

Unveiling the mechanism of bipyrimidine photoproduct formation in thio substituted DNA.

Eva Vos Estebana

Máster en Química Teórica
y Modelización Computacional



MÁSTERES
DE LA UAM
2017 - 2018

Facultad de Ciencias

Master Erasmus Mundus in Theoretical Chemistry and Computational Modelling

Unveiling the mechanism of bipyrimidine photoproduct formation in thiosubstituted DNA.

Eva Vos Esteban



Directores: Inés Corral Pérez y Jesús González Vázquez
Universidad Autónoma de Madrid, Facultad de Ciencias,
Departamento de Química.

Master Thesis. Course 2017-2018

UNIVERSIDAD AUTÓNOMA DE MADRID

DEPARTAMENTO DE QUÍMICA

**Unveiling the mechanism of
bipyrimidine photoproduct
formation in thiosubstituted DNA.**

Eva Vos Esteban

Supervised by:

Inés Corral Pérez and Jesús González Vázquez

Madrid, July 2018

Contents

Acknowledgement	i
Abstract	iii
Resumen	v
1 Introduction	1
2 Theoretical Methods	9
2.1 Wave-function based methods	12
2.1.1 Hartree-Fock Theory	12
2.1.2 Post Hartree-Fock methods	16
2.2 Density Functional based methods: Density Functional Theory	22
2.2.1 Local Density Methods	26
2.2.2 General Gradient Approximation	26
2.2.3 Hybrid Methods	27
2.2.4 Density funtional corrections	27
2.3 Basis Sets	29
3 Results	33
3.1 Singlet 64-PP Mechanism	37
3.2 Triplet 64-PP Mechanism	45
4 Conclusions	57
References	66

Acknowledgement

I should start by thanking the entities that have made possible this work. In first place, the master program, European Master in Theoretical Chemistry and Computational Modeling (EMTCCM) coordinated with much dedication and effort by the professor Manuel Yáñez Montero. Also, to the Erasmus Mundus program of the European Union that together with the Ministerio de Educación Cultura y Deporte have financed my Master and PhD studies. I also would like to thank Prof. Manuel Alcamí Pertejo who agreed to supervise this Research Project. I also want to thank the Centro de Computación Científica de la Universidad Autónoma de Madrid for all the calculation time necessary for this work.



I am really grateful to Martial Boggio-Pascua who kindly received me as a one more integrant of his group in the University of Paul Sabatier in Toulouse.

In a more personal way, I would like to thank to professor Otilia Mó and Manuel Yáñez to take me under their protective wings and welcome me to their enormous family that is the Chemistry Department. The tireless work that Wilson and Viviana carry out to deal with the arduous bureaucratic issues that get ahead the Department.

Of course, this work would not have been possible with the unconditional support of my supervisors Inés Corral and Jesús González who each day have to cope my independent and quiet character. Jesús' optimism always helps to see the problems from a better perspective. The constant work that Inés does, including "the" articles missing in my english redaction is priceless. Thank you so much, to be there, unfortunately for you but very luckily for my, we have 3 more years ahead.

All this amazing and hard path, would not be the same without extraordinary people such as Rubén with his adorable pedantry, Iñigo and his every morning bad mood, Michael the most unlucky latine person in France and even I-ting with her odd insanity. Dario

has become more than a simple friend, he has been the main support in these 3 amazing months in Toulouse that I will never forget as *The Room*, a movie that has changed my life.

I cannot forget my family, but a special mention goes to my cousin Ana or as I call her "my personal writing assistant". Who despite her low english level and her reduced knowledge on chemistry, reads everything that I write, checking my redaction and trying to understand the scientific purpose behind it. She really does not know, how much I appreciate her effort and compromise with everything that I do. Muchas gracias Anita, por todo y añadiré una palabra para que la incorpores a tu cuadernito de inglés staunch.

¡Muchas gracias!

Abstract

The simple substitution of an oxygen of a carbonyl group by a sulphur atom in nucleobases can dramatically alter their photophysical and photochemical properties as compared to natural nucleobases. For this reason, drugs based on thionucleobases have been used for years, as immunosuppressants and anti-inflammatories. An unavoidable consequence of the treatment with this kind of drugs, is their accumulation in DNA. Although, these compounds are not considered intrinsically dangerous for DNA, their combination with non-lethal doses of ultraviolet radiation produces long-lived excited states, which could interact with environmental molecules, such as molecular oxygen or adjacent canonical nucleobases, leading to other species that may pose a threat to DNA and other biomolecules. In this way, several studies have concluded that there exists an indisputable relationship between thiobases prescription and the development of certain types of cancer, especially skin cancer.

The aim of the present work is to offer a general overview on the mechanism behind dimerization processes in thiosubstituted DNA upon UVA light absorption, being the two main photoproducts cyclobutane pyrimidine dimer and pyrimidine-(6,4)-pyrimidone (64-PP). The theoretical study has been focused on a deep analysis of the mechanism of 64-PP formation. For the modelling of this photoproduct, a dinucleoside monophosphate including a thymine and 4-thiothymine has been used.

With the purpose of clarifying the mechanism for the dimerization reactions and determining the potential implication of the excited states, not only singlet but also triplet multiplicity pathways have been described using the time dependent formalism of DFT and the multiconfigurational methods, CASSCF and CASPT2.

A detailed understanding at molecular level of the mechanism leading to DNA damage or to the generation of cytotoxic species is crucial for the improvement or design of new drugs for photochemotherapeutic applications.

Resumen

Durante décadas, los medicamentos basados en tionucleobases han sido utilizados por sus propiedades antiinflamatorias e inmunodepresoras. La sustitución de un grupo carbonilo por un tiocarbonilo en la molécula provoca un cambio drástico en sus propiedades fotofísicas y fotoquímicas, justificando que estos compuestos hayan adquirido un gran protagonismo, en los últimos tiempos, por su potencial uso como agentes en fotoquimioterapia.

Tas la absorción de luz, estas tiobases, ya se encuentren en disolución o bien incorporadas en la hebra de ADN, son capaces de generar estados excitados de vida largas, que pueden interaccionar con moléculas del entorno, tales como oxígeno molecular o nucleobases naturales adyacentes en la hebra del ADN. En ambos casos, los fotoproductos finalmente generados provocan efectos citotóxicos en las células del organismo.

En este trabajo, se pretende determinar un posible mecanismo para una de las dos fotolesiones más relevantes registradas hasta el momento en la molécula de ADN, el 6,4 fotoproducto (64-PP). Esta lesión, junto con el dímero ciclobutano, son el resultado de la exposición del ADN tiosustituido a la luz ultravioleta-A, por lo que se ha llevado a cabo un detallado análisis de los mecanismos no sólo que implican los estados excitados singlete, sino también tripletes para el mecanismo de formación del dímero 64-PP. Para llevar a cabo el estudio de los perfiles de energía potencial de dicha reacción se ha recurrido al método DFT y su formalismo dependiente del tiempo, TD-DFT, así como los métodos multiconfiguracionales, CASSCF y CASPT2, utilizando el modelo molecular, dinucleósido monofosfatado de timina-4-tiotimina.

La presente investigación a nivel molecular pretende contribuir al esclarecimiento de la generación de especies citotóxicas debido a la combinación de la luz con las tiobases, procurando en último término colaborar en el desarrollo del diseño de nuevos fármacos con aplicaciones fotoquimioterapéuticas.

1. Introduction

For almost five decades, thiosubstituted DNA nucleobases, like ones show in Figure 1.1, also known as thionucleobases, have been prescribed as immunosuppressant, anti-inflammatory and also anticancer drugs [1]. Their powerful pharmacological effectiveness is the result of a carbonyl by a thiocarbonyl group substitution in the structure of the natural bases. This apparently simple modification provokes surprising changes in the photophysical and photochemical properties of the canonical nucleobases, altering not only its ground state potential energy surface, but also their excited states [2].

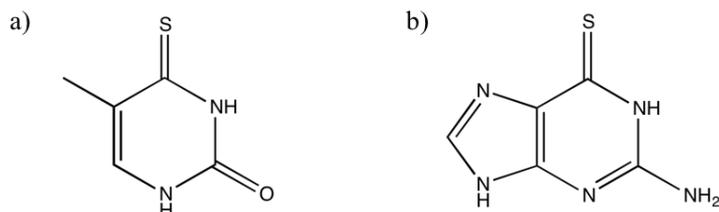


Figure 1.1: Molecular structure of a thiopyrimidine, 4-Thiothymine (a) and a thiopurine, 6-Thioguanine (b).

Drugs based on thiobases are classified as prodrugs [3], as the final products of their metabolisms consist of nucleic acid precursors, that will be ultimately incorporated into the genetic material of cells. In this way, a long treatment with this kind of drugs leads, inevitably, to the accumulation of thiobases in DNA strands [4].

As already mentioned, the efficiency of the thionucleobase drugs is related to the sulphur presence in the structure of the nucleobases. However, the enhanced chemical reactivity of these species can trigger undesirable secondary effects, as the generation of unwanted side reactions which could cause important damages in cell biomolecules [4]. In fact, several studies have investigated the correlation between the long term use of thiobases and the risk of developing various malignancies [4], concluding that there exists an indisputable relationship between thiobases prescription and the development of certain types of cancer such as myeloid leukemia or non-Hodgkin lymphoma, being par-

ticularly high the probability of developing skin cancer in these patients when exposed to light [4]. Although, the immunosuppressant effect of these drugs has been suggested as the causative factor in the development of these malignancies, the incorporation of the thionucleobases in the patients' DNA, seems to be the real responsible of the occurrence of these chromosomal disorders [4]. Actually, the International Agency for Research on Cancer has classified thiobases as a human carcinogens [4]. Contrary to what one might think, these compounds are not intrinsically responsible for their cytotoxic effect [5], but its combination with non-lethal doses of ultraviolet radiation produces a harmful impact in the organism.

Unlike canonical nucleobases, which present a maximum of absorption in the ultraviolet C region of the electromagnetic spectrum (100 - 260 nm), sulphur incorporation into these biomolecules, provokes an important bathochromic shift in their absorption properties [6], as it can be seen in Figure 1.2.

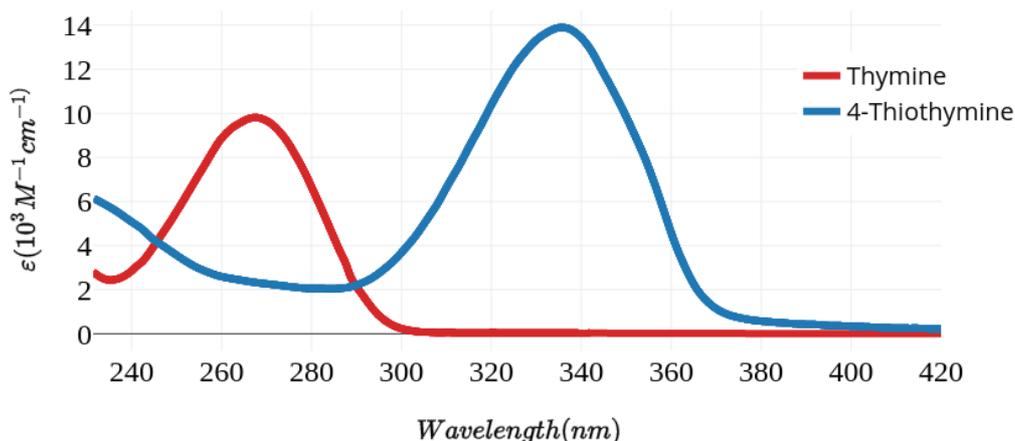


Figure 1.2: Absorption spectrum of canonical nucleobase, thymine, and their sulphur analogue, 4-thiothymine. (Adapted from Ref.[4])

Due to the shift, caused by the substitution of an oxygen for a sulphur, thiobases are considered UV-A chromophores. This effect can be explained attending to the values of the ionization potential of each element. While oxygen presents an ionization potential of 13.6 eV, sulphur ionization potential has a value of 10.4 eV, requiring this latter a lower energy to induce electron promotion. As it is well known, UV-A radiation constitutes around 90% of the total ultraviolet incident solar light [6]. Continuous exposure of natural DNA, to these wavelengths, does not mean a serious problem, since natural nucleobases hardly absorb in this region. However, the incorporation of thionucleobases in DNA strands results in light absorption and hence in the photoactivation of the sulphur

substituted biomolecule. In addition to the differences in the optical properties, canonical and thionucleobases also considerably differ in the topography of their excited states hypersurfaces. Comparing the intrinsic deactivation mechanism of both systems, it is more than evident that they present very different photophysics. Figure 1.3 shows schematic potential energy profiles for both families of systems.

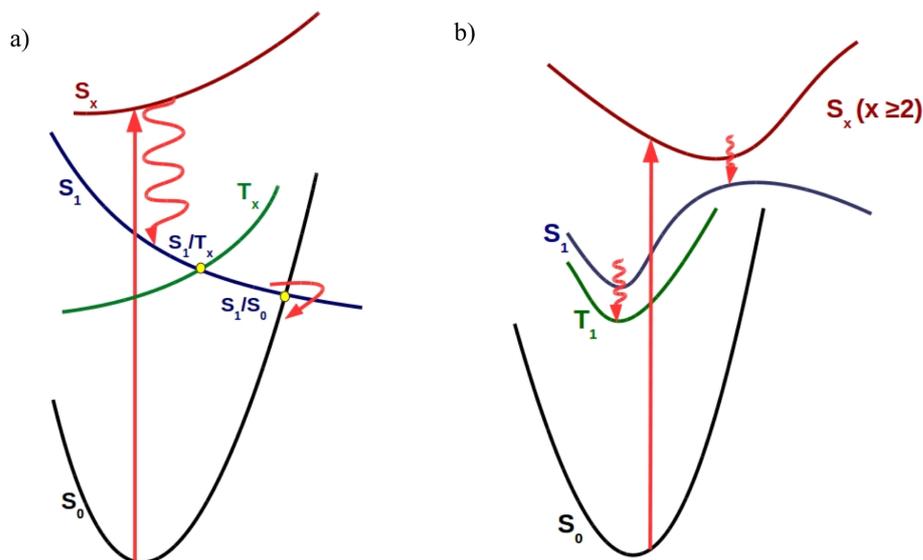


Figure 1.3: Scheme of the potential energy profiles of canonical nucleobases (a) and thiosubstituted nucleobases (b).

The topography of the potential energy profile of canonical nucleobases, Figure 1.3 a, is characterized by a pronounced descending slopes. Once the system is excited, a fast deactivation from the excited to the ground state takes place. This characteristic profile is inherent to the natural DNA and RNA nucleobases, and allows the system to electronically relax in an ultrafast time-scale, avoiding potential reactions occurring in the excited manifold. This phenomenon leads the compounds to their original state and confers them photostable properties [7]. A really different scenario is observed for thiobases, instead. The corresponding excited states present several minima, in the singlet and also in the triplet manifold, where the system can be trapped once light absorption occurs. Another important consequence of sulphur substitution, is due to the heavy atom effect. Introducing sulphur atoms in the molecule translates into an increment of the population transfer between singlet and triplet states, due to large values of spin orbit coupling. This feature, as well as the characteristic topography of the excited states, enhance intersystem crossing (ISC) between singlet and triplet states in thiobases, leading to the population of the triplet manifold [8]. Contrary to natural nucleobases, which are consider photostable systems, thionucleobases, as a photoactivated compounds, have been proposed as potential photosensitizers. The photosensitizers are compounds which after energy absorption are able to originate long-lived excited states. The excess of energy can

be lost by interaction with environmental molecules, through charge transfer or Type I reactions and energy transfer or Type II mechanisms [9].

Type I mechanisms involve electron transfer or proton abstraction, Scheme 1. Initially, the photosensitizer in its ground state, (^1PS), suffers a photo absorption and triplet excited species are generated, (^3PS). These unstable compounds would eventually interact with an environmental molecule, transferring an electron and producing free radicals, which could react with molecular oxygen to form reactive oxygen species (ROS), such as superoxide anion (O_2^-) or hydroxyl radicals ($\cdot\text{OH}$). [9]



Scheme 1: General scheme for Type I reactions of photosensitizers.

In contrast, in Type II reactions an energy transfer takes place between the excited photosensitizer and the molecular oxygen in its fundamental state ($^3\text{O}_2$), in this way, singlet oxygen ($^1\text{O}_2$) is generated as shown in Scheme 2.



Scheme 2: General scheme for Type II reactions of photosensitizers.

The resulting products in both scenarios are able to provoke serious damage in DNA and other biomolecules, which eventually concludes in cell death [10]. Although, ROS are generated along the usual oxidative metabolism, the organism has several potent mechanisms to block these dangerous species. However, even when exposed to low UV-A doses, thionucleobases, incorporated in DNA strands, are potential centers of production of ROS. Thus, the risk of producing oxidative DNA damage is particularly high in sulphur substituted DNA.

All the side effects inherent to these prodrugs can be profited for the design of new photochemotherapeutic agents. Photochemotherapy is a medical procedure which combines UV radiation and photoactive drugs in order to produce a cytotoxic effects in malign cells. As it has already been explained, despite neither UV-A radiation nor thionucleobases drugs, are able to induce a harmful effect on their own, the combination of both generates ROS, among other cytotoxic species, through Type I reactions. Therefore, the combination of localized radiation with an appropriate wavelength and a photosensitizer, can cause the death of specific cells, becoming a really powerful technique for the treatment of malignancies such as cancer [11]. Owing to the red-shifted absorption spectrum of thionucleobases, Figure 1.2, the use of this kind of these prodrugs in photochemotherapy relies in the use of UV-A light. Since this radiation belongs to the low energy region of the UV electromagnetic spectrum, it is able to penetrate deeper into the tissues, facil-

itating its photochemotherapeutic applications [2]. This is the reason why this kind of systems are the focus of several recent studies that try to clarify their particular reactivity, mechanism and also the possible secondary effects motivated by their use.

Besides being precursors of $^1\text{O}_2$ and ROS upon UV-A exposure, thiobases, incorporated in DNA strand, can interact with neighboring natural nucleobases, as a consequence of the generation of long-lived excited states. [12]

Several experimental and theoretical studies have determined the UV light exposure as the direct responsible for the structural damage in natural DNA [13], being the most common photoproducts, pyrimidine dimers, Figure 1.4. Photoexcited pyrimidines can react with adjacent pyrimidines nucleobases of the same strand in their ground state, leading to the formation of cyclobutane pyrimidine dimers (CPD), Figure 1.4a, and pyrimidine-(6-4)-pyrimidone photoproduct (64-PP), Figure 1.4b, which are the most common photolesions in DNA [14].

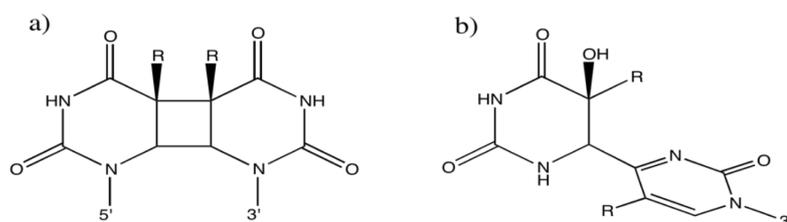
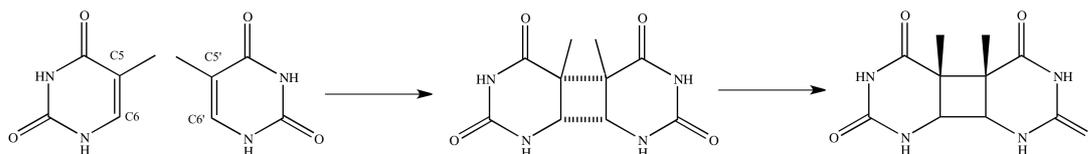


Figure 1.4: Molecular structure of the two main DNA photoproducts after UV light exposure, cyclobutane pyrimidine dimer (a) and (6-4) pyrimidine-pyrimidone adduct (b).

These covalent interactions between pairs of nucleobases are themselves a serious structural damage, but they are also associated with the inhibition of relevant enzymes involved in DNA replication and transcription [15]. Thus, photoinduced dimers are considered mutagenic agents and are correlated with development of cancer [13, 16, 17].

In order to understand in detail the formation of each photolesion, it results completely necessary to carry out an exhaustive exploration of the PES of simple DNA models relevant to the formation of CPD and 64-PP dimers. Despite the fact that upon UV direct irradiation of DNA, CPD is the most abundant photoproduct, followed by the 64-PP [18, 19], it is surprising that their formation mechanisms continue to be controversial. Several factors affect the generation of these dimers, among them it could be cited the relative configurational orientation of the nucleobases in the DNA strand [20], the kind of nucleobases involved [21] or the irradiation wavelength used [18]. Depending on these factors, the distribution of DNA bipyrimidine photoproducts can be significantly different.

Cyclobutane pyrimidine dimer reaction is the result of [2+2] cycloaddition between adjacent pyrimidines, such as thymine and cytosine, where the C-C double bonds of each nucleobases interact to generate a cyclobutane ring, as it can observe in Scheme 5.



Scheme 3: Reaction scheme for CPD formation.

Due to the high activation energy that this reaction requires, it cannot be thermally initiated, being the photochemical path the one energetically feasible [22]. According to theoretical results [23], this reaction is driven by a concerted mechanism, since the formation of the 2 newly $C - C$ bonds occurs simultaneously. The formation of this particular dimer in natural DNA, upon UV-B light absorption, has been suggested to be an ultrafast process [24] occurring via barrierless mechanism along the singlet excited state (S_1) and mediated via a conical intersection (CI) with the ground state (S_0), $(S_1/S_0)_{CI}$, Figure 1.5 [23, 25].

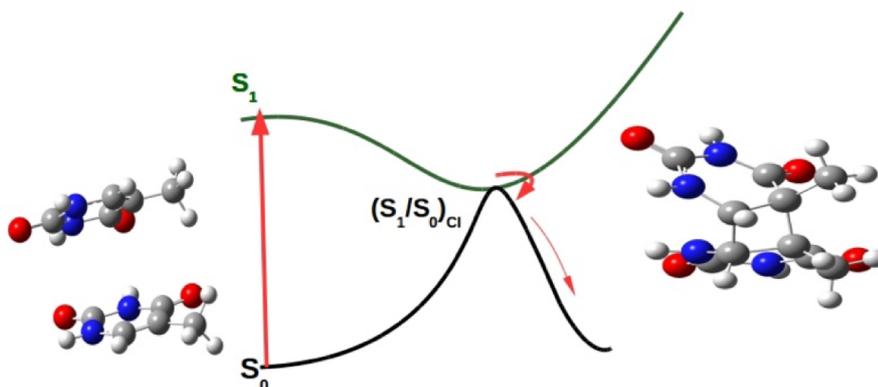
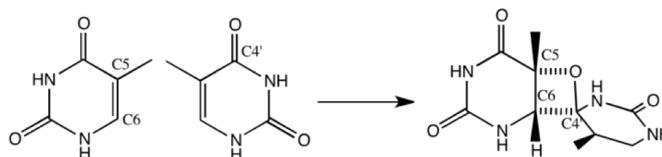


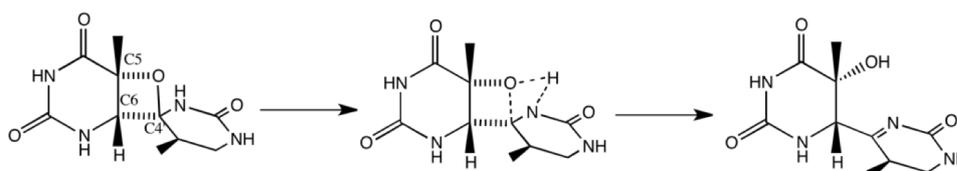
Figure 1.5: Potential energy profile for the formation of the cyclobutane dimer along singlet manifold. Adapted from Ref.[15]

Although the quantum yield for 64-PP is registered one order of magnitude lower than that of CPD [14, 18, 21], it is the second DNA photoproduct more common after the direct exposure to UV radiation. As CPD, the 64-PP has been also related with the cellular mutation and lethality, being considered its mutagenic effect even more harmful than that of CPD [26]. Experimental [19] and theoretical studies [27] have determined that 64-PP is formed via an oxetane intermediate. Oxetane is the result of a [2+2] cycloaddition between the carbonyl group of an excited pyrimidine and the double $C - C$ bond of the neighboring one in this ground state, as it can see in Scheme 4. The above chemical reaction between a carbonyl compound and alkene is known as the Paternÿ-Büchi reaction [28, 29], which can take place either in the triplet and single excited state [30]. Once the oxetane intermediate has been formed, a thermal ring opening reaction occurs spontaneously, probably as a consequence of the inherent ring strain and the loss



Scheme 4: Reaction scheme of the first step in the complete 64-PP reaction. The formation of the oxetane intermediate.

of aromatic stabilization [31], to give the final photoproduct, 64-PP, Scheme 5. The



Scheme 5: Reaction scheme of 64-PP formation.

mechanism behind photoinduced 6,4 adduct formation has not been totally clarified. In fact, very little is known about it. Despite a singlet excited pathway was accepted at the beginning for the description of the 64-PP mechanism [12, 25, 26], nowadays the role of the triplet excited states in the generation of this type of DNA lesion, is still under debate, Figure 1.6 [32, 33].

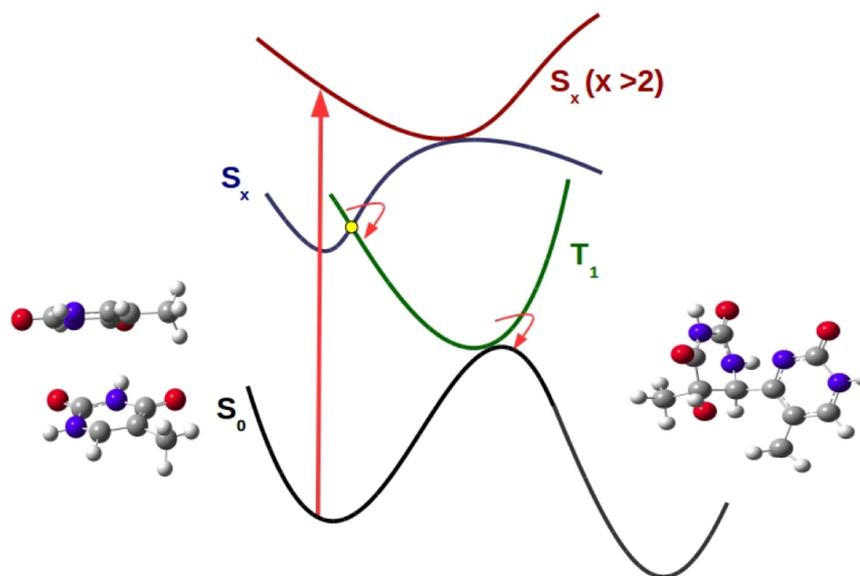


Figure 1.6: Potential energy profile of 64-PP formation reaction. Adapted from Ref.[33]

The initial step, after light absorption, would consist in the population of the single excited states. From here, the deactivation is initialized and the system might decay to lower energy states of the same multiplicity through internal conversions or conical intersection. Along the relaxation pathway, the population of the singlet states can be transferred to

the triplet manifold by suitable intersystem crossing (ISC) processes $(S_x/T_x)_{ISC}$. Following this path, the lowest-lying triplet state (T_1) would increase its population [34], from where the 64-PP dimerization could start [27]. The high energetic barriers, that this particular pathway presents, make, however, this reaction path rather improbable. The most feasible route involves a new ISC with the ground state (S_0), along which the oxetane intermediate is generated and which presents no other energy barriers until 64-PP formation [33]. In contrast, recent studies [35] have been concluded that the contribution of the triplet states to the [2+2] cycloaddition, implied in the formation of the intermediate, is less than 10%, indicating that the mechanism of oxetane formation takes place via singlet excited states along a barrierless path, as CPD generation.

Considering that the previously described photolesions are triggered by UV-B light absorption and that this radiation just constitutes 5% of the total ultraviolet solar light reaching the Earth surface, it can be really interesting to study the influence of UV-A light in the generation of these dimers. While CPD formation has been registered after UV-A irradiation, in the absence of UV-B, no evidences of 64-PP have been found [18]. This scenario suggests that the formation of CPD, under UV-A conditions, differs from the above path initiated with UV-B. In this way, some studies have tried to clarify the mechanism by which CPDs are formed by direct exposure of DNA to UV-A light. In fact, recent studies have determined that double-stranded DNA structure increases the capacity of DNA bases to absorb UVA photons [36]. Alternatively, it has been suggested that a photoactivated sensitizer, would generate long-lived excited triplet state, which could transfer the excess of energy to DNA pyrimidines leading to CPDs [37].

The use of the thionucleobases incorporated in DNA strands can be proposed as a model to understand and clarify the role of the triplet excited states in the photochemistry of natural nucleobases, as sulphur incorporation to the DNA structure leads to similar dimers as the photolesions occurring in the canonical bases. Moreover, the heavy atom effect help to elucidate the potential deactivation paths along excited states. Understanding the mechanism behind thiodimers formation pretends to build a clearer picture of the dimerization reaction in natural DNA, providing simultaneously insight into the lethal effect that incorporating thionucleobases in natural DNA has in cells.

2. Theoretical Methods

Unlike every single element within our visual field which is governed under classical mechanics laws, atoms are constituted by extremely small subparticles such as protons, neutrons and electrons, whose behavior is determined by quantum mechanics rules. In this way, it results fundamental to have a precise idea about the principles of this awesome and complicate scene, in order to apply it in chemical problems.

In quantum mechanics the state of a system is described by a function, known as wave function (Ψ), which is expressed in terms of coordinates (\bar{x}) and time (t). The wave function owns all the possible information of the system, allowing to determine several of its properties. Nevertheless, due to the particular statistical nature of quantum mechanics, it is completely impossible to determine exactly several observables simultaneously, in concordance with the Heisenberg uncertainty principle, being the concept of probability the only predictable knowledge in quantum mechanics. In 1926, a revolutionary discovery, made by Erwin Schrödinger [38], gave place to a complete quantum theory, concluding in Time-Dependent Schrödinger Equation,

$$\frac{\partial\Psi(\bar{x}, t)}{\partial t} = -\frac{i}{\hbar}\hat{H}\Psi(\bar{x}, t) \quad (2.1)$$

Equation (2.1) describes how the wave function changes dynamically in time and offers the possibility of predicting its future behavior and properties.

In this context, the application of quantum mechanics principles with the objective of describing the nature and properties of chemical systems, such as atoms or molecules, opened the field of Quantum Chemistry. Although the Time-Dependent Schrödinger equation provides an evolving description of the system with time, if the Hamiltonian is not changing with time, the Equation (2.1) can be simplified, upon a phase change, leading to the Time-Independent Schrödinger Equation, (2.2)

$$\hat{H}\Psi = E\Psi \quad (2.2)$$

where, \hat{H} , represents the Hamiltonian operator, which is related to the total energy of

the system, E . From an atomistic perspective and just considering electrostatic forces, the Hamiltonian can be defined in terms of positions of the nuclei (R) and electrons (r), as the sum of the nuclei and electron kinetic energy, \hat{T}_{nuclei} and \hat{T}_{elect} respectively, the term of interaction energy between nucleus and electrons, $\hat{V}_{nuclei-elect}$, which is determined by the Coulomb attraction, and two more terms related to the repulsion energy between electron-electron, $\hat{V}_{elect-elect}$ and the nuclei-nuclei, $\hat{V}_{nuclei-nuclei}$. In this way, the Hamiltonian is expressed, in atomic units, in Equation (2.3),

$$\begin{aligned}\hat{H} &= \hat{T}_{nuclei} + \hat{V}_{nuclei-nuclei} + \hat{T}_{elect} + \hat{V}_{nuclei-elect} + \hat{V}_{elect-elect} \\ &= -\frac{1}{2} \sum_{a=1}^M \frac{1}{m_a} \nabla_a^2 + \sum_{a=1}^M \sum_{b>a}^M \frac{Z_a Z_b}{R_{ab}} - \sum_{i=1}^N \frac{1}{2} \nabla_i^2 - \sum_{i=1}^N \sum_{a=1}^M \frac{Z_a}{r_{ia}} + \sum_{i=1}^N \sum_{j>i}^N \frac{1}{r_{ij}}\end{aligned}\quad (2.3)$$

where M and N are the numbers of nuclei and electrons, respectively, m_a the nuclear ratio of the mass with respect the electrons, and Z the atomic number.

The main aim of Quantum Chemistry is to solve the Schrödinger equation, (2.2), considering the Hamiltonian given in (2.3). One of the most difficult aspects that the resolution of this eigenvalue problem presents, is a consequence of the coupling between nuclei and electrons coordinates in the $\hat{V}_{nuclei-elect}$ term. Since nuclei are much heavier than electrons, a good approximation could be to consider that the electrons of the molecule are moving in a potential energy field, caused by the fixed nuclei. This approximation based on the size difference of the atomic elements is known as Born-Oppenheimer approximation, and allows the decoupling between nuclei and electron motion. Consequently, $\hat{V}_{nuclei-elect}$ will just depend parametrically on the nuclear positions and the Hamiltonian of Equation (2.3) can be, thus, divided into electronic and nuclear Hamiltonian.

$$\hat{H}_{total} = \hat{H}_{elect} + \hat{H}_{nuclear}\quad (2.4)$$

Just focusing on the electronic part of the Hamiltonian (2.4), the eigenvalue problem to solve is given by,

$$\hat{H}_{elect} \Psi_{elect} = E_{elect} \Psi_{elect}\quad (2.5)$$

where electronic energy, E_{elect} , is an eigenvalue of \hat{H}_{elect} . The total energy (E) will be the sum of the electronic energy E_{elect} and the potential $\hat{V}_{nuclei-nuclei}$, which now is a constant for a given nuclear configuration.

The electronic Schrödinger equation (??, even in the electronic case, can only be solved analytically for very simple cases, the mono-electronic systems, where the electron repulsion term, $\hat{V}_{elect-elect}$, can be removed. For this reason, Quantum Chemistry is completely focused on developing approximated methods that allow to find the most accurate solution to this problem. In order to solve the Equation (2.5), several procedures can be used, that are usually classified as methods based on the wave function optimization,

i.e. Hartree-Fock and post Hartree-Fock methods and those which are focused on the optimization of the electronic density, Density Function Theory (DFT).

Since the principal objective of Quantum Chemistry is to solve the Schrödinger equation, (2.2), it is necessary to understand the concept of the two main approaches to solve eigenvalue problems in quantum mechanics.

In the following section, it will present the fundamentals of Variational Principle and Perturbation Theory, which are two powerful tools to find accurate solutions for the Schrödinger equation.

Variational Principle

Variational principle demonstrates that it is always possible to find an approximate wave function whose energy will be always above or equal to the exact energy. Giving a normalized wave function, the expected value of the Hamiltonian is an upper bound to the exact ground state energy,

$$\langle \Psi | \hat{H} | \Psi \rangle \geq E_0 \quad (2.6)$$

In this formalism, the energy can be considered an indication of the accuracy, and the lower the energy, the better the Ψ . Therefore, the condition $\langle \Psi_1 | \hat{H} | \Psi_1 \rangle \leq \langle \Psi_2 | \hat{H} | \Psi_2 \rangle$, confirms that Ψ_1 is a better wave function than Ψ_2 . In consequence, if Ψ_1 and Ψ_2 depend on a set of parameters, it is always possible to say that the best set of values for the parameters will be those which minimize the value of the energy.

Perturbation Theory

The central aspect of Perturbation theory is the possibility of considering the exact and unknown Hamiltonian (\hat{H}) as a sum of a principal Hamiltonian (\hat{H}_0), whose eigenfunctions and eigenvalues are known, and a perturbation (\hat{H}') which minimally contributes to \hat{H} . Assuming that the perturbation is gradually applied by a λ factor, the total Hamiltonian can be expressed as,

$$\hat{H} = \hat{H}_0 + \lambda \hat{H}' \quad (2.7)$$

Introducing the modified Hamiltonian, (2.7) in Schrödinger equation (2.2),

$$(\hat{H}_0 + \lambda \hat{H}') \Psi_i = E_i \Psi_i \quad (2.8)$$

At this point, the goal will be to relate the unknown eigenfunctions and eigenvalues of the perturbed system (2.7) with the familiar ones of \hat{H}_0 . In this way, it is expected that the eigenfunctions and energies of the \hat{H}_0 improve and become closer and closer to those of \hat{H} . According to Equation (2.7), \hat{H} depends on λ and therefore, their eigenfunctions and eigenvalues will do also. For this reason, both the energy and Ψ can be expanded in

a Taylor series with respect λ ,

$$E_i = E_i^{(0)} + \lambda E_i^{(1)} + \lambda^2 E_i^{(2)} + \dots \quad (2.9)$$

$$\Psi_i = \Psi_i^{(0)} + \lambda \Psi_i^{(1)} + \lambda^2 \Psi_i^{(2)} + \dots \quad (2.10)$$

Substituting these equations in the expression (2.8), the Schrödinger equation is transformed,

$$(\hat{H}_0 + \lambda \hat{H}')(\Psi_i^{(0)} + \lambda \Psi_i^{(1)} + \lambda^2 \Psi_i^{(2)} + \dots) = (E_i^{(0)} + \lambda E_i^{(1)} + \lambda^2 E_i^{(2)} + \dots)(\Psi_i^{(0)} + \lambda \Psi_i^{(1)} + \lambda^2 \Psi_i^{(2)} + \dots) \quad (2.11)$$

Then, it is possible to group the terms with identical power of λ and obtain the following expressions in the base of the order of perturbation.

$$\hat{H}_0 \Psi_i^{(0)} = E_i^{(0)} \Psi_i^{(0)} \quad (2.12)$$

$$\hat{H}_0 \Psi_i^{(1)} + \hat{H}' \Psi_i^{(0)} = E_i^{(0)} \Psi_i^{(1)} + E_i^{(1)} \Psi_i^{(0)} \quad (2.13)$$

$$\hat{H}_0 \Psi_i^{(2)} + \hat{H}' \Psi_i^{(1)} = E_i^{(0)} \Psi_i^{(2)} + E_i^{(1)} \Psi_i^{(1)} + E_i^{(2)} \Psi_i^{(0)} \quad (2.14)$$

From these equations, the energy and function of any order can be calculated, but considering that the perturbation is small, the development of the first terms would be enough to achieve a good energy and wave function.

The above formalism, called Rayleigh-Schrödinger Perturbation Theory [38, 39], can be applied for the theoretical study of atoms and molecules. The application of this approximation in chemical systems was introduced by Møller and Plesset in 1934 [40] and it is known as Møller-Plesset perturbation theory (MPPT).

2.1 Wave-function based methods

2.1.1 Hartree-Fock Theory

The intuitive concept in which the electrons are occupying a specific region of the space, attempts to simplify the complex N -electron problem and replace it by an one-electron problem, considering the electron-electron interaction in an average way. Even though, this simple picture is very extended in the chemistry community, it is necessary to be aware that it is, in fact, an approximation. The Hartree-Fock approximation plays an important role in the quantum study of chemical systems, not only setting the basis of *ab initio* calculations but also being a great starting point for many other methods, which offer a more accurate description of the problems [41, 42].

As it has been mentioned before, electronic Schrödinger equation, (2.5) cannot be analytically solved for polyelectronic systems. One of the very first steps in the modeling of a particular system will be to obtain an approximate wave function for these complex

systems.

In theory, any arbitrary wave function can be expressed as a linear combination of spatial mono-electronic functions $\phi_i(\bar{r})$. However, in order to describe exactly the wave function, an infinite set of mono-electronic functions should be employed, what results totally impossible in practice. In addition to the spatial component, the spin part, α or β , must be considered to obtain a good description of the desired poly-electronic wave functions. In this way, the combination of the spatial distribution of the electron and its characteristic spin is called spin orbital.

$$\chi_i(\bar{x}) = \phi_i(\bar{r})\alpha(\omega) \quad \text{or} \quad \chi_i(\bar{x}) = \phi_i(\bar{r})\beta(\omega) \quad (2.15)$$

Each spin orbital describes specifically one single electron, thus the product of different spin orbitals could be taken as a great starting point for the poly-electronic function. The wave function of a poly-electronic systems can be expressed through a Slater determinant, which for an N -electron system present the generalized form of the Equation (2.16),

$$\Phi(\bar{x}_1, \bar{x}_2, \dots, \bar{x}_N) = (N!)^{-\frac{1}{2}} \begin{vmatrix} \chi_i(\bar{x}_1) & \chi_j(\bar{x}_1) & \dots & \chi_k(\bar{x}_1) \\ \chi_i(\bar{x}_2) & \chi_j(\bar{x}_2) & \dots & \chi_k(\bar{x}_2) \\ \vdots & \vdots & & \vdots \\ \chi_i(\bar{x}_N) & \chi_j(\bar{x}_N) & \dots & \chi_k(\bar{x}_N) \end{vmatrix} \quad (2.16)$$

where N is the total number of electrons in the system and $(N!)^{-\frac{1}{2}}$ is a normalization factor. In this determinant form, (2.16), the exchange between two electron coordinates implies an exchange between two rows, which leads to a change in the sign of the determinant, satisfying the antisymmetric principle. On the other hand, Pauli exclusion principle is kept, as two fermions occupying the same spin orbital, translates into two equal rows and thus the value of the determinant would be zero. Consequently no more than one electron can occupy a spin orbital. Usually, with the objective to simplify the notation of the Slater determinant, only the diagonal elements are indicated,

$$|\Phi(\bar{x}_1, \bar{x}_2, \dots, \bar{x}_N)\rangle = |\chi_i(\bar{x}_1)\chi_j(\bar{x}_2)\dots\chi_k(\bar{x}_N)\rangle \quad (2.17)$$

In the present context, one of the most important approximations of the Hartree-Fock Theory is the use of a single Slater determinant to describe the ground state of an N -electron system,

$$\Psi_0 \approx |\Phi(\bar{x}_1, \bar{x}_2, \dots, \bar{x}_N)\rangle \quad (2.18)$$

An other relevant aspect of this theory is the approximate electronic Hamiltonian used, which does not explicitly consider the electron-electron interaction. However, it assumes that each particle is subject to a mean field created by the other electrons. Thus, the N -electron problem is treated in an average way. In this approximation, the one electron

Hamiltonian is defined by Fock operator,

$$\hat{F}(i) = \hat{h}(i) + \nu_{HF}(i) \quad (2.19)$$

The first term of Equation (2.19) is a electronic core-Hamiltonian, $\hat{h}(i)$, which includes the electron kinetic operator and the electron-nuclei attraction,

$$\hat{h}(i) = -\frac{1}{2}\nabla_i^2 - \sum_M \frac{Z_a}{r_{ia}} \quad (2.20)$$

The remaining term in Equation (2.19), called Hartree-Fock potential ($\nu_{HF}(i)$), incorporates the effect of the "electron interactions",

$$\nu_{HF}(i) = \sum_{i \neq j} (J_j(i) - K_j(i)) \quad (2.21)$$

where the first component is the Coulomb operator $\hat{J}(i)$, which represents the repulsion energy that electron i suffers by all the rest electrons j , and the second element is the exchange term $\hat{K}(i)$, which is a consequence of the antisymmetric nature of the wave function. Substituting the Fock operator in the Schrödinger equation (2.2), the Hartree-Fock equation for a single electron is obtained.

$$\hat{F}\Psi_i = E_i\Psi_i \quad (2.22)$$

It is important to notice that \hat{F} is an one electron operator, whose eigenfunctions are the spin orbitals and its eigenvalues are the energies of these particular spin orbitals, so that instead of solving a coupled N particle equation, it is only necessary to solve N equations of single particle, being computationally much cheaper. Since the electron-electron contribution, $\nu_{HF}(i)$, is included for each particle, the total energy of a N -electron system is not the sum for each spin orbital. It results necessary to remove half of these terms to avoid their double inclusion in the total energy. Thus the Hartree-Fock energy is expressed as,

$$E_{HF} = \sum_i E_i - \frac{1}{2} \sum_i^N \sum_j^N \langle \Psi_i | \hat{J}_j - \hat{K}_j | \Psi_i \rangle = \sum_i E_i - \frac{1}{2} \sum_i^N \sum_j^N \langle ij | ij \rangle \quad (2.23)$$

In order to solve the Hartree-Fock equation, (2.22), it is important to take into account the dependence of the Fock operator on the spin orbitals, as the operator depends on the solutions. For this reason, the Hartree-Fock equation must be solved in an iterative way. Starting with an initial guess of spin orbitals, the Hartree-Fock potential is calculated. Once the Fock operator is known, the energy can be determined as the eigenvalue of \hat{F} and a new better set of spin orbitals is generated as eigenfunctions. Again, with the new set of orbitals a new value of $\nu_{HF}(i)$ is obtained and the process continues until achieving

an insignificant variation in the energy. This procedure is known as Self-Consistent Field method (SCF).

The above explained Hartree-Fock formalism becomes more complex for molecules, for which it does not exist a practical procedure for their numerical solution. An interesting transformation over the Hartree-Fock equation is to expand the spin orbitals in a linear combination of atomic orbitals ψ_p ,

$$\Phi_i = \sum_{p=1} c_{pi} \psi_p \quad (2.24)$$

where c_{pi} are the corresponding coefficients for each atomic orbital. In this way, the problem of calculating molecular orbitals in the Hartree-Fock equations is simplified to a set coefficient calculation by matrix techniques [43, 44].

Attending now to the spin function, two different types of spin orbitals can be defined. When the same spatial function can be used for the two spin functions the Restricted Hartree-Fock method is invoked [43, 44]. On the contrary, if each spin part has different spatial functions, the proper method would be the Unrestricted Hartree-Fock [45]. The treatment of each method is different and the use of one of them depends on the problem to solve.

Limitations of Hartree-Fock Theory

Regarding every commented aspect of the Hartree-Fock Theory, it is apparently clear the importance of choosing an appropriate basis set for the initial point of SCF. The use of a complete or an infinite basis set would offer the exact E_{HF} , however the use of an infinite basis set is totally impracticable. Thus, it is expected that the more extended basis set is, the lower Hartree-Fock energy will be obtained, due to the fact that the wave function will be better described. According to variational principle, lower energy are translated into better wave functions and, hence, the accuracy of the calculation can be systematically improved. This situation is, however, not always observed within the Hartree-Fock framework. In fact, the energy achieves an upper limit, despite increasing the size of the basis, this value of the energy is known as the Hartree-Fock energy limit.

In addition, one of the most significant approximations of the Hartree-Fock Theory is the way in which the electronic interaction is treated. It is hoped that not including strictly the electron-electron repulsion but consider it in an average way, might suppose a great deficiency of the method. Thus, it can define the difference between the non-relativistic exact energy (E_{exact}) of the system and the calculated Hartree-Fock energy limit as electron correlation,

$$E_{corr} = E_{exact} - E_{HF} \quad (2.25)$$

In order to overcome the limitations of the Hartree-Fock Theory, post Hartree-Fock

methods are used. These methods are able to introduce electronic correlation, providing a more realistic picture of the problem.

2.1.2 Post Hartree-Fock methods

Post Hartree-Fock methods are a collection of procedures which pretend to improve the Hartree-Fock scenario. Although Hartree-Fock wave function is able to account for up to 99% of the total energy, usually for the accurate description of a chemical phenomena the remaining percentage results essential. For this reason, every single electron correlated method will be focused on introducing, in one way or another, this part of the total energy.

Configuration Interaction

The Hartree-Fock method determines the best one determinant wave function and thus a qualitatively correct description of the one electron systems. Nevertheless, several problems in chemistry cannot be just defined by a single determinant. If more than one electronic configuration has a significant importance in the system, Hartree-Fock method would not provide an accurate solution. Since Hartree-Fock method offers the best wave function using a single Slater determinant, a reasonable starting point to improve it, will be the incorporation of more Slater determinants to describe the wave function. The Configuration Interaction method (CI) is based on the variational principle. In this method, the wave function can be expressed as a linear combination of all the determinants which are constructed from the promotion of electrons from occupied to empty orbitals,

$$\Psi_{CI} = c_0\Psi_0 + \sum_{ir} c_i^r\Phi_i^r + \sum_{\substack{i<j \\ r<s}} c_{ij}^{rs}\Phi_{ij}^{rs} + \sum_{\substack{i<j<k \\ r<s<t}} c_{ijk}^{rst}\Phi_{ijk}^{rst} + \dots \quad (2.26)$$

where i , j and k represent doubly occupied spin orbitals, while r , s and t empty ones. Then, Equation (2.26) describes all the possible single, double, triplet and so on excitations from occupied to virtual orbitals. Equation (2.26) is known as *full CI* and recovers the total electron correlation, within the frame established by the limitations of the use of a finite basis set of functions and also the initial Born-Oppenheimer approximation. The aim of the CI method is to optimize the coefficients of each electronic excitation (c_i) in Equation (2.26) by using the variation principle, calculating the wave function which minimizes the energy of the system. However, for medium to large systems the complete incorporation of all the possible excitations is an immeasurable amount of determinants. In these cases, the number of determinants has to be reduced in order to develop a tractable model. The truncated version of this method is called *Truncated CI*, which depends on the number of excitations included, in this way different levels for this method can be defined. Unlike the *full CI* method, which is size consistent, when the truncated procedure is applied, the method loses this relevant property and some corrections are used to

overcome the size inconsistent problem [46]. The use of Coupled Cluster methods would be a reasonable way to avoid the size inconsistent problem generated by the truncation of *full CI*, however they are non variational methods.

Multiconfigurational Self Consistent Field

Multiconfigurational Self Consistent Field methods or MCSCF can be considered as a derivative of the CI method, and hence an extension of the monoconfigurational approaches. The MCSCF wave functions are constructed by a set of Slater determinants Φ , Equation (2.16), depicting different electronic configurations i ,

$$\Psi_{MCSCF} = \sum_i c_i |\Phi_i\rangle \quad (2.27)$$

where c_i are the expansion coefficients for each Slater determinant. More precisely, the MSCF method, not only optimizes the CI expansion coefficients, but also the spin orbital functions, which have been used to construct the determinants.

The mathematical development of MCSCF method can be performed using the second quantization formalism. Second quantization allows to treat N -electron problems, transferring the property of antisymmetric principle from the Slater determinant formalism to algebraic operators. This elegant mathematical way to describe the N -electron wave function, it is based on the creation (a_i^\dagger) and annihilation operators (a_i), which can act on arbitrary Slater determinant, adding and removing, respectively, an electron from the i spin orbital.

$$a_i^\dagger |\chi_k \dots \chi_l\rangle = (-1)^p |\chi_i \chi_k \dots \chi_l\rangle \quad (2.28)$$

$$a_i |\chi_i \chi_k \dots \chi_l\rangle = (-1)^p |\chi_k \dots \chi_l\rangle \quad (2.29)$$

In this way, the independent spin Hamiltonian can be expressed in terms of these operators as,

$$\hat{H} = \sum_{ij} \langle i | \hat{h} | j \rangle a_i^\dagger a_j + \frac{1}{2} \sum_{ijkl} \langle ij | kl \rangle a_i^\dagger a_k^\dagger a_l a_j \quad (2.30)$$

where first term is related with one-electron integrals, whereas the other component defines the bi-electron ones. Considering the excitation operators (\hat{E}_{ij}) as,

$$\hat{E}_{ij} = \sum_{s=\alpha,\beta} a_{is}^\dagger a_{js} = a_{i\alpha}^\dagger a_{j\alpha} + a_{i\beta}^\dagger a_{j\beta} \quad (2.31)$$

Equation (2.30) can be transformed into,

$$\hat{H} = \sum_{ij} \hat{h}_{ij} \hat{E}_{ij} + \frac{1}{2} \sum_{ijkl} g_{ijkl} (\hat{E}_{ij} \hat{E}_{kl} - \delta_{jk} \hat{E}_{il}) \quad (2.32)$$

whose \hat{h}_{ij} term includes the kinetic energy of electrons and the attraction nucleus-electron,

whereas g_{ijkl} represents the repulsion integrals of molecular orbitals i, j, k and l . Solving Schrödinger equation, (2.5), with the Hamiltonian of Equation (2.32) in terms of the MCSCF wave functions (2.27), the energy of MCSCF is,

$$\begin{aligned}
 E = \langle \Psi | \hat{H} | \Psi \rangle &= \sum_{ij} \hat{h}_{ij} \sum_{mn} c_m^* \langle \Phi_m | \hat{E}_{ij} | \Phi_n \rangle c_n + \frac{1}{2} \sum_{ijkl} g_{ijkl} c_m^* \langle \Phi_m | \hat{E}_{ij} \hat{E}_{kl} - \delta_{jk} \hat{E}_{il} | \Phi_n \rangle c_n \\
 &= \sum_{ij} \hat{h}_{ij} \sum_{mn} c_m^* D_{ij}^{mn} c_n + \sum_{ijkl} g_{ijkl} \sum_{mn} c_m^* P_{ijkl}^{mn} c_n \\
 &= \sum_{ij} \hat{h}_{ij} \mathbf{D}_{ij} + \sum_{ijkl} g_{ijkl} \mathbf{P}_{ijkl}
 \end{aligned}
 \tag{2.33}$$

\mathbf{D}_{ij} and \mathbf{P}_{ijkl} are called first and second order density matrices, respectively and contain the coefficients of MCSCF wave function. In addition, all information related to the orbitals is contained in the mono and bi electronic integrals, \hat{h}_{ij} and g_{ijkl} , respectively. The optimization of MCSCF coefficients and orbitals is not an easy procedure, therefore there are several techniques to carry out these optimizations. First order methods are based on the calculation of the energy and its first derivative, the gradient, with respect to the parameters, the orbitals and their coefficients. Second order MCSCF methods consider the gradient and also the hessian or the second derivatives of the energy. Newton-Raphson method [47] is probably the second order method more popular, whose energy is expanded in a Taylor series in terms of the parameters. MCSCF is an iterative method, as SCF and its convergence and also the computational cost of the calculations depends closely on the number of configurations included in the MCSCF wave function. Similarly to the *full CI*, MCSCF method is limited by very large numbers of configurations. Probably, the major problem in MCSCF methods is to properly select the configurations, which are necessary to describe the problem at hand.

Nowadays, one of the most popular MCSCF methods is Complete Active Space SCF (CASSCF) [48–51]. In the CASSCF method, the spin orbitals can be divided into three main different sets, inactive, active and virtual orbitals. Inactive orbitals contain doubly occupied orbitals. In contrast, virtual space consist of empty orbitals, with a zero occupation value. Active space is integrated by orbitals with an intermediate occupations, between 0 and 2, being the most relevant subspace in this method. While the inactive orbitals are treated at Hartree-Fock level of theory, a *full CI* is performed for active space as Figure (2.1) depicts. Thus, within the active space, all the possible excitations involving the active electrons and orbitals are considered in order to built the CASSCF wave function. Although there is not a systematic procedure to choose the correct active space, the selection of these orbitals is essential and determines the accuracy of the results. As a general rule, the active space should include the orbital whose occupation varies significantly of 2 and 0, that way orbitals with occupation values close to 2 and 0

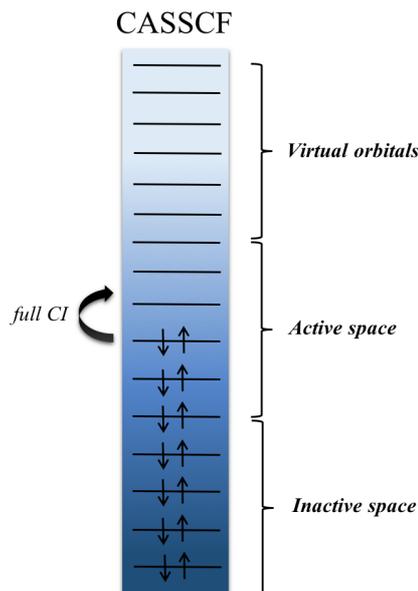


Figure 2.1: Representation of active space for CASSCF method.

must be treated as a virtual orbitals. The energetic highest occupied orbitals and lowest empty ones, calculated at Hartree-Fock level, are usually taking part into the active space. Essentially, the orbitals which are important for describing the chemical problem should be included in this relevant region, so chemical intuition plays an important role in the CASSCF method. In spite of the limitation of the number of active space orbitals, the *full CI* calculation performed in the frame of CASSCF might become extremely large even for small active spaces.

A very extended and interesting version of the CASSCF method, is State Average CASSCF method (SA-CASSCF). Until now, all the problems have been focused on the description of the ground state. However, in description of photoinduced excited state problems, it is expected that electronic excited states acquire an essential role. In these situations, the energy function to be optimized is expressed as the average of all the considered states energies,

$$E_{average} = \sum_i \omega_i E_i \quad (2.34)$$

where ω_i is the relative weight of each state i . The final result of SA-CASSCF is a unique set of average spin orbitals, which is optimized for the number of state included in the calculation. Although, these average orbitals cannot perfectly describe every single state separately, this procedure allows to solve the "root flipping", consisting in the exchange between states, which is a common problem in single state CASSCF calculations.

As already highlighted, the selection of an adequate active space is the key of CASSCF methods, therefore it cannot be considered a black box method like Hartree-Fock.

Multiconfigurational methods allow to describe the wave function in terms of several different configurations. Moreover, they offer the possibility that the orbitals to become partially occupied instead of forcing their occupation, adding an indispensable flexibility at wave function of the system. In this way, CASSCF method is mainly introduced for recovering part of the total electron correlation energy, which is denominated static or non-dynamical correlation. However, an important part of the electron correlation, related with the electron motion is not considered by MCSCF approaches. This remaining electron correlation, known as dynamical correlation, must be necessarily included in order to obtain accurate energies. One of the most common way to introduce this energy, is thought a second order perturbation treatment using CASSCF wave function as a reference, given place to Complete Active Space Second order Perturbation Theory (CASPT2) method.

Complete Active Space Second Order Perturbation Theory

Complete Active Space Second Order Perturbation Theory (CASPT2) method [52, 53] is framed within Perturbation Theory. Therefore, the Hamiltonian to solve, is divided in an unperturbed Hamiltonian (H_0) and a small perturbation (H'). In standard Møller-Plesset Theory, the zeroth order Hamiltonian is determined by the sum of Fock operator over each orbital for close shell systems, and electron correlation is considered the perturbation. The CASPT2 method is equivalent to Møller-Plesset method, but in this particular case, the function of reference is multiconfigurational, concretely CASSCF wave functions. In this way, the Fock operator has to be formulated to operate over them,

$$\hat{F} = \sum_{ij} \hat{f}_{ij} \hat{E}_{ij} = \sum_{ij} \left(\hat{h}_{ij} + \sum_{kl} D_{kl} \left[\langle ij|kl \rangle - \frac{1}{2} \langle il|kj \rangle \right] \right) \hat{E}_{ij} \quad (2.35)$$

As a consequence, the zeroth order Hamiltonian H_0 can be expressed using the projection, $P^i = |\Psi_i\rangle\langle\Psi_i|$, of the four subspaces in which the CASSCF space is divided, over the Fock operator. Particularly, these four subspaces are defined by the expanded space for the state reference function, Ψ_i^{CASSCF} , the orthogonal space restricted to *full CI* subspace of the CASSCF method Ψ^k , the space correlated with single and double excitations, Ψ^{SD} , and finally the rest of space constitutes by higher order excitations, Ψ^X , these two latter over Ψ^{CASSCF} . Therefore, H_0 presents the following form,

$$\hat{H}_0 = P^{CASSCF} \hat{F} P^{CASSCF} + P^k \hat{F} P^k + P^{SD} \hat{F} P^{SD} + P^X \hat{F} P^X \quad (2.36)$$

In the Rayleigh-Schrödinger perturbation theory formalism, the functions, which are implied in the first order correction to the reference wave function, compound the first order interacting space. Due to the fact that in CASPT2 only single and double excited configurations interact with the CASSCF wave function, these functions will form the first

order interacting space, and thus these excitations are the uniques included in the first order correction wave function.

$$\Psi^{(1)} = \sum_{ijkl} c_{ijkl} \left(\hat{E}_{ij} \hat{E}_{kl} | \Psi_{CASSCF} \rangle \right) \quad (2.37)$$

With that wave function, the second order correction to the energy can be calculated according to the expression,

$$E_i^{(2)} = \langle \Psi^{(0)} | \hat{H}' | \Psi^{(1)} \rangle = \sum_i \frac{|\langle \Psi_0^{(0)} | \hat{H}' | \Psi_i^{(0)} \rangle|^2}{E_i^0 - E_0^{(0)}} \quad (2.38)$$

Finally, the CASPT2 wave function can be written as,

$$\Psi_{CASPT2} = c_0 | \Psi_{CASSCF} \rangle + c_1 | \Psi^{(1)} \rangle \quad (2.39)$$

where c_0 and c_1 are called weight coefficients, and inform about the contribution of the reference and the perturbation wave function over Ψ_{CASPT2} . The reference weight, c_0 , is commonly used as a evaluation criterion of the proposed perturbation theory. Since the perturbation might contribute minimally to the total wave function, the weight of the reference should be close to one. However, in certain situations, a state of the first order interacting space can be really close in energy with the reference, causing that the state of reference is degenerated. When this fact occurs, the coefficient of the perturbations, c_1 , presents values considerably high and, hence, the c_0 will be low, which suggests the presence of intruder states in the active space of CASPT2 calculation. Intruder states are characterized by having an eigenvalue for the zeroth order Hamiltonian (H^0) similar to the zeroth order of the reference function, considering the second order energy of Equation (2.38), the degeneration causes that the denominator tends to zero. In these situations, perturbation treatment is not valid anymore. The best solution to this problem, it is to expand the active space and include this intruder state in it. This option, however, is not always possible and it is necessary to increase the difference between the eigenvalues. A small shift (ε) can be added at the zeroth order Hamiltonian [54–56].

$$\hat{H}^0 + \varepsilon P^{SD} \quad (2.40)$$

The second order corrected energy solutions of Equation (2.40) is obtained,

$$\tilde{E}_i^{(2)} = - \sum_{i=1} \frac{|\langle \Psi_0^{(0)} | \hat{H}' | \Psi_i^{(0)} \rangle|^2}{E_i^0 - E_0^0 + \varepsilon} \quad (2.41)$$

From Equation (2.41), it can be concluded that the level shift would avoid the vanishing of the denominator, allowing the perturbation treatment for these intruder states. However, the second order energy will depend on ε .

The CASPT2 method is not suitable for the description of the chemical problems when the CASSCF wave function is not a proper reference state. In these situations, it is necessary to use an extension of CASPT2 method, known as Multi State CAPT2 (MS-CASPT2) [57]. The MS-CASPT2 procedure takes in account the coupling via electron correlation, between several CASSCF wave functions of different states. In this context, an effective Hamiltonian (\hat{H}_{eff}) is built, considering the CASSCF energy (E_i) and the interaction between $\Psi^{(1)}$ and Ψ^{CASSCF} for all included states i ,

$$(\hat{H}_{eff})_{ij} = \delta_{ij}E_i + \langle (\Psi_{CASSCF})_i | \hat{H} | \Psi_j^{(1)} \rangle \quad (2.42)$$

The solution of the Schrödinger equation including the above \hat{H}_{eff} , results in the MS-CASPT2 energies and also the MS-CASPT2 wave function, which can be defined as,

$$\Psi_i^{MS-CASPT2} = \sum_p c_{ip} |\Psi_{CASSCF}\rangle + \Psi_i^{(1)} \quad (2.43)$$

Additionally, the mathematical formulation of CASPT2 tends to underestimate the energy of chemical process in which the number of paired electron changes, such as dissociation or excitation energies. The IPEA shift [58] enables to correct this error, caused by an underestimation of the energy of open shell systems, including a shift parameter in the zeroth order Hamiltonian.

2.2 Density Functional based methods: Density Functional Theory

The methods based on the optimization of the electron density ($\rho[\bar{r}]$) are an alternative way to solve the electronic Schrödinger equation, (2.5), besides the conventional methods explained before based on the optimization of the wave function (Ψ). Unlike Ψ , which depends on $4N$ coordinates for a N -electron system, $\rho[\bar{r}]$ is just expressed in terms of four coordinates, three spatial coordinates, (\bar{x}) and an additional one related to the spin (ω). For this reason, while the complexity of wave function increases with the number of electrons, the electron density depends on the same number of variables, independently of the system size, becoming these methods an effective tool for the study from medium to large size molecular systems.

From a wave function of a N -electron system, $\Psi(\bar{x}_1, \bar{x}_2, \dots, \bar{x}_N)$, the probability to find electron 1 , in a particular region of the space, between x_1 and $x_1 + dx_1$, is defined by the square of Ψ . In a similar way, and since electrons are indistinguishable particles, the density function ($\rho[\bar{x}_1]$) is determined by the probability to find electron 1 independently of the positions of the rest N -electron.

$$\rho(\bar{x}_1)d\bar{x}_1 = N d\bar{x}_1 \int \Psi(\bar{x}_1, \bar{x}_2, \dots, \bar{x}_N) \Psi^*(\bar{x}_1, \bar{x}_2, \dots, \bar{x}_N) d\bar{x}_2 \dots d\bar{x}_N \quad (2.44)$$

Integrating Equation (2.44) with respect the spin coordinate, $d\omega_1$, the electron density ($\rho[\bar{r}_1]$) is obtained,

$$\rho(\bar{r}_1) = \rho(\bar{x}_1)d\omega_n = N \int \Psi(\bar{x}_1, \bar{x}_2, \dots, \bar{x}_N)\Psi^*(\bar{x}_1, \bar{x}_2, \dots, \bar{x}_N)d\omega_1 d\bar{x}_2 \dots d\bar{x}_N \quad (2.45)$$

The basis of Density Functional Theory or DFT is the use of $\rho[\bar{r}]$, to calculate the electronic energy of the system, instead of Ψ . Hohenberg and Kohn [59] established the foundations of this theory, in which proposed that every single observable of the ground state can be written in terms of the electron density of this particular ground state. According to the first Hohenberg-Kohn theorem, the number of electrons in the system and also the electron-nuclear potential is determined by $\rho[\bar{r}]$. In this way, the Hamiltonian operator will be defined. Consequently, the wave function of the ground state and, hence, all the properties of the system, such as the ground state electronic energy would be completely specified by the electron density. Therefore, the first Hohenberg-Kohn theorem defends a direct correlation between the electronic density and the wave function,

$$\rho(\bar{r}) \Leftrightarrow \Psi \quad (2.46)$$

which makes possible to express the energy of the system as a functional of the electron density, $E[\rho]$.

$$E[\rho] = T[\rho] + V_{nuclei-elect}[\rho] + V_{elect-elect}[\rho] + V_{nuclei-nuclei}[\rho] \quad (2.47)$$

Due to the Born-Oppenheimer approximation, $V_{nuclei-nuclei}[\rho]$ is treated as a constant. In addition, external potential created by the positive charge of the nuclei, ν_{ext} , can be used to express $V_{nuclei-elect}[\rho]$,

$$V_{nuclei-elect}[\rho] = \int \nu_{ext}(\bar{r})\rho(\bar{r})d(\bar{r}) \quad (2.48)$$

Inserting this above expression in $E[\rho]$ transforms it into $E_\nu[\rho]$. As the terms of the kinetic energy, $T[\rho]$, and electron-electron interaction, $V_{elect-elect}[\rho]$, from Equation (2.47) do not depend on the external potential ($\nu_{ext}(\bar{r})$), they can be included in the functional $F_{HK}[\rho]$. Thus, Equation (2.47) can be rewritten as,

$$E_\nu[\rho] = \int \rho(\bar{r})\nu_{ext}(\bar{r})d(\bar{r}) + F_{HK}[\rho] \quad (2.49)$$

Similarly to the variational principle in the wave function formalisms, the second Hohenberg-Kohn theorem evidences that an approximate electron density always provides an energy, which is an upper bound to the real ground state one. Therefore, in order to obtain a

more accurate density for the ground state, $E_\nu[\rho]$ must be minimized.

$$\left[\frac{\delta E_\nu[\rho]}{\delta \rho} \right] = 0 \quad (2.50)$$

Introducing the energy from Equation (2.49) and considering the N -representability, the minimization of the energy functional with respect to the density is,

$$\left[\frac{\delta E_\nu[\rho]}{\delta \rho} \right] = \nu_{ext}(\bar{r}) + \frac{\delta F_{HK}[\rho]}{\delta \rho(\bar{r})} \quad (2.51)$$

Equation (2.51) allows minimizing the energy and, hence, determines the ground state density. However, as the expression which relates $F_{HK}[\rho]$ with electron density is unknown, it is not possible to use this variational procedure. Kohn and Sham [60] proposed a different method to calculate the energy of the system from the electron density. Assuming a non-interacting N -electron system over a constant external potential ($\nu_{KS}[\rho]$), the Hamiltonian will be,

$$\hat{H}_{KS} = -\frac{1}{2}\nabla^2(\bar{r}) + \nu_{KS}[\rho](\bar{r}) \quad (2.52)$$

$\nu_{KS}[\rho]$ is denominated the Kohn-Sham potential and generally pretends to be equal to the potential of the real system.

$$\nu_{KS}[\rho](\bar{r}) = \nu_{ext}(\bar{r}) + \nu_J[\rho](\bar{r}) + \nu_{XC}[\rho](\bar{r}) \quad (2.53)$$

where $\nu_{ext}(\bar{r})$ is the external field and $\nu_J[\rho](\bar{r})$ is the contribution of the electron-electron repulsion energy, defining by the Coulomb interaction. The remaining component, $\nu_{XC}[\rho](\bar{r})$, is the exchange-correlation potential. The solution of the corresponding Schrödinger equation for the Kohn-Sham hamiltonian, Equation (2.52),

$$\hat{H}_{KS}\phi_{KS} = E_{KS}\phi_{KS} \quad (2.54)$$

The orbitals presents in Equation (2.54), are called Kohn-Sham orbitals and they allow calculating the electron density immediately.

$$\rho(\bar{r}) = \sum_{i=1}^N \phi_i(KS)(\bar{r})\phi_i^*(KS)(\bar{r}) \quad (2.55)$$

As in Hartree-Fock method, the coefficients of the Kohn-Sham orbitals are optimized through an iterative procedure. Starting with a guess set of orbitals, the electron density is evaluated, and this density is used to calculate the $\nu_{KS}[\rho]$. With all the terms known, the Schrödinger equation, (2.54) can be solved, obtaining a new set of orbitals, which allows to recalculate the electron density until convergence is achieved.

Although, above demonstration is just valid for the non-degenerated ground states,

lately Levy extended these theorems for degenerated ground states [61].

The resolution of the Equation (2.54) provides the ground state electronic energy,

$$E[\rho] = T_S[\rho] + \int \rho(\bar{r})\nu_{ext}(\bar{r})d(\bar{r}) + \frac{1}{2} \int \int \frac{\rho(\bar{r}_1)\rho(\bar{r}_2)d(\bar{r}_1)d(\bar{r}_2)}{|\bar{r}_1 - \bar{r}_2|} + E_{XC}[\rho] \quad (2.56)$$

Although, the kinetic energy component, T_S , in Equation (2.52), is exact for a non-interacting electron system, the real one will be slightly different. Thus, the energy expression should include not only the kinetic energy term for the non-interacting systems ($T_S[\rho]$), but also the difference between this energy and the real one, ($T[\rho] - T_S[\rho]$). This factor is included in the exchange-correlation energy $E_{XC}[\rho]$.

According to their electronic configuration, systems can be classified in close and open shell, and their corresponding treatment will be different. Unrestricted KS-DFT (UKS-DFT) is applied fundamentally to open shell and excited state systems, which present an odd number of electrons when one or more electrons are unpaired. Unrestricted DFT allows to assign different spatial functions to the two spin functions, α and β . In this way, electrons of α spin can be described by a set of spatial functions and electrons of β spin by other completely different spatial orbitals. In this particular case, the interactions of one α electron are not only the result of the Coulomb and exchange-correlations interactions with all the other α electrons, but also the Coulomb interaction with the electrons of β spin. Therefore, both α and β spin present different electron densities and their equations must be evaluated separately.

$$\begin{aligned} E_\nu[\rho^\alpha] = & T_S[\rho^\alpha] + \int \rho^\alpha(\bar{r})\nu_n(\bar{r}) + \frac{1}{2} \int \int \frac{\rho^\alpha(\bar{r}_1)\rho^\alpha(\bar{r}_2)}{|\bar{r}_1 - \bar{r}_2|} d(\bar{r}_1)d(\bar{r}_2) \\ & + \frac{1}{2} \int \int \frac{\rho^\alpha(\bar{r}_1)\rho^\beta(\bar{r}_2)}{|\bar{r}_1 - \bar{r}_2|} d(\bar{r}_1)d(\bar{r}_2) + E_{XC}[\rho^\alpha, \rho^\beta] \end{aligned} \quad (2.57)$$

An equivalent equation could be also derived for β spin. In this case, the kinetic and the exchange energies depend exclusively on the Kohn-Sham orbitals with α spin, whereas the Coulomb repulsion and the correlation energy present a dependency on both spin orbitals. Therefore, the determination of the energy of each spin function is intrinsically related to the other set of electron with different spin. In conclusion, they cannot be solved independently and must be solved by a simultaneous iterative process.

All the terms of Equation (2.53) were well known with the exception of the $\nu_{XC}[\rho]$ component. This fact provokes that the Kohn-Sham method cannot offer an exact solution for the electronic problem. Unfortunately, $E_{XC}[\rho]$ must be approximated and it is usually divided into an exchange and a correlation part,

$$E_{XC}[\rho] = E_X[\rho] + E_C[\rho] \quad (2.58)$$

The corresponding exchange-correlation potential, $\nu_{XC}[\rho]$, is given as the derivative of

the $E_{XC}[\rho]$ with respect to the density. The closer the $\nu_{XC}[\rho]$ to the real one, the more accurate will be the electron density. Several exchange-correlation potentials have been developed, in fact the selection of the functionals which represent the exchange and correlation contribution is essential in DFT calculations.

2.2.1 Local Density Methods

The Local Density Approximation (LDA) considers the electron density locally, as in a uniform electron gas. According to this approximation, the electron density is finite and constant and the exchange-correlation functional only depends on it.

$$E_{XC}^{LDA}[\rho] = E_X^{LDA}[\rho] + E_C^{LDA}[\rho] \quad (2.59)$$

For the exchange functional, E_X^{LDA} , of the Equation (2.59), the Dirac formula for a model electron gas with a homogeneous density is used,

$$E_X^{LDA}[\rho] = -\frac{3}{4^{1/3}} 3\pi \int \rho^{4/3} d\bar{r} \quad (2.60)$$

Contrary, the correlation functional, E_C^{LDA} , has to be approximate by using for instance, Monte Carlo simulations from a uniform electron gas [62]. This method performs reasonably well for systems whose density is maintained almost constant through the whole structure, such as a periodic metallic surface. However, it does not provide a good description when the system has a highly non-uniform electron density, for instance in molecular systems. Despite the simplicity of the approximation, it provides surprisingly good results in geometry optimizations and vibrational frequencies. Nevertheless, it usually overestimates the bond energy, not being an appropriate method for systems with weak bonds.

The already mentioned UKS-DFT can be treated by Local Spin Density Approximation (LSDA), which applies different LDA functionals for α and β electronic spins.

2.2.2 General Gradient Approximation

Since LDA functionals incorporate the correlation-exchange effect in a local way, which is far from a realistic picture, a logical improvement is to introduce density gradients in the description of this effect. The General Gradient Approximation (GGA) takes into account not only the electron density in a certain point, but also its variation around this point.

$$E_X^{GGA}[\rho] = E_X^{LDA}[\rho] + E_X^{NLDA}[\rho] \quad (2.61)$$

where $E_X^{NLDA}[\rho]$ is defined as,

$$E_X^{NLDA}[\rho] = \int \rho^{4/3} f^{NLDA}(x) d\bar{r} \quad (2.62)$$

Depending on the function, f^{NLDA} , included in Equation (2.62), different exchange corrections are obtained. One of the most popular was proposed by Becke in 1988, B88, [63]. Attending to the correlation part, the function more used is the proposed by Perdew [64] and Lee, Yang and Parr [65], LYP, whose parameters were determined by fitting data for the helium atom. This latter correlation functional combined with the exchanged Becke, results in BLYP functional. The GGA functionals improve the calculation of geometries and frequencies compared to calculations using LDA functionals. However, they tend to underestimate chemical reaction barriers.

Beyond GGA functionals, meta-GGA functionals can be found. These functionals, additionally, introduce the second derivative of the density, i.e. its Laplacian (∇^2). The most relevant functionals from this group are Minnesota functionals of Truhlar and coworkers [66].

2.2.3 Hybrid Methods

These functionals combine a small fraction of the exact exchange from the Hartree-Fock solution, with other correlation-exchange fractions from semi-empirical or ab initio methods. The exchange energy can be determined using Kohn-Sham orbitals, which intrinsically define the electron density. In order to connect a Hartree-Fock calculation for non-interacting system, with that of a real system, the adiabatic connection must be used. Possibly, the most extended adiabatic method is the three parameters functional of Becke [67],

$$E_{XC} = E_{XC}^{LSDA} + a_0(E_X^0 - E_C^{LSDA}) + b\Delta E_X^{B88} + c\Delta E_C^{PW91} \quad (2.63)$$

which associated with the LYP correlation functional, result in the popular B3LYP functional with the parameters $a = 0.20$, $b = 0.72$ and $c = 0.81$ [63, 65].

Nowadays, hybrid functionals, which combine KS-DFT with wave function theory, are the most widely used functionals in quantum chemistry, due to their remarkable accuracy.

2.2.4 Density functional corrections

Although, DFT allows accounting for the electron correlation, it is unable to describe dispersive interactions, which are a pure long-range electron correlation effect. The accurate description of weak interactions, such as Van der Waals forces or $\pi - \pi$ interactions, is essential to describe current realistic chemical problems. A practical way to circumvent this deficiency in the methods, is to add an empirical potential to the DFT energy. Thus, the electronic energy can be considered as the sum of two energy parts,

$$E_{total} = E_{DFT} + E_{disp} \quad (2.64)$$

where E_{DFT} is the DFT energy and E_{disp} is an empirical dispersion correction given by

$$E_{disp} = -s_6 \sum_{a=1}^{M-1} \sum_{b=a+1}^M \frac{C_6^{ab}}{R_{ab}^6} f_{dmp}(R_{ab}) \quad (2.65)$$

where M is the numbers of atoms in the system, C_6^{ab} the dispersion coefficient for the atom pair ab , s_6 is a global scaling factor and R_{ab} the interatomic distance. Finally, f_{dmp} describes different damped functions which represent the interatomic potential. This approach to account for dispersion interaction is known as Grimme dispersion correction [68]. In addition, to the limitation that DFT offers to determine the dispersion interactions, other relevant effects are not well described by ordinary DFT functionals. In practical DFT calculations some problems have been found, such as the underestimation of the reaction energy barriers, the overstabilization of high-angular momentum orbitals or a poor polarizabilities of long chains [69]. As the mentioned problems in DFT calculations seems to be related to the description of the long-range electron-electron distance, long-range correction (LC) [70] pretends to improve the exchange functional for calculating this particular contribution. In this scheme, the electron repulsion operator can be divided into a short-range and a long-range part of the exchange interaction. While the local density approximation is employed for small electron-electron distance contributions, the Hartree-Fock exchange integrals are incorporate to describe long-range interelectronic distances. The use of LC allows overcoming the problems caused by the standard exchange functionals [69].

One of the biggest drawbacks of DFT methods, however, is that it does not exist a systematic way to improve the errors. In order to ensure the accuracy of a particular functional, the DFT results must be compared with experimental data or high level wave function methods, that are used as benchmark.

Time-Dependent Density Functional Theory

Although several interesting properties, such as molecular geometries or vibrational modes, can be inferred from the study of the ground state, other relevant features require a detailed knowledge of, not only the ground state, but also of the excited states of the system.

The basis of Time-Dependent Density Functional Theory or TD-DFT was proposed by Runge and Gross [71], who demonstrated that there exists an unique relationship between time-dependent densities and the changing external potentials with time. This theorem, analogue to that of Hohenber-Kohn, implies that any property of the system can be written as a functional of the time-dependent density [71]. Similarly to Kohn-Sham DFT, the initial external potential in TD-DFT corresponds to that of a non-interacting particle system. Thus, this potential is also called time-dependent Kohn-Sham potential, $\nu_{KS}[\rho, t]$, and has the same time-dependent density as the interacting system of interest.

Similarly to DFT, $\nu_{KS}[\rho, t]$ can be separated into different contributions,

$$\nu_{KS}(\bar{r}, t) = \nu_{ext}(\bar{r}, t) + \nu_J(\bar{r}) + \nu_{XC}(\bar{r}, t) \quad (2.66)$$

In this particular case, the external potential changes with time, therefore the evolution with time of the external potential variation ($\Delta\nu_{KS}$), is treated as a perturbation over the ground state KS orbitals,

$$\phi_i(\bar{r}, t) = e^{iE_0\Delta\nu_{KS}}\phi_i^{GS} \quad (2.67)$$

The perturbation can be directly propagated employing the Time-Dependent Schrödinger Equation, (2.1)

$$i\frac{\partial}{\partial t}\phi_i(\bar{r}, t) = -\frac{1}{2}\nabla^2\phi_i(\bar{r}, t) + \nu_{KS}(\bar{r}, t)\phi_i(\bar{r}, t) \quad (2.68)$$

Again, the lack of information of the exchange functional $\nu_{XC}[\rho, t]$ requires to approximate it. Consequently, it might be better to look for an alternative theory, as Time-Dependent Density Functional Response Theory [72].

As DFT, its time dependent version presents significant problems in some particular situations, i.e., the poor estimation of Rydberg and charge-transfer excitation energies and also oscillator strengths, employing conventional functionals. These inconveniences may be related to the insufficient exchange interaction between distant orbitals in the common exchange functionals rather than incorrect far-nucleus asymptotic behavior of them. The incorporation of the Long-range correction in TD-DFT calculations improves the accuracy of excitation energies and oscillator strengths [73].

2.3 Basis Sets

All calculations, without exception, require a basis set to expand spin orbitals functions in terms of a known set of functions, such as exponentials, Gaussians, polynomials or plane waves. For a complete basis, formed by an infinite number of functions, the expansion will not be an approximation. In practice, the use of an infinite basis set is impossible and resorting to a finite basis set is an inherent approximation in all *ab initio* methods. However, as expected, the more flexible the basis set used to expand the molecular orbitals, the better representation of them will be obtained. In this way, the increase of the basis set size is correlated with a higher accuracy. Besides the number of functions employed, their mathematical form is also a really important factor to achieve accurate results without an exceeding computational effort. In other words, the functions should be able to describe correctly the physics of the problem. In this respect, it is convenient that these functions tend to zero when the distance between the nuclei and electrons becomes infinite. Also important, the selected functions have to make it easy to calculate all the

required integrals. In chemistry, the two more common type of functions used are Salter Type Orbitals (STO) and Gaussian Type Orbitals (GTO). Slater type orbitals [74] take the form,

$$\chi_{\zeta,n,l,m}^{STO}(r, \theta, \psi) = NY_{l,m}(\theta, \psi)r^{n-1}e^{-\zeta r} \quad (2.69)$$

where N is the normalization constant, $Y_{l,m}$ are spherical harmonic functions, ζ is the nuclear effective charge and n , l and m are the principal, the angular momentum and magnetic quantum numbers, respectively. The exponential part, which depends on the nuclear-electron distance (r), represents exactly hydrogenlike orbitals, providing an accurate description of the behavior of the electrons at short and also at long distances from the nuclei. Although STOs ensure a quite rapidly convergence and offer a pretty good representation of the features of the molecular orbitals, an important disadvantage is found in the calculation of bi-electronic integrals. Despite these integrals can be analytically evaluated for diatomic molecules, for polyatomic systems only numerical solutions are possible, which translates into a great computational demand. For this reason, the use of Gaussian functions is widely extended.

In contrast to STOs, Gaussian type orbitals [75], with the form of Equation 2.71 allow calculating more efficiently two electron integrals. However, they provide a poorer description for electronic structure.

$$\chi_{\zeta,n,l,m}^{GTO}(r, \theta, \psi) = AY_{l,m}(\theta, \psi)r^{2n-2-l}e^{-\zeta r^2} \quad (2.70)$$

Due to the exponential part, $e^{-\zeta r^2}$, GTOs cannot correctly reproduce the electron density close to the nuclei. In addition, these functions decay off more quickly than STO, thus the electronic behavior at long distance is neither correctly described. Therefore, Gaussian functions are not the optimum basis functions, as its behavior rather differs to the molecular orbitals. In order to solve this problem, a linear combination of Gaussian functions, Equation (2.71) can be used to emulate a Slater type orbital, Equation (2.69). In this way, the basis will be adequate to describe the physics of the system and the simple evaluation of the bi-electronic integrals is preserved.

$$\chi^{CGTO} = \sum_i^k a_i \chi_i^{GTO} \quad (2.71)$$

where χ_i^{GTO} are primitive Gaussian type orbitals (PGTO) and the resulting functions χ^{CGTO} , contracted GTOs (CGTOs). The larger the number of PGTO in the expansion, the better fitting will be. Nevertheless, the best compromise between accuracy and computational economy is achieved by using 3 Gaussians to fit the Slater function behavior, as it shows Figure (2.2), this basis set is known as STO-3G. As already mentioned, the other important factor, which determines the accuracy of the results, is the flexibility of the basis set. The smallest possible basis set, known as *minimum basis set* [76], only

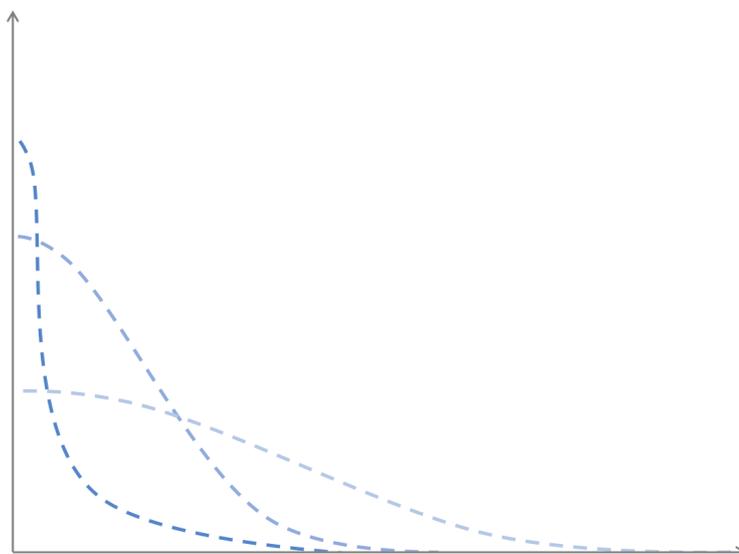


Figure 2.2: Representation of fitting a Slater orbital type using 3 Gaussian type orbitals.

considers enough orbitals to contain all the electrons. With the purpose of improving the *minimum basis set*, the number of functions included in this small basis can be duplicated, triplicated or multiplied by cardinal number X , leading to Double Zeta (DZ), Triple Zeta (TZ) basis set, or XZ respectively [77]. The increase in the number of basis functions confers a major flexibility to the molecular orbitals, allowing a much better description of the bonding in molecules, situations where the electron distribution is different in each direction. Usually, core orbitals do not contribute in a large extend to the description of the chemical bond, being the valence orbitals more relevant. Therefore, it is quite common that extra added functions are just incorporated to the valence orbitals. The different flexibility of the basis depending on the orbitals considered is referred as split valence basis and are represented by VXZ , where X is a cardinal number which defines the degree of the increment in the basis. Even though split valence basis set introduce an essential flexibility to the basis, in certain situations these basis are not enough to completely describe the chemical problem. For instance, the changes in the electron density occurring when a bond is formed or broken, require the addition of higher angular momentum functions, known as polarization functions. In the same way, other systems present weak interactions, for which special functions are needed for their proper description, such as non-covalent, Van der Walls interactions or hydrogen bonds. In such circumstances, diffuse functions, functions with a small exponent, should be included in the basis, in order to represent in a proper way charges far away from the nuclei.

Several basis set have been developed to solve different problems, but there are two main basis precisely devised to account for correlation effects, i.e., Atomic Natural Orbital basis set (ANO) [78] and Correlation Consistent basis set ($cc-BS$). In this work, ANO

and correlation consistent basis set have been used.

The Atomic Natural Orbitals basis set is formed by natural orbitals, which are calculated by using correlated method for each free atom, generally at CISD level. Natural orbitals are the eigenvectors of the density matrix, whose eigenvalues correspond to orbital occupation numbers. When a correlated wave function is used, the occupation numbers can take any value between 0 and 2, thus the *ANO* would select the important primitive GTOs to represent these occupation numbers. A disadvantage of this basis set is the great amount of PGTOs which are necessary for reaching the basis set limit. With the goal of solving this drawback, Dunning proposed the correlated consistent basis set [79, 80]. Its name makes reference to the fact that all functions in the basis, which have similar contribution, are included at the same stage, independently on function type. Many different sizes of *cc-BS* are available, *cc-pVDZ*, *cc-pVTZ*, *cc-pVQZ*. This basis set is systematically improved by increasing the number of basis functions towards the basis set limit, which can be calculated also by an extrapolation procedure. This basis already includes polarization functions, but diffuse functions must be included by an extra functions with a small exponent for each angular momentum. This extended basis is known as augmented correlation consistent basis set, for instance *aug-cc-pVDZ* [81].

3. Results

Photoactivated thiobases included in DNA strands can interact with neighboring canonical nucleobases upon light absorption, leading to intermolecular dimers formation. Although, any DNA nucleobases could be implicated in dimerization reactions, several studies [18] have concluded that in natural DNA reactions involving two adjacent pyrimidines occur with the highest yield, being cyclobutane pyrimidine and pyrimidine-(6,4)-pyrimidone photoproducts the two most common dimers. Since, the incorporation of the sulphur atom in the structure of natural nucleobases can help to trigger this kind of reactions, the thionucleobases dimerization reactions will be the central point in this work, contributing to clarify and get a deeper knowledge on the mechanism of formation of these dimers, not only in natural, but also in thiosubstitued DNA.

The present work is focused on the analysis of the generation of pyrimidine-(6,4)-pyrimidone adduct. Particularly, the dimer formed between a natural thymine and a thiothymine has been considered in the dimerization study. With the purpose of obtaining more realistic results, the organic chain composed by the sugar and the phosphate group that links nucleobases in the DNA strand, is also included in the model. Thus, the thymine-4-thiothymine dinucleoside monophosphate (T-4T-DNMP), Figure 3.1, will be the used model in the dimerization reaction study.

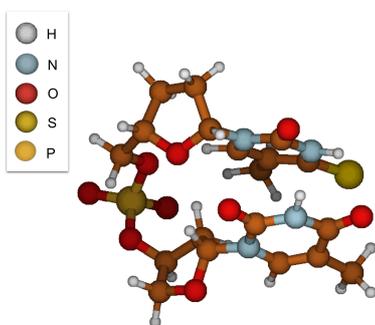


Figure 3.1: Dinucleoside monophosphate of thymine-4-thiothymine model in stacked position, considered in the simulations of the dimerization reactions.

Along all the text, the atoms including in the thymine nucleobase will be indicated as X , while the atoms concerning to the 4-thiothymine, as it depicts in the original structure of Figure 3.1, will represent adding a prime to the atom symbol, X' . Additionally, in order to facilitate the description of the mechanism, the numeration of the atoms in the used model will follow the the ordinary numeration of the natural thymine molecule, as shown in Scheme 4.

The study will have as general goal to elucidate the mechanisms for 64-PP formation for T-4T-DNMP, determining the potential role that excited states could play in it.

The generation of 64-PP can occur via single and also via triplet state mechanism, so the present work has been divided into these two alternative paths. For each of them, the mechanism has been carefully described, characterizing all the stationary points which take part into the reaction, taking as a reference the structures reported in Ref.[33]. The geometry optimization has been carried out using the DFT method as implemented in *ORCA* program [82], Version 3.0.3. Despite the dinucleoside is linked through the sugar-monophosphate chain, a dispersion corrected functional is necessary to describe correctly $\pi - \pi$ interaction between the aromatic rings of both nucleobases. Specifically, B3LYP-D3 [68] has been used in the optimizations as a functional.

Considering the large size of the molecular model, the Resolution of the Identity approximation [83] (RI) has been included in the calculations. The RI decomposes the usual four centers two electrons integrals of into three center integrals expanding the bielectronic integrals in an atom-centered Gaussians auxiliary basis set. Particularly, the RIJCOSX approximation [84] has been applied for the geometry optimization. This approximation allows speeding up calculations significantly, while a very small error is introduced.

An augmented correlation consistent basis set, *aug-cc-pVDZ* [81, 85], has been used throughout this work, which introduces polarization and also diffuse functions. For the auxiliary basis set required in the RI approximation, the *cc-pVDZ/J* [85] has been applied as well.

The vibrational frequencies of the optimized geometries have been also evaluated, with the purpose of ensuring the nature of the stationary points. Within the Born-Oppenheimer approximation, in which this study is framed, the procedure followed determines the electronic wave function of the system where the nuclei are considered as point charges. Adding the classical nuclear-nuclear repulsion ($\hat{V}_{nuclei-nuclei}$), the total energy is obtained as a function of a set of nuclear coordinates. The calculation of the total energy of the system of interest has been systematically repeated for different nuclear coordinates, resulting in the potential energy surface (PES). In this way, the region of the PES, relevant to the dimerization mechanism, can be built using the optimized geometries, which consist in the set of nuclear coordinates which minimize the total energy. In order to obtain a more detailed description of the PES, linear interpolation in internal coordinates (LIIC) calculations between the stationary points have been carried out. Even though,

the resulting LIIC provides no more than an upper bound of the minimum energy path, this procedure can offer an useful vision of the mechanism, being an important tool to provide relevant information of the excited state participation. Once the potential energy surface has been well described at B3LYP-D3/*aug-cc-pVDZ* level, the role of the excited states has been determined by employing different theoretical methods, such as TD-DFT [71], CASSCF [50] and CASPT2 [52].

Using *Gaussian09* program [86], the energies of the first states of singlet and triplet multiplicities have been calculated at TD-CAMB3LYP/*cc-pVDZ* level. The Coulomb-attenuating method (CAM) incorporates long-range correction, with two extra parameters. One of them for the Hartree-Fock exchange and the other for DFT counterpart, both over the whole range of electron distances. The extra flexibility, introduced by these parameters, allows appreciating the importance of the HF and DFT exchange contributions for short and long-range interactions, respectively. The use of CAMB3LYP [87] as functional in TD-DFT calculations, ensures a correct description of the charge transfer excitations between the two fragment of the dinucleoside.

The obtained results by DFT and TD-DFT calculations have been verified using multiconfigurational methods. CASSCF and CASPT2 calculations have been executed with the *MOLCAS* program [88]. For these methods an *ANO-RCC* [78] basis, contracted to *VDZ*, has been used, and *RI* approximation in a Cholesky Decomposition, RICD [89] was introduced to reduce the computational cost. Probably the key point in a CASSCF calculation, is the active space selection. According to the electronic structure of the dinucleoside monophosphate model, it is expected that the electronic excitations are localized in the nucleobases. Therefore, the optimal active space should include all the π and lone pair (n) electrons and orbitals of the two nucleobases included in our model. For each of the nucleobases, thymine and 4-thiothymine molecules, the complete π space should consist of 10 electrons and 8 orbitals, (10,8), that makes all together a total of 20 electrons and 16 orbitals. Even for a little active space, the *full CI* calculation that is implicit in the CASSCF and CASPT2 method, can be computationally very expensive, making this procedure unworkable. This limitation in the CASSCF method causes that the active space has to be reduced. Taking as a reference a previous theoretical study on thymine dimers, [23], the four first active orbitals have a quite high occupation, so they could be removed from the active space. This would result in a (12,12) active space. For our particular case, the lone pair of the sulphur atom in the thionucleobase is expected to play a really important role along the reaction, so this orbital and its corresponding two electrons should be included, leading to an active space of (14,13). Finally, as the breaking of the $C'_4 - S'$ bond occurs in the last step of the reaction, the corresponding σ and its respective σ^* orbitals need to be included in the active space as well. In summary, the reduced active space, which represents properly the dimer formation between a dinucleoside of thymine and 4-thiothymine, would be formed by a total of 16 electrons

and 15 orbitals.

After some preliminary calculations with an active space of (16,15), it has been observed that some orbitals with high and low occupation do not seem to contribute specially to the excitations. In this way, four electrons and four orbitals have been removed, generating an ultimate active space of (12,11) which is illustrated in Figure 3.2.

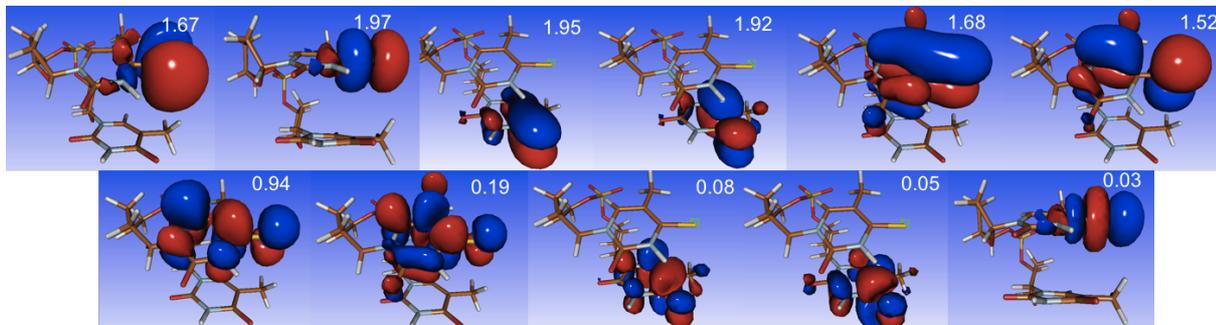


Figure 3.2: Representation of the final (12,11) active space for the multiconfigurational methods of the used model in stacking position. Calculated at CASSCF(12,11)/*ANO-RCC-VDZ* for seven roots.

Due to the significant reorganization that the structures will undergo along the reaction path, the active space can be affected. The LIIC calculations allow the smooth evolution of the active space between the different stationary points.

CASSCF energies with the optimal (12,11) active space have been performed on the stationary points optimized at B3LYP-D3/*aug-cc-pVDZ* level. In order to improve the CASSCF picture, which does not include part of the total electron correlation, it is necessary to move to MS-CASPT2, which includes the dynamical correlation. As before, *MOLCAS* program is used to carry out these calculations, including *RI* approximation and *ANO-RCC-VDZ* as a basis set. The use of perturbation theory could lead to unreliable results due to the effect of intruder states, which forces, in our case, to add a level shift of 0.3 to avoid this undesired situation, in our case. Besides CASPT2 energies, the spin-orbit coupling (SOC), which provides relevant information about the effectiveness of the transfer population between states of different multiplicity, has been evaluated.

Following the above procedure, we have calculated the reaction mechanism of the 64-PP dimerization along the equilibrium coordinates of lowest-lying singlet and triplet state. Although, each pathway has been analyzed individually, it is important to notice that at every moment there exists a correlation and dependence between them, which, in last term, will allow generating the final product 64-PP.

3.1 Singlet 64-PP Mechanism

In this first part of our work, the ground and the excited states will be studied considering the stationary points of the dimerization mechanism along the lowest singlet state S_0 . The optimized geometries for singlet mechanism at DFT level, using B3LYP-D3 functional and the *aug-cc-pVDZ* basis set are collected in Figure 3.3.

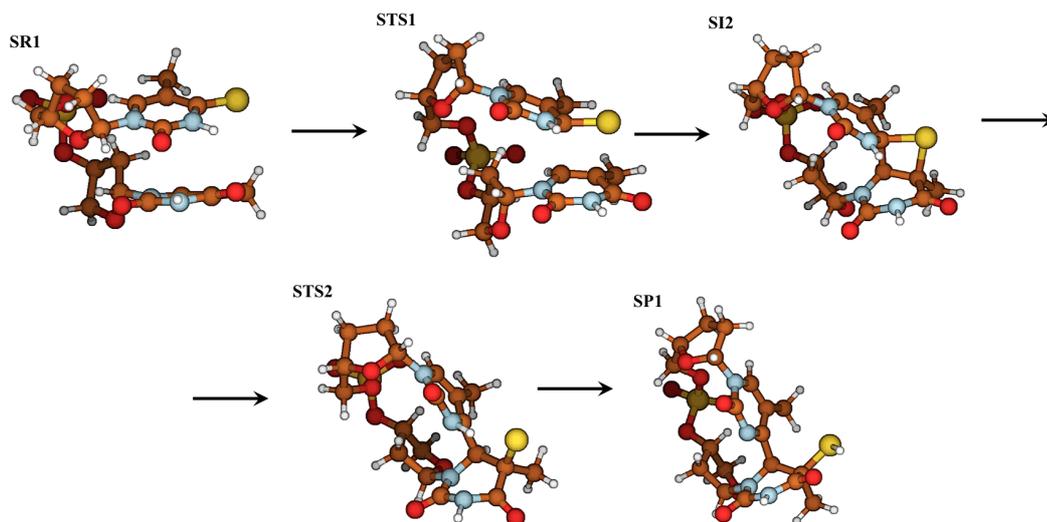


Figure 3.3: Stationary points of the mechanism for the 64-PP formation in S_0 manifold, optimized at B3LYP-D3/*aug-cc-pVDZ* level.

The complete mechanism of 64-PP formation along S_0 presents four different stages, including a total of 5 stationary points, i.e. minima and transitions states, Figure 3.3.

Starting from the initial geometry, where the two nucleobases are in a stacked position (SR1), the first part of the mechanism would consist in the approach between the two aromatic rings, in order to facilitate the intermediate (SI2) formation. Once the distance between 4-thiothymine and thymine is suitable, a [2+2] cycloaddition takes place between the $C_5 - C_6$ double bond of thymine and the $C'_4 - S'$ bond of the thiocarbonyl group of 4-thiothymine. The numeration of the atoms have been established by Scheme 4 and they are also indicated in atomic structures Figure 3.4. This step results in an intermediate, called thietane (SI2). Since this intermediate is formed in one single step, the reaction of 64-PP formation along the stationary points in S_0 is considered driven by a concerted mechanism. Thietane presents an important instability, due to the high stress caused by the four membered ring which provokes the opening of the close structure. The $C'_4 - S'$ bond dissociation causes the transfer of the sulphur atom from one nucleobase to the other. The last step of the reaction is the hydrogen migration between nucleobases from N'_3 to the S , following the numeration of Figure 3.4, resulting in the final 64-PP (SP1).

As shown in Figure 3.3, each of the three mentioned minima stationary points are

connected by transition states, STS1 and STS2, which present an imaginary frequency whose normal modes links the two corresponding minima. To confirm the correlation between stationary points along the potential energy profile, we have undertaken IRC calculations.

Different potential energy profiles have been calculated at different levels of theory. In particular, the energies have been determined at B3LYP-D3/*aug-cc-pVDZ*, CASSCF(12,11)/*ANO-RCC-VDZ* and CASPT2(12,11)/*ANO-RCC-VDZ* level, collected in Figure 3.4.

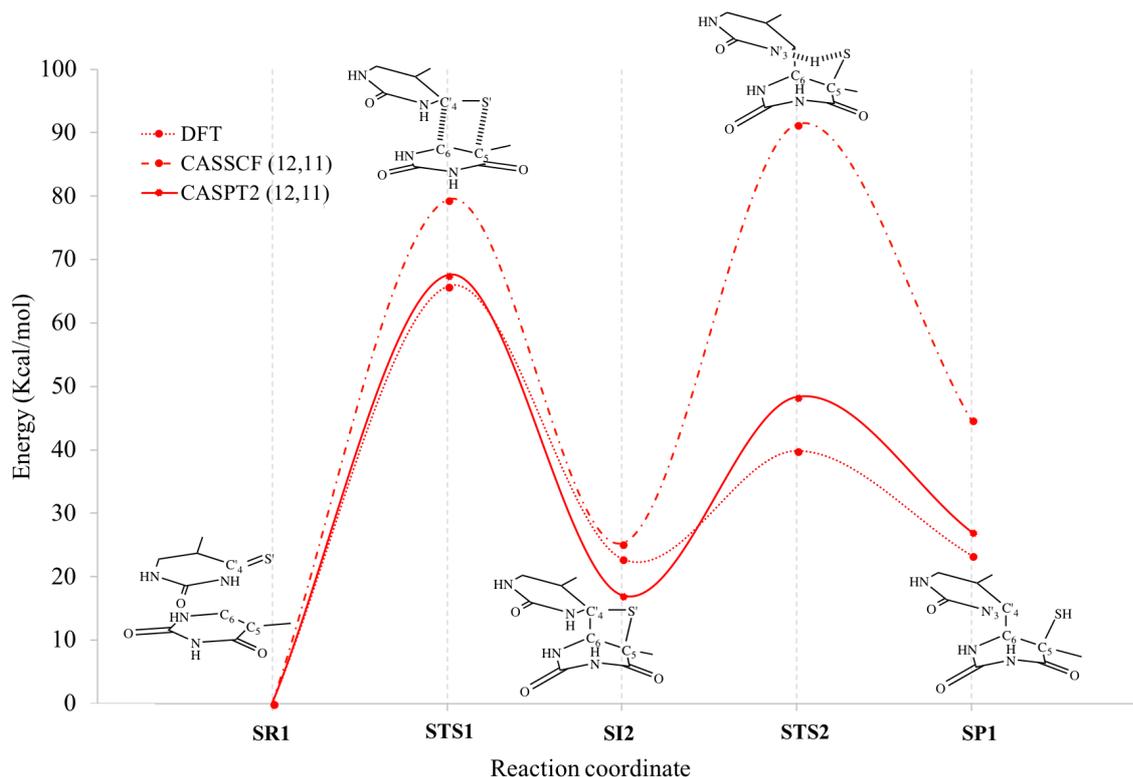


Figure 3.4: Potential energy profile for the mechanism of 64-PP formation in S_0 , at B3LYP-D3/*aug-cc-pVDZ* (red dashed line), CASSCF(12,11)/*ANO-RCC-VDZ* (red dotted line) and CASPT2(12,11)/*ANO-RCC-VDZ* (red solid line) level. Energies in Kcal/mol relative to the SR1 minimum.

As Figure 3.4 illustrates, the different methods provide slightly different potential energy profiles for same optimized geometries. It is important to notice, that while DFT and CASPT2 introduce a great amount of the total correlation energy, the CASSCF method does not include the so called "dynamical correlation", which causes that this method differs significantly from the rest, overestimating the energy barrier of the reaction. Although DFT calculation underestimates the energy of the second barrier, surprisingly, this method, incorporating dispersion corrections, delivers a profile very similar to CASPT2. It is expected that the best results are provided by CASPT2 method, as not it only incorporates the dynamical correlation energy through perturbation theory, but also because of the multiconfigurational character of this method.

Due to the extremely high barriers that the mechanism presents along the coordinates of the optimized stationary points in S_0 state, a thorough study of the excited state potential energy profiles is essential to determine the feasibility of these mechanisms when the system is activated with light. Considering the optimized stationary points of Figure 3.3, vertical excitation energy calculations for the lowest-lying states of singlet and triplet multiplicity have been carried out at TD-CAMB3LYP/ $cc-pVDZ$ level. Although, it is more than probable that the most stable path in the ground state does not coincide with the minimum energy path (MEP) in the excited states, these profiles are expected to shed some light into the excited state mechanism for 64-PP formation. Indeed, this procedure has been considered as a very first approximation to elucidate the mechanism to lead to 64-PP involving the excited states. The obtained excited potential energy profiles at TD-CAMB3LYP/ $cc-pVDZ$ level of theory are illustrated in Figure 3.5.

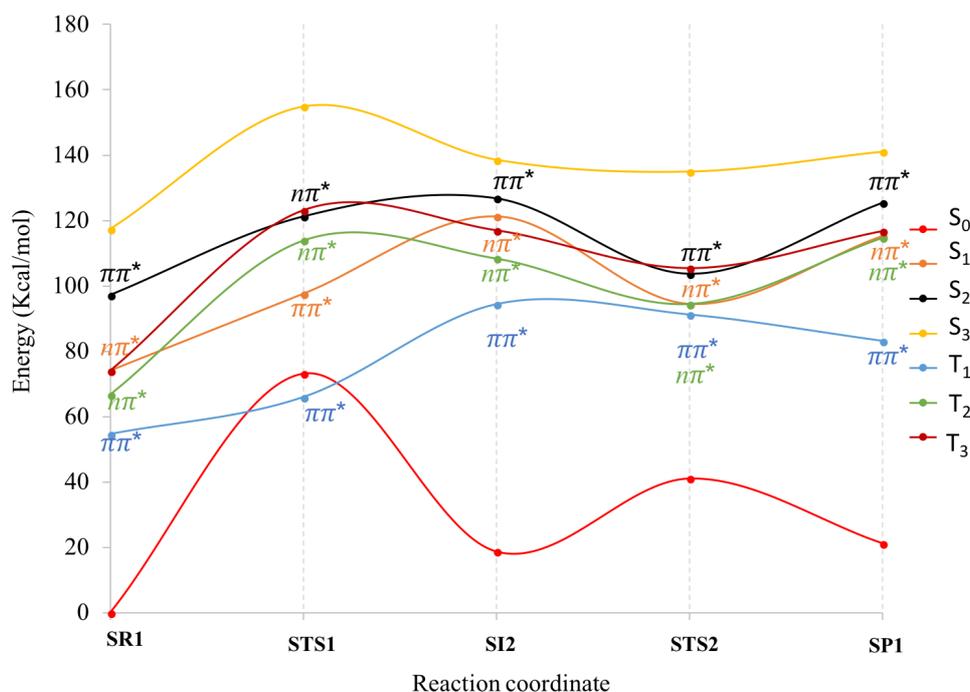


Figure 3.5: Ground state and singlet and triplet excited state potential energy profiles for the mechanism of 64-PP formation in the S_0 at TD-CAMB3LYP/ $cc-pVDZ$ level of theory. Energies in Kcal/mol relative to the ground state equilibrium minimum SR1.

The spectroscopic state is determined by its large oscillator strength. The values of the oscillator strength are directly related with the intensity of the experimental absorption bands. In this particular case, the $\pi\pi^*$ state with the third lowest energy, S_2 state represented in Figure 3.5 by a black line, presents the largest oscillator strength. From the Franck-Condon region, SR1, the system from the ground state, S_0 in red, absorbs energy to reach the lowest spectroscopic state, S_2 . The necessary energy to activate the

photosystem is 97 Kcal/mol. Once the system has been excited to the S_2 state, it is our aim to propose a potential deactivation path, which would conclude in 64-PP formation.

A quick look to Figure 3.5 reveals that the potential energy profile of the excited states do not follow the behavior of that of ground state, confirming that these reaction coordinates are not the ones leading to the MEP for the excited states and thus that they are not the most adequate to analyze the decay of the system from the S_2 state. The upward slopes of all excited states, shown in Figure 3.5, hamper the elucidation of the mechanism for deactivation from S_2 to S_0 . Despite the fact that other coordinates would be more appropriate to interpret the relaxation mechanism from the S_2 , these results can also provide some preliminary relevant information.

From the analysis of the character of the different states in each stationary point, it can be observed an exchange between S_2 and S_1 , in orange. While in the Franck-Condon region the S_2 and S_1 states present $\pi\pi^*$ and $n\pi^*$ natures, the situation is the opposite at STS1 point. The exchange in the character of these states suggests the existence of a point, between SR1 and STS1 of the PES, in which these states become degenerate. In other words, it would be possible to locate a conical intersection (CI) between S_2 and S_1 , $(S_2/S_1)_{CI}$, which would be responsible for the transfer of the population from the S_2 to the S_1 .

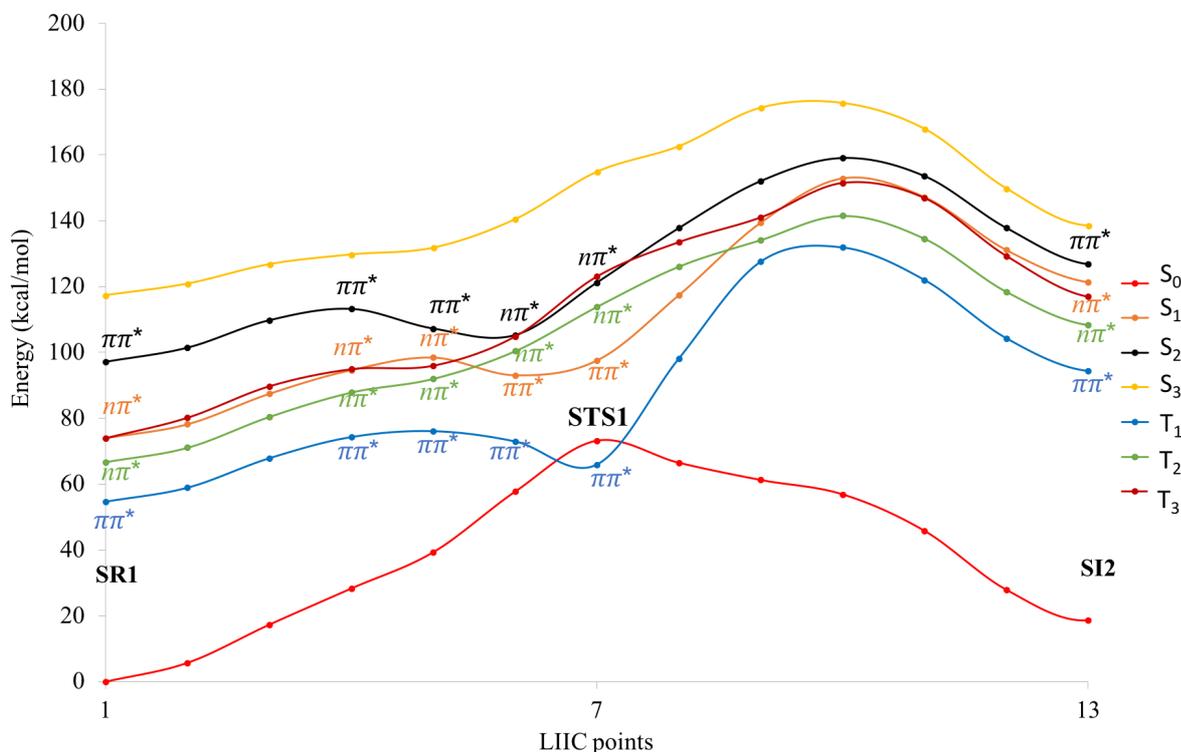


Figure 3.6: Singlet and triplet potential energy profiles for the ground and excited states of the T-4T-DNMO model along the LIIC points between SR1 and SI2 of the mechanism leading to 64-PP in the S_0 at TD-CAMB3LYP/*cc-pVDZ* level of theory. Energies relative to the ground state equilibrium geometry SR1 in Kcal/mol.

A simple approach to check the occurrence and obtain an approximated geometry to be optimized of the $(S_2/S_1)_{CI}$, is to interpolate linearly the internal coordinates of these two interesting points. In this way, Figure 3.6 depicts the corresponding generated points of the LIIC from SR1 to SI2, including the transition state STS1. As it can be seen at point 5 in Figure 3.6, S_1 and S_2 states approach. Attending to the character of S_2 and S_1 at the previous and following point, 4 and 6, respectively, we expect the occurrence of a CI around point 5. Considering the presence of the $(S_2/S_1)_{CI}$ at this point, the population would be able to decay to singlet states of lower energy. Undoubtedly, the evaluation of a MEP from the Franck-Condon position S_2 could offer a slightly different scenery. In this way, a different relevant points could be found, providing a more reliable information of the relaxation path. However, this just would be achievable through a MEP calculation.

Another interesting feature in the PES, which could contribute to understand the 64-PP formation mechanism, is the STS1 position. At this particular point, the lowest-lying triple state, T_1 in blue, exhibits lower energy than the ground state. Unlike the STS1 structure, which is a transition state in S_0 state, in T_1 this region is practically flat, Figure 3.5 even a minimum can be appreciated in LIIC of Figure 3.6. Interestingly, the TD-DFT calculations of the ground and excited states and DFT for the ground state predicts the energy of the T_1 to be more stable than those of the S_0 . The picture, however, changes when the DFT approach is employed for the characterisation of the S_0 and T_1 , Figure 3.7.

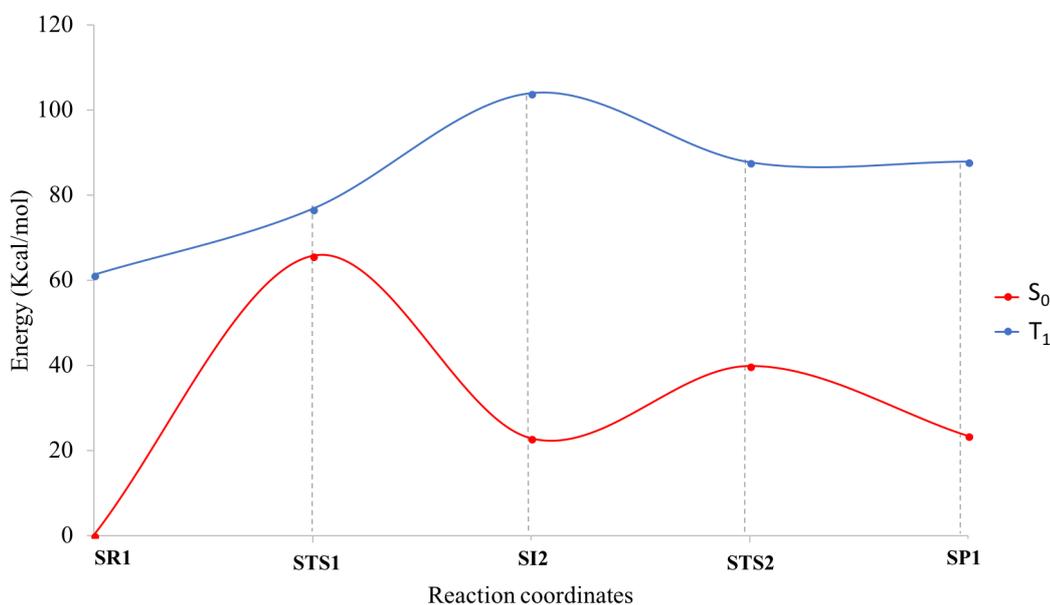


Figure 3.7: Potential energy profiles for the S_0 and T_1 states along the stationary points of the S_0 at B3LYP-D3/aug-cc-pVDZ level of theory. Energies calculated in Kcal/mol and relative to SR1 geometry of the S_0 state.

The energy difference at the STS1 point for the T_1 between these methods is of 10.7

Kcal/mol. Despite the functional used in each method is different, the discrepancy of the results obtained with the two methods is quite significant, indicating that the TD-DFT results of excited states must be taken with caution. In any case, this region of the PES is energetically accessible for the system and the proximity between the S_0 and T_1 at this point of the PES suggests that this point might be considered as a relevant point in the deactivation path.

In order to improve the accuracy of our results, we will move to a multiconfigurational method, such as CASPT2, much more appropriate in the description of the excited state problem.

Figure 3.8 collects the CASPT2 potential energy profiles of the ground and excited states along stationary points optimized for the ground state. As for TD-DFT, upon light absorption, the system allows populating the lowest spectroscopic state, S_2 , which presents an energy of 100 Kcal/mol, close to the TD-DFT result.

In contrast to TD-DFT, the lowest triplet state is not below the S_0 at any point of the potential profile as shown Figure 3.8. However, similarly to the TD-DFT method, the potential energies of the excited states show an ascending profile, not being possible to reveal what is the actual deactivation mechanism for this system when excited to the S_2 .

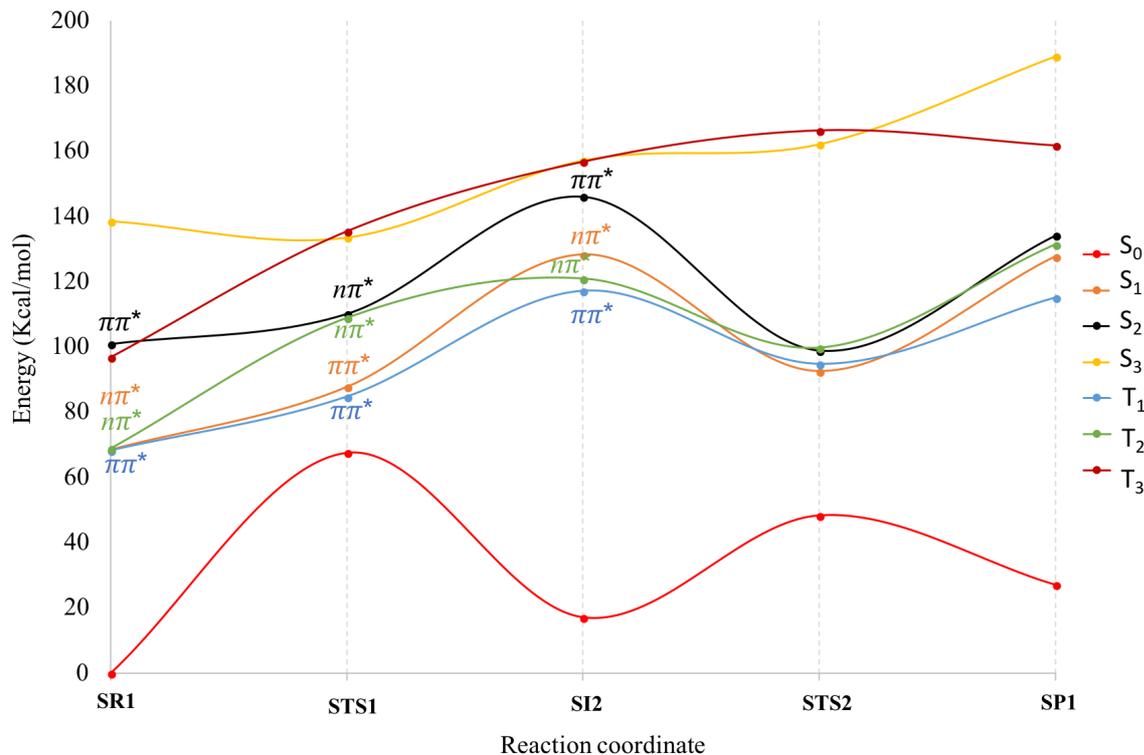


Figure 3.8: Singlet and triplet potential energy profiles for the ground and excited states along the S_0 stationary points involved in the mechanism for the formation of the 64-PP at CASPT2(12,11)/ANO-RCC-VDZ level. Energies in Kcal/mol relative to the ground state equilibrium geometry SR1.

From the potential energy profile of Figure 3.8, it is possible to extract some important information. Although, the excited states show an upward behavior, the proximity between the triplet and the singlet states is evident along all the reaction coordinates. At the SR1 point in Figure 3.8, i.e. the Franck-Condon region, the energy value of the S_1 is 68.4 Kcal/mol, the energy difference between S_1 and T_1 is 0.2 Kcal/mol, having the lowest triplet state a energy of 68.2 Kcal/mol. For S_1 and T_2 (68.8 Kcal/mol), the $\Delta E(S_1/T_2)$ is 0.4 Kcal/mol. The SOC calculated at CASPT2(12,11)/*ANO-RCC-VDZ* level are very high. For instance, at the Franck-Condon region, the SOC between S_1 and T_1 is 148.6 cm^{-1} , also $\text{SOC}(S_1/T_2)$ is 59.7 cm^{-1} . The heavy atom effect, which is caused by the introduction of sulphur atom, is responsible for such high values of SOCs. However, this factor itself does not ensure the transfer population through an intersystem crossing (ISC). It is the combination of high SOCs and the proximity between states that suggests the occurrence of ISC between states of different multiplicity, being able to cause the population transfer from the single to the triplet manifold. With the purpose of trying to obtain more relevant information about the relaxation mechanism along the reaction coordinate of the S_0 state, a LIIC for the first steps of the reaction has been generated. The level of theory employed for this LIIC calculation is CASPT2(12,11)/*ANO-RCC-VDZ*. The resulting LIIC is showed in Figure 3.9.

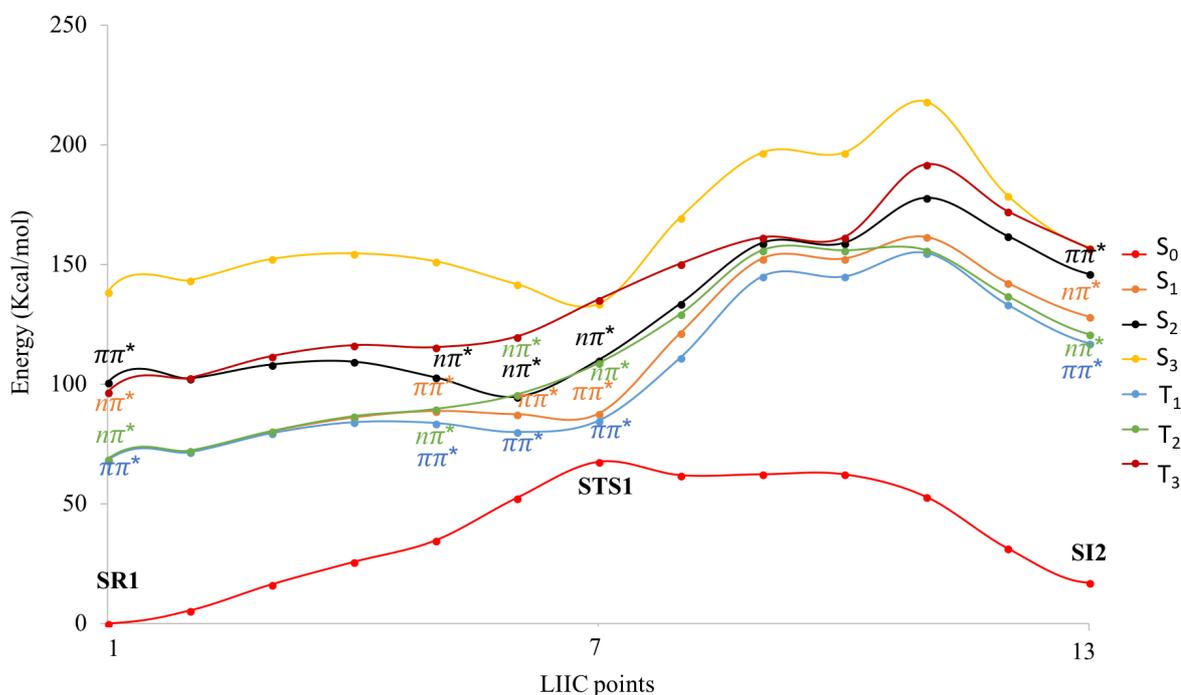


Figure 3.9: Singlet and triplet potential energy profile for ground and excited states at CASPT2(12,11)/*ANO-RCC-VDZ* level along the LIIC points interpolated between the SR1 and SI2 structures of the 64-PP formation reaction mechanism. Energies in Kcal/mol relative to the SR1 geometry of the dinucleoside in the ground state.

According to these energy profiles, we can see that the characters of the S_2 and S_1 state, $\pi\pi^*$ and $n\pi^*$ respectively, exchange between point 6 and 7. This fact could suggest a $(S_2/S_1)_{CI}$, which would be energetically accessible from the Franck-Condon region. Thus, a decay of the system to the lowest singlet state seems to be plausible. Furthermore, the first part of the LIIC between the SR1 and STS1 points, Figure 3.9, makes evident the proximity in energy of some triplet and singlet states. The more significative points are found at 4, 5 and 6 position, where the energy differences and SOCs have been collected for the S_1 , S_2 , T_1 and T_2 states in Table 3.1.

States	LIIC points								
	4			5			6		
	E_{total} (Kcal/mol)	ΔE (Kcal/mol)	SOC (cm^{-1})	E_{total} (Kcal/mol)	ΔE (Kcal/mol)	SOC (cm^{-1})	E_{total} (Kcal/mol)	ΔE (Kcal/mol)	SOC (cm^{-1})
S_1/T_2	86.0/86.6	0.6	18.1	88.8/89.6	0.8	13.1	87.4/95.6	8.2	133.9
S_1/T_1	86.0/84.1	1.9	164.1	88.8/83.8	5.00	164.2	87.4/79.9	7.5	24.1
S_2/T_2	109.4/86.6	22.8	125.2	102.9/89.6	13.3	126.0	94.7/95.6	0.9	3.9

Table 3.1: Energy differences in Kcal/mol and SOC values in cm^{-1} between states of different multiplicity at relevant points of the LIIC of Figure 3.9

As it can observe in Figure 3.9 and also in Table 3.1, the proximity and the high SOC values between different singlet and triplet states suggest the tendency to transfer the population from the single to the triplet states, in any of the regions of the potential energy profile analyzed above. Considering that there are evidences that could indicate the transfer of population to the lowest triplet state at these points, it is logical to assume that the system could decay to the lowest lying triplet state, T_1 , at the SR1 position. Point 7 of Figure 3.9 requires special attention. Assuming that the system is able to relax to the S_1 , at point 7, this state with an energy of 87.7 Kcal/mol is close in energy to the lowest-lying triplet state which presents an energy value of 84.7 Kcal/mol. The $\Delta E(S_1/T_1)$ is equal to 2.9 Kcal/mol and also to the ground state (67.5 Kcal/mol), the $\Delta E(S_1/S_0)$ is 20.1 Kcal/mol. The value of $SOC(S_1/T_1)$ is 47 cm^{-1} and $SOC(S_0/T_1)$ is 64 cm^{-1} . Although, the values of the corresponding SOCs are not as high as in other regions of the PES, see Table 3.1 the possibility that the population from S_1 is transferred to the T_1 at STS1 position and from here to ground state through a second consecutive ISC cannot be rejected. Moreover, a direct population transfer could occur from S_1 to the ground state, via a singlet mechanism. In any of the latter situations, the population could decay to ground state, from where the system might evolve to SI2 intermediate or return to the initial structure SR1.

As already mentioned, the above analysis of the excited states has been carried out

along the coordinate of the S_0 reaction mechanism, which is not the most appropriate. In fact, there should exist other more energetically favorable coordinates for the deactivation from the S_2 , which could reveal other relevant points of the PES where the population transfer would be facilitated.

Owing to the fact that there is a non-negligible possibility that the triplet manifold becomes populated, the study of the triplet path is indispensable to create a complete picture of the 64-PP formation mechanism. At the same time, considering the unfavorable upward potential energy profile of the excited states along the reaction coordinate of the S_0 mechanism, an analysis of the states along the triplet reaction mechanism could shed more light in the deactivation mechanism.

3.2 Triplet 64-PP Mechanism

As already discussed above, triplet states are expected to have a really relevant role in the 64-PP formation reaction. Because of their importance, an individual analysis of the 64-PP dimerization along the triplet mechanism will be made in the following part of the work. As in the above singlet mechanism, in order to refer to the atoms involved in the reaction, the numeration of them follows Scheme 4. Moreover, Figure 3.11 and 3.16 recollect the structures for each particular mechanism, indicating the numeration used in the description.

The mechanism of the 64-PP generation along the lowest triplet state is defined by a total of 12 stationary points, considering minima and transition states. Unlike what has been found for the singlet mechanism, which consists in 4 simple steps and corresponds to a concert mechanism, the reaction through triplet state is more complicated. The formation of thietane, occurs in two independent steps, indicating that the reaction via triplet manifold is not concerted anymore. Attending to the sequence that the mechanism follows to form the intermediate thietane, two different scenarios for the formation of the 64-PP are possible and they will be studied separately. The first possibility describes the generation of the thietane as the result of the formation bond between the C_6 of one thymine and the C'_4 of the thiothymine, following by the closing of the ring due to the formation bond $C_5 - S'$. Since the $C - C'$ bond formation along this path occurs in a first place, this path has been called CC-CS Triplet mechanism. In contrast, the other proposed mechanism is based on the formation of the $C_5 - S'$ bond in a first stage, then the $C_6 - C'_4$ bond occurs. This mechanism will be referred as CS-CC Triplet mechanism. These two paths differ in the first stationary points that compose the mechanism. However, the final route is common for both mechanisms.

CC-CS Triplet Mechanism

Figure 3.10 shows the complete mechanism for the 64-PP formation along CC-CS triplet path. The CC-CS Triplet mechanism, Figure 3.10, begins with the two nucleobases in a stacked position, (R1). This initial structure is similar to that calculated for the singlet path, SR1. In a first step, these two nucleobases approach in the direction of the $C_6 - C'_4$ bond, ending in the generation of the intermediate $I1_{CC}$. The subsequent formation of the bond between $C_5 - S'$ results in the intermediate thietane, (I2).

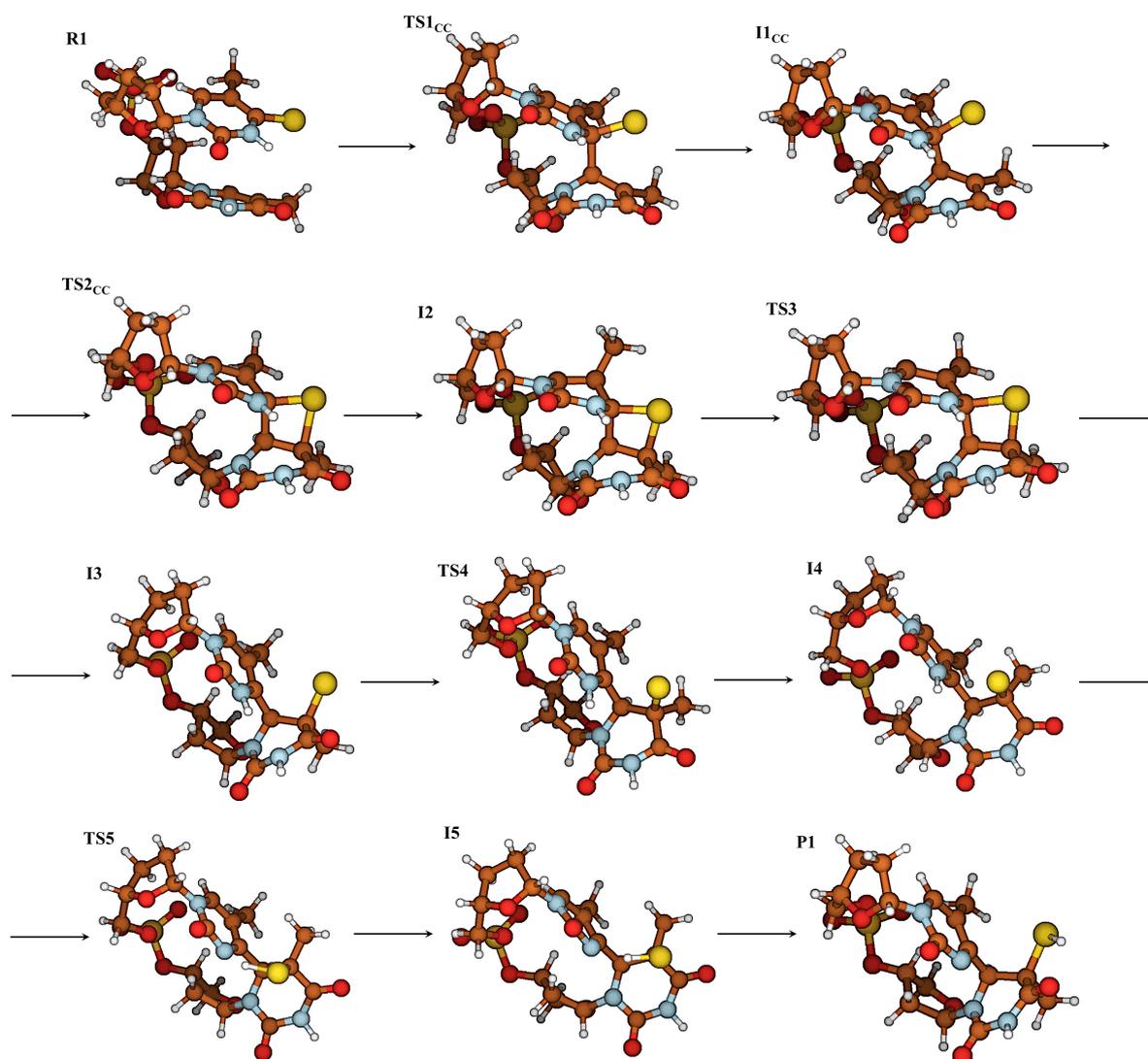


Figure 3.10: Stationary points for the mechanism of the 64-PP formation along the lowest triplet state for the CC-CS triplet mechanism, optimized at B3LYP/*aug-cc-pVDZ* level.

From this point, the route to the final 64-PP is shared between CC-CS and CS-CC triplet mechanisms. The instability of thietane causes the opening of the ring, which does not occur in a unique step as in the singlet mechanism commented above. The thietane

opening takes place, through the breaking $C'_4 - S'$ bond (I3). Once the ring is broken, the hydrogen from the nitrogen atom, N'_3 of thiothymine migrates to the S , which has been transferred to the thymine after the breaking $C'_4 - S'$ bond. In order to facilitate the final step of the mechanism, the migration of the proton, the compound undergoes a dihedral rotation to approach the S atom to the N'_3 of the other monomer (I4). Finally, the hydrogen is transferred to the S (I5). To conclude in the 64-PP (P1), the group SH rotates.

All the minima in the mechanism were connected with the next one through a transition state with the exception of the I5 and P1. Between these two minima, there might exist a transition state, which has not yet been found. With the geometries included in Figure 3.10 calculated at B3LYP-D3/*aug-cc-pVDZ* level, the potential energy profile for T_1 has been built at same level of theory, and also considering the computational protocols CASSCF(12,11)/*ANO-RCC-VDZ* and CASPT2(12,11)/*ANO-RCC-VDZ*.

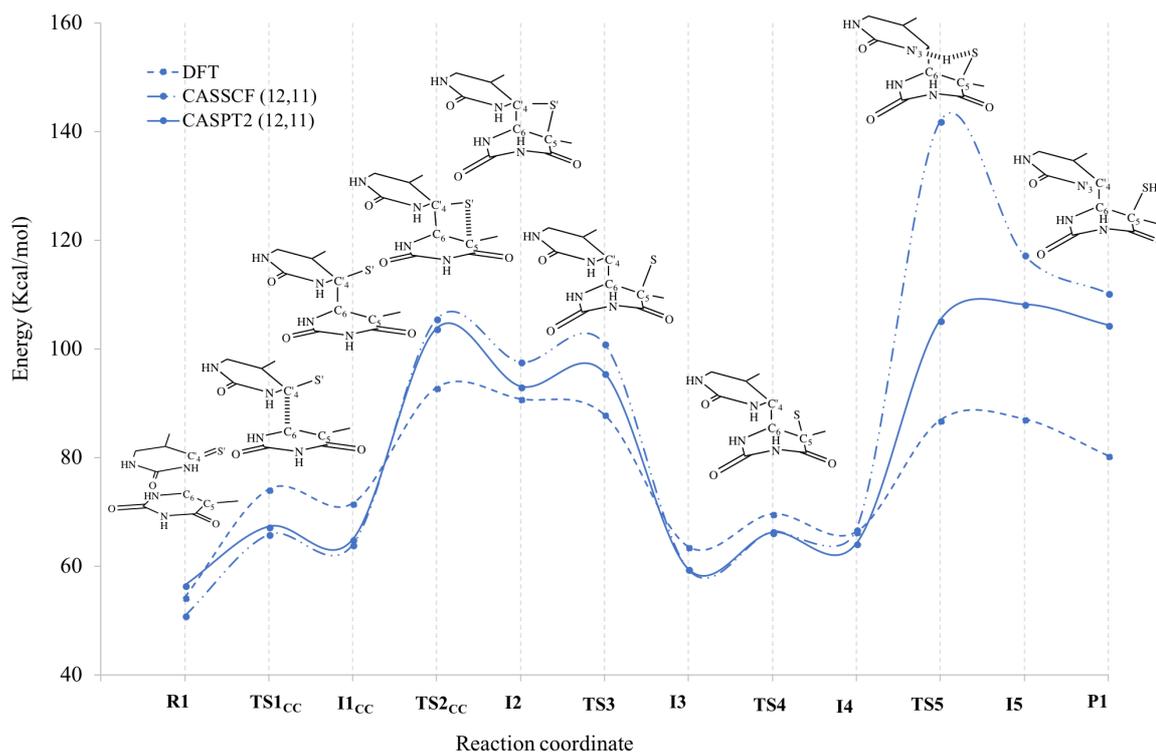


Figure 3.11: Triplet potential energy profile for 64-PP formation along lowest triplet mechanism calculated at B3LYP-D3/*aug-cc-pVDZ* (blue dashed line), CASSCF(12,11)/*ANO-RCC-VDZ* (blue dotted line) and CASPT2(12,11)/*ANO-RCC-VDZ* (blue solid line) level. Energy in Kcal/mol relative to the most stable singlet minimum SR1 in S_0 state.

As it can be seen in Figure 3.11, the profiles obtained with these methods slightly differ from each other. The CASSCF method calculates the highest energy barriers. This overestimation is specially important for the barrier related with the migration of

the hydrogen, TS5, whose difference of energy with the CASPT2 method is almost 50 Kcal/mol. The significant discrepancy is due to the impossibility of the CASSCF method to account for dynamical correlation. The B3LYP-D3 results, are rather good compared to the CASPT2 ones, but this method tends to underestimate some energy barriers, see for instance the TS5 point in Figure 3.11. The more reliable data can be considered those calculated by CASPT2 method. Nevertheless, similarly to the other levels of theory, it predicts barriers extremely high, preventing the reaction to exclusively occur in the T_1 state.

Considering the stationary points of the triplet mechanism optimized at B3LYP-D3/*aug-cc-pVDZ* level, vertical excitation energy calculations would allow evaluating the excited states behavior along this reaction coordinate. For this purpose the TD-CAMB3LYP/*cc-pVDZ* method has been applied, obtaining the results show in Figure 3.12.

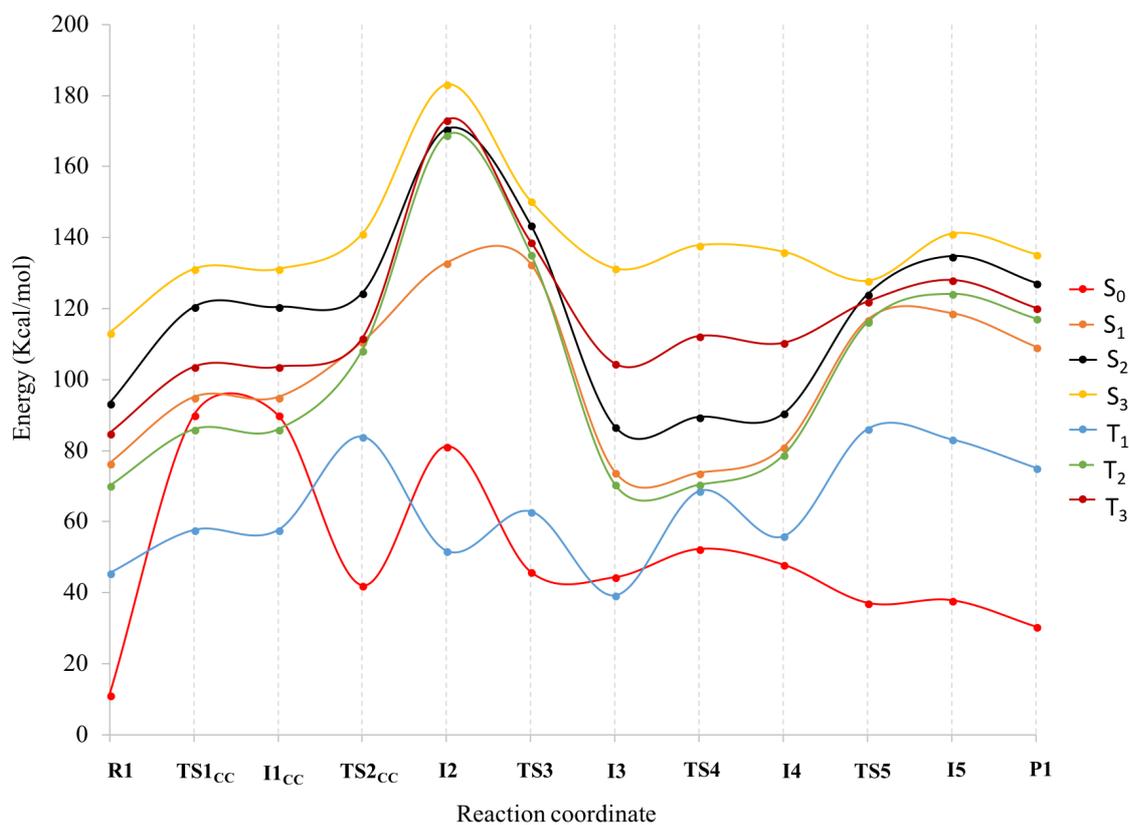


Figure 3.12: Energetic profiles of the ground and excited states of the stationary points along the 64-PP formation CC-CS triplet mechanism at TD-CAMB3LYP/*cc-pVDZ*. Energies in Kcal/mol relative to the SR1 geometry of S_0 .

Figure 3.12 shows the upward profiles of the singlet and triplet excited states along the optimized coordinate leading to the 64-PP along the T_1 state. As for the profiles examined in Figure 3.5, the global coordinate does not correspond to those defining the minimum energy path for the relaxation of the excited states.

Attending to the results obtained for the singlet mechanism, there are non-negligible possibilities that the triplet manifold becomes populated. Assuming that, internal conversion (IC) would occur in the triplet manifold and that the T_1 would eventually collect to the population. According to the conclusions obtained from the analysis of the profiles along the global coordinate in the S_0 , the T_1 state could be populated in the proximity of two different S_0 stationary points, SR1 or STS1. Due to the similarity of the structures of SR1 and R1, if the population is transferred to triplet state at the SR1 point, the system would be expected to be retained in the minimum R1. In contrast, the geometry of STS1 does not present an analogue in T_1 manifold. Thus, the population is transferred at the vicinity through the STS1 point, expecting to populate the R1 minimum as well. Once the system is established in the minimum R1 of T_1 , it would have enough energy, to surpass the $TS1_{CC}$ barrier and to reach the following minimum $I1_{CC}$. Analyzing Figure 3.12, it is very remarkable that the lowest triplet state has lower energy than ground state along certain points of the reaction path. This situation is notable in the region between R1 and $I1_{CC}$, where the transferred population to triplet state is expected to occur. The crossing between these states in this region suggests a potential $(T_1/S_0)_{ISC}$, leaking the population to the S_0 reaction mechanism, from where the reaction would carry on.

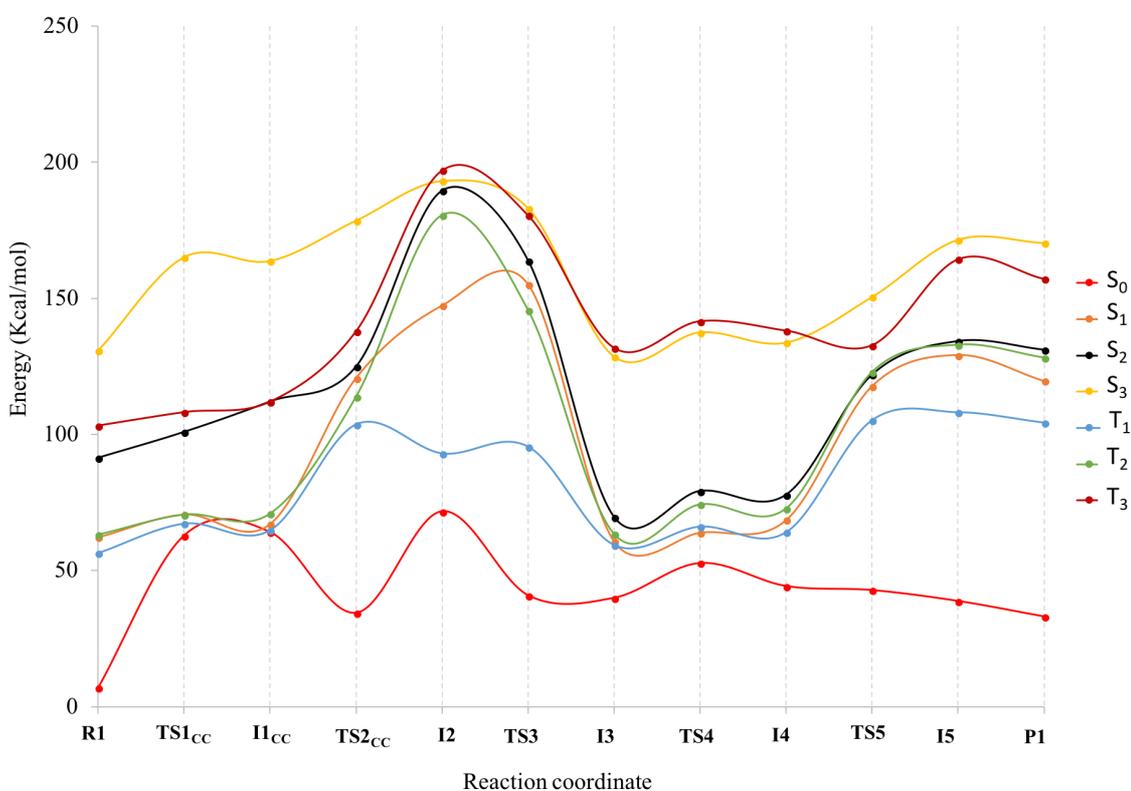


Figure 3.13: Potential energy profile of ground state and singlet and triplet excited states for the stationary points of CC-CS triplet mechanism leading to 64-PP, at CASPT2(12,11)/ANO-RCC-VDZ level. Energies in Kcal/mol relative to the SR1 geometry of S_0 .

In order to secure a more reliable description, the energy of the stationary points of Figure 3.10 and also the SOC values have been evaluated at CASPT2(12,11)/ANO-RCC-VDZ level, Figure 3.13.

The analysis of Figure 3.13 reveals that all the excited states present an upward slope along this reaction coordinate. The comparison of Figure 3.13 and Figure 3.12 allows appreciating important differences between TD-DFT and CASPT2 methods. It must be noticed that the energy difference between T_1 and S_0 in the region R1-I1_{CC} is significantly high for the case of the TD-DFT calculation, whereas the difference is much lower at CASPT2 level. This fact could indicate a wrong description of the states by the TD-CAMB3LYP method. Despite the excited states of the CASPT2 picture present an ascending potential energy profiles, in certain areas the CASPT2 slopes are much less pronounced than those of Figure 3.12. In this way, the region between R1 and I1_{CC} shows a rather flat profiles, with several states close in energy and high SOC values as Table 3.2 collects, which would support the idea of possible ISCs between states of different multiplicity.

Observing the relevant data of Table 3.2, the population transfer to triplet manifold could be possible. Moreover, the I1_{CC} point of the T_1 potential energy profile could be considered a really important point in the CC-CS triplet mechanism, a potential ISC funnel, due to the high values of the SOC values and also the evident degeneracy of the states.

States	Stationary points					
	TS1 _{CC}			I1 _{CC}		
	E _{total} (Kcal/mol)	ΔE (Kcal/mol)	SOC (cm ⁻¹)	E _{total} (Kcal/mol)	ΔE (Kcal/mol)	SOC (cm ⁻¹)
S ₁ /T ₂	70.7/70.6	0.1	24.1	67.2/70.9	3.8	149.6
S ₁ /T ₁	70.7/67.3	3.4	167.6	67.2/64.9	2.3	81.6
S ₀ /T ₂	62.9/70.6	7.7	167.7	64.1/70.9	6.9	82.1
S ₀ /T ₁	62.9/67.3	4.4	27.9	64.1/64.9	0.8	148.7

Table 3.2: Energy differences in Kcal/mol between singlet and triplet states and the corresponding SOC values in cm⁻¹ at TS1_{CC} and I1_{CC} stationary points of the CC-CS triple mechanism leading to 64-PP reaction.

Although, the possibility that an ISC occurs between any triplet or singlet state, in this region of the PES, is non-negligible, from excited triplet states the population would be transferred to T_1 through IC. Whereas in Figure 3.12 the energy of the lowest triplet state was, in many points, lower than that of the ground state, this situation is not observable anymore in Figure 3.13, which shows T_1 and S_0 state are practically degenerate at TS1_{CC} and I1_{CC}, see Table 3.2. The degeneration and also the high SOC values that these states presents would facilitate the ISC, with the transfer of the population to the ground state.

Assuming that ISC to the ground state takes place at $I1_{CC}$, the reaction to generate 64-PP formation would continue from $SI2$ through the singlet mechanism as shown in Figure 3.8. Figure 3.14 pretends to offers a global picture of both reaction mechanisms, singlet and triplet. It depicts how the system would evolve from $I1_{CC}$ of the CC-CS triplet mechanism at S_0 state (red dotted line) to $SI2$ of single mechanism in S_0 state (red line).

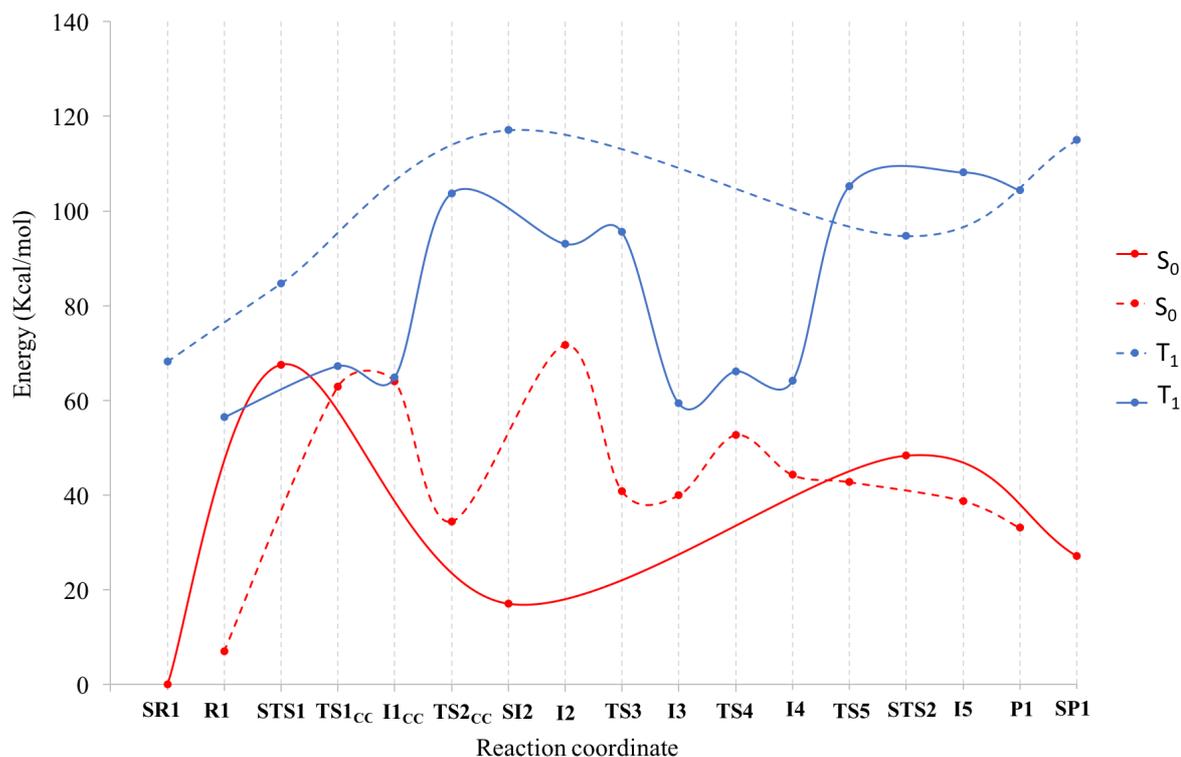


Figure 3.14: S_0 and T_1 potential energy profile along the global coordinates defined by the ground state mechanism (red solid line and blue dotted line, respectively). S_0 and T_1 potential energy profile along the global coordinates defined by the T_1 mechanism (red dotted line and blue solid line, respectively). Energies calculated at CASPT2(12,11)/ANO-RCC-VDZ level and expressed in Kcal/mol relative to the most stable geometry of S_0 mechanism, SR1.

Figure 3.14 reveals the energetic degeneration at $I1_{CC}$ point, between T_1 (blue solid line) and S_0 (red dotted line) along the coordinates of the CC-CS triplet mechanism. Also the fact that $I1_{CC}$ is a minimum in T_1 would contribute to retain the population at this point of the PES, enhancing the transfer to ground state and reaching the final photoproduct along S_0 mechanism, represented by a red solid line. Considering the energy of the $I1_{CC}$ point in the CC-CS triplet mechanism in blue solid line, the energy barrier at STS2 point could be surpassed.

CS-CC Triplet Mechanism

The CS-CC triplet mechanism just differs from the previous one in the first stationary points, as in this case the $C_5 - S'$ bond formation occurs in first place. Figure 3.15 collects

the different stationary points involved in the CS-CC triplet mechanism in T_1 state. As in previous analysis, the same numeration for the atoms of the T-4T-DNMP model has been employed, following the Scheme 4, which can be seen in Figure 3.16.

Along this mechanism, the initial stacked dimer (R1) starts to reorganize facilitating the approach of the S' atom of thiothymine to the C_5 of the adjacent thymine.

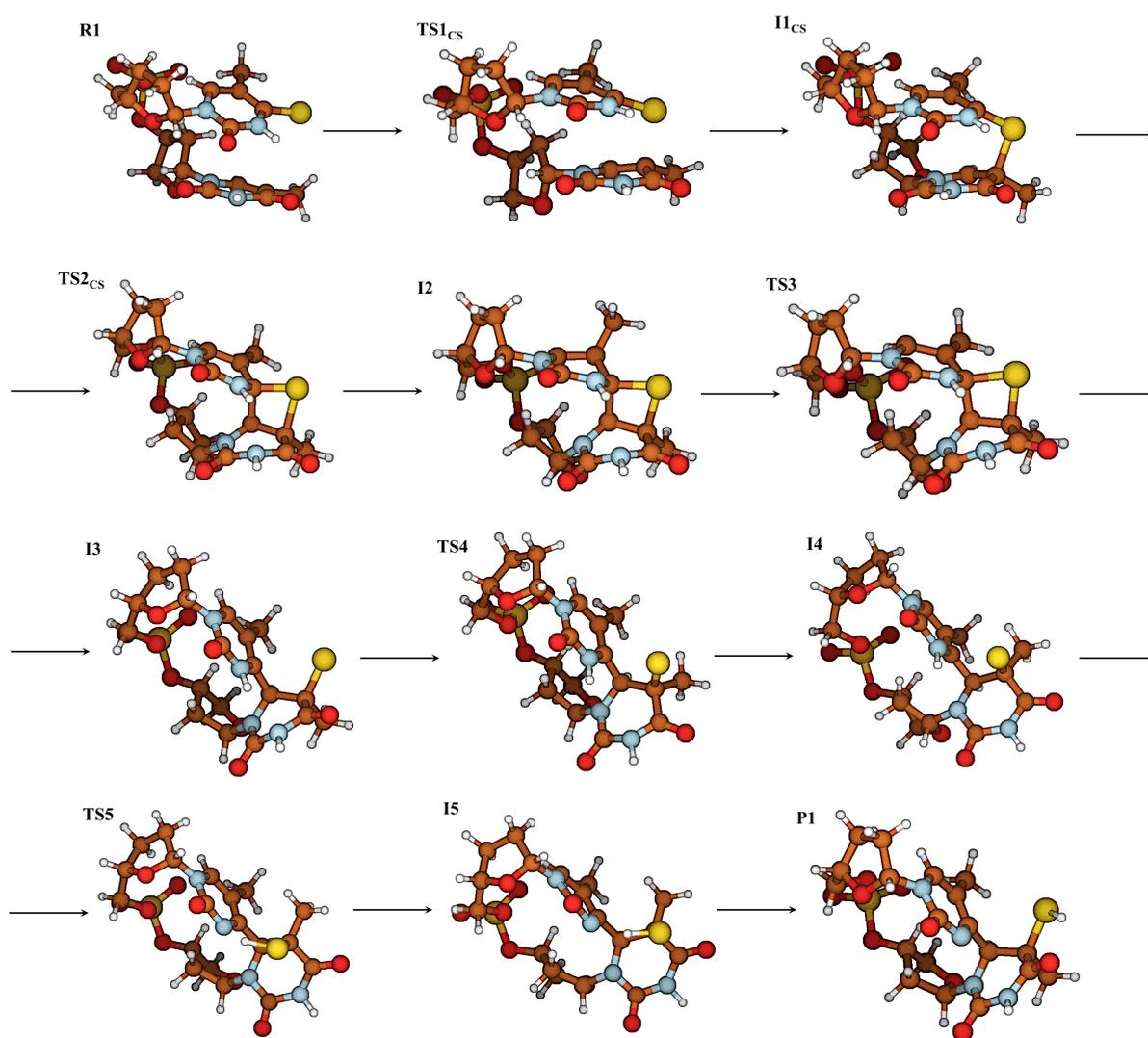


Figure 3.15: B3LYP-D3/*aug-cc-pVDZ* stationary points for 64-PP formation mechanism along lowest triplet excited state for the CS-CC triplet mechanism.

The approach concludes in the formation of the bond between these atoms ($I1_{CS}$). In a second step the $C_6 - C'_4$ bond takes place, resulting in the thietane intermediate ($I2$). In summary, the thiethane generation along the present mechanism is initialized by the $C - S'$ bond formation, and then the ring is closed by the formation of the bond between the C_6 of thymine and the C'_4 of the thiothymine. The differences with the CC-CS triplet mechanism just affects the first steps of the reaction, once the thietane, $I2$, is generated,

the two mechanisms follow the same route.

The potential energy profile for the CS-CC triplet mechanism at B3LYP-D3/*aug-cc-pVDZ* level are depicted in Figure 3.16. This figure also collects the equivalent energy profiles calculated at CASSCF(12,11)/*ANO-RCC-VDZ* and also CASPT2(12,11)/*ANO-RCC-VDZ* level of theory.

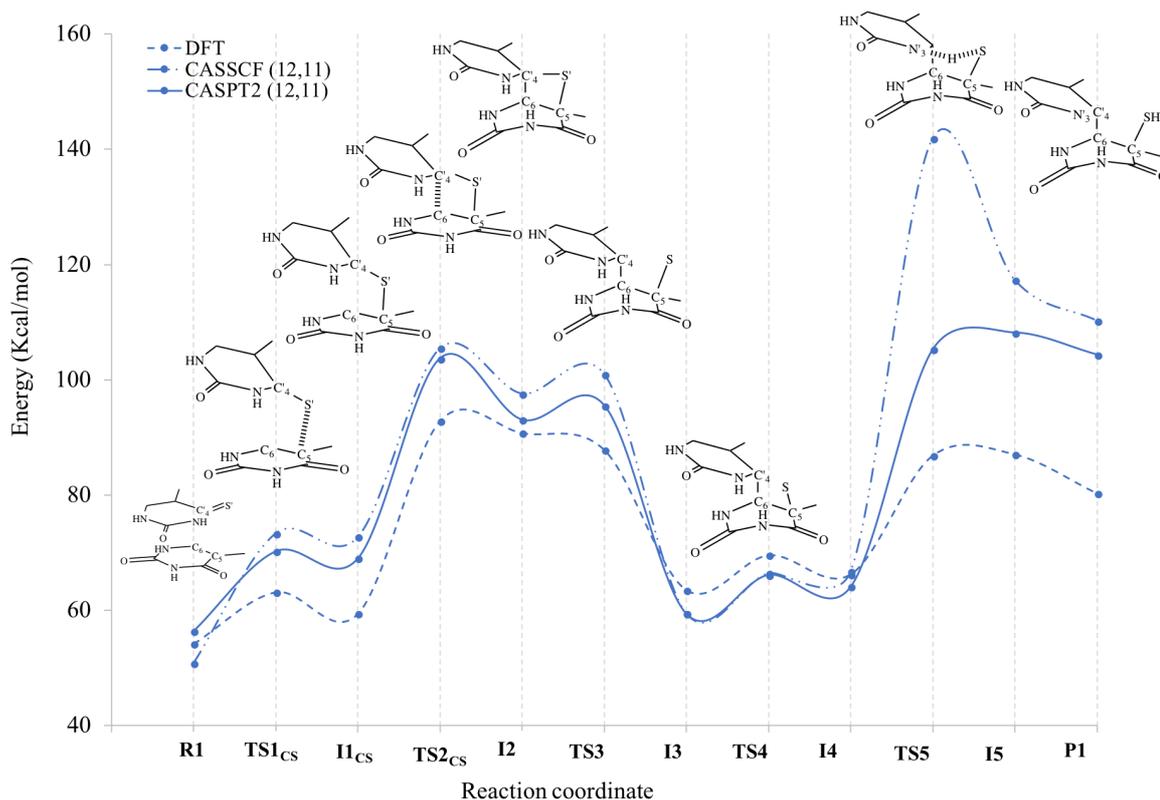


Figure 3.16: Potential energy profiles of T_1 state for the CS-CC triplet mechanism at B3LYP-D3/*aug-cc-pVDZ* (blue dashed line), CASSCF(12,11)/*ANO-RCC-VDZ* (blue dotted line) and CASPT2(12,11)/*ANO-RCC-VDZ* (blue solid line) level of theory. Energies in Kcal/mol relative to equilibrium geometry SR1 in S_0 .

The comparison of the CS-CC (Figure 3.16) and CC-CS (Figure 3.11) potential energy profiles, reveals some important differences in the energetics of thietane formation. While the value of the energy barrier to form the $C_5 - S'$ bond in CS-CC triplet mechanism is 8.9 Kcal/mol at B3LYP-D3/*aug-cc-pVDZ*, the value of the energy barrier to form the $C - C$ bond in CC-CS triplet mechanism amounts to 19.9 Kcal/mol at same level. However, the situation is completely opposite attending to the CASPT2 results. In this case, the energy barrier to form the $C - S$ bond in CS-CC triplet path is 13.8 Kcal/mol at CASPT2(12,11)/*ANO-RCC-VDZ*, while for the $C - C$ bond in CC-CS triplet mechanism is 10.8 Kcal/mol. While in TD-DFT calculation the formation of the thietane along CS-CC triplet mechanism is energetically more favored, the CASPT2 method provides that

the formation of the $C - C$ bond in CC-CS triplet mechanism is lower in energy than $C - S$ in CS-CC path. The disagreement in the energetics of the mechanisms of 64-PP formation along the lowest excited triplet paths using different methods, indicates the necessity to take into account the multiconfigurationa nature of the excited states.

Since the last part of Figure 3.16 is exactly equal than Figure 3.11, the conclusion drawn are the same. The DFT method provides good results for the optimizations in the lowest triplet state, despite underestimating certain energy barriers. The worst results are obtained by CASSCF (12,11) method, as it does not include a large part of the correlation energy. The high energy barriers of the triplet mechanism prevents the completion of the 64-PP reaction just taking into consideration this state.

The study of the excited state profiles along the coordinate defined by the CS-CC triplet mechanism could provide relevant information about the decay mechanism which would conclude in the 64-PP formation. Considering the optimized geometries of Figure 3.15, the vertical excitations including states of same and different multiplicity have been calculate using TD-CAMB3LYP/*cc-pVDZ* method, Figure 3.17.

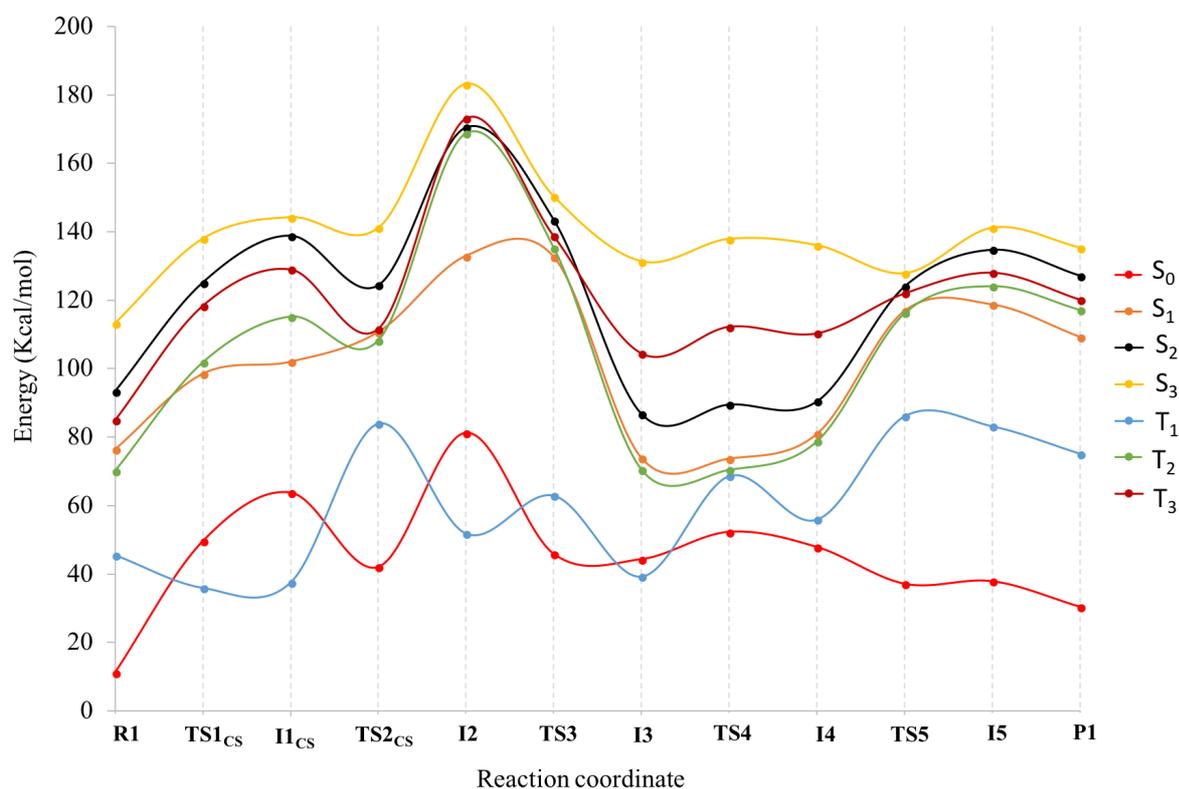


Figure 3.17: TD-CAMB3LYP/*cc-pVDZ* potential energy profiles of ground state and excited states calculated along the coordinate of the CS-CC triplet reaction mechanism leading to 64-PP. Energies in Kcal/mol relative to the equilibrium minima SR1 in S_0 .

The excited state profiles present the particular ascending behavior already observed throughout the study. This fact indicates again that the coordinates selected for the

analysis of the feature of the PES of the excited states paths are not the appropriated ones, as the excited states have a completely different decay route compared to the lowest triplet state or the ground state. A special feature of the PES appears in Figure 3.17. Attending to the energies of the T_1 optimized stationary points in Figure 3.16, the $TS1_{CS}$ transition state presents a higher energy than the two minima which it connects. The situation is different to the TD-CAMB3LYP calculation, Figure 3.17, where the transition state $TS1_{CS}$ shows a lower energy than the initial minimum R1 and the next point $I1_{CS}$. This surprising behavior suggests that the TD-DFT method does not offer reliable results. Additionally, with the TD-DFT, the T_1 state takes lower energy values than the ground state, in some regions of the PES. The deficiencies of TD-CAMB3LYP in the study of this mechanism, make impossible to extract reliable information from these results.

It is expected that CASPT2(12,11)/ANO-RCC-VDZ method provides more reliable results. Figure 3.18 compiles the potential energy profiles from the ground state and the excited states calculated along the coordinate defined by the CS-CC triplet mechanism at CASPT2 method.

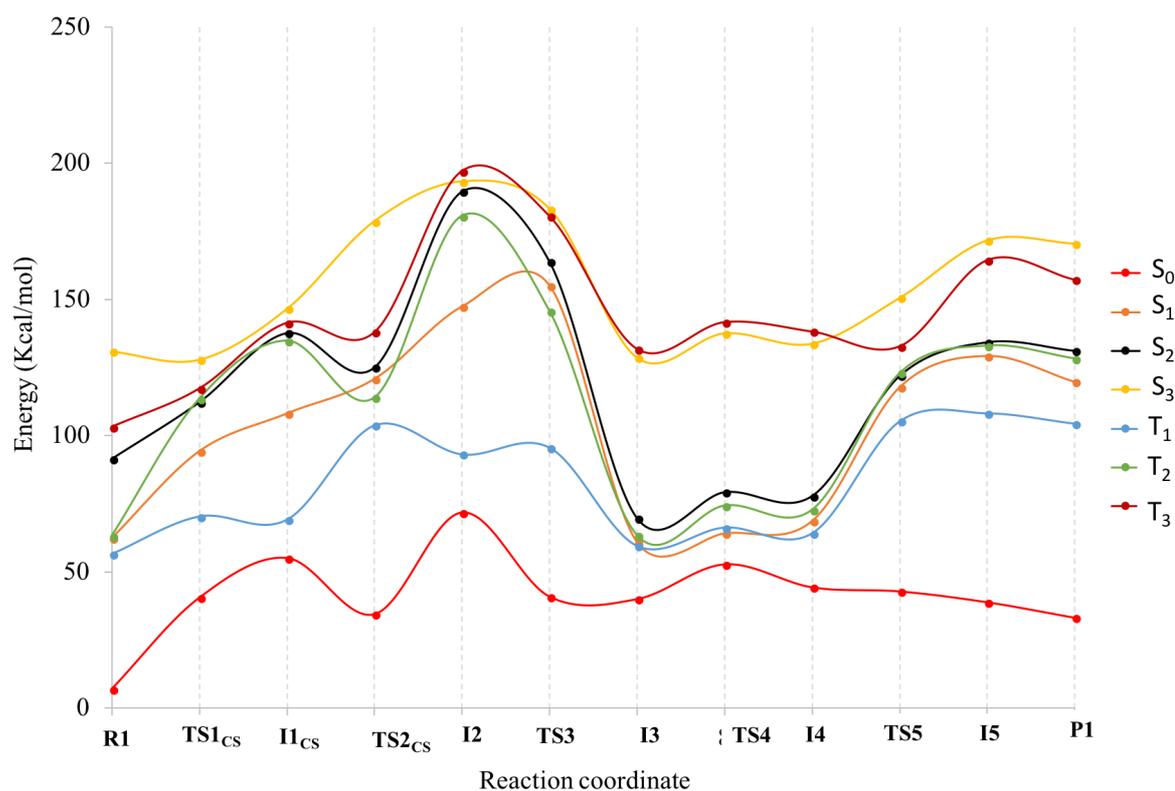


Figure 3.18: CASPT2(12,11)/ANO-RCC-VDZ potential energy profile of ground, singlet and triplet excited states calculated along the global coordinate defined by the CS-CC T_1 mechanism. Energy in Kcal/mol relative to the most stable geometry of S_0 mechanism, SR1.

Comparing the CASPT2 and the TD-DFT results, it is evident that the two methods

predict a different behavior for the lowest triplet state. At CASPT2 level, the T_1 state lies always above the ground state. In addition, the $TS1_{CS}$ has a higher energy than the two contiguous minima. Similarly to TD-DFT, however, the CASPT2 also predicts potential energy profiles for the excited states that grow in energy throughout the pathway, indicating that other coordinates might be much more appropriate for the correct description of excited states. Considering that the population of triplets is possible from the singlet manifold, an interesting point to analyze is $I1_{CS}$. Similarly to the $I1_{CC}$, this stationary point is a minimum for T_1 potential and apparently a transition state for S_0 , this fact would facilitate the transfer of the population to ground state. However, this situation seems to be more favored in the CC-CS triplet mechanism. From the analysis of Figure 3.18, it is appreciated that the state T_1 and S_0 are energetically further away in energy, being the energy difference between them of 14 Kcal/mol. Moreover, the value of the SOC in this point is rather low 6.6 cm^{-1} . All these factors turn less probable the transfer of the population at this particular point of the PES, compared to the $I1_{CC}$ point in the CC-CS triplet mechanism. However, it cannot be discarded that there exists other points in the PES where ISC might be efficient. Since the probability of reaching the ground state through ISC with triplet state seems to be very small along CS-CC triplet mechanism, and the development of the reaction along the lowest triplet state is impossible due to the high energy barrier, see Figure 3.16, the formation of the 64-PP via CS-CC triplet path appears to be more complicated than through the CC-CS triplet mechanism.

4. Conclusions

The present work has been focused on the mechanistic study of one of the most common crosslinking photolesions produced by DNA light exposure, the pyrimidine-(6,4)-pyrimidone dimer (64-PP). With the aim of contributing to understand the 64-PP generation mechanism, the thymine-4-thiothymine dinucleoside monophosphate has been used as a model. The incorporation of the thionucleobase to the model has been a key point of the study, not only because of the alteration in the photophysical and photochemical properties that the substitution of the carbonyl group by a thiocarbonyl provokes in the molecule, but also attending to the relevant role that excited states will have in the reaction due to the heavy atom effect.

The complete mechanism of the 64-PP formation cannot take place spontaneously along the ground state, due to the extremely high barriers that along this path have been found at B3LYP-D3/*aug-cc-pVDZ* level, but also at CASSCF(12,11)/*ANO-RCC-VDZ* and CASPT2(12,11)/*ANO-RCC-VDZ* levels of theory. Neither possible, is the development of the reaction only through the lowest triplet state, which presents lower energy barriers than S_0 , but in any case too high to be surpassed. The study of the PESs for the 64-PP formation reaction with different methods provides a way to assess the performance of different theoretical approaches in the description of these mechanisms. Despite some energy barriers are underestimated by DFT, the results obtained with this method for the ground and also for the lowest triplet state are in a good agreement with those calculated at CASPT2 level. No doubt, the use of multiconfigurational methods allows the possibility of improving the DFT results, introducing to the wave function significant flexibility. However, the CASSCF method overestimates, to a large extent, the energetic barriers between minima as this method does not include an important component of the total electron correlation. The discrepancies with CASPT2 indicate that the incorporation of dynamical correlation, through second order perturbation theory, results indispensable to achieve an accurate description of the potential energy profiles.

Considering the unfeasibility of the S_0 and T_1 the 64-PP, as a first approach to the problem we have analyzed the potential energy profiles of excited states along the global

coordinates defined by the S_0 and T_1 mechanism. To this purpose the time-dependent formalism of DFT has been used, in first place, to determine a potential excited state deactivation paths. The disagreement found between TD-DFT and DFT approaches for the ground and lowest triplet state confirm that the study of the excited states using TD-DFT can provide unreliable results. Therefore, multiconfigurational methods are really necessary to get accurate results which can be used to determine a potential relaxation pathway from the spectroscopic state.

Similarly to TD-DFT, the calculations at CASPT2(12,11)/*ANO-RCC-VDZ* evaluates the S_2 as the lowest spectroscopic state with an energy of 100 Kcal/mol. Unfortunately, both CASPT2 and TD-DFT show upward potential energy profiles for the lowest lying singlet and triplet states along the coordinates defined by the S_0 and T_1 mechanisms. With these results, it can be concluded that the selected coordinates are not the appropriate ones to get information about relaxation path along the excited states. In fact, the use of these coordinates, has made it impossible to determinate an potential decay mechanism for the dimer generation.

The proximity in the energy, that several states of different multiplicities present in both paths, could be considered an indirect proof that demonstrates that it is energetically possible to populate triplet states from singlet excited states. In addition, the significant values of the spin orbit coupling, that can amount to 160 cm^{-1} , support the idea that ISC is possible from singlet excited states.

These findings justify the study of the 64-PP formation along the T_1 potential where two different pathways were investigated, the CC-CS and the CS-CC mechanism. While the formation of thiethane along the S_0 takes place through a concerted [2+2] cycloaddition, the mechanism along the T_1 is stepwise. The analysis of the stationary points of both triplet paths at CASPT2(12,11)/*ANO-RCC-VDZ* level, indicates that the formation of the $C_6 - C_4'$ bond followed by the $C_5 - S'$ bond formation, that is the CC-CS mechanism, is energetically favored. Moreover the calculated SOCs along this pathway are also larger. For these very preliminary evidences, it could be concluded that the 64-PP formation would be favored along the CC-CS triplet path. However, this conclusion only holds along the coordinates considered so far, that is, the global coordinates connecting the stacked dimer with the final 64-PP along the T_1 potential. Undertaking minimum energy path calculations from the spectroscopic state will provide more reliable mechanistic insight for the dimerization reaction.

Besides the possible deactivation path along triplet excited states, a decay just involving singlet states cannot be discarded. The analysis of the potential energy profiles point to the existence of energetically accessible conical intersections that would allow the decay of the population to the S_0 . From the lowest singlet state the population could be transferred to the ground state through STS1 position, where a bifurcation of the system population would occur towards initial reactants or the intermediate thiethane.

Outlook

Despite the results presented in this work do not offer a complete explanation of the 64-PP formation mechanism, they provide very useful information to continue the exploration of 64-PP generation mechanism. First, the optimized geometries at B3LYP-D3/*aug-cc-pDVZ* will be recalculated using multiconfigurational method, i.e. CASSCF method. Although, the given geometries at DFT level expected to be reliable, the multiconfigurational character of CASSCF might affect to the structures. However, due to the large size of the molecule of interest, these multiconfigurational calculations can turn out to be computationally very expensive. For this reason, the use of another type of approximation like truncated Coupled Cluster (CC2) or Algebraic Diagram Construction (ADC2) could be an alternative to multiconfigurational methods which would be interesting to investigate. In addition, relevant deactivation funnels such as conical intersection and intersystem crossing areas, which have been determined in this work, will be carefully studied, optimizing them at multiconfigurational level and analyzing their potential as deactivation channels.

As a first approach to the problem we have analyzed the potential energy profiles along the S_0 and T_1 mechanism, easier to characterize with standard DFT. In the future, minimum energy path calculations will carry out from excited states. With these MEPs, a reliable exploration of the excited state PES accurately determines the most probable deactivation pathways activated with light. The thorough examination of the PES through minimum energy path might be useful to reveal many other alternative deactivation funnels including conical intersection or intersystem crossing, which will need to be studied in detail.

A similar mechanistic study as the one performed for 64-PP will be performed also for the cyclobutadiene dimer, recurring at the same molecular model than in this work. The comparison of both mechanisms will allow explaining the very different of quantum yields recorded for that the two photolesions occurring in natural DNA. As a more extensive overlook, the static study will complemented with a time resolved study, using molecular dynamics, which will give accurate information of the different potential decay paths.

Our simulations on the DNA damaging species generated from long-lived excited states upon UVA light exposure are intended to contribute to the developing investigations of novel prospectives of these drugs for photochemotherapeutic applications.

References

1. Karran, P. Thiopurines, DNA damage, DNA repair and therapy-related cancer. *British medical bulletin* **79-80**, 153–70. ISSN: 0007-1420 (2006).
2. Pollum, M., Jockusch, S. & Crespo-Hernández, C. E. 2,4-Dithiothymine as a potent UVA chemotherapeutic agent. *Journal of the American Chemical Society* **136**, 17930–3. ISSN: 1520-5126 (2014).
3. Aarbakke, J., Janka-Schaub, G & Elion, G. B. Thiopurine biology and pharmacology. *Trends in pharmacological sciences* **18**, 3–7. ISSN: 0165-6147 (1997).
4. Karran, P. & Attard, N. Thiopurines in current medical practice: molecular mechanisms and contributions to therapy-related cancer. *Nature reviews. Cancer* **8**, 24–36. ISSN: 1474-1768 (2008).
5. Massey, A., Xu, Y.-Z. & Karran, P. Photoactivation of DNA thiobases as a potential novel therapeutic option. *Current Biology* **11**, 1142–1146. ISSN: 0960-9822 (2001).
6. Martínez-Fernández, L., González, L. & Corral, I. An ab initio mechanism for efficient population of triplet states in cytotoxic sulfur substituted DNA bases: the case of 6-thioguanine. *Chemical Communications* **48**, 2134. ISSN: 1359-7345 (2012).
7. Serrano-Andrés, L. & Merchán, M. Are the five natural DNA/RNA base monomers a good choice from natural selection? *Journal of Photochemistry and Photobiology C: Photochemistry Reviews* **10**, 21–32. ISSN: 13895567 (2009).
8. Pollum, M., Martínez-Fernández, L. & Crespo-Hernández, C. E. in (eds Barbatti, M., Borin, A. C. & Ullrich, S.) 245–327 (Springer International Publishing, Cham, 2015). ISBN: 978-3-319-13371-3.
9. DeRosa, M. C. & Crutchley, R. J. Photosensitized singlet oxygen and its applications. *Coordination Chemistry Reviews* **233-234**, 351–371. ISSN: 00108545 (2002).
10. Harada, Y., Suzuki, T., Ichimura, T. & Xu, Y.-Z. Triplet Formation of 4-Thiothymidine and Its Photosensitization to Oxygen Studied by Time-Resolved Thermal Lensing Technique. *The Journal of Physical Chemistry B* **111**, 5518–5524. ISSN: 1520-6106 (2007).
11. Reelfs, O., Karran, P. & Young, A. R. 4-thiothymidine sensitization of DNA to UVA offers potential for a novel photochemotherapy. *Photochemical photobiological sciences Official journal of the European Photochemistry Association and the European Society for Photobiology* **11**, 1–7. ISSN: 1474-9092 (2012).
12. Wondrak, G. T., Jacobson, M. K. & Jacobson, E. L. Endogenous UVA-photosensitizers: mediators of skin photodamage and novel targets for skin photoprotection. *Photochem. Photobiol. Sci.* **5**, 215–237. ISSN: 1474-905X (2006).

13. Pfeifer, G. P. Formation and processing of UV photoproducts: effects of DNA sequence and chromatin environment. *Photochemistry and photobiology* **65**, 270–83. ISSN: 0031-8655 (1997).
14. Cadet, J & Vigny, P. The Photochemistry of Nucleic Acids. In Bioorganic Photochemistry; Morrison, H., Ed.; John Wiley & Sons. Inc.: New York **1**, 1 (1990).
15. Durbeej, B. & Eriksson, L. A. On the Formation of Cyclobutane Pyrimidine Dimers in UV-irradiated DNA: Why are Thymines More Reactive? *Photochemistry and Photobiology* **78**, 159–167. ISSN: 00318655 (2007).
16. Sage, E. Distribution and repair of photolesions in DNA: genetic consequences and the role of sequence context. *Photochemistry and photobiology* **57**, 163–74. ISSN: 0031-8655 (1993).
17. Warren, M. A., Murray, J. B. & Connolly, B. A. Synthesis and characterisation of oligodeoxynucleotides containing thio analogues of (6-4) pyrimidine-pyrimidinone photo-dimers. *Journal of molecular biology* **279**, 89–100. ISSN: 0022-2836 (1998).
18. Perdiz, D. *et al.* Distribution and repair of bipyrimidine photoproducts in solar UV-irradiated mammalian cells. Possible role of Dewar photoproducts in solar mutagenesis. *Journal of Biological Chemistry* **275**, 26732–42. ISSN: 00219258 (2000).
19. Marguet, S. & Markovitsi, D. Time-Resolved Study of Thymine Dimer Formation. *Journal of the American Chemical Society* **127**, 5780–5781. ISSN: 0002-7863 (2005).
20. McCullagh, M. *et al.* Conformational Control of TT Dimerization in DNA Conjugates. A Molecular Dynamics Study. *The Journal of Physical Chemistry B* **114**, 5215–5221. ISSN: 1520-6106 (2010).
21. Douki, T. & Cadet, J. Individual Determination of the Yield of the Main UV-Induced Dimeric Pyrimidine Photoproducts in DNA Suggests a High Mutagenicity of CC Photolesions. *Biochemistry* **40**, 2495–2501 (2001).
22. Durbeej, B & Eriksson, L. A. Reaction mechanism of thymine dimer formation in DNA induced by UV light. *J. Photochem. Photobiol., A* **152**, 95 (2002).
23. Boggio-Pasqua, M., Groenhof, G., Schäfer, L. V., Grubmüller, H. & Robb, M. A. Ultrafast deactivation channel for thymine dimerization. *Journal of the American Chemical Society* **129**, 10996–7. ISSN: 0002-7863 (2007).
24. Schreier, W. J. *et al.* Thymine dimerization in DNA is an ultrafast photoreaction. *Science* **315**, 625 (2007).
25. Blancafort, L. & Migani, A. Modeling thymine photodimerizations in DNA: Mechanism and correlation diagrams. *Journal of the American Chemical Society* **129**, 14540–14541. ISSN: 00027863 (2007).
26. Cadet, J., Mouret, S., Ravanat, J.-L. & Douki, T. Photoinduced Damage to Cellular DNA: Direct and Photosensitized Reactions. *Photochemistry and Photobiology* **88**, 1048–1065. ISSN: 00318655 (2012).
27. Bo Yang, Z., Eriksson, L. A. & bo Zhang, R. A Theoretical Rationale for Why Azetidine Has a Faster Rate of Formation Than Oxetane in TC(6-4) Photoproducts. *The Journal of Physical Chemistry B* **115**, 9681–9686. ISSN: 1520-6106 (2011).
28. Paterno, E & Chieffì, G. Syntheses in organic chemistry using light. Note II. Compounds of unsaturated hydrocarbons with aldehydes and ketones. *Gazz. Chim. Ital.* **39**, 341–361 (1909).

29. Büchi, G., Inman, C. G. & Lipinsky, E. S. Light-catalyzed Organic Reactions. I. The Reaction of Carbonyl Compounds with 2-Methyl-2-butene in the Presence of Ultraviolet Light. *Journal of the American Chemical Society* **76**, 4327–4331. ISSN: 0002-7863 (1954).
30. Palmer, I. J., Ragazos, I. N., Bernardi, F., Olivucci, M. & Robb, M. A. An MC-SCF Study of the (Photochemical) Paterno-Buchi Reaction. *Journal of the American Chemical Society* **116**, 2121–2132. ISSN: 0002-7863 (1994).
31. Kim, S. T., Malhotra, K, Smith, C. A., Taylor, J. S. & Sancar, A. Characterization of (6-4) photo-product DNA photolyase. *The Journal of biological chemistry* **269**, 8535–40. ISSN: 0021-9258 (1994).
32. Giussani, A., Serrano-Andrés, L., Merchán, M., Roca-Sanjuán, D. & Garavelli, M. Photoinduced formation mechanism of the thymine-thymine (6-4) adduct. *The journal of physical chemistry. B* **117**, 1999–2004. ISSN: 1520-5207 (2013).
33. Bo Yang, Z., bo Zhang, R. & Eriksson, L. A. A triplet mechanism for the formation of thymine-thymine (6-4) dimers in UV-irradiated DNA. en. *Physical chemistry chemical physics : PCCP* **13**, 8961–6. ISSN: 1463-9084 (2011).
34. Serrano-Pérez, J. J., Gonzalez-Luque, R., Merchan, M. & Serrano-Andrés, L. On the intrinsic population of the lowest triplet state of thymine. *The journal of physical chemistry. B* **111**, 11880–3. ISSN: 1520-6106 (2007).
35. Banyasz, A. *et al.* Electronic Excited States Responsible for Dimer Formation upon UV Absorption Directly by Thymine Strands: Joint Experimental and Theoretical Study. *Journal of the American Chemical Society* **134**, 14834–14845. ISSN: 0002-7863 (2012).
36. Mouret, S. *et al.* UVA-induced cyclobutane pyrimidine dimers in DNA: a direct photochemical mechanism? *Organic & Biomolecular Chemistry* **8**, 1706. ISSN: 1477-0520 (2010).
37. Francisco Bosca, Virginie Lhiaubet-Vallet, M. Consuelo Cuquerella, Jose V. Castell, and Miguel A. Miranda. The Triplet Energy of Thymine in DNA. doi:10.1021/JA060651G (2006).
38. Schrödinger, E. An Undulatory Theory of the Mechanics of Atoms and Molecules. *Physical Review* **28**, 1049–1070. ISSN: 0031-899X (1926).
39. Rayleigh, J. W. S. & Lindsay, R. B. *The theory of sound*, ISBN: 0486602923 (Dover, 1945).
40. Møller, C & Plesset, M. S. Note on an Approximation Treatment for Many-Electron Systems. *Phys. Rev.* **46**, 618 (1934).
41. Hartree, D. R. The Wave Mechanics of an Atom with a Non-Coulomb Central Field. Part I. Theory and Methods. *Mathematical Proceedings of the Cambridge Philosophical Society* **24**, 89. ISSN: 0305-0041 (1928).
42. Fock, V. Näherungsmethode zur Lösung des quantenmechanischen Mehrkörperproblems. *Zeitschrift für Physik* **61**, 126–148. ISSN: 1434-6001 (1930).
43. Roothaan, C. C. J. New Developments in Molecular Orbital Theory. *Reviews of Modern Physics* **23**, 69–89. ISSN: 0034-6861 (1951).
44. Hall, G. G. The Molecular Orbital Theory of Chemical Valency. VIII. A Method of Calculating Ionization Potentials. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences* **205**, 541–552. ISSN: 1364-5021 (1951).
45. Pople, J. A. & Nesbet, R. K. Self Consistent Orbitals for Radicals. *The Journal of Chemical Physics* **22**, 571–572. ISSN: 0021-9606 (1954).

46. Langhoff, S. R. & Davidson, E. R. Configuration interaction calculations on the nitrogen molecule. *International Journal of Quantum Chemistry* **8**, 61–72. ISSN: 0020-7608 (1974).
47. Siegbahn, P. E. M., Almlöf, J., Heiberg, A. & Roos, B. O. The complete active space SCF (CASSCF) method in a Newton-Raphson formulation with application to the HNO molecule. *The Journal of Chemical Physics* **74**, 2384–2396. ISSN: 0021-9606 (1981).
48. Werner, H.-J. in, 1–62 (Wiley-Blackwell, 2007). doi:10.1002/9780470142943.ch1.
49. Shepard, R. in, 63–200 (Wiley-Blackwell, 2007). doi:10.1002/9780470142943.ch2.
50. Roos, B. O., Taylor, P. R. & Siegbahn, P. E. A complete active space SCF method (CASSCF) using a density matrix formulated super-CI approach. *Chemical Physics* **48**, 157–173. ISSN: 0301-0104 (1980).
51. Roos, B. O. The complete active space SCF method in a fock-matrix-based super-CI formulation. *International Journal of Quantum Chemistry* **18**, 175–189. ISSN: 00207608 (2009).
52. Andersson, K., Malmqvist, P. A., Roos, B. O., Sadlej, A. J. & Wolinski, K. Second-order perturbation theory with a CASSCF reference function. *The Journal of Physical Chemistry* **94**, 5483–5488. ISSN: 0022-3654 (1990).
53. Andersson, K., Malmqvist, P. & Roos, B. O. Second-order perturbation theory with a complete active space self-consistent field reference function. *The Journal of Chemical Physics* **96**, 1218–1226. ISSN: 0021-9606 (1992).
54. Roos, B. O. & Andersson, K. Multiconfigurational perturbation theory with level shift - the Cr 2 potential revisited. *Chemical Physics Letters* **245**, 215–223 (1995).
55. Roos, B. O. *et al.* Applications of level shift corrected perturbation theory in electronic spectroscopy. *Journal of Molecular Structure: THEOCHEM* **388**, 257–276. ISSN: 0166-1280 (1996).
56. Forsberg, N. & Malmqvist, P.-Å. Multiconfiguration perturbation theory with imaginary level shift. *Chemical Physics Letters* **274**, 196–204. ISSN: 0009-2614 (1997).
57. Finley, J., Malmqvist, P.-Å., Roos, B. O. & Serrano-Andrés, L. The multi-state CASPT2 method. *Chemical Physics Letters* **288**, 299–306. ISSN: 0009-2614 (1998).
58. Ghigo, G., Roos, B. O. & Malmqvist, P.-Å. A modified definition of the zeroth-order Hamiltonian in multiconfigurational perturbation theory (CASPT2). *Chemical Physics Letters* **396**, 142–149. ISSN: 0009-2614 (2004).
59. Hohenberg, P. & Kohn, W. Inhomogeneous Electron Gas. *Physical Review* **136**, B864–B871. ISSN: 0031-899X (1964).
60. Kohn, W. & Sham, L. J. Self-Consistent Equations Including Exchange and Correlation Effects. *Physical Review* **140**, A1133–A1138. ISSN: 0031-899X (1965).
61. Levy, M. Electron densities in search of Hamiltonians. *Physical Review A* **26**, 1200–1208. ISSN: 0556-2791 (1982).
62. Ceperley, D. M. & Alder, B. J. Ground State of the Electron Gas by a Stochastic Method. *Physical Review Letters* **45**, 566–569. ISSN: 0031-9007 (1980).
63. Becke, A. D. Density-functional exchange-energy approximation with correct asymptotic behavior. *Physical Review A* **38**, 3098–3100. ISSN: 0556-2791 (1988).
64. Perdew, J. P., Burke, K & Ernzerhof, M. Generalized Gradient Approximation Made Simple. *Phys. Rev. Lett.* **77**, 3865 (1996).

65. Lee, C., Yang, W. & Parr, R. G. Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density. *Physical Review B* **37**, 785–789. ISSN: 0163-1829 (1988).
66. Zhao, Y. & Truhlar, D. G. A new local density functional for main-group thermochemistry, transition metal bonding, thermochemical kinetics, and noncovalent interactions. *The Journal of Chemical Physics* **125**, 194101. ISSN: 0021-9606 (2006).
67. Becke, A. D. Density-functional thermochemistry. III. The role of exact exchange. *The Journal of Chemical Physics* **98**, 5648. ISSN: 00219606 (1993).
68. Grimme, S., Antony, J., Ehrlich, S. & Krieg, H. A consistent and accurate ab initio parametrization of density functional dispersion correction (DFT-D) for the 94 elements H-Pu. *The Journal of chemical physics* **132**, 154104. ISSN: 1089-7690 (2010).
69. Iikura, H., Tsuneda, T., Yanai, T. & Hirao, K. A long-range correction scheme for generalized-gradient-approximation exchange functionals. *The Journal of Chemical Physics* **115**, 1. doi:10.1063/1.1383587 (2001).
70. Leininger, T., Stoll, H., Werner, H.-J. & Savin, A. Combining long-range configuration interaction with short-range density functionals. *29 August 1997 Chemical Physics Letters* **275**, 151–160 (1997).
71. Runge, E. & Gross, E. K. U. Density-Functional Theory for Time-Dependent Systems. *Physical Review Letters* **52**, 997–1000. ISSN: 0031-9007 (1984).
72. Casida, M. E. Time-dependent density-functional response theory for molecules.
73. Tawada, Y., Tsuneda, T., Yanagisawa, S., Yanai, T. & Hirao, K. A long-range-corrected time-dependent density functional theory. *The Journal of Chemical Physics* **120**, 8425–8433. ISSN: 0021-9606 (2004).
74. Slater, J. C. Note on Hartree's Method. *Physical Review* **35**, 210–211. ISSN: 0031-899X (1930).
75. Boys, S. F. Electronic Wave Functions. I. A General Method of Calculation for the Stationary States of Any Molecular System. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences* **200**, 542–554. ISSN: 1364-5021 (1950).
76. Wiberg, K. B. Ab Initio Molecular Orbital Theory by W. J. Hehre, L. Radom, P. v. R. Schleyer, and J. A. Pople, John Wiley, New York, 548pp. Price: \$79.95 (1986). *Journal of Computational Chemistry* **7**, 379–379 (1986).
77. Ditchfield, R., Hehre, W. J. & Pople, J. A. Self-Consistent Molecular-Orbital Methods. IX. An Extended Gaussian-Type Basis for Molecular-Orbital Studies of Organic Molecules. *The Journal of Chemical Physics* **54**, 724–728. ISSN: 0021-9606 (1971).
78. Pierloot, K., Dumez, B., Widmark, P.-O. & Roos, B. O. Density matrix averaged atomic natural orbital (ANO) basis sets for correlated molecular wave functions. *Theoretica Chimica Acta* **90**, 87–114. ISSN: 0040-5744 (1995).
79. Woon, D. E. & Dunning, T. H. Gaussian basis sets for use in correlated molecular calculations. IV. Calculation of static electrical response properties. *The Journal of Chemical Physics* **100**, 2975. ISSN: 00219606 (1994).
80. Woon, D. E. & Dunning, T. H. Gaussian basis sets for use in correlated molecular calculations. V. Core-valence basis sets for boron through neon. *The Journal of Chemical Physics* **103**, 4572–4585. ISSN: 0021-9606 (1995).

81. Kendall, R. A., Dunning, T. H. & Harrison, R. J. Electron affinities of the first-row atoms revisited. Systematic basis sets and wave functions. *The Journal of Chemical Physics* **96**, 6796–6806. ISSN: 0021-9606 (1992).
82. Neese, F. The ORCA program system. *Wiley Interdisciplinary Reviews: Computational Molecular Science* **2**, 73–78 (2012).
83. Vahtras, O., Almlöf, J. & Feyereisen, M. Integral approximations for LCAO-SCF calculations. *Chemical Physics Letters* **213**, 514–518. ISSN: 0009-2614 (1993).
84. Neese, F., Wennmohs, F., Hansen, A. & Becker, U. Efficient, approximate and parallel Hartree-Fock and hybrid DFT calculations. A 'chain-of-spheres' algorithm for the Hartree-Fock exchange. *Chemical Physics* **356**, 98–109. ISSN: 03010104 (2009).
85. Dunning, T. H. Gaussian basis sets for use in correlated molecular calculations. I. The atoms boron through neon and hydrogen. *The Journal of Chemical Physics* **90**, 1007–1023. ISSN: 0021-9606 (1989).
86. Et al Frisch, M. *Gaussian Inc* 2009.
87. Yanai, T., Tew, D. P. & Handy, N. C. A new hybrid exchange-correlation functional using the Coulomb-attenuating method (CAM-B3LYP). *Chemical Physics Letters* **393**, 51–57. ISSN: 00092614 (2004).
88. Karlström, G. *et al.* MOLCAS: a program package for computational chemistry. *Computational Materials Science* **28**, 222–239. ISSN: 0927-0256 (2003).
89. Aquilante, F., Lindh, R. & Bondo Pedersen, T. Unbiased auxiliary basis sets for accurate two-electron integral approximations. *The Journal of Chemical Physics* **127**, 114107 (2007).

