}_e + \hat{V}_{NN} + \hat{V}_{ee} + \hat{V}_{Ne} \tag{3}$$

In Equation (2), Ψ_i and ε_i are the wave function and the energy of the state i respectively and $\hat{H}$ is the Hamiltonian operator. As it is shown in Equation (3) the $\hat{H}$ of a molecule contains the kinetic energy operators $\hat{T}$ of the nuclei (N) and the electrons (e) and the potential energy operators $\hat{V}$, that account for the different attractive and repulsive interactions.

Unfortunately, the Schrödinger equation only has analytical solution for the hydrogen atom. For larger systems the use of some approximations to determine (approximately) Ψ_i and ε_i is required. The first approach that can be applied is the Born-Oppenheimer (BO) approximation, which is based on the mass difference between the nuclei and the electrons. Since the mass of the nuclei is much higher than the mass of the electrons, the kinetic energy of the nuclei is negligible in comparison to the kinetic energy of the electrons. This is why the $\hat{T}_N$ term of the $\hat{H}$

can be removed achieving a definition of the Schrödinger Equation (6) that can be solved for fixed nuclei yielding the electronic wave function:

$$\hat{H}_{BO} = \hat{T}_e + \hat{V}_{NN} + \hat{V}_{ee} + \hat{V}_{Ne} \quad (4)$$

$$\hat{H}_{elec} = \hat{T}_e + \hat{V}_{ee} + \hat{V}_{Ne} \quad (5)$$

$$\hat{H}_{elec}\Phi_{elec} = \varepsilon_{elec}\Phi_{elec} \quad (6)$$

Where Φ_{elec} is the electronic wavefunction and ε_{elec} the function eigenvalue, which gives values for ground and excited states. Then, the total energy of the system is the sum of the electronic energy calculated with Equation (6) and the nuclear energy, which is constant for a given position of the nuclei.

$$E_{Total} = \varepsilon_{elec} + \varepsilon_{nucl} \quad (7)$$

Because of the electron repulsion term $\hat{V}_{ee}$ the electronic Schrödinger equation is not separable. Omitting this term, the $\hat{H}_{elec}$ can be expressed as a sum of the other monoelectronic terms and the approximate Φ_{elec} can be obtained as a result of a linear combination of Hartree products. These expressions can be written as a determinant, called Slater determinant, where each row represents an electron and each column a spin orbital. The Slater determinants satisfy the Pauli antisymmetry and exclusion principles and for this reason are good Φ_{elec} of the system. In a compact form, the Φ_{elec} can be expressed as the diagonal of the Slater determinant:

$$|\Phi_{elec}\rangle = |X_1 X_2 \dots X_n\rangle \quad (8)$$

For a molecular system calculation, the molecular orbitals can be expressed as linear combination of atomic orbitals (X). At this point, the effect of the $\hat{V}_{ee}$ term can be introduced as electron correlation and following an iterative process the minimum energy configuration can be found. This procedure is known as Hartree-Fock (HF) method. In general, HF gives satisfactory results in the description of the ground state but fails in other cases like excited states. In order to solve this problem, there are a variety of more accurate methods that can be used.

A better approximation to solve the Schrödinger equation is the Full Configuration Interaction (FCI) method. This method uses a combination of all the possible configurations with their different occupations as wave function. As a result of this, FCI provides the exact solution of the Schrödinger equation for a given

atomic orbital basis set. However, the computational cost of this calculation is too high to be carried out in almost all systems. For this reason the most feasible manner to study excited states is to use an intermediate method between HF and FCI, such as CASSCF.

In the CASSCF method (Complete Active Space Self Consistent Field) a reduced set of orbitals and electrons (Active space) are treated at the FCI level and the remaining part (occupied and virtual orbitals) is treated similar to HF. This method provides a better representation of the chemically relevant orbitals and allows the description of the electronic state mixing.[11] For all these reasons, the CASSCF method is useful in excited state and CoI studies, which will be carried out in this work.

3.1.2. Molecular Mechanics/ Force Field Methods

Molecular mechanics or force field methods use classical mechanics to predict the structure and the potential energy of the system. In this approach the electrons are ignored and the energy is obtained as a parametric function of the nuclear coordinates. The potential energy function is expressed as the sum of different terms, depending on the nature of the potential energy.

$$E_{Total} = \varepsilon_{Stretching} + \varepsilon_{bending} + \varepsilon_{Torsional} + \varepsilon_{Van\ der\ Waals} + \varepsilon_{Electrostatic} \quad (9)$$

Each term depends on two, three or four atoms and has one or more associated parameters. These parameters depend on the atom type and can be obtained experimentally or theoretically with *ab initio* methods. UFF or AMBER are examples of these force fields. Both can be used to get optimized geometries of the DNA systems but the latter is preferred because it was primarily developed for the study of proteins and nucleic acids, since this allows the study of systems with charged atoms.[12] However, as it has been explained in the section 1.3, the MM methods have important drawbacks for the study of chemical reactions and the best alternative in this work is the use of a hybrid method like ONIOM.

3.1.3. The ONIOM Scheme

The ONIOM method is an extrapolation scheme that allows the combination of more than one computational technique in a single calculation. The two-layer ONIOM method divides the whole system (real system) in two regions: the model part (QM

region), which is the chemically active region and the surrounding environment part (MM region), as it is shown in Figure 5. The energy of the entire system can be obtained as an extrapolation by adding the MM energy of the real part and the QM energy of the model and subtracting the MM energy of the model part, which gives the following equation:

$$E_{ONIOM\ (QM\ high:MM\ low)} = E_{MM,real} + E_{QM,model} - E_{MM,model} \quad (10)$$

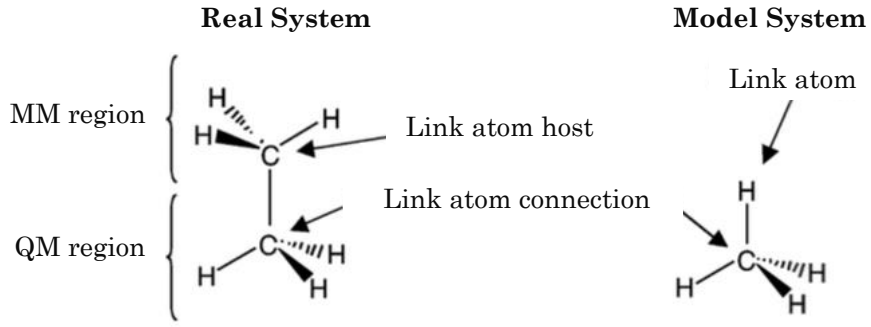


Figure 5. ONIOM terminology and system division using ethane as an example (Ref. 13).

In the ONIOM scheme there are two options to consider the electrostatic interactions between the QM and MM regions.[13] On one hand, we have the classical embedding or mechanical embedding, that is the simplest form and consists in describing the interaction between the two parts at the MM level. This approach uses only atomic point charges to describe the electrostatic interactions. On the other hand, we have the electronic embedding, where the QM wave function of the model system can be polarized by the incorporation of the partial charges of the MM region into the QM Hamiltonian. In general the electronic embedding provides a better description of the system because the electrostatic interaction between the model and the environment is treated at the QM level, and for this reason we will use it here.

3.2. Geometry Optimization

In order to study the behavior of the systems, it is necessary to find the stationary points of the wave function, i.e. minima, first-order saddle points, CoIs, etc. There are many optimization methods that can be used for finding minima, but this section is focused on the *Newton-Raphson* method.[10]

The *Newton-Raphson* (NR) method is an iterative process that can be used for function optimization. To optimize a function, $f(\boldsymbol{q})$, the second order Taylor expansion of $f(\boldsymbol{q})$ in a certain point, $\boldsymbol{q}_0$, is considered.

$$f(\boldsymbol{q}) \simeq f(\boldsymbol{q}_0) + \boldsymbol{g}^t(\boldsymbol{q}-\boldsymbol{q}_0) + \frac{1}{2}(\boldsymbol{q}-\boldsymbol{q}_0)\boldsymbol{H}^t(\boldsymbol{q}-\boldsymbol{q}_0) \tag{11}$$

Where $\boldsymbol{g}$ and $\boldsymbol{H}$ are the first and second derivatives of the function with respect to the $\boldsymbol{q}$ coordinates, also called gradient and Hessian, respectively. Requiring the general condition for finding a stationary point, $\boldsymbol{g} = 0$ of the Equation (11) produces the following steps.

$$\boldsymbol{g} = \boldsymbol{g}_0 + \boldsymbol{H}(\boldsymbol{q}-\boldsymbol{q}_0) \tag{12}$$

$$\Delta\boldsymbol{q} = (\boldsymbol{q}-\boldsymbol{q}_0) = -\boldsymbol{H}^{-1}\boldsymbol{g} \tag{13}$$

This NR formula can be used iteratively to reach the stationary point. In each step the nuclear coordinates are modified and the gradient and Hessian matrix are recalculated until they reach the minimum conditions. In the case of a molecular system optimization, the function $f(\boldsymbol{q})$ represents the energy of the system with respect to the nuclear coordinates.

3.2.1. Conical Intersections Optimization with Projected Gradients

For the optimization of conical intersections there are two conditions that must be fulfilled in order to find the MECI point. These conditions can be described as:

$$\Delta E = 0 \qquad \boldsymbol{g}_{proj.} = 0 \tag{14}$$

Where ΔE is the energy difference between the two studied states and the $\boldsymbol{g}_{proj.}$ is the projection of the gradient in the intersection space. This $\boldsymbol{g}_{proj.}$ can be obtained by applying the projector ($\boldsymbol{P}$) to the gradient of one of the degenerated states.

$$\boldsymbol{g}_{\boldsymbol{proj.}} = \boldsymbol{P}\boldsymbol{g}_{\boldsymbol{i}} = \boldsymbol{P}\frac{\partial E_i}{\partial \boldsymbol{q}} \tag{15}$$

$$\boldsymbol{P} = \mathbf{I} - \hat{\boldsymbol{x}}_1\hat{\boldsymbol{x}}_1{}^t - \hat{\boldsymbol{x}}_2\hat{\boldsymbol{x}}_2{}^t \tag{16}$$

In the Equation 16, **I** is the identity matrix and the $\hat{\boldsymbol{x}}$ notation refers to the normalized vector $\boldsymbol{x}$. x_1 and x_2 are the two vectors that break the degeneracy at the intersection (Equation 1). The projector eliminates the components of x_1 and x_2 along the optimization gradient so that steps along the directions that break the degeneracy are avoided.

There are different algorithms that can be used for the optimization of MECIs that employ projected gradients. In this section the composed gradient (CG) and the composed step (CS) methods will be explained.[14,15] The main idea of the CG scheme is to use a composed gradient for the NR optimization. This gradient is expressed as:

$$\boldsymbol{g}^{CG} = \boldsymbol{g}_{proj.} + 2\Delta E\boldsymbol{x}_1 \tag{17}$$

Each part of the $\boldsymbol{g}^{CG}$ is related to one condition of the Equation 14. The $2\Delta E\boldsymbol{x}_1$ term is used to reach the degeneracy of the two electronic states, while the $\boldsymbol{g}_{proj.}$ term is involved in the energy minimization inside the seam space. The CS scheme is based on carrying out one NR optimization for each condition of Equation 14. This gives us two optimization steps, $\Delta\boldsymbol{q}_1$ and $\Delta\boldsymbol{q}_2$, which can be grouped to obtain the final optimization step:

$$\Delta\boldsymbol{q} = \Delta\boldsymbol{q}_1 + \Delta\boldsymbol{q}_2 \tag{18}$$

For the $\Delta\boldsymbol{q}_1$ step, related to the first condition, the ΔE has to be minimized into the plane formed by the x_1 and x_2 vectors (Figure 2b) and the NR optimization step can be expressed as:

$$\Delta\boldsymbol{q}_1 = -\boldsymbol{H}_\Delta{}^{-1}\hat{\boldsymbol{x}}_1 \tag{19}$$

Where $\boldsymbol{H}_\Delta$ is the Hessian of the ΔE. The second $\Delta\boldsymbol{q}_2$ stp is based on the NR minimization in the intersection space using the projected gradient of Equation 15.

$$\Delta\boldsymbol{q}_2 = -\boldsymbol{H}_{IS}{}^{-1}\boldsymbol{g}_{proj} \tag{20}$$

$\boldsymbol{H}_{IS}$ is the Hessian of the projected gradient in the intersection space. The CS method gives a better definition of the Hessian and for this reason it will be used in this work to improve the microiteration steps procedure for CoI optimization.

3.2.2. Microiterations Method

Geometry optimization of a system in the ONIOM(QM:MM) scheme follows a particular strategy, the microiterations method. Since the ONIOM scheme works

with two or more theoretical levels the use of a hybrid optimization method, that combines both micro and macroiterations, is the best strategy. First, the environment part is optimized at MM level, with the fixed model part (microiterations). When the electronic embedding (section 3.1.3) is used in the calculation, the electronic density of the QM region is recalculated after each microiterations cycle and the MM optimization is repeated with the new QM density. This series of microiterations is carried out until the QM density and the MM geometry are consistent. Then, a QM optimization step of the model part is done (macroiterations). This procedure is repeated *n* times until the forces of QM and MM part fulfill the optimization criteria. The consideration of electrostatic interactions between the chromophore and the surroundings are important in the photochemical studies of charged systems. In particular, if the unprojected gradient is used in the microiterations, then the energy degeneracy could be lost because steps along x_1 and x_2 would be possible. For this reason, the projected gradient (Equation 15) has to be calculated also during the microiteration steps. This part is the one that will be implemented in this work.

3.3. Computational Details

3.3.1. Projected Gradient in Microiteration Steps

The use of projected gradient in the ONIOM microiteration steps has been included in the Gaussian Development version Program. For this, two subroutines of Gaussian, MMCGOp and IndMMC, were modified. In the MMCGOp subroutine the projection step of Equation (15) was included. Thus, the displacements of the low level atoms are taken into account in the optimization of the CoI. The modifications in the IndMMC subroutine are based only on the memory allocation for the x_1, x_2 and projected gradient vectors.

3.3.2. Thymine Models

To test the projected gradient implementation, three different model systems were considered: Thy in the gas phase was calculated as reference; thymidine-phosphate in gas phase, *i.e.* TMP(g), was used as an initial test system and thymidine-phosphate in solution, *i.e.* TMP(s), (TMP surrounded by a sodium cation and a sphere of more than 850 water molecules) was studied, as a possible test system for

future more demanding tests. The starting structures for the optimization of the ground state and the first excited state in solution have been obtained from a molecular dynamics simulation of the TMP(s) carried out by Dr. L. Blancafort. Our main aim is not to characterize the PES with high accuracy but to test the implementation, therefore we have used the small, computationally efficient 3-21G basis set. For the case of Thy, the search for the minimum-energy structure of S_0, S_1 and S_1/S_0-CoI was performed at the CASSCF(8,7)/3-21G level of theory. Otherwise, the optimizations in the case of TMP in the gas phase and in solution were performed at the ONIOM level with CASSCF(8,7)/3-21G as high level and UFF and AMBER as low level, respectively. In both cases thymine was treated at high level and the rest of the system at low level. In the optimizations in solution, the deoxyribose-monophosphate, the sodium ion and the water molecules within 5 Å were considered active while the rest of the system was frozen during the optimization. The optimizations of the S_0 and S_1 states in all the cases and the optimization of the S_1/S_0-CoI in the case of thymine in gas phase were carried out using the Gaussian 09 package.[8] For the optimization of the S_1/S_0-CoI of the remaining models the development version of the Gaussian package was used.[9]

The implemented CoI optimization scheme has been tested with the S_1/S_0-CoI optimization of the TMP in gas phase. To evaluate the effect of the projected gradient during the microiteration steps, the optimization was carried out using both the original and modified versions of Gaussian and comparing their results. The calculations have been carried out at the ONIOM level with CASSCF(8,7)/3-21G as a high level and AMBER as low level.

3.4. Computational Chemistry For a Sustainable Research

This section is dedicated to an important feature in scientific research, sustainability. Computational chemistry has many advantages from the sustainability point of view as it can reduce the environmental costs of research. Computational chemistry uses chemistry, mathematics and computing techniques to solve chemical problems. This allows scientists to make predictions before running experiments so that they can save resources, accelerate the research and avoid as much as possible the generation of hazardous substances. From a sustainable point of view, this project is helpful because it is based on the use and improvement of computational tools.

4. RESULTS AND DISCUSSION

4.1. Model Systems Optimization

In order to have a better understanding of the test system and get an optimal model to test the performance of the implemented CoI optimization scheme, the models introduced in Section 3.3.2 were studied. The energy results obtained for the optimized critical points (minima and CoIs) of the different models are summarized in Table 1 and Figure 6.

Table 1. QM and ONIOM energies in eV (relative to S_0-Min) of the optimized structures of S_0-Min, S_1-Min and S_1/S_0-CoI-Min for the Thy, TMP(g) and TMP(s) model systems.

Model system	Level of Theory	Relative QM Energy (eV)			Relative ONIOM Energy (eV)		
		S_0-Min	S_1-Min	S_1/S_0-CoI-Min	S_0-Min	S_1-Min	S_1/S_0-CoI-Min
Thy	CASSCF	0.000	3.479	5.437	---	---	---
TMP(g)	CAS:UFF	0.000	4.208	5.524	0.000	4.205	6.069
TMP(g)	CAS:AMBER	0.000	6.036	---	0.000	4.663	---
TMP(s)	CAS:UFF	0.000	3.782	---	0.000	4.591	---

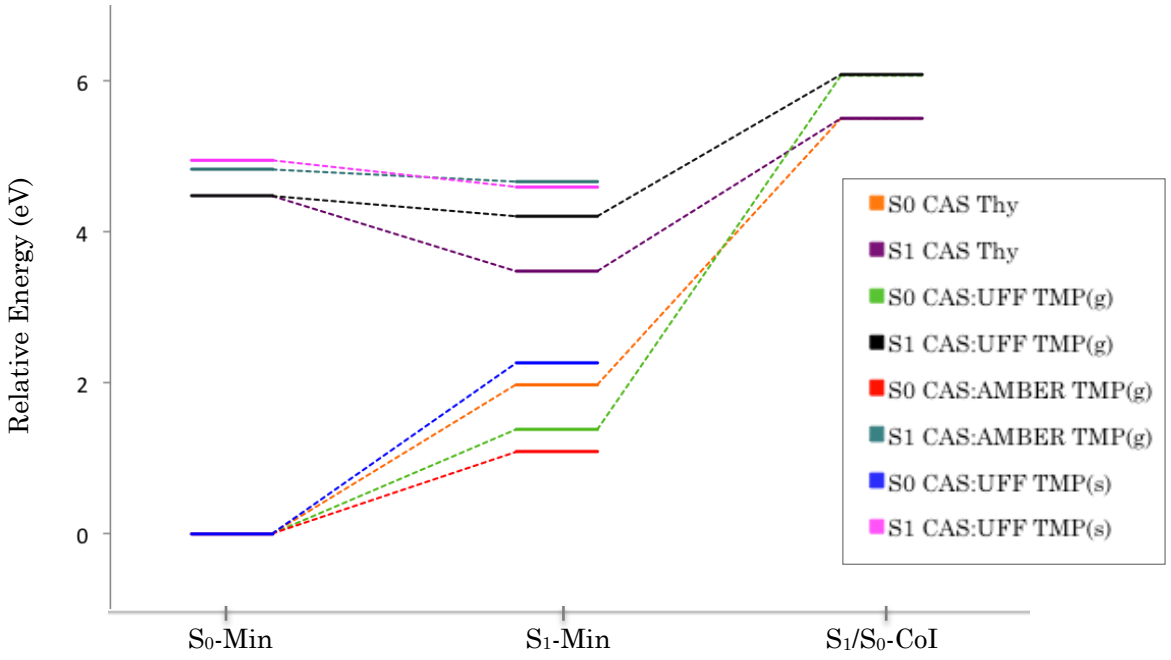
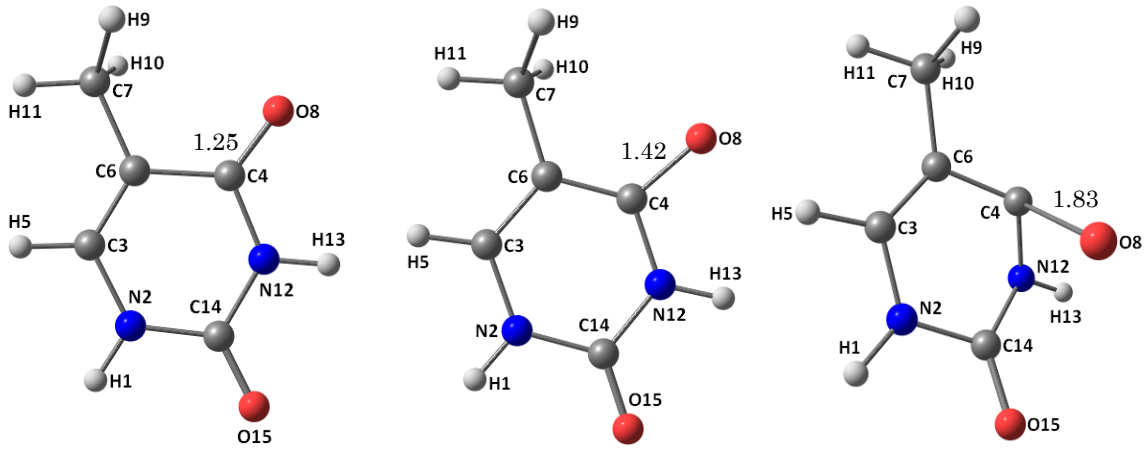


Figure 6. Energy diagram of the S_0 and S_1 states of the different thymine models, calculated with the different methods shown in the legend at the equilibrium geometries of S_0, S_1 and S_1/S_0-CoI. The energies (relative to the S_0 state) are given in eV.

The energy diagrams show that the different studied models have a similar energy profile. The excitation energies in the Franck-Condon region for the different cases are the following: 4.48 eV for Thy model, 4.48 eV for TMP(g) at the CAS:UFF level, 4.83 eV for the TMP(g) at the CAS:AMBER level and 4.95 eV for TMP(s) at the CAS:UFF level.

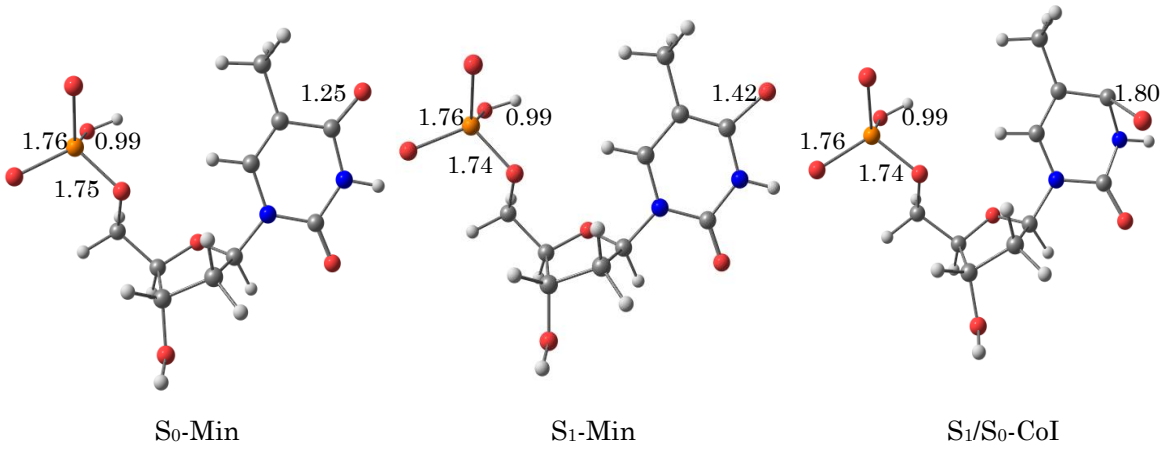
The optimized geometries of the S_0-Min, the S_1-Min and the S_1/S_0-CoI-Min were obtained as a result of the different thymine models optimization. Here the most important features of each optimized structure will be discussed.



S_0-Min **S_1-Min** **S_1/S_0-CoI**

Figure 7. Equilibrium geometries of the GS, lowest excited state and CoI between them for the Thy model system, optimized at CASSCF/3-21G level. Bond lengths are given in angstroms.

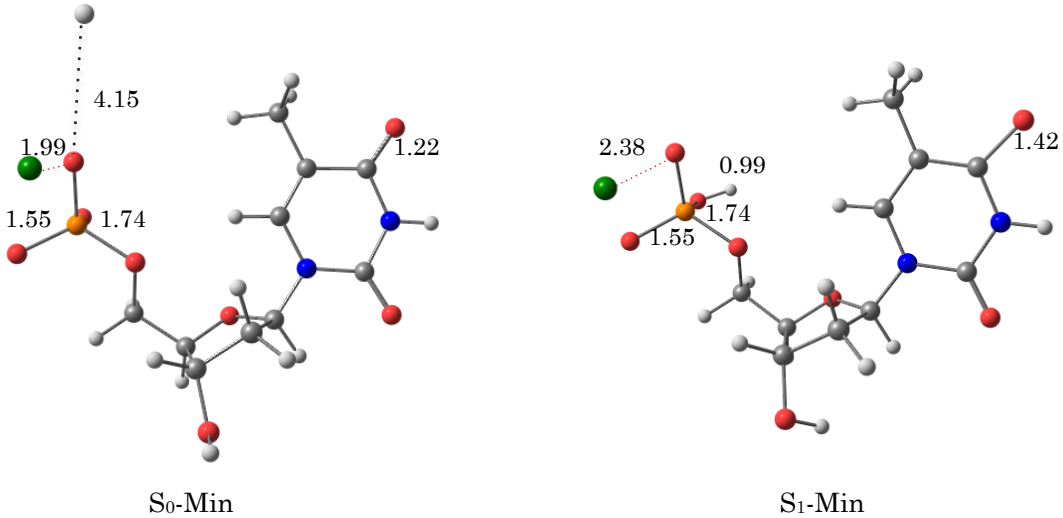
The optimized ground state, first excited estate and S_1/S_0-CoI geometries of the Thy at the CASSCF/3-21G level are shown in Figure 7. The S_0-Min and S_1-Min present planar geometry, while in the S_1/S_0-CoI an out-of-plane puckering is observed. This agrees with the structures reported in the literature.[5,16] The most relevant structural change between S_0-Min and S_1-Min is the elongation of the C_4-O_8 bond, which is larger in the S_1-Min structure. The elongation is even more pronounced in the case of the S_1/S_0-CoI structure. It must be noted that the optimized geometries were obtained using a reduced basis set, therefore all the bond length values are only qualitative.



S_0-Min S_1-Min S_1/S_0-CoI

Figure 8. Equilibrium geometries of the GS, lowest excited state and CoI between them for the TMP(g) model system, optimized at CASSCF/3-21G:UFF level. Bond lengths are given in angstroms.

The identification of S_0, S_1 and S_1/S_0-CoI in the TMP(g) gives structures that differ principally in the thymine and the phosphate group. In both, CASSCF:UFF and CASSCF:AMBER calculations the geometrical changes between the different structures in the thymine nucleobase are qualitatively the same as those observed in the single Thy optimization. Furthermore, three P-O bonds of the phosphate group are longer in the S_1 and S_1/S_0-CoI than in S_0.



S_0-Min S_1-Min

Figure 9. Equilibrium geometries of the GS and the lowest excited state for the TMP(s) model system, optimized at CASSCF/3-21G:UFF level. The water molecules were removed from the structure for a better view. Bond lengths are given in angstroms.

Finally we have optimized the S_0-Min and the S_1-Min structures of the TMP(s) at the CASSCF/3-21G:UFF level (figure 9). Here a major change in geometry of the OH of the phosphate group can be observed between the two structures. In the S_0-Min geometry the hydrogen of the P-OH moiety is bonded to a water molecule instead of the oxygen of the phosphate group, while in the S_1-Min the O-H bond is conserved. The sodium ion is close to the partially charged oxygens of the phosphate group, as it is shown in Figure 9. This shows that the sodium is closer to the phosphate group in the S_0-Min state, where the hydroxyl is not present. The rest of the phosphate group did not differ in one or another geometry. Moving to the thymine nucleobase, it can be seen that the C_4-O_8 distance is elongated in the case of the S_1-Min like in the other optimized models.

Considering the results shown above, it is confirmed that the different models obtained are adequate to test the implementation of the microiterations optimization part of the ONIOM scheme. The energies and the geometries of the different models are consistent with each other and with the literature.

4.2. Projected Gradient Subroutine

As a first step for the implementation of the projected gradient, an independent subroutine was programmed in Fortran to calculate the projected gradient (see Equation 15) from the x_1 and x_2 vectors (see Equation 1), the energy difference between the two intersecting states and the gradient of the previous optimization step.

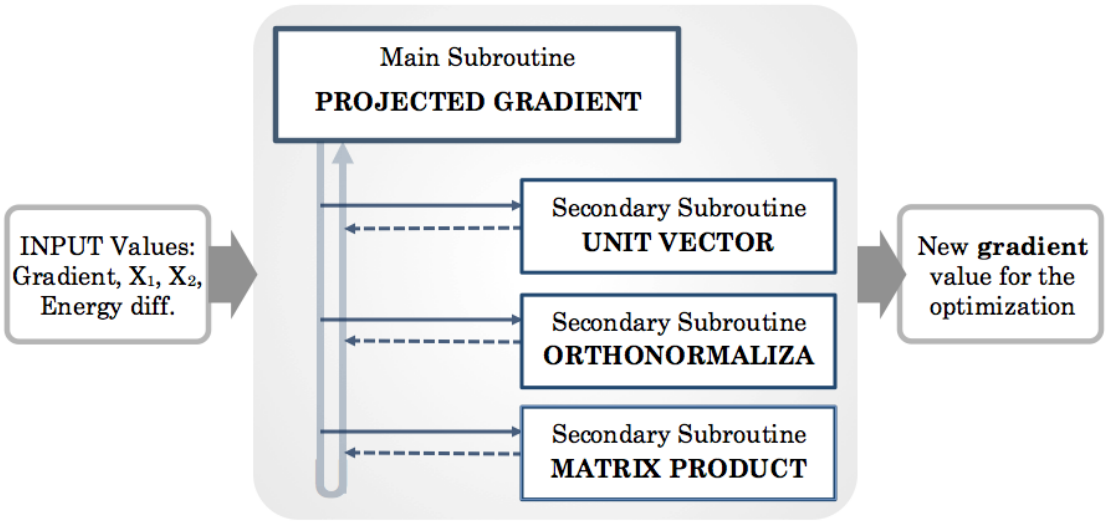


Figure 10. Schematic representation of the Fortran subroutine functioning.

The subroutine was programmed to read the input values from an external file. This subroutine follows the memory allocation strategy used in the Gaussian program, where all the variables are stored in a single vector C in the main subroutine. In the same way, all the input and output variables required in the internal subroutines are read and written from/in the C vector of the main subroutine. The position of the C vector where the variables are stored is set by pointers. As it can be seen in Figure 10, the main subroutine is composed of three secondary subroutines: The Unit Vector subroutine, which is used to normalize the x_1 and x_2 vectors; the Orthonormalization subroutine used to construct the projection matrix (Equation 16) from the normalized vectors; and the Matrix Product subroutine, which is employed do perform all the matrix and vectorial products of the other subroutines.

4.3. Microiterations in CoI Optimization in the ONIOM Scheme

The implementation of the CoI optimization with the ONIOM scheme for the electronic embedding case was done and tested with the calculation of the TMP(g) system at the CASSCF:AMBER level. With the implementation, the optimization process has worked throughout the first macroiteration step. Because of some failures of other Gaussian subroutines (not treated in this project) it has not been possible to complete the optimization of the CoI structure. The results of the first series of macroiteration cycles of the S_0-S_1-CoI optimization of TMP(g) carried out with the original and the implemented versions of Gaussian are presented in Tables 2 and 3.

Table 2. Energy values of the CoI optimization for the TMP(g) system without including the projection in the microiteration steps. The number of microiteration cycle within the macrocycle and the number of steps of each microcycle are shown. Energies in hartee.

Microiterations cycle	No. of Steps	QM Energy S_0	QM Energy S_1	Energy Diff.	ONIOM Energy
---	---	-449.85733	-449.77427	-0.08306	-448.99502
1	2317	-449.82313	-449.77125	-0.05188	
2	527	-449.82047	-449.77239	-0.04808	
3	202	-449.82032	-449.77237	-0.04795	
4	27	-449.82005	-449.77213	-0.04792	
5	1	-449.82005	-449.77213	-0.04792	-449.00267

Table 3. Energy values of the CoI optimization for the TMP(g) system including the projection in the microiteration steps. The number of microiteration cycle within the macrocycle and the number of steps of each microcycle are shown. Energies in hartee.

Microiterations cycle	No. of Steps	QM Energy S_0	QM Energy S_1	Energy Diff.	ONIOM Energy
---	---	-449.85733	-449.77427	-0.08306	-448.99502
1	1279	-449.81464	-449.7679	-0.04674	
2	1155	-449.8112	-449.76787	-0.04333	
3	504	-449.81198	-449.76907	-0.04291	
4	297	-449.81172	-449.76884	-0.04288	
5	2	-449.81166	-449.76878	-0.04288	
6	1	-449.81165	-449.76877	-0.04288	-448.9993

The first step in both tables is the same as it is just the energy of the starting point geometry. If we analyze the progress of the MECI optimization in the two cases, we can see that the energy difference (Energy Diff.) decreases during the microiterations process. The final Energy Diff. value obtained in the case where the projected gradient is used in the microiterations is 0.005 Hartrees lower than in the other case, without using it. Although the fact of including the projected gradient did not produce large changes in the optimized values of the first macroiteration step, there is an improvement over the current version. At the same time, the absolute energy (ONIOM energy) obtained with the unmodified gradient is approximately 0.003 hartree lower than the result of the optimization with the projected gradient. This is the prize to pay for retaining the energy degeneracy, *i.e.* the full optimization of the excited state energy is not possible. The environment also plays an important role in the optimization process. In order to properly test the effect of the implementation in the CoI optimization process, it would be necessary to analyze a larger number of optimization steps. We also note that the fact that the projection has only a small effect is due to the fact that the MM part (the sugar backbone) is small, and we expect a much more relevant effect when the solvent is included in the MM part. Therefore, the implementation also must be tested using the TMP(s) system. For that reason, the next goal is to find and solve the bugs of the other Gaussian subroutines to test the efficacy of the implementation in the CoI optimization inside the ONIOM scheme.

To analyze further the effect of the projection, we have calculated the angle between the gradient before the projection and the gradient difference and derivative coupling vectors (x_1 and x_2), respectively. The more these angles differ from 90º, the more important will be the effect of the projection. These angles were plotted during the first microiterations optimization cycle to see their behavior during the process (Fig. 11).

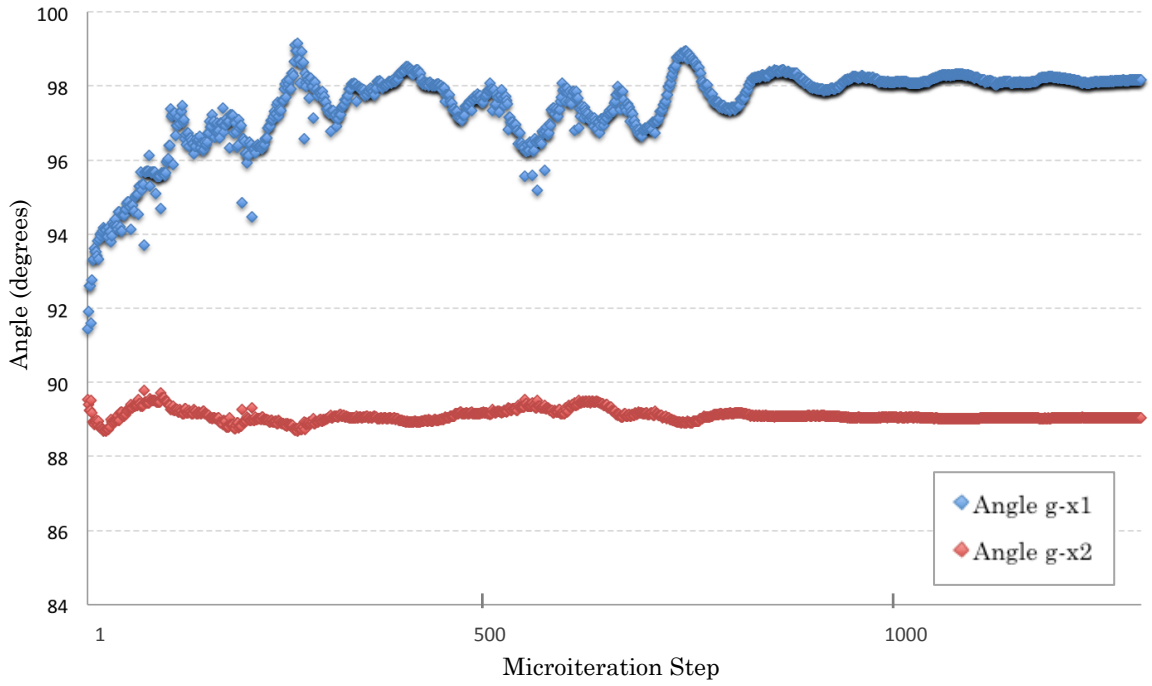


Figure 11. Change of angle values between the g before the projection and x_1 and the g before the projection and x_2 during the first microiterations cycle. The angle values are in degrees.

In general we can see that the angle value in both cases is close to 90º during the microiterations cycle. The fact that unprojected gradient is almost orthogonal to the x_1 and x_2 vectors is consistent with the low effect of the projection on the energy difference. However for x_1 there is a significant deviation from 90º that explains why the projection helps to reduce the energy difference, ant the inclusion of the projected gradient results in an improvement of the CoI optimization. As it is said, the implementation has only been tested for the TMP(g) system, which has a small environment. The real benefits of the implementation will be assessed in the future for larger systems, where the environment have a great effect.

5. CONCLUSIONS AND OUTLOOK

The potential energy surface for the Thy, TMP(g) and TMP(s) models has been characterized by optimizing the structures of S_0-Min, S_1-Min and S_1/S_0-CoI. The three structures have been optimized for Thy at the CASSCF level and TMP(g) at the CASSCF:UFF level, while for TMP(g) and TMP(s) at the CASSCF:AMBER and CASSCF:UFF levels, respectively, we have only optimized S_0-Min and S_1-Min. Overall, the minima and intersections obtained with different methods on the different models are similar. Therefore, we conclude that the TMP(g) and TMP(s) systems are suitable test systems for the implementation, as they show similar photophysical features compared to the Thy reference.

A Fortran subroutine for the calculation of the composed gradient was programmed and it has served as a trial to understand the functioning of the Gaussian subroutines.

The gradient projection step was implemented in the microiterations optimization part of the ONIOM scheme for the MECI location. This allows the optimization of the CoI taking into account the electrostatic contribution of the low level part. We have carried out a first analysis of the implementation on TMP(g). The implementation works correctly, and the projection has a positive effect on the optimization as it helps to reach a lower energy difference between the intersecting states. The small contribution of the projection in this case is due to the reduced size of the environment part. In the future it will be desirable to study the performance of the implementation on the more demanding TMP(s) system, where we also expect a larger effect of the surroundings on the energy difference.

6. BIBLIOGRAPHY

1. Suppan, P., *Chemistry and light*, The Royal Society of Chemistry, Cambridge, **1994**.
2. Zimmermann, J.; Zeug, A.; Röder, B. *Phys. Chem. Chem. Phys.*, **2003**, *5*, 2964–2969.
3. Kang, H.; Lee, T. K.; Jung, B.; Ko, Y. J.; Kim, S. K. *J. Am. Chem. Soc.* **2002**, *124*, 12958-12959.
4. Bernardi, F.; Olivucci, M.; Robb, M. A. *Chem. Soc. Rev.* **1996**, *25*, 321–328.
5. Improta, R.; Santoro, F.; Blancafort, L. *Chem. Rev.* **2016**, *116*, 3540–3593.
6. Dapprich, S.; Komáromi, I.; Byun, K. S.; Morokuma, K.; Firsch, M. J. *J. Mol. Struct. (THEOCHEM)* **1999**, *461–462*, 1–21.
7. Ruiz-Barragan, S.; Morokuma, K.; Blancafort, L. *J. Chem. Theory Comput.* **2015**, *11*, 1585–1594.
8. Frisch, M. J.; *et al. Gaussian 09*, *Rev. A.02*, Gaussian, Inc., Wallingford CT, 2009.
9. Frisch, M. J.; *et al. Gaussian Development version*, *Rev. H.13*, Gaussian, Inc., Wallingford CT, 2010.
10. Jensen, F., *Introduction to computational chemistry*, John Wiley & Sons, Ltd., Chichester, **2007**.
11. *CASSCF, GAUSSIAN 09 USER'S REFERENCE. **2009-2016,** Gaussian, Inc.* [visited: 22/05/2016] http:// http://www.gaussian.com/g_tech/g_ur/k_casscf.htm
12. Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, Jr., K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. *J. Am. Chem. Soc.* **1995**, *117*, 5179-5197.
13. Vreven, T.; Byun, K. S.; Komáromi, I.; Dapprich, S.; Montgomery, Jr., J. A.; Morokuma, K.; Firsch, M. J. *J. Chem. Theory Comput.* **2005**, *2*, 815–826.
14. Bearpark, M. J.; Robb, M. A.; Schlegel, H. B. *Chem. Phys. Lett.* **1994**, *223*, 269–274.
15. Ruiz-Barragan, S.; Robb, M. A.; Blancafort, L. *J. Chem. Theory Comput.* **2013**, *9*, 1433–1442.
16. Asturiol, D.; Lasorne, B.; Robb, M. A.; Blancafort, L. *J. Phys. Chem. A* **2009**, *113*, 10211-10218.