Alma Mater Studiorum – Università di Bologna

DOTTORATO DI RICERCA IN

CHIMICA

Ciclo XXXI

Settore Concorsuale di afferenza: 03/C1

Settore Scientifico disciplinare: CHIM/06

Computational Investigations of Catalyzed Organic Reactions:
Carbocatalysis, Biocatalysis, Metal and Organo Catalysis

Presentata da: **Tainah Dorina Marforio**

Coordinatore Dottorato   **Relatore**

Chiar.mo Prof. Aldo Roda   **Chiar.mo Prof. Andrea Bottoni**

Esame finale anno 2018

Abstract

Catalysis is the key for efficiency in organic synthesis, where resource efficiency maximization, limitation of toxic chemicals and waste minimization are sought. A deep comprehension of catalyzed processes is the base to develop new and efficient catalytic technologies. Because of the complexity of many catalytic chemical processes, experimental techniques may produce ambiguous results for a given reaction mechanism. To supply complementary insights to significant chemical related problems, the research activity carried out during this PhD course focused on studying catalyzed reaction mechanisms by means of computational methods, employing Quantum-Mechanical (QM) or hybrid Quantum-Mechanical/Molecular Mechanical (QM/MM) approaches. The reaction paths of several unconventional (carbon nanoparticles) and conventional (enzymes, metal and proline) catalyzed reactions were investigated. In the former case, metal-free carbon-based catalysts can be considered, from an industrial standpoint, environmentally friendly and resource-saving. Oxidized carbon nanotubes (o-CNTs) were proved to perform as excellent catalysts in the oxidative dehydrogenation of ethylbenzene to styrene, and we unveiled, by means of QM computations, the direct role of oxygen-containing functionalities on the tube surface in the dehydrogenation of the substrate. CNTs can also be exploited as molecular *hosts* to confine and manipulate *guest* organic reactions. To disclose the catalytic and regioselective role of the cavity of CNTs, we investigated, at QM/MM level, the bromination of *N*-phenylacetamide inside a CNT, finding that van der Waals interactions are the key factors that determinate the acceleration of the confined reaction. More conventional catalysts are enzymes, and one of the roles ascribed to these systems is to provide the cell with energy. This function can be completed by breaking glycosidic bonds: glycogen and starch are among the most abundant polysaccharides and the enzymes responsible for the cleavage of the glycosidic bond are glycoside hydrolases (GHs). However, this cleavage can be carried out also by α-1,4-glucan lyases (GLases). Using a QM approach, we highlighted that the reaction proceeds *via* three stages: glycosylation, deglycosylation-elimination and keto/enol tautomerism and that the tautomerization must occur inside the enzyme before leaving the active site, rather than in the aqueous solution. Another polysaccharide involved in essential physiological functions, from regulating cell growth to blocking coagulation, is heparan sulfate (HS). The biosynthetic path of HS introduces various degrees of sulphonation by the action of specialized sulfotransferases such as 2-*O*-sulfotransferase (2OST), where the sulfuryl group is transferred from the donor PAPS (3'-phosphoadenosine-5'-phosphosulfate) to the sugar unit. By means of QM/MM computations we corroborated that 2OST prefers to perform sulfonation on iduronic acid rather than on glucuronic unit by a S_N2-like mechanism. Polysaccharides are also ideal candidates in the route towards a carbohydrate bio-based economy. Through pyrolysis, plant biopolymers can be broken apart into lower molecular weight compounds. Among the various products, the cellulose pyrolytic specie (1*R*,5*S*)-1-hydroxy-3,6-dioxabicyclo[3.2.1]octan-2-one (LAC) is a versatile scaffold towards novel substances. We unveiled a mechanistic scenario, by means of QM computations, where LAC can be obtained following two reaction paths: in both cases the rate-determining step of the process is the initial keto-enol isomerization. Also, we demonstrated that

water, which is present in the reaction mixture, can catalyse the reaction by assisting the proton transfers present in all the steps of the process. These results highlighted the importance of water in catalysing secondary pyrolysis reactions of cellulose leading to potential chemical platforms. Water also plays a crucial role in proline-based catalysis. Investigating the Hajos-Parrish-Sauder-Wiechert (HPESW) reaction we revealed the importance of water molecules that participate in the key transformations undergone by the reaction intermediates. Also, the water is the key factor for some relevant enzymatic processes, such as hydration and dehydration reactions. In conventional organic synthesis, hydration/dehydration of olefins/tertiary alcohols rely on using highly acid or basic chemical reagents. Because of the harsh conditions, the scope of candidates is limited if the substrate presents acid- or base-sensitive functionalities. The linalool dehydratase isomerase (LinD) reversibly catalyses the selective hydration of the monoterpenoid β-myrcene to the tertiary alcohol (*S*)-linalool and its further conversion into geraniol, suggesting that LinD can be exploited in chemoenzymatic approaches. Our preliminary QM and QM/MM results suggest that the reaction is more likely to proceed via the formation of an acid-base catalysis carried out by the active site residues rather than through the formation of a covalent or carbocationic intermediate. We also verified that the protonation state of cysteines in the active site strongly affects the course of the reaction. For specific reactions the chemo-enzymatic synthesis is a valid and sustainable alternative, but in many cases the use of metal is a judgment call. When this is the case, it is important to minimize the environmental impact of the sub-products and employ low-impact starting materials. The decarboxylative cross-coupling is a valid synthetic approach that attempts to limit toxic wastes by producing CO_2 as a by-product. A-oxocarboxylic acids (inexpensive and widely available substrates), in presence of Cu(OAc)$_2$, were reported to undergo decarboxylative cross-coupling in presence of isatin (1*H*-indole-2,3-dione). The reaction affords simultaneously the formation of the C-N and C-O bonds in the coupling product, 4*H*-[3,1]benzoxazin-4-one, an important aromatic amine scaffold in the synthesis of quinazolinone analogues. Our DFT computations clarified the reaction mechanism, which proceeds through four steps, where the decarboxylation process represents the rate-determining step of the whole reaction. Alongside with the study of catalytic processes, computational bio-nanotechnology investigations were carried out, employing Molecular Mechanics (MM) approaches. Hybrid assemblies, formed by carbon-based materials and biological systems, are now arising as innovative platforms for medical, nanotechnologies and material science applications. Computational design offers the opportunity of studying in detail these assemblies. An *in-silico* mutagenesis approach was adopted to engineer the lysozyme-fullerene C60 interface and understand the driving forces of the binding process. Computational classical approaches can also be exploited to supply insights for drug development, by evaluating the tendency of a candidate drug to bind a target receptor. An interesting example is the development of carborane-based drugs which are valid candidates to those drugs for which resistance has been developed. We investigated the energetic contributions that drive the binding of a carborane-based calcitriol analogue to the Vitamin D receptor (VDR), finding that the hydric hydrogens of carborane cage have such an active role in the binding process that VDR preferably binds the carborane cage rather than to the secosteroidal core of calcitriol.

A Mamma e Papà

Preface

This thesis comprises most of the computational results obtained during the three year Chemistry Ph.D course, during which the research activity mainly focused on investigating catalyzed reaction mechanisms. Catalysis is a vast phenomenon and this dissertation does not presume to cover its extent, nor to be a mere list of computational investigations. The purpose is to stress out that computational organic chemistry is a tool to be exploited to explain *how* and *why* some reactions prefer to cover one path rather than another by interpreting computation outcomes. Among the immense world of catalysts, the activity of carbon nanoparticles, enzymes, metal complex and proline in catalyzed organic reactions were investigated. Commercially available software for molecular computations were used to carry out the investigations. The thesis is divided in few parts; the former (Part I) provides some insights into the computational methods used during the PhD activity where Quantum-Mechanics (QM), Molecular-Mechanics (MM) and hybrid QM/MM methods are briefly described.

Parts II-V gather the results obtained during the PhD activity. In Part II, two examples of carbo-catalyzed reactions are reported, while Part III and IV collect the computational evidences achieved by analysing the reaction mechanism of enzymatic and metal- (or organo-) catalyzed reactions respectively. The last Part (V) includes two, among the many, side works that were carried out during the PhD course. A résumé of the computational results is reported in Part VI.

Contents

Introduction

Catalysis has been used for millennia long before the concept of catalysts was rationalized by Berzelius in 1835, which stated that catalysts are "*simple or compound bodies, soluble and insoluble, [that] have the property of exercising on other bodies an action very different from chemical affinity. By means of this action they produce, in these bodies, decompositions of their elements and different recombination of these same elements to which they remain indifferent*". Today's definition of a catalyst can be summarized in *a specie that enhances the rate of a chemical transformation without undergoing a permanent change*.

The importance of catalysis is tremendous, since food, energy, fuel, chemical and pharma productions rely on catalyzed transformations. Catalysis is the key for efficiency in organic synthesis, where resource efficiency maximization, limitation of toxic chemicals and waste minimization are sought. Nevertheless, there are still industrial chemical processes that have high E-factors (mass of waste/mass of product)[1], so the development of new and efficient catalytic technologies to overcome environmental issues urges, along with a deep comprehension of how catalyzed processes take place. Because of the complexity of many catalytic reactions, experimental techniques may produce ambiguous information about the reaction mechanism. Insights into unclear aspects, such as structures and reaction mechanisms, can be supplied by computational methods, which are a powerful tool at chemist's disposal to furnish complementary insights to experimental evidences. The focus of computational chemistry is to solve chemically related problems by calculations. A computational organic chemist approaches the study of chemical reactivity by modelling and computing the structures and the properties of the species that are transformed during the chemical process of interest. We can exploit computational methods to investigate and understand catalyzed reactions, supplying details at molecular level that can be the basic guideline for new and improved catalytic systems. The development of powerful computers enables us to increase the accuracy of the theoretical models and, consequently, improve the computational results. Nevertheless, when we study a reactive process using a computational approach, we have to consider that a computer only generates numbers and some text-strings, and the solution of the chemical problem relies on the ability of the chemist that interprets and rationalizes those numbers on the computer screen; this is the reason why a strong organic and physical chemistry backgrounds are essential for approaching this research area. As Jensen wrote "*Computers don't solve problems, people do*".[2]

Only few systems can be computed *exactly*, but approximated models are available to evaluate almost every aspect of a chemical transformation (see Part I). The results of a computational investigation also help chemists to make prediction before running experiments, in which the reactants may result to be too expensive or the reaction set up is too difficult.

The computational investigations reported in this dissertation are addressed to supply valid basis for the comprehension of significant chemical problems where the catalytic aspects are still unclear. To this purpose, a *modus operandi* common to each problem has been adopted: (i) selection of a chemical transformation and analysis of the experimental conditions (reactants, catalyst, ligands, cofactors, solvent, pH); (ii) detailed study of the literature and formulation of reaction mechanism's hypothesis to design the computational-model which comprises all the reaction partners; (iii) choice of the computational level, *i.e.* theoretical level and accuracy: Molecular Mechanics approach, Quantum-Mechanical methods with different basis set, hybrid QM/MM method. The choice of the theoretical level is determined by the size of the model-system and the nature of the problem (structural investigations or reactivity study). Importantly, we must remember that in case of a QM approach the higher the accuracy (extended basis sets), the larger is the required CPU time and the longer computations take to converge; (iv) launch computations to investigate the potential surface and locate the critical points that chemically correspond to reactant complexes, transition states, intermediates, and products; (v) disclosure of the reaction path and corroboration/disapproval of the initial hypothesized mechanism.

References

[1] Shedon, R.A., *Green Chem.* **2017**, *19*, 18-43.
[2] Jensen, F. *Introduction to Computational Chemistry*, Wiley, **2007**.

Part I

Computational Methods

1. Potential Energy Surface

1.1. Critical Points

One of the main tasks of an organic computational chemist is to determine the structure and energy of the molecules and the transition states that are involved in the reaction. The *potential energy surface* (PES) is a surface where the points represent the geometries and energies of a collection of nuclei, *i.e.* the PES is a *Born-Oppenheimer surface*. The reaction path is defined by line that connects the points on the PES, which corresponds to the *Minimum Energy Path* (MEP).

For a non-linear polyatomic molecule with N atoms, the potential energy is a function of $3N$-6 internal coordinates (3 translational and 3 rotational coordinates are removed). The internal coordinate system can be obtained by choosing $3N$-6 linearly independent coordinates that coincide with bond lengths, angles and dihedrals between the atoms. The PES would therefore be a multi-dimensional surface (*hypersurface*), and its representation in a unique graph is unfeasible. To ease the visualization and the comprehension, it is possible to extract a cross-section of the whole surface, *i.e.* a two-dimensional plot where potential energy is plotted against the *reaction coordinate*, which represents the key geometric parameters that change during the course of the reaction. Reagents, reaction intermediates, products and transition states are called *critical* or *stationary points* and, mathematically, they correspond to the points where the first derivative of the potential energy E is zero

$$\frac{\delta E}{\delta q_1} = \frac{\delta E}{\delta q_2} = \cdots = 0 \qquad \textbf{Eq.1.1}$$

The partial derivative reminds that each derivative is respect to just one of the variables q of which E is a function. Stationary points can be *minima* (global or local) or *saddle points* (of various orders), and as chemists we are interested into minima and first order saddle points, where the formers are reagents, reaction intermediates and products, and the latter are transition states (TSs). Considering

the second derivative of the energy, minima are described by a second derivative of the energy greater than zero for all q

$$\frac{\delta^2 E}{\delta q^2} > 0 \qquad\qquad \text{Eq.1.2}$$

While for a transition state

$$\frac{\delta^2 E}{\delta q^2} > 0 \qquad\qquad \text{Eq.1.3}$$

For all q except along the *reaction coordinate*, where

$$\frac{\delta^2 E}{\delta q^2} < 0 \qquad\qquad \text{Eq.1.4}$$

1.2. Geometry optimization

To find a *stationary point* on a PES, a geometry optimization is carried out on an input structure, which is a guess of the geometry of the critical point. If the exact (analytical) shape of the PES would be knows, the *stationary points* would be rather easily obtained. However, the analytical expression of the PES is usually unknown, and an approximate representation of the surface is needed. Generally, the potential energy is approximated to a Taylor series around a critical point, so that the surface takes the shape of a parabola (harmonic or quadratic approximation).

For the easiest case, *i.e.* a diatomic molecule, to move along the two-dimensional PES and reach the nearest minimum, one simply changes the interatomic bond length till finding the lowest energy.

For a polyatomic molecule of N atoms, the geometry optimization is described by more complicated algorithms, that take into consideration all the 3N-6 internal coordinates. Starting from a guess input geometry, we arrange input's derivatives into a matrix formalism. The first derivatives are collected into the gradient vector

$$\mathbf{g}_i = \begin{pmatrix} \left(\frac{\delta E}{\delta q_1}\right)_i \\ \left(\frac{\delta E}{\delta q_2}\right)_i \\ \vdots \\ \left(\frac{\delta E}{\delta q_{3N-6}}\right)_i \end{pmatrix} \qquad\qquad \text{Eq.1.5}$$

And the second derivatives into the *Hessian* **H**, which corresponds to the force constant matrix, (the second derivative of the potential energy with respect to the geometric displacement is the force constant),

$$\mathbf{H} = \begin{pmatrix} \left(\dfrac{\delta^2 E}{\delta q_1 q_1}\right)_i & \left(\dfrac{\delta^2 E}{\delta q_1 q_2}\right)_i & \cdots & \left(\dfrac{\delta^2 E}{\delta q_1 q_{3N-6}}\right)_i \\ \left(\dfrac{\delta^2 E}{\delta q_2 q_1}\right)_i & \left(\dfrac{\delta^2 E}{\delta q_2 q_2}\right)_i & \cdots & \left(\dfrac{\delta^2 E}{\delta q_2 q_{3N-6}}\right)_i \\ \vdots & \vdots & \vdots & \vdots \\ \left(\dfrac{\delta^2 E}{\delta q_{3N-6} q_1}\right)_i & \left(\dfrac{\delta^2 E}{\delta q_{3N-6} q_2}\right)_i & \cdots & \left(\dfrac{\delta^2 E}{\delta q_{3N-6} q_{3N-6}}\right)_i \end{pmatrix} \qquad \textbf{Eq.1.6}$$

The geometry coordinates matrices for the initial and output structures are

$$\mathbf{q}_i = \begin{pmatrix} q_{i1} \\ q_{i2} \\ \vdots \\ q_{i3N-6} \end{pmatrix} \qquad \textbf{Eq.1.7}$$

and

$$\mathbf{q}_o = \begin{pmatrix} q_{o1} \\ q_{o2} \\ \vdots \\ q_{o3N-6} \end{pmatrix} \qquad \textbf{Eq.1.8}$$

The geometry optimization methods differ on whether they employ only the gradient or both the gradient and the *Hessian* matrix. The *steepest descend* method is an algorithm that performs the geometry optimizations by considering only the negative gradient direction, since the gradient always points in the direction where the function increases most. By following the opposite direction, we can perform a minimization of the function. Because of its nature, the steepest descend method can only locate minima. The main problem is that this method "forgets" the history of the performed steps. *Conjugate gradient* (GC) method tries to overcome this problem by taking into account some information about the previous minimization step. This method is based uniquely on the gradient and works well only for near-quadratic surfaces.

The approaches based on the *Newton-Raphson* (NR) method employ both the gradient and the Hessian matrix to perform the geometry optimization. Within a NR algorithm, the matrix equation to obtain the internal coordinates $\mathbf{q}_o$ of the *critical point* is

$$\mathbf{q}_o = \mathbf{q}_i - \mathbf{H}^{-1}\mathbf{g}_i \qquad \textbf{Eq.1.9}$$

The NR geometry optimization protocol can be summarized as follows: from the input structure, the algorithm generates the matrix of the geometric coordinate $\mathbf{q}_i$, from which calculates the initial gradient in the shape of $\mathbf{g}_i$ and the *Hessian* matrix $\mathbf{H}$. Generally, the first step of the minimization does not bring us directly to the minimum, rather we arrive at $\mathbf{q}_1$, the first computed geometry. From $\mathbf{q}_1$, the calculation "moves" on the surface searching for the minimum using the new gradient and

updating the Hessian. The process continues until the geometry (or the gradient) does not change significantly from the previous step (a threshold values is set). Transition state optimizations demand more complicated algorithms, that can be divided into two general categories, those based on interpolation between two minima, and those using only local information. The former is suitable when the reactant and product geometries are knows and approximates that the transition state is located between these two end-points. The local methods can be used when very little is known about the TS geometry.

1.3. Normal Mode Vibrations and Zero-Point Energy

Once we minimize a geometry, we need to check whether it is a minimum or a saddle point, by calculating the vibrational frequencies. To a first approximation, the vibrations are described by a harmonic motion, where the vibrational energy is quadratic (parabola) with respect to the displacement of the nuclei of the molecule. A better description and better results can be achieved by using a more accurate potential, like the one described by the Morse functional form.

A non-linear polyatomic molecule possesses $3N$-6 independent normal modes which frequency $\tilde{\nu}$, in wavenumber [cm^{-1}] is

$$\tilde{\nu} = \frac{1}{2\pi c} \left(\frac{k}{\mu}\right)^{\frac{1}{2}} \qquad \textbf{Eq.1.10}$$

Where c is the speed of light, k is the force constant of the vibration and μ is the reduced mass (which is easily calculated from the atomic masses). We are interested in the force constant, which represents the rigidity of the vibrational mode: the higher k, the harder is to stretch or bend the molecule, and *vice versa*. For a stationary point, we can obtain k by diagonalizing the *Hessian* matrix. In fact, the Hessian's eigenvalue matrix comprises the force constants for all the $3N$-6 vibrations. When all the Hessian eigenvalues are positive, the point is a minimum, when all but one Hessian eigenvalues are positive, we have found a 1st order saddle point (transition state). The computation of normal mode vibrations also allows to evaluate the *zero-point energy* (ZPE). This energy contribution, which needs to be added to the total energy of the molecule to give a reliable description of the system, represents that the nuclei, vibrate incessantly about the equilibrium position (x_o), even at the absolute zero.

The ZPE term is obtained $E_{ZPE} = h\nu/2$, ν are the harmonic vibrational frequencies, correlated to the wavenumber by $\tilde{\nu} = \nu c$.

2. Quantum Mechanics

2.1. The Schrödinger equation

Chemical bond breaking and forming involve electrons. To understand what governs a reaction mechanism, we have to face the quantum mechanical theory.

In classical mechanics, the word *state* means the simultaneous specification of position and velocity of a particle. In quantum mechanics, we have to consider the *Heisenberg uncertainty principle*: we are not able to know the exact position and velocity contemporaneously of a subatomic particle. Therefore, in quantum mechanics we rely on a less complete prediction of the future motion of the particle. A function of the particles' coordinates is therefore postulated. This function, which is called *state function* or *wave function* Ψ, describes the probability of finding the particle in a determined spot and contains all the possible information about the system, even its dependency of time. Having Ψ, we know everything about the present state. Classically, thanks to the Newton's second law, we predict where the system is addressed. In quantum mechanics we have at our disposal a fancy formulation: the *time-dependent Schrödinger equation* (TDSE). For a one-particle, one-dimensional system, this equation is expressed as

$$-\frac{\hbar}{i}\frac{\partial \Psi(x,t)}{\partial t} = -\frac{\hbar^2}{2m}\frac{\partial^2 \Psi(x,t)}{\partial x^2} + V(x,t)\Psi(x,t) \qquad \textbf{Eq.2.1}$$

Where the constant $\hbar$ is defined as

$$\hbar \equiv \frac{h}{2\pi} \qquad \textbf{Eq.2.2}$$

If we have the wave function at t_0, the Schrödinger equation allows us to calculate the future state of the system at any time. Of course, we can not obtain the exact position of the particle (*Heisenberg's principle*): what we get from Ψ is the probability, at time t, of finding the particle in a region lying between x and $x+dx$. In fact, we can count on the Born postulate, which states that $|\Psi(x,t)|^2$ is the probability density for finding the particle at various points on the x axis.

Reminding that quantum mechanics is statistical, we are only able to predict the probability of an experimental result and not its exact solution. For many applications in the chemistry field, we can use a simplified version of the TSDE. First, we can consider that the potential energy V depends only on the position x, so that

$$-\frac{\hbar}{i}\frac{\partial \Psi(x,t)}{\partial t} = -\frac{\hbar^2}{2m}\frac{\partial^2 \Psi(x,t)}{\partial x^2} + V(x)\Psi(x,t) \qquad \textbf{Eq.2.3}$$

Now, we can take in consideration only those solution of **Eq.2.3** that can be written as

$$\Psi(x,t) = f(t)\psi(x) \qquad \text{Eq.2.4}$$

where the lowercase *psi* is used for the factor that depends only on x. Taking partial derivatives of **Eq.2.4** and then substituting them with **Eq.2.3**, we obtain

$$-\frac{\hbar}{i}\frac{df(t)}{dt}\psi(x) = -\frac{\hbar^2}{2m}f(t)\frac{d^2\psi(x)}{dx^2} + V(x)f(t)\psi(x) \qquad \text{Eq.2.5}$$

$$-\frac{\hbar}{i}\frac{1}{df(t)}\frac{df(t)}{dt} = -\frac{\hbar^2}{2m}\frac{1}{\psi(t)}\frac{d^2\psi(x)}{dx^2} + V(x) \qquad \text{Eq.2.6}$$

In **Eq.2.6**, the left side is independent of x, while the right side in independent of t. The function to which each side is equal must be independent of t and of x, so it has to be a constant. This constant, that we call E, has the dimension of energy.

We obtain, by simple manipulations, that the left side, that depends on t, equals to

$$f(t) = e^{-\frac{iEt}{\hbar}} \qquad \text{Eq.2.7}$$

Equating the right side of **Eq.2.6** to E, we get the *time-independent Schrödinger equation* of single particle of mass m that is moving in one dimension.

$$\left[-\frac{\hbar^2}{2m}\frac{d^2}{dx^2} + V(x)\right]\psi(x) = E\psi(x) \qquad \text{Eq.2.8}$$

We can write that, substituting **Eq.2.7** into **Eq.2.4**, wave functions of the following form exist

$$\Psi(x,t) = e^{-\frac{iEt}{\hbar}}\psi(x) \qquad \text{Eq.2.9}$$

And these Ψs correspond to states of constant energy E. By substituting **Eq.2.9** into the *Born postulate*, we obtain that

$$|\Psi(x,t)|^2 = |\psi(x)|^2 \qquad \text{Eq.2.10}$$

For states described by equation **Eq.2.9**, the probability density does not depend on the time. These of states are called *stationary*. We will call $\psi(x)$ as wave function, but properly, the real wave function is $\psi(x)$ multiplied by the exponential $e^{\wedge}(-iEt/\hbar)$ as in **Eq.2.9**. Saying that a state is *stationary* means that the probability is stationary, not the particle itself. In fact, if the particle would be stationary, it would be fixed and, we would be able to define its position and momentum simultaneously.

The entity in squared brackets of equation **Eq.2.8** is the Hamiltonian operator **H**. The operator acts on the wave function giving back the operator multiplied by the constant E. For conservative systems, the Hamiltonian operator is basically the quantum-mechanical counterparty of the macroscopic total energy. More precisely, the eigenvalues of the Hamiltonian are the possible values of the system's energy. The *Schrödinger equation* is, indeed, an eigenvalue equation, where the wavefunction Ψ is an eigenfunction of the Hamiltonian operator, and the energies are the

eigenvalues. For a given molecule there is a set of wavefunctions Ψ_i with associated eigenvalues E_i that are orthonormal, which means that the functions are normalized and mutually orthogonal.

If we want to consider a poly-atomic system in three-dimensions we can "expand" our system over the coordinates *(x, y, z)* of each particle *i=1, 2, … n*.

The Hamiltonian is described by:

$$\mathbf{H} = -\frac{\hbar^2}{2}\sum_\alpha \frac{1}{m_\alpha}\nabla_\alpha^2 - \frac{\hbar^2}{2m_e}\sum_i \nabla_i^2 + \sum_\alpha \sum_{\beta > \alpha}\frac{Z_\alpha Z_\beta e'^2}{r_{\alpha\beta}} - \sum_\alpha \sum_i \frac{Z_\alpha e'^2}{r_{i\alpha}} + \sum_j \sum_{i>j}\frac{e'^2}{r_{ij}}$$

Eq.2.11

Where α and β run over the nuclei and *i* and *j* refer to the electrons, m_e is the mass of the electron, Z is the atomic number and r_{ij} is the distance between the particles *i* and *j*. The term ∇^2 is the Laplacian operator, and, in cartesian coordinates, takes the following form

$$\nabla_i^2 \equiv \frac{\partial^2}{\partial x_i^2} + \frac{\partial^2}{\partial y_i^2} + \frac{\partial^2}{\partial z_i^2}$$

Eq.2.12

Let's see in detail the terms that compose the Hamiltonian in equation **Eq.2.11**:

i. $-\frac{\hbar^2}{2}\sum_\alpha \frac{1}{m_\alpha}\nabla_\alpha^2$: kinetic operator of the nuclei;

ii. $-\frac{\hbar^2}{2m_e}\sum_i \nabla_i^2$: kinetic operator of the electrons;

iii. $-\sum_\alpha \sum_{\beta > \alpha}\frac{Z_\alpha Z_\beta e'^2}{r_{\alpha\beta}}$: potential repulsion operator between the nuclei;

iv. $\sum_\alpha \sum_i \frac{Z_\alpha e'^2}{r_{i\alpha}}$: potential attractive operator between the nuclei and the electrons;

v. $\sum_j \sum_{i>j}\frac{e'^2}{r_{ij}}$: potential repulsion between the electron.

The Hamiltonian in **Eq.2.11** contains attractive and repulsive terms between the particles and some of these interactions make next to impossible to compute the *Schrödinger equation* for systems with more than two electrons. One electron moves depending of the motion of all the other particles, *i.e.* the electrons are *correlated*. In a molecule we can have thousands of particles, and each particle is subject to the motion of the thousand-*1* particles.

The *Born-Oppenheimer* (BO) approximation allows us to decouple the motion of electrons from the motion of nuclei. This assumption is valid since the nuclei are moving much, much slower than the electrons. If we want a classical comparison, we can tell that during the time of an electronic motion, the nuclei are basically immobile. In this view, we can assume that the correlation in the attractive electron-nuclear potential energy (iv) can be eliminated, and the repulsive nuclear-nuclear potential energy term (iii) takes the shape of a constant (V_{NN}) for a given geometry.

At this point we have a Hamiltonian that considers only the electrons and we can, under the approximations made, rewrite it as a purely electronic Hamiltonian H_{el}

$$\mathbf{H}_{el} = -\frac{\hbar^2}{2m_e}\sum_i \nabla_i^2 - \sum_\alpha \sum_i \frac{Z_\alpha e'^2}{r_{i\alpha}} + \sum_j \sum_{i>j} \frac{e'^2}{r_{ij}} = \mathbf{T} + \mathbf{V}_{Ne} + \mathbf{V}_{ee} \qquad \text{Eq.2.13}$$

Where $\mathbf{T}$ is the kinetic operator, $\mathbf{V}_{Ne}$ the nuclear-electron attraction operator and $\mathbf{V}_{ee}$ the electron-electron repulsion operator. The *Schrödinger equation* based on the electronic motion, where the potential repulsion between the nuclei is added as V_{NN}, takes the shape of

$$(\mathbf{H}_{el} + V_{NN})\psi_{el} = U\psi_{el} \qquad \text{Eq.2.14}$$

We here wrote the energy as U, that is the electronic energy including internuclear repulsion, where it is the sum of the pure electronic energy and the nuclear repulsion:

$$U = E_{el} + V_{NN} \qquad \text{Eq.2.15}$$

The only variables in equation **Eq.2.14** are the electronic coordinates. The nuclear repulsion V_{NN} is independent on the electrons and it is a constant for a given nuclear configuration. The effect of a constant inside the Hamiltonian is just an increase of the energy eigenvalue by C, while the omission of it from the Hamiltonian is known not to affect the wave function. We can therefore delete V_{NN} from **Eq.2.14** and obtain the *purely electronic Schrödinger equation*

$$\mathbf{H}_{el}\psi_{el} = E_{el}\psi_{el} \qquad \text{Eq.2.16}$$

For convenience, the subscript el will be omit in all the following equations.

2.2. The Hartree and Hartree-Fock Self Consistent Field Methods

Back in 1928, Hartree suggested that the electronic wavefunction ψ could be expressed as a product of functions ϕ, *i.e. molecular orbitals* (MO), that depend only on the coordinates of one electron (r_n)

$$\psi\,(r_1, r_2, \dots r_n) = \phi_1(r_1)\phi_2(r_2)\dots\phi_n\,(r_n) \qquad \text{Eq.2.17}$$

And the Hamiltonian could be separated into one-electron Hamiltonian operators $\boldsymbol{h}_i$, where N is the number of the electrons of the system

$$\mathbf{H} = \sum_{i=1}^N \boldsymbol{h}_i \qquad \text{Eq.2.18}$$

We can define $\boldsymbol{h}_i$ as

$$\boldsymbol{h}_i = -\frac{1}{2}\nabla_i^2 - \sum_I^{nuclei} \frac{Z_I}{r_{Ii}} + V_i^{H} \qquad \text{Eq.2.19}$$

Where the interaction potential between the nuclei and the electrons is considered by V_i^{H}

$$V_i^H = \sum_{i \neq i}^{N} \int \frac{\rho_j}{r_{ij}} \qquad \text{Eq.2.20}$$

We can now express the Schrödinger equation for each electron in the *Hartree equation* formulation

$$\left(-\frac{1}{2} \nabla_i^2 - \sum_{I}^{nuclei} \frac{Z_I}{r_{Ii}} + V_i^H \right) \phi_i = \varepsilon_i \phi_i \qquad \text{Eq.2.21}$$

To solve these equations, we can employ the *Hartree method*, which is an iterative procedure: starting from a set of functions $(\phi_1, \phi_2, ... \phi_n)$, we generate a set of V_i^H from which we obtain better $(\phi_1, \phi_2, ... \phi_n)$. From these new electronic functions, we update V_i^H and so on, until the functions ϕ_i no longer changes. This method results to be a self-consistent field (SCF) procedure.

We have a problem with the *Hartree product* in equation **Eq.2.17**. The *Pauli exclusion principle* states that the total wavefunction must be antisymmetric with respect to exchange of two particles. If we swap two functions in **Eq.2.17** the total $\psi \, (r_1, r_2, ... r_n)$ will not be affected.

Fock recognized this problem and suggested that the best way to build the total wavefunction is by using a Slater determinant Φ_{SD}:

$$\Phi_{SD} = \psi \, (r_1, r_2, r_3, ... r_n) = \frac{1}{\sqrt{n!}} \begin{vmatrix} \phi_1(e_1) & \phi_2(e_1) & ... & \phi_n(e_1) \\ \phi_1(e_2) & \phi_2(e_2) & ... & \phi_n(e_2) \\ \vdots & \vdots & \vdots & \vdots \\ \phi_1(e_n) & \phi_n(e_2) & ... & \phi_n(e_n) \end{vmatrix} = |\phi_1, \phi_2, ... \phi_n| \qquad \text{Eq.2.22}$$

which satisfies the exclusion principle, since the resulting wavefunction is antisymmetric. In the Slater determinant Φ_{SD}, the elements in a given column involve the same function, whereas elements in the same raw involve the same electron.

Under the assumption that a single Slater determinant is sufficient to describe the wavefunction of the system and the mean-field approximation, Fock suggested to employ the *Hartree-SCF* procedure to Φ_{SD}, developing the so-called *Hartree-Fock Self-consistent field* (HF-SCF) method. The HF-MOs can be, again, determined as being the eigenfunctions of a set of one-electron operators, that in the HF-SCF are referred to as Fock operators $\boldsymbol{f}_i$. The difference with the one-electron operator $\boldsymbol{h}_i$ of **Eq.2.19** is that the electron exchange contribution is included in the mean potential term V_i^{HF}, which comprises the Coulomb operator $\boldsymbol{J}_i$, which defines the electron-electron repulsion energy due to each of the two electrons in the jth orbital, and the exchange operator $\boldsymbol{K}_i$ which describes the electron exchange contribution due to the antisymmetry of the total wavefunction.

For each electron, we can write a Fock operator as

$$\boldsymbol{f}_i = -\frac{1}{2} \nabla_i^2 - \sum_{I}^{nuclei} \frac{Z_I}{r_{Ii}} + V_i^{HF} \qquad \text{Eq.2.23}$$

2.3. LCAO approximation and Basis Set

How do we obtain the *molecular orbitals* ϕ_i that are present in the Slater determinant? Each ϕ_i can be expanded in terms of a set of k pre-defined one-electron basis functions χ. When *atomic orbitals* (AO) are used as basis functions this approach is referred to as *linear combination of atomic orbitals* (LCAO):

$$\phi_i = \sum_{\mu}^{k} c_{i\mu}\chi_{\mu} \qquad \textbf{Eq.2.24}$$

The μ spans all the AOs χ of each atom in the molecule. The set of AOs is called *basis set*.

To solve the HF-SCF procedure, we have to specify the AOs. Ideally, an infinite set of AOs, we will obtain E_{HF}, which corresponds to the lowest possible energy within the HF-SCF procedure (*HF limit*). Because of the omission of the *electron correlation*, the *HF limit* does not correspond to the real energy of the molecule. Only few AOs have been described exactly, since the Schrödinger equation is solvable only for hydrogen and hydrogen-like atoms. So, we chose some mathematical function, placed at the nuclei of the atoms, that mimes the known hydrogen function. For the sake of clarity, we can speak of *linear combination of basis functions*, instead atomic orbitals. *Slater-type orbitals* and *Gaussian-type orbitals* are the most common hydrogenic-functions that take the place of χ in **Eq.2.22**. The *Slater type orbitals* (STOs) mimic the exact solution for the single electron atom (hydrogenic orbital) and take the form of:

$$\chi_j^{STO} = N \cdot x^l y^m z^n e^{-\zeta(r-\boldsymbol{R})} \qquad \textbf{Eq.2.25}$$

Where $\boldsymbol{R}$ is the position vector of the nucleus where the function is centred, and N is the normalization constant. The value of ζ changes for each element, and it is obtained by minimizing the atomic energy with respect to ζ. STOs are suited in describing the electronic behaviour near and far from the nucleus, but when constructing the Fock matrix with STOs, many integrals need to be solved by using an infinite series and the truncation of the latter results in significant errors. To overcome this problem, Boys and later Shavit and Karplus decided to use a combination of gaussian functions to mimic the STO. These gaussian functions are known as *Gaussian Type Orbitals* (GTOs) have a similar expression of STOs, but the exponential dependence is quadratic

$$\chi_j^{GTO} = N \cdot x^l y^m z^n e^{-\alpha(r-\boldsymbol{R})^2} \qquad \textbf{Eq.2.26}$$

Generally, to simulate the behaviour of one STO, a combination of GTOs is used. If this is the case, the χ_j^{GTO} of equation (2.26) takes the name of *primitive* GTO, and the linear combination of *primitive* GTOs is referred to as *contracted* GTO. The *contracted* GTOs are, by far, the most employed functions to build the *basis set*.

But how many basis functions should we use to describe our molecule? The lowest number of functions, which is turn corresponds to the minimum level of accuracy, equals the number of occupied orbitals. For the carbon atom $1s^2 2s^2 2p^2$, we have two s-type functions and p_x, p_y and p_z functions, for a total of five functions. This is called *minimal* or single-zeta (SZ) basis set, where the term zeta reflects that each GTO emulates one STO function, which is defined by the exponent ζ. The most popular *minimal basis set* is STO-*N*G, where *N* refers to the number of Gaussian functions needed to approximate one STO. The choice of using only one function for each occupied orbital usually gives a poor description of the system: the valence electrons are not allowed to delocalize, while the core electrons can not get close enough to the nucleus. The solution is to uses *extended basis sets*, that promise to overcome the problems of the minimal basis set by doubling, or tripling (and so on) the SZ basis set, obtaining the double-zeta (DZ) or triple-zeta (ZT) basis sets respectively. We, as chemists, are interested in the processes of breaking and forming bonds, where the protagonists are the valence electrons. Pople suggested to treat differently the core from the valence atoms; in this view, he developed the *split-valence* basis set, where the core is treated with a SZ while the valence region with a DZ and the two sets are separated by a dash. In the DZ region each function is divided into an inner and an outer sphere.

To improve the basis flexibility, polarization functions can be also added and are referred to by an asterisk "*". These basis functions aim to take into account that when an atom approaches, an atom's orbital tends to polarize. A single asterisk corresponds to adding a set of *d* GTOs to carbon, a double asterisk adds also *p* GTOs to the hydrogen atoms. When a lone pair is present on an atom, especially in negatively charged atoms *i.e.* anions, the electrons tend to be held far away from the nucleus. To describe in a reasonable way this behaviour, we need a function able to expand in a larger volume: diffuse functions, labelled as "+", are generally added when very electronegative atoms or anion are present in the molecular system.

2.4. Correlation Energy and Post-HF methods

The inaccuracy in the HF procedure, due the one-electron's nature of the Fock operator, the mean-field and the single-determinantal approximations, results in the *electron correlation* which corresponds to the error made by the HF technique respect to the real energy of the system E_{real}:

$$E_{corr} = E_{real} - E_{HF} \qquad \textbf{Eq.2.27}$$

This error amounts only to ~1% of the total energy of the system but for a chemist this error is very large, since its magnitude is the same of many observables that are investigated by QM

computations. To better understand and evaluate the *correlation energy*, it is possible to separate it into a *dynamic* and a *static* contribution. The mean-field approximation in the HF approach, by which the repulsion energy between two electrons is approximated to the energy between one electron and the average electron density, is responsible for the *dynamic correlation*. This type of correlation is greater for electrons that occupy the same orbital, but it is also present between all the electrons pairs in a system, since all the electrons interact with each other.

The *static* (or *non-dynamical*) *correlation* is due to the inappropriateness of the Hartree-Fock model to describe the degeneration o quasi-degeneration of two or more electronic configurations because of the single-determinantal nature of HF wavefunction.

To include the *dynamical* and *static correlation* into the wavefunction, various procedures have been developed which are known as *post Hartree-Fock* methods. Among them, the Configuration Interaction (CI) method assumes that the system's wave function is as linear combination of diverse Slater determinants, obtained from the ground state determinant by substituting every spin-orbital function with a virtual one; while the Moller-Plesset (MP) model approaches a perturbative calculation which gives, as a result, the different order correction to the energy calculated by the Hartree-Fock method.

A different strategy to the solution of the Schrödinger equation is being evolved within the Density Functional Theory (DFT) framework; in the next paragraph we give a detailed description of the DFT method, since it is the model we adopted for most the calculations presented in this thesis.

2.5. Density Functional Theory

The complicated N-electron wavefunction can not be probed experimentally, and it depends on $4N$ variables, three spatial and one for the spin for each of the N electrons. Can the energy be obtained by using a simpler variable, that may be an observable?

The electron density $\rho(r)$ is a useful physical variable, since, integrated over all space, it gives the total number of electrons N

$$\int \rho(\mathbf{r}) \, d\mathbf{r} = N \qquad \textbf{Eq.2.28}$$

How can we correlate the electron density to the energy of the system?

The first Hohenberg-Kohn theorem states that "*the external potential $v_{ext}(\mathbf{r})$, and hence the total energy, is a unique functional of the electron density $\rho(\mathbf{r})$*". The *external potential $v_{ext}(\mathbf{r})$* is the DFT counterpart of the potential energy of interaction found in the electronic Schrödinger equation (second term of the right side of **Eq.2.13**). A *reductio ad absurdum* (here not reported) proves that the ground state

electron density ρ_0 determines the external potential $v_{ext}(\mathbf{r})$. So, the ground state electron density uniquely determines the external potential, so the Hamiltonian, and thus, the energy and all the properties of ground state of the system

$$\rho_0(\mathbf{r}) \Rightarrow \{v_{ext}(\mathbf{r}), N, Z_A, R_A\} \Rightarrow \mathbf{H} \Rightarrow \Psi_0 \Rightarrow \mathrm{E}_0 \qquad \textbf{Eq.2.29}$$

The energy of the system E_0 is therefore a functional of the function $\rho_0(\mathbf{r})$, so that

$$\mathrm{E}_0 = \mathrm{E}_v \left[\rho_0\right] \qquad \textbf{Eq.2.30}$$

The subscript v emphasizes that the ground state energy depends on the external potential $v_{ext}(\mathbf{r})$. The energy of the ground state can be expressed as a functional of the electron density and separated into the various contributions, which are in turn functionals of the electron density, as

$$\mathrm{E}_0 = \mathrm{E}_v \left[\rho_0\right] = T[\rho_0] + \mathrm{V}_{ee}\left[\rho_0\right] + \mathrm{V}_{Ne}\left[\rho_0\right] \qquad \textbf{Eq.2.31}$$

The $\mathrm{V}_{Ne}\left[\rho_0\right]$ is known $(= \int \rho_0(\mathbf{r})\, v_{ext}(\mathbf{r})d\mathbf{r})$ and it is the system dependent contribution, while the kinetic functional $T[\rho_0]$ and $\mathrm{V}_{ee}\left[\rho_0\right]$, which is the energy functional due to the electron-electron interaction are independent of the system in hand. $T[\rho_0]$ and $V_{ee}\left[\rho_0\right]$ can be collected into the Hohenberg-Kohn functional $\mathrm{F}_{HK}\left[\rho_0\right]$

$$\mathrm{F}_{HK}\left[\rho_0\right] = T[\rho_0] + V_{ee}\left[\rho_0\right] \qquad \textbf{Eq.2.32}$$

so that we can rewrite the expression for the energy functional as

$$\mathrm{E}_0\left[\rho_0\right] = \mathrm{F}_{HK}\left[\rho_0\right] + \int \rho_0(\mathbf{r})\, v_{ext}(\mathbf{r})\, d\mathbf{r} \qquad \textbf{Eq.2.33}$$

that means that if we know the exact form of $\mathrm{F}_{HK}\left[\rho(\mathbf{r})\right]$ we would be able to solve the Schrödinger equation exactly. Unfortunately, the explicit form of functional for the kinetic energy and for the electron-electron interaction are unknown. For what it concerns $\mathrm{V}_{ee}\left[\rho\right]$, we can separate it into the classical Coulomb part $J[\rho(\mathbf{r})]$ and the non-classical $\mathrm{E}_{ncl}\left[\rho(\mathbf{r})\right]$

$$\mathrm{V}_{ee}\left[\rho(\mathbf{r})\right] = J[\rho(\mathbf{r})] + \mathrm{E}_{ncl}\left[\rho(\mathbf{r})\right] \qquad \textbf{Eq.2.34}$$

Where $\mathrm{E}_{ncl}\left[\rho\right]$ contains the self-interaction correction and the exchange and Coulomb correlation. The main challenge is to find an explicit expression for $T[\rho]$ and $\mathrm{E}_{ncl}\left[\rho\right]$, since the explicit form of $J[\rho]$ is known.

The second Hohenberg-Kohn theorem states that "*The ground state energy can be obtained variationally: the density that minimises the total energy is the exact ground state density*". Following the first Hohenberg-Kohn theorem, for a trial ground-state electron density $\tilde{\rho}$ exists an associated external potential $v_{ext}(\mathbf{r})$. From $v_{ext}(\mathbf{r})$ one can construct a Hamiltonian $\tilde{\mathbf{H}}$ and its ground-state wavefunction $\tilde{\Psi}$. Applying the function $\tilde{\Psi}$ as a trial function in the variational principle leads to:

$$\langle \tilde{\Psi}|\tilde{\mathbf{H}}|\tilde{\Psi}\rangle = \mathrm{F}_{HK}\left[\tilde{\rho}\right] + \int \tilde{\rho}(\mathbf{r})\, v_{ext}(\mathbf{r})\, d\mathbf{r} = \mathrm{E}[\tilde{\rho}] \geq \mathrm{E}_v\left[\rho_0\right] = \langle\Psi_0|\mathbf{H}|\Psi_0\rangle \qquad \textbf{Eq.2.35}$$

2.5.1 The Kohn-Sham Approach. The first and second theorems of Hohenberg-Kohn tell us that the density determines the external potential, which determines the Hamiltonian, which determines the wavefunction, so all the properties of the system. What these theorems do not supply is the procedure to calculate E_0 from ρ_0, and how to find ρ_0 without the wavefunction. In 1965 Kohn and Sham realized that they could considered a (fictious) *non-interacting* system of *n non-interacting* electrons that experience the same external potential $v_s(\mathbf{r}_i)$, which makes the fictious (or reference) electron density $\rho_s(\mathbf{r})$ to be equal to the exact ground state density $\rho_0(\mathbf{r})$. In this specific *non-interacting* system, the Hamiltonian is the sum of one-electron Kohn-Sham operators $\mathbf{h}_i^{KS}$ and the ground state wavefunction $\psi_{s,0}$ of the reference system corresponds to the Slater determinant of the Kohn-Sham spin-orbitals u_i^{KS} , where the spatial part $\chi_i^{KS}(\mathbf{r}_i)$ is eigenfunction of the operators $\mathbf{h}_i^{KS}$

$$\mathbf{h}_i^{KS} \chi_i^{KS} = \varepsilon_i^{KS}\chi_i^{KS} \qquad \textbf{Eq.2.36}$$

where ε_i^{KS}'s are the Kohn-Sham orbital energies.

Kohn and Sham wrote the functional form of the reference system's energy as

$$E_v[\rho] = \int \rho\,(\mathbf{r})\,v(\mathbf{r})\,d\mathbf{r} + T_s[\rho] + \frac{1}{2}\iint \frac{\rho\,(\mathbf{r}_1)\rho\,(\mathbf{r}_2)}{r_{12}}d\mathbf{r}_1 d\mathbf{r}_2 + \Delta\bar{T}[\rho] + \Delta\bar{V}_{ee}[\rho] \qquad \textbf{Eq.2.37}$$

Where the zero subscript on ρ is omitted in this an in the subsequent equations. The $\Delta\bar{T}[\rho]$ term corresponds to the difference in the average ground-state electronic kinetic energy between the molecule and the reference system, as define previously; while the $\Delta\bar{V}_{ee}[\rho]$ is defined by

$$\Delta\bar{V}_{ee}[\rho] \equiv \bar{V}_{ee}[\rho] - \frac{1}{2}\iint \frac{\rho\,(\mathbf{r}_1)\rho\,(\mathbf{r}_2)}{r_{12}}d\mathbf{r}_1 d\mathbf{r}_2 \qquad \textbf{Eq.2.38}$$

We can sum up the $\Delta\bar{T}[\rho]$ and $\Delta\bar{V}_{ee}[\rho]$ functionals, which are unknown, into the *exchange-correlation* energy functional $E_{XC}[\rho]$

$$E_{XC}[\rho] \equiv \Delta\bar{T}[\rho] + \Delta\bar{V}_{ee}[\rho] \qquad \textbf{Eq.2.39}$$

The first three terms of the right side of **Eq.2.35** can be easily evaluate, while $E_{XC}[\rho]$ can only be approximated. An accurate DFT computation relies on a good approximation of the *exchange-correlation* energy functional. In the Kohn-Sham formalism, the kinetic energy functional is

$$T_s[\rho] = -\frac{1}{2}\sum_i^N \int \chi_i^{KS*}(\mathbf{r})\,\nabla^2\,\chi_i^{KS}(\mathbf{r})d\mathbf{r} \qquad \textbf{Eq.2.40}$$

While the nuclei-electron functional energy $V_{Ne}[\rho]$ is defined as

$$\int \rho\,(\mathbf{r})\,v(\mathbf{r})\,d\mathbf{r} = -\sum_\alpha Z_\alpha \int \frac{\rho\,(\mathbf{r}_1)}{r_{1\alpha}} \qquad \textbf{Eq.2.41}$$

Taking in consideration Eq. 2.40 and 2.41, Eq. 2.35 can be rewritten as

$$E_v[\rho] = \sum_\alpha Z_\alpha \int \frac{\rho\,(\mathbf{r_1})}{r_{1\alpha}} d\mathbf{r_1} - \frac{1}{2}\sum_i^N \int \chi_i^{KS*}(\mathbf{r})\,\nabla^2\,\chi_i^{KS}(\mathbf{r})d\mathbf{r}$$
$$+ \frac{1}{2}\iint \frac{\rho\,(\mathbf{r_1})\rho\,(\mathbf{r_2})}{r_{12}} d\mathbf{r_1}d\mathbf{r_2} + E_{XC}[\rho]$$

<div align="right">Eq.2.42</div>

Where the Kohn-Sham orbitals χ_i^{KS} minimize the molecular ground state energy and satisfy:

$$\left[-\frac{1}{2}\nabla_i^2 + v_s(\mathbf{r}_i) - \sum_\alpha Z_\alpha + \int \frac{\rho\,(\mathbf{r_2})}{r_{12}} d\mathbf{r_2} + v_{XC}\right]\chi_i^{KS} = \varepsilon_i^{KS}\chi_i^{KS}$$

<div align="right">Eq.2.43</div>

Where v_{XC} is the *exchange-correlation* potential, defined as the derivative of the *exchange-correlation*:

$$v_{XC} \equiv \frac{\delta E_{XC}[\rho(\mathbf{r})]}{\delta\rho(\mathbf{r})}$$

<div align="right">Eq.2.44</div>

The Kohn-Sham one-electron operators h_i^{KS} correspond to the term in squared brackets of **Eq. 2.43**:

$$h_i^{KS} = -\frac{1}{2}\nabla_i^2 + v_s(\mathbf{r}_i) - \sum_\alpha Z_\alpha + \int \frac{\rho\,(\mathbf{r_2})}{r_{12}} d\mathbf{r_2} + v_{XC}$$

<div align="right">Eq.2.45</div>

Which is the same of the one-electron Fock operator in equation **Eq. 2.23** except the exchange term is replaced by v_{XC}, so that both the exchange and the correlations are considered.

2.5.2 Exchange-correlation functionals. Everything but v_{XC} (and so E_{XC}) is known, so in literature various models exists that aim to calculate the exchange and correlation functional. These functionals are based on diverse approximations for E_{XC} and following the Jacob's ladder,[1] we report the functionals' families available in literature:

i) *local-density approximation* (LDA), where the system is divided into infinitesimal volumes among which the electron density is considered to be constant; the exchange part has an analytical formulation, while the correlation energy is fitted to the many body free-electron gas data;

ii) *generalised gradient approximation* (GGA), takes into account that the electron density varies rapidly, and that a constant value for the density by itself is not suitable to take into account the modification of the electron density. The GGA-functionals (also referred to as "*gradient-corrected*") include the scalar density and functions of the gradient of the density. A popular GGA functional combines the Becke exchange expression with the Lee, Yang Parr correlation formulation,[2] giving the commonly used BLYP functional;

iii) *meta-GGA*, where the density of spin-up and spin-down electrons a and the sum of the kinetic energy density of the KS orbitals are also added;

iv) Both LDA and GGA methods overestimate atomization energies while HF is known to underestimate them; so, *hybrid* functionals have been developed. In such functionals, the

LDA and GGA exchange and correlation functionals are mixed with a fraction of the "exact exchange" calculated from the HF.

Among the *hybrid* functionals, the most widely used in computational chemistry is B3LYP,[3] but it presents some disadvantages, such as it underestimates the activation barrier heights and it is imprecise in evaluating the interactions dominated by the middle-radius correlation energies, as the van der Waals, the aromatic π-stacking interactions and the isomerization energy of alkanes. Because of these limitations, a new *hybrid meta-GGA exchange-correlation functionals* has been developed by Zhao and Truhlar.[4,5] Among the Minnesota family, the M06 functional is parametrized including both transition metals and non-metals, whereas the M06-2X functional is a highly-non-local functional with double the amount of non-local exchange (2X), and it is parametrized only for non-metals.[5] These functionals are among the best at the moment for the study of organometallic and inorganometallic thermochemistry, non-covalent interactions and long-range interactions.

The choice of the functional is, therefore, always not trivial. Careful considerations should be made about the chemical system we are studying, and according to its characteristics we chose the most suitable functional. In this thesis different systems are presented. We decided to adopt the M06-2X for those systems that presented no metallic atoms, while the M06 for the organometallic section.

References

[1]   Perdew, J.P.; Schmidt, K. *AIP Conference Proceedings*, **2001**, *577*, 1.
[2]   Becke A.D. *Phis Rev.*, **1988**, *A38*, 3098.
[3]   Becke, A.D. J. *Chem. Phys.,* **1993**, *98*, 5648.
[4]   Zhao, Y.; Truhlar, D.G. *J. Phys. Chem. A*, **2006**, *110*, 13126-13130.
[5]   Y. Zhao; D.G Truhlar, *Theor. Chem. Acc.* **2008**, *120*, 215-241.

3. Solvent effects

Liquid solutions play a dominant role in chemistry and biochemistry, and the solvent affects enzymatic activity and organic reactions. In a general view, we have a solute molecule embedded into a solvent, made up of many molecules. If the solvent is polarized, it generates an electric field (*i.e. reaction field*) at each solute molecule. The presence of the *reaction field* polarizes and produces an induced dipole moment on the solute, and since the solvent molecules fluctuate, the induced dipoles fluctuate in turn. This type of polarization is a crucial contribution that has to be taken into account when studying solvated systems.

How do we treat the solvent and the interactions between the solute and the solvent? The easiest way is to consider the solvent molecules explicitly. In a MM approach, this is done by using force fields developed *ad hoc* for the solvent molecules and, generally, this procedure is not so computationally expensive, especially when the solvent particles are approximated by a molecular model. One of the most used models is the triangulated water (TIP3P) where the angle between the hydrogens is kept fixed in order to delete the contributions to the energy due to the water proton vibrations. While for MM computations explicit solvent computations are, most of the time, feasible, it is nearly impossible to treat the all the solvent molecules at QM level. This approach is practicable only if we consider few solvent molecules that are located in the first solvation sphere (*microsolvation*).

3.1. Continuum Medium Models

A much more simple and expensive approach, known as the *continuum medium* method, is to approximate the solvent to its *dielectric constant* ε_r which value can be obtained experimentally. Within this approximation, we compute the interaction between the solute and the solvent, which is considered as a uniform polarizable medium, with a fixed *dielectric constant*. In QM computations, the solute charge distribution $\sigma(r_s)$, which is affected by the dielectric medium, is obtained by the wavefunction of the system, so that the equation must be solved iteratively. This procedure is known as *self-consistent reaction field* SCRF.

To simulate the behaviour of a solution, a cavity of proper shape needs to be cut out from the medium to accommodate the solute: the simplest approach (**Fig.3.1a**) creates a spherical or ellipsoidal cavity, but with this description many solute-solvent interactions are lost. More accurate

models reproduce the molecular shape of the solute, by centring a sphere, with a particular radius, on each atom of the solute molecule. The most common radius is the van der Waals (VDW) or a scaled VDW radii, by which VDW-surfaces (VDWS) are created **Fig.3.1.b**. This description still has some deficiencies since the valleys and pockets of the external cavity may not be sufficiently wide to accommodate at least one molecule of the solvent. An extension of the VDWS approach is the generation of the so-called *solvent-accessible surface* (SAS), where the surface (and hence the cavity) is obtained by rolling a probe of the size of the solvent along the VDWS (**Fig.3.1.c**).
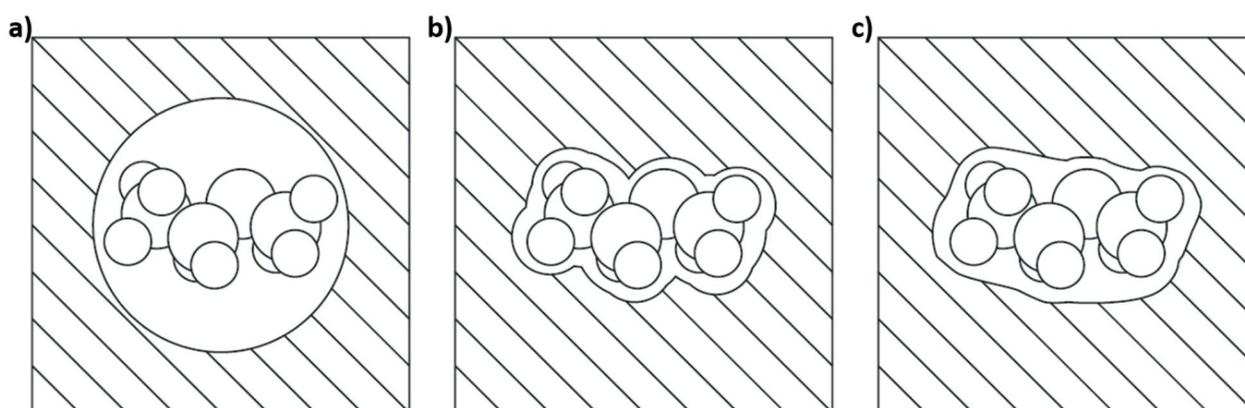

Fig.3.1. Schematic representation of the cavity shapes formed in the medium to accommodate the solute. In particular, the surface described by **a**) spherical cavity, **b**) VDW radii **c**) solvent-accessible surface.

The choice of the cavity's shape and size, the description of the charge distribution of the solute and how the dielectric medium is described are the features that differentiate the various *continuum medium* models that are available in literature. The *Osanger model* employs a spherical (or ellipsoid) cavity and the charge distribution of the solute is expressed in terms of a dipole (or multipole) expansion. The Tomasi's *polarized continuum model* (PCM)[1] has a cavity based on the VDWS, where the VDW radius is scaled by a value of 1.2 and the electrostatic interactions are obtained iteratively. Various versions of the PCM are available; in particular the *integral equation formalism* version of PCM (IEF-PCM) is the implicit solvent model adopted for all the QM computations reported in this thesis that required to take into consideration the solvent effect. IEF-PCM is also the default implicit solvent model employed in the optimization algorithm of Gaussian09,[2] which is the program used to carry out all the QM and QM/MM investigations reported in this thesis. To estimate free energy of solvated molecules level, implicit solvent approximations ease the computations, since the solvent's degrees of freedom are taken into account implicitly. We can consider that the energy of a molecule in a solvent is

$$\text{E}_{tot} = \text{E}_{vac} + \Delta G_{solv} \qquad\qquad\qquad \textbf{Eq.3.1}$$

where the solvation free energy ΔG_{solv} can be roughly described by

$$\Delta G_{solv} = \Delta G_{elec} + \Delta G_{nonpolar} \qquad \textbf{Eq.3.2}$$

The ΔG_{elec} accounts for the polar and electrostatic interaction, and represents the most time-consuming term, while $\Delta G_{nonpolar}$ adds the contributions due to the "hydrophobic" interactions, which are the favourable van der Waals attraction between the solute and the solvent, and the unfavourable cost of removing some solvent-solvent interactions when creating the solute's cavity. The two contributions that form $\Delta G_{nonpolar}$ are approximated to be proportional to the SAS *area* (SASA). To obtain the ΔG_{elec} one can solve the *Poisson equation o*f classical electrostatic, which provides the electrostatic potential $\phi(\boldsymbol{r})$ which is generated by the molecule charge distribution $\rho(\boldsymbol{r})$. If salt effects are included, the *Poisson-Boltzmann* (PB) equation are used, resulting in

$$\nabla \cdot [\varepsilon(\mathbf{r})\nabla\phi(\mathbf{r})] = -4\pi\rho(\mathbf{r}) + \kappa^2\varepsilon(\mathbf{r})\phi(\mathbf{r}) \qquad \textbf{Eq.3.3}$$

where $\varepsilon(\mathbf{r})$ is the dielectric constant of the solvent that depends on the position, and κ is the squared root of the salt concentration. Besides the rigorous physical basis of the PB model, this approach is quite expensive in terms of time. An approximation of the PB equation is the *generalized Born* (GB) model which computes the electrostatic solvation free energy ΔG_{elec} as a sum of the pairwise interaction terms between atomic charges. The key parameters in the GB computation are the B*orn radii* of the interacting atoms, which represent how much the atoms is buried within the solute.

References

[1]   Gaussian 09, Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; et al Gaussian, Inc., Wallingford CT, **2009**.
[2]   Tomasi, J.; Mennucci, B.; Cammi, R. *Chem. Rev.* **2005**, *105*, 2999–3093.

4. Molecular Mechanics

4.1. Force Field

The methods based on classical mechanics allow us to compute energy and properties of rather big molecular systems. In this view, since the electrons are not treated explicitly, the atoms are handled as charges spheres that follow the *Newton's* law of motion.

The dynamical behaviour of the molecular systems is described by *Newton's* second law, $F = ma$ where F is the force exerted on the particle, m is the particle's mass and a corresponds to its acceleration. The force exerted on each atom corresponds to the derivative, with opposite sign, of the potential energy (V) with respect to the position of the atom r_i :

$$F_i = \frac{\delta V}{\delta r_i} \frac{\delta^2 r_i}{\delta t^2} \qquad \text{Eq.4.1}$$

The potential energy is defined as a function of the coordinates of each of the atoms,

$$V = V_{bond} + V_{angle} + V_{dihedral} + V_{improper} + V_{Coulomb} + V_{LJ} \qquad \text{Eq.4.2}$$

Where V_{bond}, V_{angle}, $V_{diehedral}$ and $V_{improper}$ represent the covalent interaction, while the non-covalent are described by $V_{Coulomb}$ and V_{LJ}. These contributions are described by the following equations:

$$V_{bond} = \sum_{i=1}^{N_b} \frac{1}{2} k_i^b (r_i - r_{0,i})^2$$

$$V_{angle} = \sum_{i=1}^{N_\vartheta} \frac{1}{2} k_i^\vartheta (\theta_i - \theta_{0,i})^2$$

$$V_{improper} = \sum_{i=1}^{N_\xi} \frac{1}{2} k_i^\xi (\xi_i - \xi_{0,i})^2 \qquad \text{Eq.4.3}$$

$$V_{dihedral} = \sum_{i=1}^{N_\psi} \frac{1}{2} k_i^\phi \cos\left((\phi_i - \phi_{0,i})\right)$$

$$V_{Coulomb} = \sum_{i<j} \frac{1}{4\pi\varepsilon_0 \varepsilon_r} \frac{q_i q_j}{r_{ij}}$$

$$V_{Coulomb} = \sum_{i<j} \frac{1}{4\pi\varepsilon_0 \varepsilon_r} \frac{q_i q_j}{r_{ij}}$$

The parameters used for describing V may vary, and we refer to the set of parameters as the *force field*. The simulations performed in this thesis were carried out with the AMBER[1] force fields, which are suitable for studying proteins, nucleotides and lipids. For non-standard residues (organic, pharmaceutical, carbon-nanostructures) we employed the General AMBER force field (GAFF) as implemented in the AMBER package.

4.2. Molecular Dynamics

4.2.1. Integration Algorithms. The position of each particle of the system i evolves in time according to the *Newton's* second law:

$$\boldsymbol{F}_i(t) = m\boldsymbol{a}_i(t) \qquad\qquad \textbf{Eq.4.4}$$

For a *N*-particle system, only numerical solutions to **Eq.4.4** are available and can be obtained by integration methods (*leap-frog*, *velocity Verlet*, *Beeman's*), which allow to solve numerically the motion equation. The cited integration methods all derive from the *Verlet* algorithm, which assumes that the positions $\boldsymbol{r}$, velocities $\boldsymbol{v}$ and accelerations $\boldsymbol{a}$ can be approximated by a Taylor series expansion, such as

$$\boldsymbol{r}(t + \Delta t) = \boldsymbol{r}(t) + \boldsymbol{v}(t)\Delta t + \frac{\boldsymbol{a}(t)}{2}\Delta t^2 + \frac{\boldsymbol{b}(t)}{6}\Delta t^3 + \cdots \qquad \textbf{Eq.4.5}$$

$$\boldsymbol{v}(t + \Delta t) = \boldsymbol{v}(t) + \boldsymbol{a}(t)\Delta t + \frac{\boldsymbol{b}(t)}{2}\Delta t^2 + \cdots \qquad \textbf{Eq.4.6}$$

$$\boldsymbol{a}(t + \Delta t) = \boldsymbol{a}(t) + \boldsymbol{b}(t)\Delta t + \frac{\boldsymbol{c}(t)}{2}\Delta t^2 + \cdots \qquad \textbf{Eq.4.7}$$

We can also write

$$\boldsymbol{r}(t - \Delta t) = \boldsymbol{r}(t) - \boldsymbol{v}(t)\Delta t + \frac{\boldsymbol{a}(t)}{2}\Delta t^2 - \frac{\boldsymbol{b}(t)}{6}\Delta t^3 + \cdots \qquad \textbf{Eq.4.8}$$

And summing Eq.**4.5** and **4.8**, we obtain

$$\boldsymbol{r}(t + \Delta t) = 2\,\boldsymbol{r}(t) - \boldsymbol{r}(t - \Delta t) + \boldsymbol{a}(t)\Delta t^2 \qquad \textbf{Eq.4.9}$$

Substituting the acceleration $\boldsymbol{a}$ with $\frac{F}{m}$, we obtain the description of the position of the particles at time $t + \Delta t$

$$\boldsymbol{r}(t + \Delta t) = 2\,\boldsymbol{r}(t) - \boldsymbol{r}(t - \Delta t) + \frac{\boldsymbol{F}_i(t)}{m}\Delta t^2 \qquad \textbf{Eq.4.10}$$

The major drawback of the *Verlet* algorithm is that velocities of the particles are not defined explicitly. In *velocity Verlet*, which solves the problem of the *Verlet* algorithm, the velocities are updated as

$$\boldsymbol{v}(t + \Delta t) = v_i(t) + \frac{\boldsymbol{F}_i(t + \Delta t) + \boldsymbol{F}_i(t)}{2m}\Delta t \qquad \textbf{Eq.4.11}$$

So that we can have both the positions and the velocities defined at the same moment. In this way, the quantities that depend on the velocities and those that depend only on the position (kinetic and potential energy for example) are defined simultaneously. The choice of the timestep Δt is crucial, since it has to be small enough so that the total energy of the system is conserved.

Applying the integration algorithm, we are able to predict the future position of the particle and to describe the dynamical behaviour of the system in time. What about the properties of the system?

4.2.2. **Ensembles.** *Molecular Dynamics* (MD) is a microscopic simulation during time, so that we have a description of the different states that the system can visit. For statistical mechanics, the

macroscopic properties depend on the averages of the microscopic details, so that they are not heavily dependent on the exact motion of every single particle.

This is the basis of the *ensemble* concept used in statistical mechanics, where a representative collection of states (i.e. the *ensemble*) is used to describe a macroscopic property of the system. In an *ensemble*, all the states share common thermodynamics properties (volume, number of particles, pressure, total energy). Depending on the set of thermodynamic properties chosen to be maintained constant during a simulation, various types of *ensembles* are available in MD programs.

If the system has no changes in the number of moles (*N*), volume (*V*) and energy (*E*), the microcanonical ensemble (*NVE*) describes an adiabatic process where the total energy is conserved (but potential and kinetic energy are allowed to vary). No pressure or temperature control is done.

In an *NVT* (*N*, *V* and *T* are constant) *ensemble* (canonical), the temperature is modulated by a thermostat, which ensures the average temperature of the system to be the chosen one by acting as a heat bath on the system. The heat bath, which is like an infinite reservoir of energy, exchanges energy with the system, by adding/removing energy from the boundaries of a MD simulation.

Various thermostats are available and, in the works presented in this thesis, where MD simulations were carried out, Anderson and Berendsen thermostats were used. In Anderson thermostat,[2] the coupling system-bath is obtained by stochastic collisions that take place occasionally act on randomly selected particles, and at the end all the particles are subject to this collision. Berendsen[3] is a weak coupling method, where the coupling adds or removes energy to the system to keep the temperature constant.

In an isothermal–isobaric (*NPT*) ensemble, the simulation is carried out under constant pressure, allowing the volume to adjust by employing, in addition to a thermostat, a barostat which maintain the pressure constant.

4.2.3 Solvent and long-range interactions. The MD approach can provide detailed information on the structure and dynamical behaviour, however the treatment of the solvent and the computation of long-range interactions present some limitations. MD software packages (such as AMBER) employ some techniques that aim to enhance the computational efficiency with marginal effects on the accuracy of the potential energy calculation:

i) *Periodic boundary conditions* (PBC) are a MD "trick" that consists in placing the system in a cubic/hexagonal cell containing a relatively small number of solvent particles. When an atom moves out of the cell, it is replaced by an image particle that enters from the opposite side of

the cell.

ii) *Explicit solvent models* where the solvent particles are approximated by various molecular models in order to reduce the degrees of freedom of the system: one of the most used models is the triangulated water (TIP3P)[1] where the angle between the hydrogens is kept fixed in order to delete the contribution to the energy due to the water proton vibrations.

iii) *Cut-off* distances can be applied such that all the atoms beyond the cut-off value are ignored when computing long-range non-bonded (dispersion and repulsion) interactions that comprise all possible combinations of atom pairs in the system;

iv) *Ewald summation method* eases the computation of the electrostatic contributions by replacing the direct summation of interaction energies between point particles with two summations, so that a single slowly convergent sum can be transformed into two rapidly convergent sums.

4.3. Molecular Mechanics and Poisson-Boltzmann Surface Area

The prediction of binding affinities by means of computational methods has become a powerful tool in biochemical research. These calculations can be essential for computer-aided drug design since they assist and simplify the identification of new candidates in early stages of drug discovery projects. The analysis of the binding affinity by which a *ligand* binds to its *receptor* can be evaluated thermodynamically, considering that the binding process is driven by the minimization of the Gibbs free energy G:

$$G \equiv U + pV - TS \equiv H - TS \qquad \textbf{Eq.4.12}$$

Here U is the internal energy, p the pressure, V the volume, T the temperature, S the entropy and H the enthalpy of the system. Only if the difference in the Gibbs free energy between the free reactants (*ligand* and *receptor*) and their *complex* is negative, the binding process occurs spontaneously. If this is the case, then the difference between the potential of the bound system (*complex*) and free *ligand* and *receptor*, ΔG_{bind} can be calculated as:

$$\Delta G_{bind,solv} = G_{complex} - G_{ligand} - G_{receptor} \qquad \textbf{Eq.4.13}$$

The value of ΔG_{bind} provides a measure of the strength of the binding: the more negative the stronger the attraction between the free species.

The *Molecular Mechanics-Generalized Born Surface Area* (MM-GBSA) and *Molecular Mechanics-Poisson-Boltzmann Surface Area* (MM-PBSA)[1] are computational methods that calculate binding free energies using molecular mechanics and continuum solvent models. These two methods compute the

absolute binding free energy by calculating average free energies of the *complex*, the *ligand* and the *receptor* as:

$$\Delta G_{bind,solv} = \langle G_{complex} \rangle - \langle G_{ligand} \rangle - \langle G_{receptor} \rangle \qquad \text{Eq.4.14}$$

The average values in $\langle ... \rangle$ are calculated by post-processing a series of representative snapshots obtained from an explicit solvent MD trajectory. While the simulation of the dynamic behaviour of the complex is carrier out in an explicit solvent, the MM-PBSA(or GBSA) approach employs a continuum medium to simulate the solution. This approximation is necessary when computing binding free energies, since otherwise in explicit solvent the energy contributions due to solvent-solvent interactions would dominate the $\Delta G_{bind,solv}$ term. Since the straightforward computation of $\Delta G_{bind,solv}$ is unfeasible, the MM-PBSA(or GBSA) method employs a thermodynamic cycle where the absolute binding free energy in solution $\Delta G_{bind,solv}$ is obtained by computing it as a sum of the binding affinity in gas-phase $\Delta G_{bind,gas}$ plus the solvation free energy ΔG_{solv}:

$$\Delta G_{bind,solv} = \Delta G_{bind,gas} + \Delta G_{solv} \qquad \text{Eq.4.15}$$

Where $\Delta G_{bind,gas}$ and ΔG_{solv} comprises the contribution of the complex, ligand and receptor as **Eq. 4.6**. The binding free energy $\Delta G_{bind,gas}$ can be partitioned into various contributions *i.e.* internal, electrostatic and van der Waal classical interactions:

$$\Delta G_{bind,gas} = \Delta G_{int} + \Delta G_{vdW} + \Delta G_{elec} \qquad \text{Eq.4.16}$$

The calculation of ΔG_{solv} is related to moving the solute (*complex, ligand* or *receptor*) from the vacuum state into an explicit solvent environment. The $\Delta G_{solv,X}$ contribution for each component *X* (*complex, receptor or ligand*) is a sum of polar and non-polar contributions:

$$\Delta G_{solv,X} = \Delta G_{polar,X}^{solv} + \Delta G_{non-polar,X}^{solv} \qquad \text{Eq.4.17}$$

The polar solvation term $\Delta G_{polar,X}^{solv}$ is calculated by solving the Poisson-Boltzmann (**Eq.3.3**) or the Generalized-Born equations which consider the solvent as a high dielectric constant medium and the solute as a set of fixed point-charges embedded in a dielectric continuum.

References

[1]   Case, D. A.; Cheatham, T. E.; Darden, T.; Gohlke, H.; Luo, R.; Merz, K. M.; Onufriev, Jr. A.; Simmerling, C.; Wang, B.; Woods, R. *J. Comput. Chem.*, **2005**, *26*, 1668-1688.
[2]   Andersen, H.C. *J. Chem. Phys.*, 1980, **72**, 2384.
[3]   Berendsen, H.J.C. *J. Chem. Phys., 1984,* **81**, 3684.

5. Hybrid QM/MM Methods

There are cases where we are interested in studying the reactivity of such large systems (*i.e.* enzymes, surfaces, …) that treating all the atoms by a QM method is an unfeasible approach.

In these molecular systems, the reaction generally takes place in a delimited region, while the rest of the atoms are not directly involved in the chemical process. A promising approach to speed up the computations, is to portion the whole system (called *real*) in two regions as schematically depicted in **Fig.5.1**: a small reactive region, containing the atoms that participate into the chemical process (called *model*), which we can describe at QM level, and a bigger section that comprises the remaining atoms that are treated classically to take into account the influence of the environment on the reactive layer. This approach, which employs two levels of theory, in called *hybrid* QM/MM.
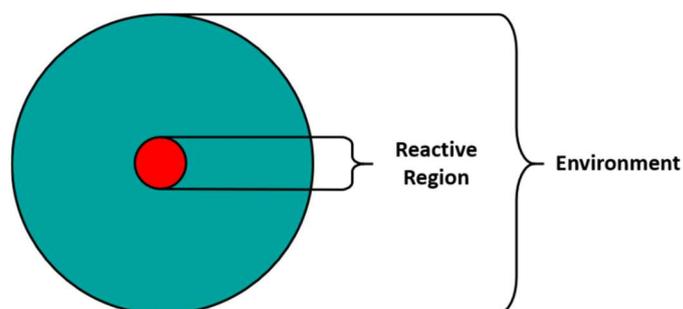
Fig.5.1. Schematic representation of the portioning of a large chemical system.

Two schemes exist to compute the energy of the total system: i) the *addictive scheme*, where the energy is described as a sum of the energies of the two layers plus the interaction between the two regions:

$$E_{QM/MM} = E_{QM,Model} + E_{MM,Real} + E_{QM/MM,Real} \qquad \textbf{Eq.5.1}$$

And ii) the *subtractive* (or *extrapolative*) *scheme*

$$E_{QM/MM} = E_{MM,Real} + E_{QM,Model} - E_{MM,Model} \qquad \textbf{Eq.5.2}$$

Where the energy of the *real* system is computed at MM level, then the QM energy of the isolate*d model* subsystem is added. In these two steps, we include twice the interactions between the layers, so the last term corrects for including these interactions twice by subtracting the MM energy of the *model*. The QM/MM computations that are presented in this thesis (see Chapter III) were carried out using the multi-layer *ONIOM* (*Our own N-layer Integrated molecular orbital and molecular Mechanics*)[1] scheme, which was developed by Morokuma and implemented in Gaussian09 series of program.

The ONIOM model employs a subtractive scheme, and it is not restricted to two layers. A three-layer

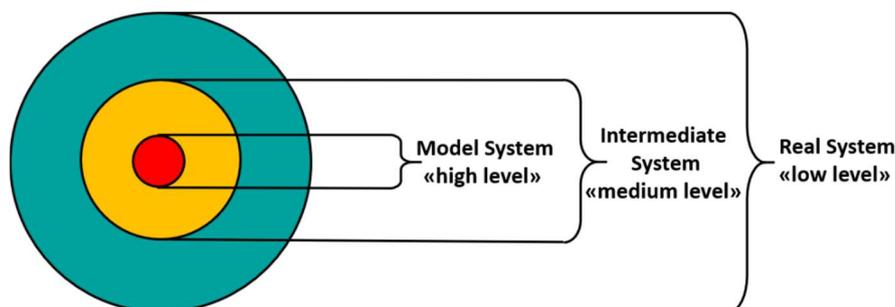ONIOM3 method has been developed by Svensson[2] and schematically depicted in **Fig.5.2**:

Fig.5.2. Schematic representation of the ONIOM3, where the three layers are identified.

As depicted in **Fig.5.2**, the system is divided into three subsystems, which are the *real*, described at a low accuracy level (MM), the *intermediate* system at a medium level and the model, and the *model* system which is described at a high level (QM). Because of the computational accuracy, the levels are sometimes also called *low*, *medium* and *high*. In this thesis only two-layers ONIOM approaches were adopted, consisting only of *high* (DFT) and *low* (MM) levels.

5.1. The Boundary region

In the easiest case, the QM and MM regions do not share a covalent bond: this is the case of a solute immersed in the solvent, or, as in our case, a reactant system that undergoes a chemical reaction insider a carbon nanotube. The situation becomes more difficult when one or more covalent bonds tight the QM and MM layers: this is the typical case of an enzymatic reaction, where the catalytic residues have their backbones with the MM region. A straightforward cut of one (or more) "bridging" bonds would generate an open valence on the QM atom, and this would distort the QM wavefunction, leading to senseless results. To overcome this problems, two strategies have been developed: i) the introduction of a *link atom*, a hydrogen, along the σ-bond vector that connects the QM and the MM atoms, which is present only in the QM subsystem, and it is invisible for the MM atoms, or ii) the introduction of a double occupied molecular orbital (*modified orbital method*) where the chemical bond connects the two regions, assuming that the electronic structure of the bond is insensitive to the changes in the QM region.

5.2. Mechanical and Electrostatic embedding

The various regions of a system described by a QM/MM model interact by both bonded and non-

bonded interactions (electrostatic and van der Waals) and various schemes have been developed to treat the coupling between the subsystems. In the simplest case, all the interactions between the layers are modelled at MM level, where the covalent bonds that connect the two regions are modelled by the force field. This approach, called *mechanical embedding* (ME) performs QM computations for the *high* layer in the absence of the *low*, so that ME does not consider that the MM environment can induce polarization of the electron density in the QM region. An improved scheme of the ME has been developed, known as *electrostatic* embedding (EE), where the polarization due to the MM environment is introduced by computing the *high* level in the presence of the *low*, so that in the QM Hamiltonian a term that describes the electrostatic interaction between the two layers is added.

References

[1]   a) Vreven, T.; Morokuma, K.; Farkas, Ö.; Schlegel, H. B.; Frisch, M. J. *J. Comp. Chem.* **2003**, *24*, 760-769. b) Vreven, T.; Byun, K. S.; Komaromi, I.; Dapprich, S.; Montgomery, J. A. Jr.; Morokuma, K.; Frisch, M. J. *J. Chem. Theory. Comp.* **2006**, *2*, 815-826

[2] Svensson, M.; Humbel, S.; Froese, R. D. J.; Matsubara, T.; Sieber, S.; Morokuma, K. *J. Phys. Chem.* **1996**, *100*, 19357.

Part II

Carbocatalysis

1. Introduction

Carbon, because of its capability of generating different bonds with other carbon atoms, forms different allotropes to which we refer as carbon nanomaterials (CNMs) or simply as carbons (**Fig. 1.1**). Carbons have nanoscale dimension and show interesting physicochemical and mechanical properties, such as large surface area, outstanding electron conductivity, corrosion resistance and thermal stability. These materials are employed as catalysts in clean and sustainable energy and electronics as nanostructured electrodes, but in the last decades the use of carbons has been progressively extended to chemical synthesis area, where they are exploited as heterogeneous catalysts and referred to as *carbocatalysts*.
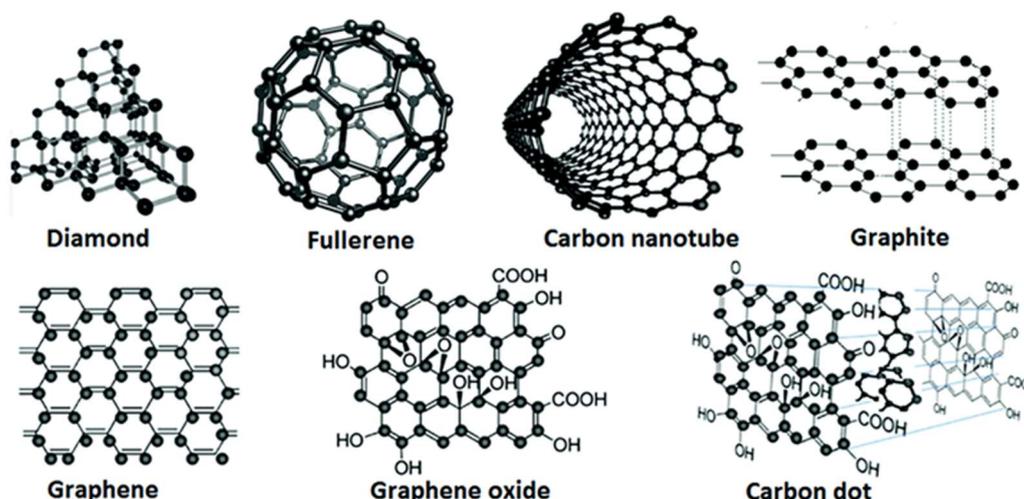
Fig.1.1 Schematic illustration of some carbon nanoparticles.

Actually, the use of *carbocatalysts* can be traced back to the beginning of XX century, when charcoal was reported to catalyze the aerobic oxidation of oxalic acid[1] and activated charcoal was demonstrated to catalyze the aerobic oxidation of ferrocyanide to ferricyanide.[2]

Nowadays, carbo-catalysis is explored in synthesis because these nanoparticles promise to be sustainable and economically convenient with respect to metal catalysts, which suffer from low availability, high costs and detrimental effects on the environment.

Indeed, *carbocatalysts* are emerging as green alternatives to conventional catalysts[3] and are being exploited to design novel synthetic technologies or to improve the performance of existing processes. Various examples of *carbocatalyzed* reactions are reported in literature,[4,5] and the development of efficient processes to synthetize carbons is widening the scope of available catalysts. Furthermore, the possibility of tailoring chemical properties during the synthesis (chemical functionalization, doping with heteroatoms and decoration with metallic nanoparticles and enzymes) opens new scenarios to the investigation and applications of these nanomaterials.[3]

Carbocatalyzed processes usually take place in gas phase, but recently liquid-phase reactions have been developed in order to avoid high temperatures and broad the scope of the substrates.[5] In these heterogenous processes, the separation of the desired product and the catalyst is eased and the nanocarbon can be reused. The catalytic performance of carbo-catalysts depends on the availability of the active sites, which are responsible for the catalytic effect. The active sites are found on the carbon nanomaterial surface and can be defects, heteroatoms and/or oxygen-containing functional groups, as depicted in **Fig.1.2**.
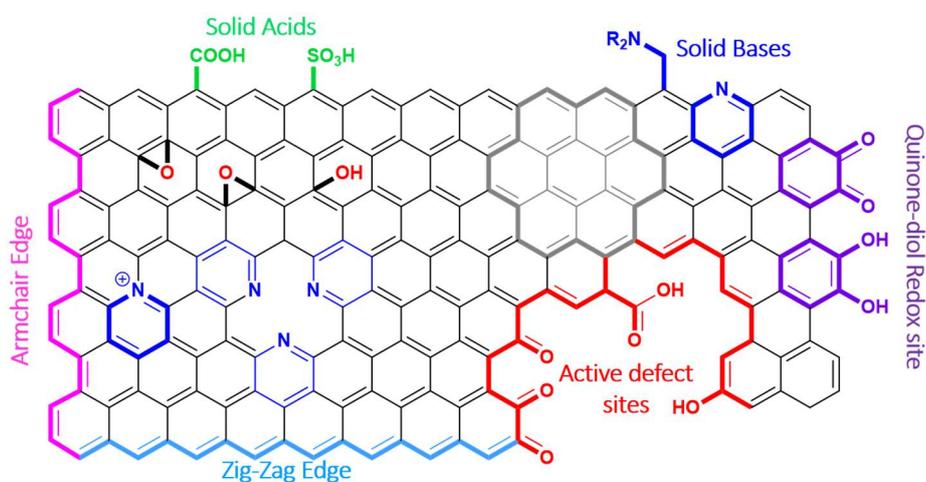

Fig.1.2 Schematic illustration of active sites found on carbocatalysts.

Despite the significant advances in characterization techniques of nanocarbon materials, the identification of active sites and the reaction mechanisms is generally challenging due to intrinsic experimental difficulties. The improvement of theoretical modellings and simulations, aimed to unveil the energetics and reactivity of carbon-catalysis, is disclosing information that are not evident by performing, solely, experiments. Quantum-Mechanical (QM) and hybrid QM/MM investigations on *carbocatalyzed* reactions are emerging as valid strategies to explore the possible catalytic sites and simulate the reaction mechanism.

Almost all the allotropes of carbon have been applied used for catalytic purposes[6]6 but their use can be limited due to the high costs required for mass-producing them, except for carbon nanotubes, which can now be obtained in large amounts with competitive costs. Because of the availability of these structures, CNTs have displayed roles as both heterogenous catalytic supports and active catalyst, exploiting their wide surface area and the presence of active sites on the surface. In fact, most of the literature on CNT-based catalysis, reports processes that occur on the external surface of the tube wall. An innovative approach to manipulate and catalyse organic reaction is the confinement of guest molecules inside molecular hosts. A CNT, with its hollow structure, represents a perfect guest and the variable size of the CNT cavity can be used to control the orientation of the guest and, consequently, the rate and the regioselectivity of a given reaction.[7] The interaction between the *host* and the *guest* governs the rate of the reaction, by stabilizing transition states and reducing reaction rates. The confined space can stabilize one product over another, modulating the thermodynamics of the reaction.

In the following Chapter, the results obtained during the PhD activity on the oxidative dehydrogenation of ethylbenzene over oxidized carbon nanotubes and the aromatic bromination of *N*-phenylacetamide confined inside CNTs will be presented. Because of the nature of the reaction studied, two different computational approaches have been chosen: a full QM methodology for studying the oxidation of ethylbenzene and a hybrid QM/MM scheme for the aromatic halogenation.

References

[1]   Rideal, K.; Wright, W.M.; J. Am. Chem. Soc. **1925**, *54*, 4473-4480.
[2]   Kolthoff, I. M. *J. Am. Chem. Soc.* **1932**, *54* (12), 4473–4480.
[3]   Liu, L.; Zhu, Y.; Su, M.; Yuan, Z. *ChemCatChem* **2015**, 2765–2787.
[4]   Duan, X.; Sun, H.; Wang, S. *Acc. Chem. Res.* **2018**, *51* (3), 678–687.
[5]   Su, D. S.; Wen, G.; Wu, S.; Peng, F.; Schlögl, R. *Angew. Chemie - Int. Ed.* **2017**, *56* (4), 936–964.
[6]   Su, D. S.; Perathoner, S.; Centi, G. N. *Chem. Rev.* **2013**, *113* (8), 5782–5816.
[7]   Miners, S. A.; Rance, G. A.; Khlobystov, A. N. *Chem. Soc. Rev.* **2016**, *45* (17), 4727–4746

2. Computational Investigation of Oxidative Dehydrogenation of Ethylbenzene catalyzed by o-CNTs

2.1. Introduction

The direct de-hydrogenation of ethylbenzene to styrene can be considered one of the most important industrial processes in petroleum industry. The catalysts usually employed contain Fe and K oxides in different relative amounts together with additional promoters, which modify the activity/selectivity features, the range of temperatures that are used (879 to 930 K), and the steam/hydrocarbon ratio.[1] The main problems encountered in ethylbenzene dehydrogenation (and, in general, in alkanes dehydrogenation) are the high endothermicity of the reaction, the excess of steam required, which is very energy consuming, and a slightly irreversible deactivation of the catalyst.[1]

An interesting alternative is the oxidative dehydrogenation (ODH)[1,2] of ethylbenzene. In this case the injection of a gas containing oxygen is used to directly oxidize the hydrogen generated in the course of the reaction, making the overall process exothermic.[1] Various catalysts [3-8] based on transition metal oxides[3-5] and phosphates[6] were proposed for ODH. Also, carbon-based catalysts such as pyrolyzed polymeric systems[7] were employed. In the latter case ultrahigh surface area carbon molecular sieves were found more active than inorganic oxides. These results were consistent with the hypothesis that the carbonaceous layers detected in many inorganic systems could be the really active catalyst for this reaction.[9,10]

Following this hypothesis, increasing attention has been focused during the last years on the use of carbon nanomaterials as catalysts.[11] These materials are particularly interesting and promising for a sustainable chemistry in many industrial processes. In the case of the ODH reaction, for example, they allow the use of lower temperatures without excess supply of steam and represent an attractive alternative to the conventional and toxic metal oxide-based catalysts.[12-15] A significant contribution in this field can be ascribed to the pioneering work of Schlögl and co-workers[16,17] and Pereira and co-workers. [18] Schlögl used nanofilaments[16] (characterized by high surface area) to catalyse the ODH process of ethylbenzene and demonstrated the higher catalytic activity and the better selectivity and reaction yield of these materials with respect to soot and graphite.

The structural characterization of carbon nanomaterials revealed that the catalytic activity is related to the surface features. Schlögl and co-workers[17,19] demonstrated that intact onion-like carbon (OLC)

materials, characterized by the complete absence of surface oxygen functionalities, do not show any catalytic activity. Catalytic activity increases gradually in the course of the process with the generation (due to the presence of oxygen in the reactant mixture) of basic quinoid functionalities (as indicated by XPS spectra) on the edge/kink regions of OLCs. A possible reaction mechanism suggested for ODH of ethylbenzene involves these basic adjacent (quinoid) oxygen centres that dehydrogenate ethylbenzene forming surface OH groups. Further investigations[20,21] compared the catalytic performance of activated carbon, nanodiamond and carbon nanotubes (CNT). These studies suggested that similar reaction pathways, where the breaking of the C-H bonds is kinetically relevant, involve the tetravalent (sp^3) nanodiamond structure and the trivalent (intermediate sp^2/sp^3) nanotube structure. The p-electron conjugation of graphite did not seem to be essential to promote catalytic activity. Multi-walled carbon nanotubes (MWNT) were used by Schlögl and co-workers [22] to catalyse the ODH reaction of ethane, even if in this case the carbon catalysts appeared to be less promising with respect to the same reaction of propane and butane.

Pereira also used MWNTs[18] to carry out oxidative dehydrogenation of ethylbenzene to styrene. In agreement with previous evidence, he found that pre-oxidized MWNTs show a higher catalytic activity than untreated materials providing further proof of the importance of oxygenated surface groups. In the various experiments the performances of nano-carbon catalysts were investigated at high temperatures (in the range 450-550 °C). [16,18-20] Typically, ethylbenzene was evaporated in flow of He and subsequently mixed with oxygen in the reactor.

The role of surface chemistry on the catalytic properties of carbon nanomaterials and, in particular, carbon nanotubes, has long been recognized and there is now a general consensus that the oxygen functionalities generated before or during the synthetic process are the active catalytic sites. [9,10,20,21,23] The adsorption of oxygen on carbon nanostructures leads to a rather complicated surface chemistry, which involves many different oxygen-bearing groups. They include carbonyl-, carboxyl-, quinone- and phenolic-type oxygen functionalities, often localized at the edge of the structures. Also, the adsorption of O atoms and O_2 molecules on the nanostructured surface can give rise to highly stable epoxide-like structures and four-membered ring systems, the former being significantly more stable than the latter.[10,24]

However, only recently useful quantitative information on the nature of active sites was obtained. [25-27] In particular, Su and co-workers,[26,27] using a novel *in situ* titration process, investigated the nature and quantity of active sites of nanocarbon materials in alkane ODH process and measured their intrinsic catalytic activity. They found evidence that the active sites are identical on various

nanocarbon catalysts (including graphene oxides, CNTs and OLCs) and that ketonic carbonyl groups are active sites for alkane ODH reactions.

During the last two years several important experimental papers on this subject have appeared. [28-30] In particular, Su [29] used conjugated polymers containing only carbonyl groups as model catalysts to investigate the mechanism of the nanocarbon-catalyzed ethylbenzene ODH reaction. He concluded that this process relies on the redox cycle of ketonic carbonyl/hydroxyl pairs. The same author[30] carried out an interesting study on the CNT catalyzed ODH reaction of *n*-butane and identified electrophilic oxygen species as the main culprits for combustion side reactions, which limit the selective transformation to butane. He showed that the reaction selectivity can be significantly improved by eliminating these electrophilic group by $LiAlH_4$ and annealing treatment. Quinone groups were identified by Su [31] as active sites on the CNT surface for the alkane ODH reaction. A new and effective method was suggested to generate these oxygen functionalities. Various nanostructures were examined by Choi and co-workers[32] using *n*-butane ODH as a model reaction. According with previous studies these authors found that the overall catalytic activity and selectivity increase with the increasing content of carbonyl active sites. However, the activity per site simultaneously decreases. This is attributed to the lower reducibility of each carbonyl at high C=O density.

At present, only a few examples of computational studies on alkane ODH reactions catalyzed by carbon nanomaterials are available in literature.[33-41] Compared to experimental work on this subject, theoretical investigations are still at an initial stage and limited to rather small alkanes. For instance, Su and co-workers[33] used DFT computations to investigate the ODH reaction of butane. They showed that the conversion of butane is energetically controlled by the abstraction of the first H atom. Du *et al.* [36] used a periodic density functional approach to study the oxidative dehydrogenation of propane catalyzed by CNTs and examined the origin of the catalytic activity of these nanostructures. They considered the role of active CO functional groups and suggested that the bulk CNT behaves as a reservoir which can donate or withdraw electrons from the active CO groups, thus modifying the activation barrier for the C-H bond breaking.

Dathar *et al.*[38] examined at the DFT level the ODH reaction of isobutane catalyzed by graphene. They identified dicarbonyls at the zig-zag edges and quinones at armchair edges of a graphene sheet as the most active oxygen functionalities. Su and co-workers[39] on the basis of periodic spin-polarized DFT computations, suggested that carbon atoms around the ketone groups on the CNT edge are active towards CH bond activation in the propane ODH process. This feature was the effect

of the reduced aromaticity around the oxygen function group. Very recently, Liu *et al.*[40] used a periodic DFT approach to investigate the direct dehydrogenation of isobutene catalyzed by nanodiamonds. They found that for the first C-H bond breaking, which is the rate-limiting step of the process, the activation barrier is lower for nanodiamonds with respect to CNTs. Also, they suggested that radical intermediate (isobutyl) species are involved in the process. Finally, the ODH reaction of ethane was examined by Khavruchenko and Frank [41] on CNTs of different length and diameter at the B3LYP level including Grimme's VdW corrections. These authors considered edge quinone groups and found that the CNT length does not affect the reaction barriers, while a slight decrease of reactivity is observed with the increasing diameter.

Even if these papers shed some light on the ODH reaction of small alkanes catalyzed by carbon nanostructures, a complete theoretical investigation of the reaction pathway of the oxidative dehydrogenation of ethylbenzene catalyzed by CNTs is still lacking. Many questions concerning this catalytic process, which is of great industrial interest and can be environmentally friendly, low-cost and resource-saving, are still open. In particular: (i) does the process proceed via a concerted or step-wise mechanism? (ii) what is the rate determining step? (iii) are ionic or radical species involved? (iv) is the reaction solely catalyzed by the edge quinone groups (as suggested by some experiments) or can other oxygen functionalities (for instance, on the surface of the carbon structure) be catalytically active? To answer these questions, the reaction is investigated here for the first time using a full quantum mechanical (DFT) approach. As a general model for the carbon nanomaterials that can be used as a catalyst, we chose a (6,6) CNT on which we considered various oxygen-containing functionalities.

2.2. Computational Details

The (6,6) carbon nanotube was built with the Nanotube Builder plugin, as implemented in VMD.[42] Its length and diameter were 11.10 and 8.14 Å, respectively. Carbon atoms at the two far ends of the tube were hydrogen-terminated. All DFT computations were carried out with the Gaussian09[43] series of programs and the M06-2X[44] functional. This functional properly describes the π-stacking interactions between the aromatic molecular plane of EB and the tube wall. To model the chemisorption of an oxygen molecule on CNT (oxidized CNT) we considered various possible structures: (i) two epoxy groups, (ii) a dioxoethane bridge on the tube surface, (iii) two ketone functionalities at the edge of the tube. In the first case the two epoxy groups were either *trans* or *cis*. In the second case we took into account either a 1,2 or a 1,4 dioxoethane bridge. In the third case we

built either a 1,4 diketone structure or a quinone structure. The resulting six structures were fully optimized using the 3-21G* basis set.[49] Then, to obtain more accurate relative energy values, we carried out single point computations with the 6-31G* basis sets on the optimized geometries. The various structures are depicted in **Fig. 2.1**. In **Table 2.1** the Relative energies for the various structures are also reported.
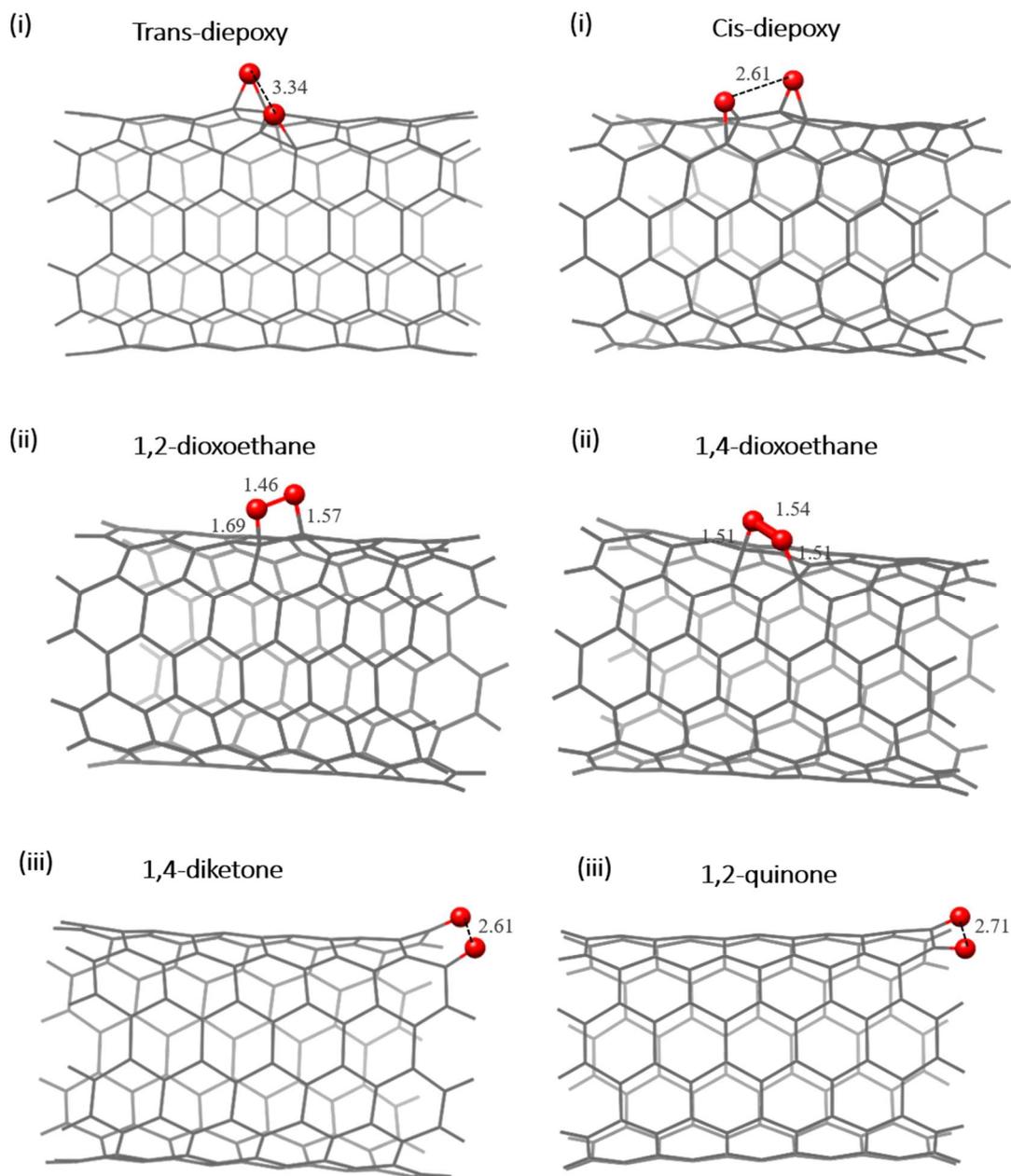

Fig.2.1 Three-dimensional representation of structures modelling O_2 chemisorption 0on a pristine CNT.

We adopted as model systems for the oxidized CNT the more stable diepoxy structure *i.e.* the *cis* structure (**Model A**) and the more stable diketone structure *i.e.* the quinone functionality (**Model B**).

We discarded both dioxoethane structures since they are much higher in energy. This choice is in agreement with the computational evidence reported in refs. 16,27.

We used a double layer basis set approach to describe the entire model-system: the 3-21G*[49] basis for the CNT and the 6-31G*[49] basis for the oxygen functionalities on the tube and the EB molecule. We employed an unrestricted formalism to take into account the possible radical nature of intermediates and transition states. To avoid the possibility of ignoring the unrestricted solution in the search of intermediates and transition states we started the calculations by reading in the triplet density, based on different α and β Kohn–Sham orbitals. In all cases we tested the stability of the unrestricted solution to be sure that it was more stable than the restricted one.

The structure of the various critical points (minima and saddle points) was fully optimized. Frequency calculations were carried out at the same level of theory and accuracy to check the nature of critical points.

Table 2.1 Stabilization energies relative to the most stable structure found for surface or rim oxidation.

	Structure	Stabilization [kcal mol^{-1}]
Surface	*cis*-diepoxy	0
	trans-diepoxy	16.1
	1,2-dioxoethane	32.2
	1,4-dioxoethane	58.8
Rim	1,2-quinone	0
	1,4-diketone	38.1

2.3. Results and Discussion

In this section we discuss the results obtained in the investigation of the singlet potential energy surface for the transformation ethylbenzene (EB) → styrene occurring on two different oxidized CNT structures: a *cis* diepoxy structure *i.e.* two adjacent epoxy groups in a *cis* configuration (**Model A**) and a quinone diketone structure (**Model B**). In the following discussion, **Rx** represents reactants, **TS** the transition states, **I** the intermediates and **Pd** the products. In brackets, the model (either **A** or **B**). The numbering is progressive along the reaction pathway.

To investigate the ODH reaction mechanism we considered the approaching of ethylbenzene to the tube surface in the vicinity of the oxygen-containing functionalities.

For both **Model A** and **Model B** the physisorbed EB on the oxidized CNT surface represents the starting reactant complex. A schematic representations of the two reactant complexes **Rx(A)** and **Rx(B)** is reported in **Fig.2.2**. In both cases the aromatic molecular plane of EB was found approximately parallel to the tube π-system. This arrangement generates strong π-stacking interactions.
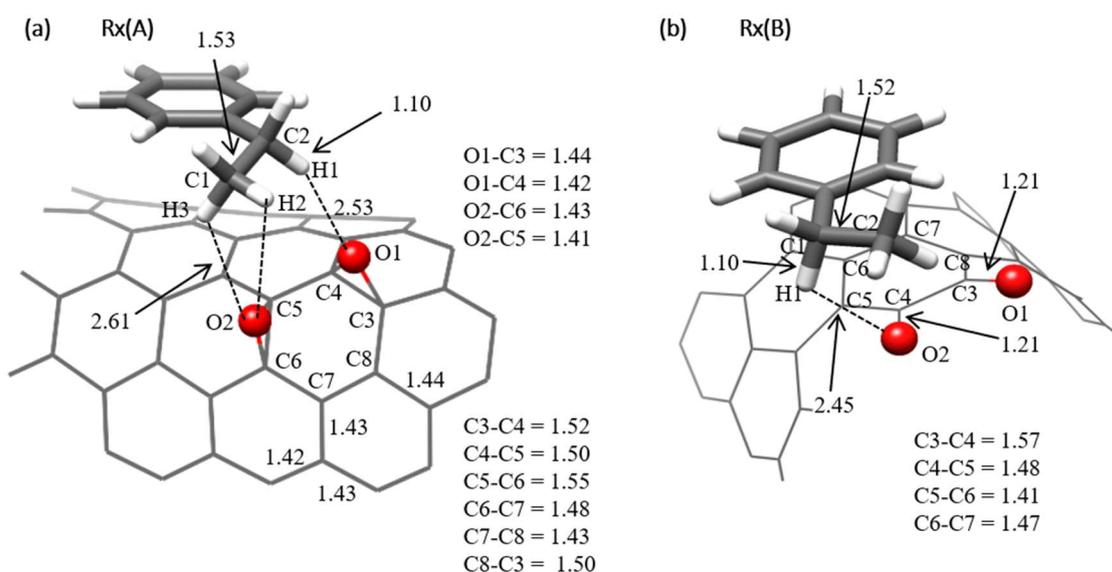

Fig.2.2 Three-dimensional representation of the two reactant complexes **Rx(A)** and **Rx(B)** obtained for **Model A** (*cis* 1,2-diepoxy-CNT + EB) and **Model B** (quinone-CNT + EB), respectively. Bond distances are in Ångstroms. For sake of clarity, only the region of the tube close the reaction site is reproduced.

An overview of the calculated energy profiles is given in **Fig.2.3**. The results obtained for the two model-systems are discussed separately.
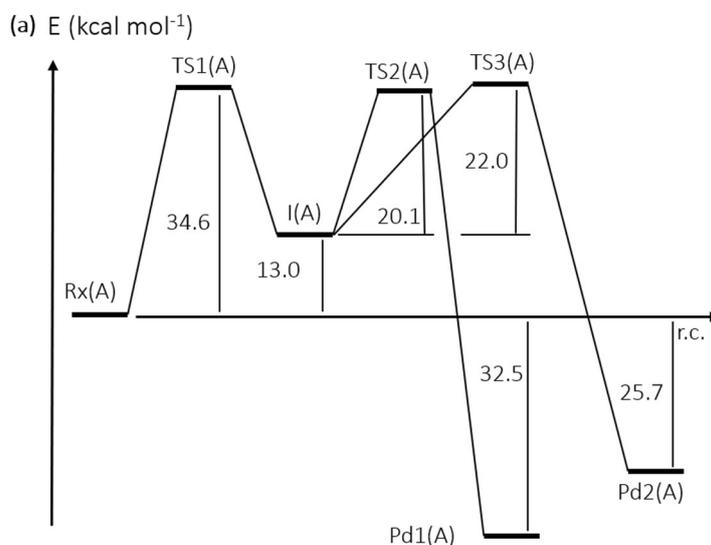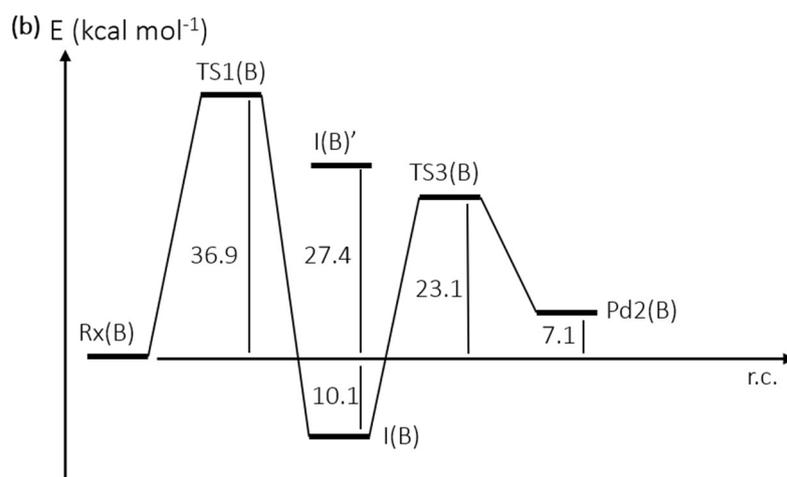
Fig. 2.3 A schematic representation of the potential energy surface for the reaction ethylbenzene → styrene obtained for (a) **Model A** and (b) **Model B**. Energy values are given in kcal mol^{-1}.

2.3.1 Model A: CNT with cis-epoxide functionalities. The energy profile is reported in **Fig.2.3a**. A schematic representation of the structure of the various critical points is given in **Fig.2.2a, 2.4** and **2.5**. In all figures, for sake of simplicity, we reproduced only a small part of the tube, *i.e.* the region around the reaction site. In reactants **Rx(A)** (**Fig.2.2a**), the well-known π-stacking interactions fasten the benzene ring to the tube wall: the two shortest C(benzene)···C(tube) distances are 3.10 and 3.31 Å. Methyl and methylene hydrogens interact with the two epoxide oxygen atoms O1 and O2. H1 and H2 form two hydrogen bonds with O1 and O2, respectively (H1···O1 distance of 2.53 Å and H2···O2 distance of 2.61 Å). Even if the CNT carbon hybridization state is intermediate between sp^2 and sp^3, a non-negligible aromatic character is present in the tube structure. As a result the typical C-C bond lengths are between 1.42 and 1.44 Å *i.e.* a little longer than the benzene CC bond (1.39 Å). The two epoxide oxygen atoms break the aromatic character of the five benzene rings sharing the tetravalent carbon atoms C3, C4, C5 and C6, as evidenced by the CC bond lengths (C3-C4 and C5-C6 are 1.52 and 1.55 Å, respectively). The hydrogen contact H1···O1 anticipates the transfer of hydrogen H1 occurring in the subsequent transition state **TS1(A)** (see Figure 3), which has an activation barrier of 34.6 kcal mol^{-1}. In **TS1(A),** hydrogen H1 moves from the methylene carbon C1 to O1 (C1···H1 and H1···O1 distances are 1.23 and 1.34 Å, respectively) causing the simultaneous breaking of the O1-C3 bond. The entire process is rather asynchronous and the computed bond distances indicate that the breaking of the epoxydic bond is early and precedes the hydrogen transfer (the O1-C3 distance is 2.21 Å). During the re-arrangement, the interaction involving O2 and one methyl hydrogen is maintained, the H3···O2 distance being 2.75 Å.

The resulting intermediate **I(A)** (see **Fig.2.4**) is 13.0 kcal mol^{-1} above **Rx(A)**. The hydrogen transfer is completed and one of the epoxide functions is replaced by a hydroxyl group. The benzene ring is still π-stacked to the tube wall. The hydroxyl O1 oxygen now mainly interacts with the remaining methylene hydrogen (H1'···O2 distance = 2.69 Å) and the methyl hydrogen H2 (H2···O2 distance = 2.66 Å), while the hydrogen contacts involving the other epoxide group (O2) have almost disappeared (the H2···O2 and H3···O2 distances are 3.52 and 4.45 Å, respectively). Crucially, the computations showed that **TS1(A)** and **I(A)** are both biradical in nature.
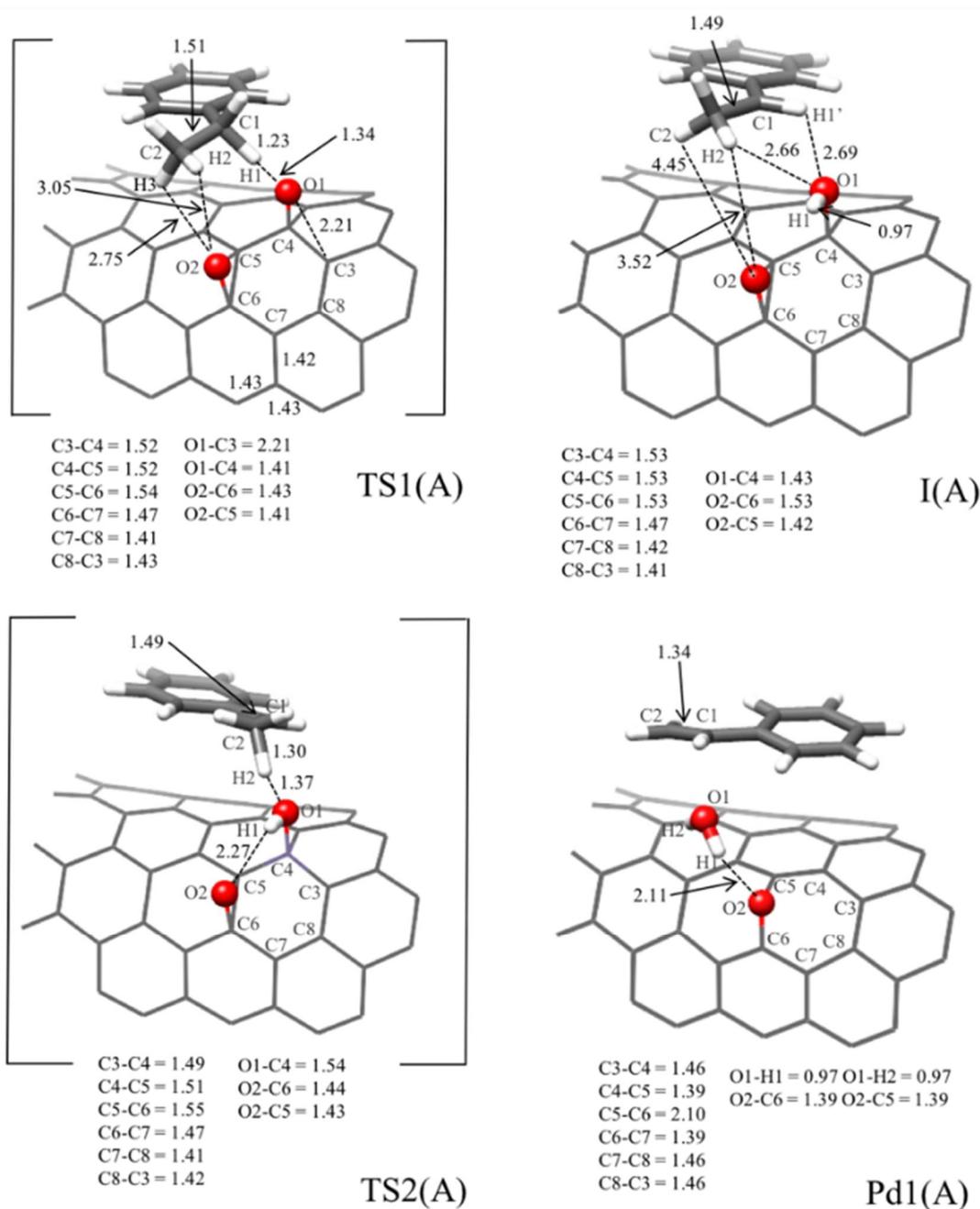
Fig.2.4 Three-dimensional representation of **TS1(A)**, **I(A)**, **TS2(A)** and **Pd1(A)**. Bond distances are in Ångstroms.

The unrestricted DFT calculations provide two singly occupied α and β Kohn–Sham orbitals characterized by a different spatial distribution: one is mainly localized on the methylene carbon C1 and the other on the two CNT carbons C3 and C7 (in **I(A)** the corresponding spin densities are 0.80 on C1 and -0.35 and -0.26 on C3 and C7, respectively). A stability-check showed that this unrestricted solution was more stable than the restricted one (closed shell electronic arrangement).

In **Model A**, the second hydrogen abstraction can involve, in principle, the same oxygen atom (O1) of the first step (now a hydroxyl function) or, alternatively, the second epoxide group (O2). Thus, from **I(A)** the reaction channel bifurcates into two paths: along one path (**path 1**, epoxide O1 involved) the second hydrogen abstraction provides directly styrene and one water molecule; along the other path (**path 2**, epoxide O2 involved) the formation of styrene is accompanied by that of two adjacent hydroxyl functions. **TS2(A)** is the transition state encountered along **path 1**. Since **TS2(A)** is 33.1 kcal mol^{-1} above the starting complex (and 1.5 kcal mol^{-1} below **TS1(A)**), a barrier of 20.1 kcal mol^{-1} must be overcome. In **TS2(A)** (depicted in **Fig.2.4**) the hydroxyl hydrogen H1 forms a hydrogen bond with O2 (the H1$\cdots$O2 distance is 2.27 Å) and H2 is approximately half way between C2 and O1 (the H2$\cdots$C2 and the H2$\cdots$O1 distances are 1.30 and 1.37 Å, respectively). The transfer of the second hydrogen to O1 causes a lengthening of the O1-C4 bond that becomes 1.54 Å (it was 1.43 Å in **I(A)**). The lengthening of O1-C4 anticipates the expulsion of a water molecule.In the product **Pd1**(A) (see **Fig.2.4**), the expelled water molecule remains anchored to the epoxide oxygen O2 via a strong hydrogen bond (H1-O2 distance = 2.11 Å). The formation of the product and the expulsion of water is accompanied by rupture of the C5-C6 bonds (C5-C6 = 2.10 Å), which determines the appearance on the tube surface of two adjacent seven-membered rings. The styrene molecule lies on the CNT surface attracted by π-π interactions that involve now both the phenyl ring system and the newly formed C1C2 π bond (C1-C2 distance = 1.34 Å). The formation of a stable water molecule (as observed in the experiment) and the partial recovery of aromaticity of CNT (only one epoxide remains on the surface), make the entire process strongly exothermic, being **Pd1(A)** 32.5 kcal mol^{-1} lower than the starting reactant complex. Along **path 2 TS3(A)** (see **Fig. 2.5**) must be overcome for the transfer of the methyl hydrogen H2 to the second epoxide group O2. **TS3(A)** is only 0.4 kcal mol^{-1} higher than **TS1(A)**. The corresponding activation barrier is 22.0 kcal mol^{-1}. When geometrically compared to **TS2(A)**, **TS3(A)** is significantly earlier and the H2 hydrogen is still strongly bonded to C2 (C2$\cdots$H2 distance = 1.17 Å). As a consequence O2 is still partially bonded to C6, the O2$\cdots$C6 distance being 1.74 Å. Since along this path we do not have the formation of a stable water molecule,

the process is less exothermic: the resulting product **Pd2(A)**, where two hydroxyl groups replace the two original epoxide groups, is 6.8 kcal mol⁻¹ higher than **Pd1(A)**.
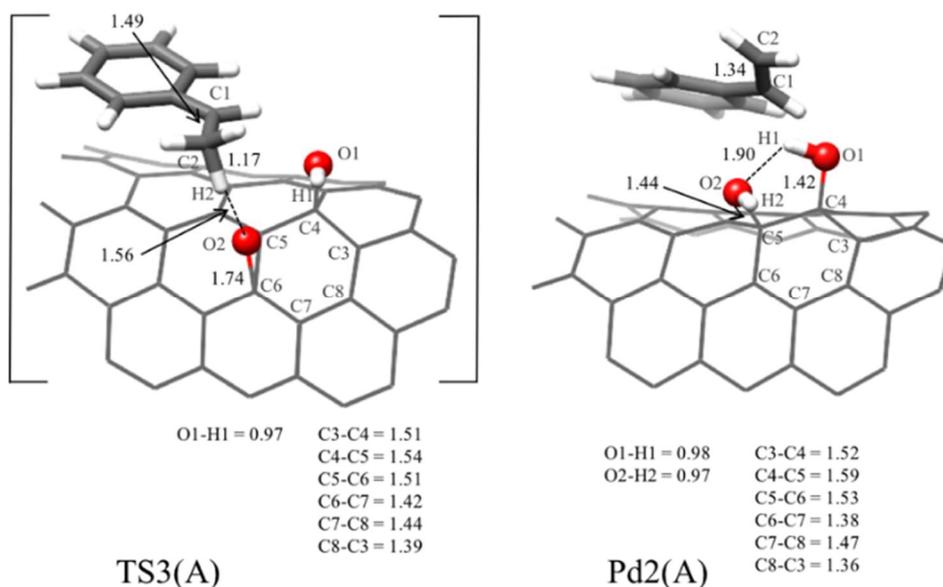

O1-H1 = 0.97	C3-C4 = 1.51
	C4-C5 = 1.54
	C5-C6 = 1.51
	C6-C7 = 1.42
	C7-C8 = 1.44
	C8-C3 = 1.39

TS3(A)

O1-H1 = 0.98	C3-C4 = 1.52
O2-H2 = 0.97	C4-C5 = 1.59
	C5-C6 = 1.53
	C6-C7 = 1.38
	C7-C8 = 1.47
	C8-C3 = 1.36

Pd2(A)

Fig. 2.5. Three-dimensional representation of **TS3(A)** and **Pd2(A)**. Bond distances are in Ångstroms.

2.3.2 Model B: CNT with quinone functionality. The energy profile is displayed in **Fig.2.3b**. A schematic representation of the structure of the various critical points is given in **Fig.2.2b** and **2.6**.

In the reactant **Rx(B),** since the two carbonyl groups are at the edge of the tube, the phenyl ring can be positioned either on the outside of the tube or point in the opposite direction above the CNT surface. The second arrangement is favoured since it allows the activation of stabilizing π-π stacking interactions between the phenyl ring and the tube wall. No minimum was found for the first orientation. The C(benzene)···C(tube) distances are similar to those computed for **Rx(A),** the two shortest C(benzene)···C(tube) distances being 3.06 and 3.31 Å. The methylene hydrogen H1 forms a hydrogen bond with the quinone oxygen O2 (H1···O2 distance = 2.45 Å). This contact precedes the hydrogen transfer occurring in the subsequent transition state. **TS1(B)** has an energy barrier of 36.9 kcal mol⁻¹. It is a rather late transition state where the new O2H1 bond is almost completed: the H1-O2 and H1···C1 distances are 1.06 and 1.62 Å, respectively.

As in **Model A**, also in the search of **TS1(B)** we adopted an unrestricted approach to compute the electron density. However, in this case the restricted solution was more stable, suggesting that a proton rather than a hydrogen atom is moving from C1 to O2. The different nature of **TS1(A)** and **TS1(B)** is confirmed by the variation of the positive charge on the migrating hydrogen: 0.27 in the former case (hydrogen transfer) and 0.41 in the latter (proton transfer).
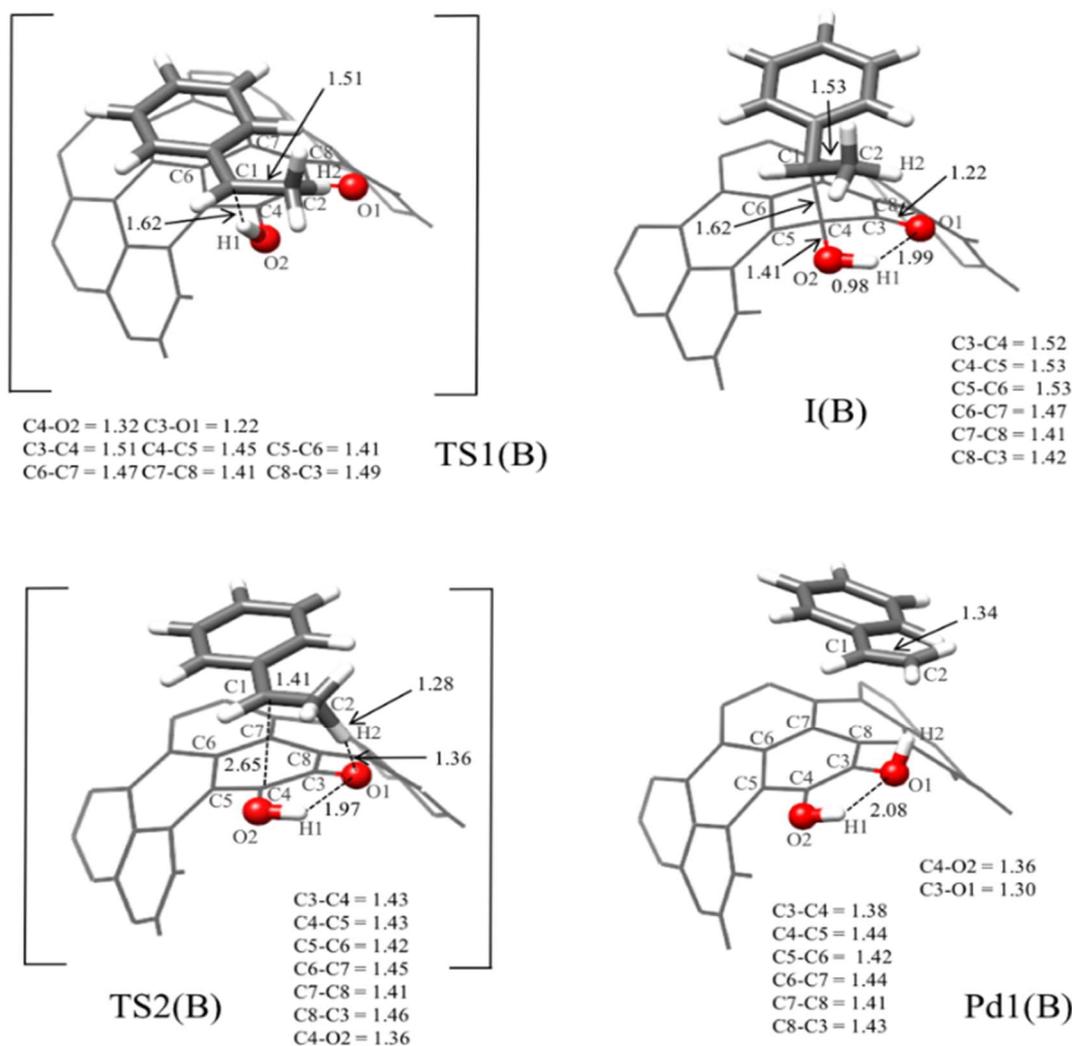
Fig. 2.6 Three-dimensional representation of **TS1(B)**, **I(B)**, **TS2(B)** and **Pd1(B)**. Bond distances are in Ångstroms.

Consequently, the nature of the resulting intermediate **I(B)** is rather different from that of **I(A)**. **I(B)** is chemisorbed on the CNT surface and an almost complete bond is formed between C1 and C4 (C1-C4 distance is 1.62 Å). **I(A)** was basically a benzyl radical intermediate, interacting with the tube via π-π stacking interactions. The charge on C1 changes from -0.38 in **TS1(B)** (where the substrate, which is losing a proton, resembles an incipient carbanion) to -0.07 in **I(B)** where the electron pair on C1 is involved in the formation of the C1-C4 bond. The formation of this bond explains the high stability of **I(B)** (compared to **I(A)**), which is 10.1 kcal mol^{-1} lower than the starting complex **Rx(B)**. Interestingly, we could also locate a biradical intermediate, similar in nature to **I(A)** and corresponding to an unrestricted solution of DFT equations. However, this biradical species **I(B)'** (depicted in **Fig.2.7**) is very high in energy (37.5 kcal mol^{-1} higher than **I(B)**) and its involvement in the catalytic process is energetically rather unlikely. Moreover, **I(B)'** is much less stable (14.4 kcal

mol[-1]) than **I(A)** if we assume as a reference point the energies of **Rx(B)** and **Rx(A)**, respectively. This is in agreement with the lower electron delocalization on the CNT surface caused by the position of the carbonyl groups at the edge of the tube.

C4-O2 = 1.33 C3-O1 = 1.25
C3-C4 = 1.47 C4-C5 = 1.41 C5-C6 = 1.43   I(B)'
C6-C7 = 1.45 C7-C8 = 1.41 C8-C3 = 1.46
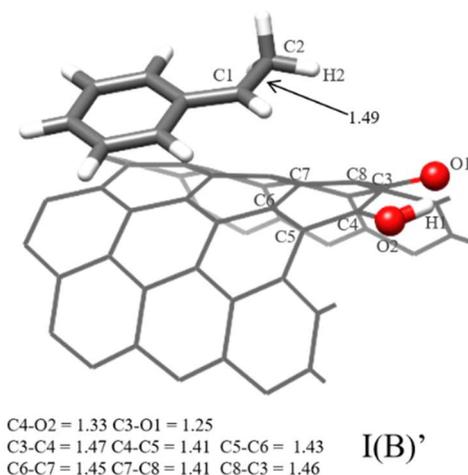
Fig. 2.7 Three-dimensional representation of **I(B)'**. Bond distances are in Ångstroms.

The subsequent transition state **TS2(B)** (**Fig. 2.6**) is 13.8 kcal mol[-1] lower than **TS1(B)**, which determines the rate of the process. In **TS2(B)** the methyl proton is about half way between C2 and O1: C2$\cdots$H2 and H2$\cdots$O1 are 1.28 and 1.36 Å, respectively. Simultaneously, the C1C2 double bond character increases (C1C2 = 1.41 Å) causing the breaking of the C1C8 bond (C1C8 = 2.65 Å).

The resulting product **Pd1(B)**, where the styrene product is stacked on the CNT surface, is 7.1 kcal mol[-1] higher than **Rx(B)**. Thus the reaction is slightly endothermic.

Reasonably, this can ascribed to the lack of water formation and the decreased delocalization caused by the replacement of the quinone functionality with two hydroxyl groups.

Furthermore, the reaction channel along which the proton transfer would involve the newly formed hydroxyl group O2H1 (thus entailing the formation of a water molecule) can be discarded on energy ground since it involves a high-energy transition state, **TS3(B)**, 61.3 kcal mol[-1] above the starting complex **Rx(B)**. A picture of **TS3(B)** is reported in **Fig. 2.8**.
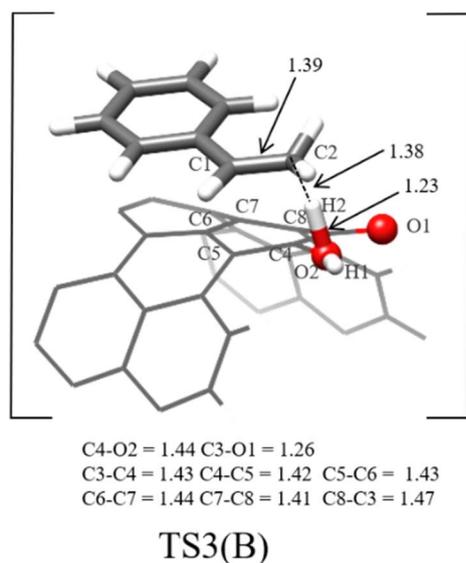
Fig. 2.8 Three-dimensional representation of **TS3(B)**. Bond distances are in Ångstroms.

The results obtained for **Model B** indicate that a single carbonyl group at the edge of the tube is not effective in catalyzing the reaction. Two adjacent carbonyl groups (quinone functionality) seems to be essential for an effective catalysis that requires an activation barrier of 36.9 kcal mol^{-1}. This barrier is similar to the energy required when one or two epoxides are involved: in that case the transition state **TS1(A)** is 34.6 kcal mol^{-1} above the reactant complex. Thus, at the working temperature of the experiment both epoxide and quinone can be effective in catalyzing the reaction.

2.4. Conclusions

We carried out a computational investigation on the mechanism of the oxidative dehydrogenation of ethylbenzene catalyzed by CNTs, a reaction with potential applications in sustainable chemistry. We used two different model-systems (**Model A** and **Model B**) to simulate the catalytic oxygen-containing groups on CNT (oxidized CNT). **Model A** contained two adjacent epoxy groups in a *cis* configuration. **Model B** consisted of a quinone functionality at the edge of the tube. Other possible functionalities, such as 1,4 carbonyl groups or dioxoethane bridges, due to the chemisorption of an oxygen molecule on CNT, were discarded because they are less stable structures.

Our computations indicate that the epoxide groups and a quinone group can effectively catalyse the ODH reaction of ethylbenzene. The activation energies required in the two cases are very similar (in the range 34-37 kcal mol^{-1}), values that can be easily surmounted at the working temperature used in the experiment (between 450 and 550 °C).

In both models, the reaction proceeds in two steps corresponding to the transfer of the two hydrogen atoms from the substrate to the oxygen atoms on the tube. A single epoxide group can catalyse styrene formation. In this case a second hydrogen abstraction can be carried out by the hydroxyl group formed on the CNT surface (transition state **TS2(A)**). A water molecule is generated *via* a highly exothermic process. Thus, the catalytic process itself (when only one epoxide group is involved) contributes to the water formation experimentally observed. Alternatively, a second adjacent epoxide group (if present) can complete the catalytic pathway by driving the second hydrogen transfer (transition state **TS3(A)**). In this case two hydroxyl groups (in relative position 1,2) form on the nanotube surface.

As suggested elsewhere, the water formation observed experimentally is probably due to the reaction between these adjacent groups or with the oxygen injected into the reactor. The energies of **TS2(A)** and **TS3(A**) (both corresponding to the second hydrogen abstraction), are only slightly lower/higher than that of transition state **TS1(A)** (first hydrogen abstraction). Interestingly, the first hydrogen abstraction involves radical species and a benzyl radical is the intermediate of the process. Two adjacent carbonyl groups (quinone functionality) are needed for an effective catalysis: the two hydrogen transfers involve both carbonyls leaving two hydroxyl groups at the edge of the tube. When a quinone functionality is involved the rate determining step clearly corresponds to the first hydrogen transfer, in agreement with recent computational evidence[33] and no radical species are involved. Unlike what found for the epoxide, the catalytic process engaging a single carbonyl group requires very high activation energy for the second hydrogen transfer and becomes very unlikely even at the high working temperature used experimentally.

2.5. Perspective

In the present work, the activation and the re-activation mechanisms of the carbon-catalyst by molecular oxygen went beyond the purpose of this computational investigation, which focused on disclosing the role of the quinone and diepoxy group in the elementary steps of oxidation of ethylbenzene.

To shed a light on the complete mechanistic scenario of ODH catalyzed by O-CNTs, it would be interesting to evaluate the mechanism of activation and re-activation of the catalysts by molecular oxygen, by means of additional computational investigations.

Two kinetic models, for the re-oxidation of the carboncatalyst, are reported in literature: the Mars van Krevelen (M-K) and the Langmuir-Hinshelwood (L-H) mechanisms. The two mechanism for

the ODH comprise the i) adsorption of ethylbenzene, ii) the hydrogen abstractions, iii) the product desorption and formation of the reduced catalyst, iv) the restoration of the active site by a molecular O_2. The two mechanisms differ in the latter step, as depicted schematically in **Scheme 2.1**.
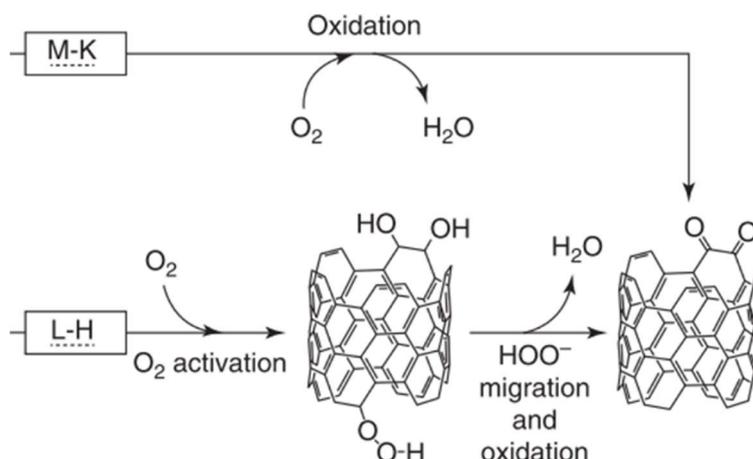
Scheme 2.1 Schematic representation of catalyst reactivation by M-K or L-H scheme.

It is been suggested that the restoration of the carbocatalyst (-OH) is more likely to proceed through the L-H scheme, where the molecular oxygen is activated in a defect-site on carbon catalysts; once the oxygen encounters the surface, it dissociates and the atomic Os diffuse on the basal plane of the carbocatalyst's surface and reaches the hydroxyl group, and the oxygen oxidizes the -OH, reactivating the active site in its carbonyl intial state.[31]

Future works may foresee the investigation of the re-activation (and activation) of the catalyst by molecular oxygen, comprising in our model some defects on the basal plane. Up to date, even if L-H mechanism is suggested to be more likely to occur, no conclusive experimental evidence have been reported in support of one or the other kinetic model.

Also, chemical modifications of the CNTs surface with borate and phospate has been proven to improve the selectivity towards the formation of the desired olefin product. Future computational investigations may be addressed to unveil the effect of doping atoms in proximity of the active sites.

References

[1] Cavani, F.; Trifirò, F. *Appl. Catal. A* **1995**, *133*, 219-239.
[2] Ballarini, N.; Cavani, F.; Cericola, A. *Catal. Today* **2007**, *127*, 113-131.
[3] Dziewçcki, Z.; Makowski, A. *React. Kinet. Catal. Lett.* **1980**, *13*, 51-54.
[4] Oganowski, W.; Hanuza, J.; Kepinski, L. *Appl. Catal. A* **1998**, *171*, 145-154.
[5] Grabowoski, R. *Catal. Rev.-Sci. Eng.* **2006**, *48*, 199-268.
[6] Emig, G.; Hofmann, H. *J. Catal.* **1983**, *84*, 15-26.
[7] Grunewald, G.C.; Drago, R.S. *J. Mol. Catal.* **1990**, *58*, 227-233.

[8]   Zhang, L.; Wu, Z.; Nelson, N.C.; Sadow, A.D.; Slowing, I.I.; Overbury, S.H. *ACS Catal.* ***2015***, *5*, 6426-6435.

[9]   Su, D.S.; Perathoner, S.; Centi, G. *Chem Rev.* **2013**, *113*, 5782-5816.

[10]   Chen, D.; Holmen, S.; Sui, Z.; Zhou, X. *Chin. J. Catal.* **2014**, *35*, 824-841.

[11]   Serp, P.; Corrias, M.; Kalck, P. *Appl. Catal. A* **2003**, *253*, 337-358.

[12]   Rinaldi, A.; Zhang, J.; Frank, B.; Su, D.S.; Hamid, S.B.A.; Schlögl, R. *ChemSusChem* **2010**, *3*, 254-260.

[13]   Su, D.S. *ChemSusChem* **2011**, *4*, 811-813.

[14]   Centi, G.; Perathoner, S. *ChemSusChem* **2011**, *4*, 913-925.

[15]   Titirici, M-M.; White, R.J.; Brun, N.; Budarin, V.L.; Su, D.S.; del Monte, F.; Clark, J.H.; MacLachlan, M.J. *Chem. Soc. Rev.* **2015**, *44*, 250-290.

[16]   Mestl, G.; Maksimova, N.I.; Keller, N.; Roddatis, V.V.; Schlögl, R. *Angew. Chem. Int. Ed.* **2001**, *40*, 2066-2068.

[17]   Keller, N.; Maksimova, N.I.; Roddatis, V.V.; Schur, M.; Mestl, G.; Butenko, Y.V.; Kuznetsov, V.L.; Schlögl, R. *Angew. Chem. Int. Ed.* **2002**, *41*, 1885-1888.

[18]   Pereira, M.F.R.; Figuerido, J.L.; Orfao, J.J.M.; Serp, P.; Kalck, P.; Kihn, Y. *Carbon* **2004**, *42*, 2807-2813.

[19]   Su, D.S.; Maksimova, N.I.; Delgado, J.J.; Keller, N.; Mestl, G.; Ledoux, M.J. *Catal. Today* **2005**, *102-103*, 110-114.

[20]   Zhang, J.; Su, D.; Zhang, A.; Wang, D.; Schlögl, R.; Hebert, C.A. *Angew. Chem. Int. Ed.* **2007**, *46*, 7319-7323.

[21]   Frank, B.; Zhang, J.; Blume, R.; Schlögl, R.; Su, D.S. *Angew. Chem. Int. Ed.* **2009**, *48*, 6913-6917.

[22]   Frank, B.; Morassutto, M.; Schomacker, R.; Schlögl, R.; Su, D.S. *ChemCatChem* **2010**, *2*, 644-648.

[23]   Pereira, M.F.R.; Figueiredo, J.L. *Catal. Today* **2010**, *150*, 2-7 and refs. therein.

[24]   Sorescu, D.C.; Jordan, K.D.; Avouris, P. *J.Phys. Chem. B* **2001**, *105*, 11227-11232.

[25]   Qi, W.; Su, D.S. *ACS Catal.* **2014**, *4*, 3212-3218

[26]   Qi, W.; Liu, W.; Zhang, B.; Gu, X.; Guo, X.; Su, D.S. *Angew. Chem. Int. Ed.* **2013**, *52*, 14224-14228.

[27]   Qi, W.; Liu, W.; Guo, X.; Schlögl, R.; Su, D.S. *Angew. Chem. Int. Ed.* **2015**, *54*, 13682-13685.

[28]   Ba, H.; Liu, Y.; Truong-Phouc, L.; Duong-Viet, C.; Nhut, J.M.; Nguyen, D.L.; Ersen, O.; Tuci, G.; Giambastiani, G.; Pham-Huu, C. *ACS Catal.* **2016**, *6*,1408-1419.

[29]   Li, B.; Su, D.S. *Chem. Eur. J.* **2014**, *20*, 7890-7894.

[30]   Li, P.; Yu, P.; Xie, J.; Liu, J.; Wang, Z.; Wu, C.; Rong, J.; Liu, H.; Su, D.S. *ACS Catal.* **2017**, *7*, 7305-7311.

[31]   Zhang, Y.; Wang, J.; Rong, J.; Diao, J.; Zhang, J.; Shi, C.; Liu, H.; Su, D.S. *ChemSusChem* **2017**, *10*, 353-358.

[32]   Kwon, H.C.; Yook, S.; Choi, S.; Choi, M. *ACS Catal.* **2017**, *7*, 5257-5267.

[33]   Zhang, J.; Liu, X.; Blume, R.; Zhang, A.; Schlögl, R.; Su, D.S. *Science* **2008**, *322 (5898)*, 73-77.

[34]   Hu, X.; Zhou, Z.; Lin, Q.; Wu, Y.; Zhanh, Z. *Chem. Phys. Lett.* **2011**, *503*, 287-291.

[35]   Khavryuchenko, O.V.; Frank, B.; Trunschke, A.; Hermann, K.; Scholgl, R. *J. Phys. Chem. C* **2013**, *117*, 6225-6234.

[36]   Du, Y.; Li, Z.; Fan, K. *Chin. J. Catal.* **2013**, *34*, 1291-1296.

[37]   Mao, S.; Li, B.; Su, D.S. *J. Mater. Chem.* **2014**, *2*, 5287-5294.

[38]   Dathar, G.K.P.; Tsai, Y.T.; Gierszal, K.; Xu, Y.; Liang, C.; Rondinone, A.J.; Overbury, S.H.; Scwarth, V. *ChemSusChem* **2014**, *7*, 483-491.

[39]   Sun, X.; Li, B.; Su, D.S. *Asian J.,* **2016**, *11*, 1668-1671.

[40]   Liu, T-F.; Ali, S.; Li, B.; Su, D.S. *ACS Catal.* **2017**, *7*, 3779-3885.

[41]   Khavryuchenko, O.V.; Frank, B. *J. Phys. Chem. C* **2017**, *121*, 3958-3962.

[42]   Humphrey, W.; Dalke, A.; Schulten, K. *J. Molec. Graphics* **1996**, *14*, 33-38.

[43]   Gaussian 09, Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; et al Gaussian, Inc., Wallingford CT, **2009**.

[44]   Zhao, Y.; Truhlar, D.G. *Theor. Chem. Acc.* **2008**, *120*, 215-241.

3. Aromatic Bromination of *N*-Phenylacetamide inside CNTs. Are CNTs real nanoreactors controlling regioselectivity and kinetics? A QM/MM investigation.[a]

3.1. Introduction

A challenging and innovative approach to reactivity, that has been receiving greater and greater attention during the last two decades, is the confinement of guest molecules inside hosts.[1-18] This approach represents a novel way for manipulating reaction energy surfaces, *i.e.* for modifying the thermodynamic properties of a reacting system and controlling the rates of reactions and their outcome.[1-18] Examples of host systems able to encapsulate reacting molecules and to behave as real nanoreactors, are cyclodextrins, calixarenes, cavitands, carcerands and zeolites.[1] More recently, specific attention as potential nanoreactors has been addressed to carbon nanotubes (CNTs), in particular single-walled carbon nanotubes (SWNTs).[12-19] These hollow nanostructures are characterized by an internal cylindrical cavity that can host molecules of different size. Depending on the CNT diameter, different reacting molecules can be selected and trapped inside the tube. An internal diameter 0.6 nm wider than the guest molecules, is often suggested for a successful encapsulation.[12-18] Thus, the variable size of the CNT cavity represents an ideal and optimal way to control the relative orientation of the guest molecules and, consequently, the regioselectivity of a given reaction. Furthermore, CNTs, that are formed by sp2 carbon atoms held together by strong covalent bonds, show great mechanical, thermal and chemical stability. These features make possible, in principle, to carry out chemical reactions inside nanotubes under many different conditions that cannot be easily achieved inside other nanoreactor systems.[12,20] In spite of the enormous amount of experimental works on CNTs that has been carried out during the last years, much of CNT-based catalysis concerns the catalytic role of the external tube wall that interacts with the reacting systems.[21-24] Because of the technical difficulties in discriminating the molecules reacting outside and insides CNTs and determining the corresponding concentrations, the usage of these nanostructures as effective nanoreactors has not been sufficiently explored yet. However, a few examples are already available in literature.[25-30] Recently, Khlobystov and co-workers[25] investigated the influence of confinement in narrow CNTs on aromatic halogenation reactions and demonstrated the possibility of regioselective control by means of SWCNTs. They examined the bromination of *N*-phenylacetamide using piridiniumdichlorobromate (PyHCl₂Br) as a brominating agent. This molecule was demonstrated to be an effective source of positive bromine in the form of

bromine chloride complex BrCl.[31] In water solution at room temperature this reaction provides a mixture (68:32) of *N*-(4-bromophenyl)acetamide and *N*-(2-bromophenyl)acetamide, *i.e.* the *para* and *ortho* regioisomers, respectively. These authors found a drastic increase of regioselectivity (up to 97% of the para product) when the same reaction was carried out inside nanotubes. Since the diameters of single-walled nanotubes are largely determined by the method of production, they decided to use nanotubes obtained by the HiPCO process[32] that should have the most suitable available inner space to contain the reaction. In the experiment, one of the reactants (*N*-phenylacetamide) was first encapsulated to make sure that the observed bromination reaction was occurring within the nanotube and not outside on its surface. Interestingly, when the reaction was carried out with closed nanotubes, no increase of the *para* product was observed with respect to the reaction in water, enforcing the hypothesis that an open SWCNT can behave like a nanoreactor. As a further check, the reaction was carried out with nanotubes obtained by CoMoCAT[25] and arc discharge (AD)[25] techniques that provide narrower and wider SWCNTs, respectively, with respect to HiPCO. In both cases no change in para selectivity was observed with respect to the solution. This result was consistent with the idea that in the former cases the internal cavity is too small to encapsulate reactants, while in the latter case, being the nanotube much larger, the reaction proceeds similarly to the solution without interaction with the internal tube wall. Also, the authors pointed out a significant acceleration of the confined reaction with respect to the solution. They suggested that the highly polarizable nanotube stabilizes the charge separation (and the associated dipole moment) occurring in the intermediate and transition state along the *para* reaction path with a consequent decrease of the activation barrier.

In the present paper we describe the results of a computational investigation on the aromatic bromination of *N*-phenylacetamide by pyridiniumdichlorobromate (PyHCl$_2$Br) in simulated water solution and inside carbon nanotubes. To this purpose we used either a full QM DFT approach based on the M06-2X functional (reaction in water) or a QM/MM approach (reaction within CNTs). The QM/MM method that we adopted is based on ONIOM technology and has been recently proven suitable to describe the chloride exchange S$_N$2 reaction of CH$_3$Cl and the Menshukin reaction inside CNTs.[33-35] Our model system was formed by the *N*-phenylacetamide substrate molecule and the Br-Cl species (originating from PyHCl$_2$Br, which is the source of positive electrophilic bromine). Before investigating the reaction in water and inside CNTs, we carried out, for the sake of comparison, a preliminary study in a non-polar aprotic solvent (ethylbenzene). The computations inside CNTs were carried out in ethylbenzene to emulate the experimental conditions where CNTs are first filled

with *N*-phenylacetamide (the dielectric constant of *N*-phenylacetamide and ethylbenzene are almost identical).

3.2. Computational Details

Nanotubes were built using the Nanotube Builder plugin, as implemented in VMD.[36] Carbon atoms at the two far ends of the tube were hydrogen-terminated. A (8,8) CNT was chosen as a model-system. This tube has an external diameter of 10.8 Å, which is close to the average diameter of the tubes used in the experiment[13] and produced with the HiPCO technique.

The length of the model-CNT (8,8) is 12.08 Å, identical to that of our previous work.[33,34] All reported computations were carried out with the Gaussian09[37] series of programs. In the full QM DFT calculations (reaction in ethylbenzene and water) we used the M06-2X[38] functional and the 6-31+G* basis set.[37] The structure of the various critical points (minima and saddle points) was fully optimized using the gradient method available in Gaussian 09. The solvent effect was evaluated with the Polarizable Continuum Model (PCM),[39] as implemented in Gaussian 09. The nature of the critical points was checked by frequency computations at the same level of theory. Additional computations with the more accurate 6-311++G** basis[37] were carried out to check the reliability of our approach.

ONIOM calculations[40] were carried out for the reaction inside CNTs considering mechanical and electrostatic embedding.[41] The inner layer (high-level layer) consisted of the reacting system, *i.e.* the N-phenylacetamide substrate molecule and the Br-Cl species. These were described at the DFT level using the M06-2X functional[38] and the 6-31+G* basis set.[37] The outer layer (low-level layer) was formed by the nanotube and was described at the molecular mechanics (MM) level based on the UFF force field.[42] In the MM calculations partial atomic (point) charges were used to compute the electrostatic interactions. The charges were calculated using the QEq formalism.[42] The structure of the various critical points was first optimized in gas-phase and frequency calculations were carried out. All critical points were re-optimized in the presence of solvent effects (ethylbenzene emulated by the implicit solvent approach PCM). [39]

3.3. Results and Discussion

3.3.1 The aromatic halogenation reaction in solution. An initial computational investigation was carried on the aromatic bromination of N-phenylacetamide in an aprotic nonpolar solvent (ethylbenzene) and in water.

Following the conventional interpretation of electrophilic aromatic substitutions reported in textbooks, these reactions should proceed via a two-step mechanism (S_EAr) involving the formation of an arenium ion (σ-complex or Wheland intermediate) followed by proton loss from the *ipso* position leading to the final product. This rather popular mechanism is schematically represented in **Scheme 3.1**.
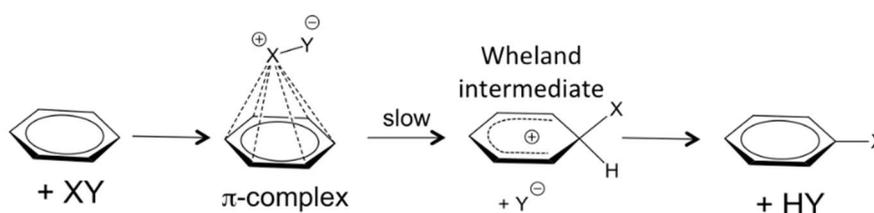
Scheme 3.1 Electrophilic aromatic substitution mechanism.

Various experimental data support the hypothesis of a σ-complex formation, at least under conditions favorable for their existence and detection.[44-47] The most common evidence is the very small k_H/k_D kinetic isotope effect measured in most cases.[44]

This finding was explained by assuming that in the first step (considered as rate-determining) one of the ring carbon atoms is involved the attack by the electrophile and changes its hybridization state (from sp^2 to sp^3) forming the σ complex, while in the faster second step the C-H bond breaks and the proton moves away.

We tried to locate the σ-complex for the *para* and *ortho* bromination occurring in ethylbenzene (the corresponding energy profiles are reported in **Fig.3.1**). We identified a Wheland intermediate only along the *ortho* pathway (left side of the figure). The intermediate **W(o)$_E$** is an ion-pair (positive arenium ion and chloride) stabilized by hydrogen contacts involving chloride and both the amidic hydrogen and *ortho* hydrogen bonded to the nearby sp^3 carbon (the Cl$\cdots$HN and Cl$\cdots$HC distances are 1.90 and 2.52 Å, respectively). These interactions give rise to a bridged structure evidenced in **Fig.3.1**. The complete energy profile reported in the same figure shows that the intermediate **W(o)$_E$** is 7.2 kcal mol^{-1} above the initial π complex (according to a partial aromaticity loss). The first transition states **TS1(o)$_E$** leading to **W(o)$_E$** has an activation energy of 28.3 kcal mol^{-1} (rate-determining step). The second transition state **TS2(o)$_E$** affording the final products is much lower in

energy, being only 10.6 kcal mol^{-1} above the π complex. In **TS2(o)$_E$** the chloride ion is removing the proton in the *ipso* position to restore the ring aromaticity: Cl strongly interacts with the hydrogen (Cl$\cdots$H distance = 1.91 Å) causing a weakening of the C-H bond (H$\cdots$C distance = 1.19 Å). **TS1(o)$_E$**, **TS2(o)$_E$**, the starting π complex and the final *ortho* substitution product are all depicted in **Fig.3.1**.
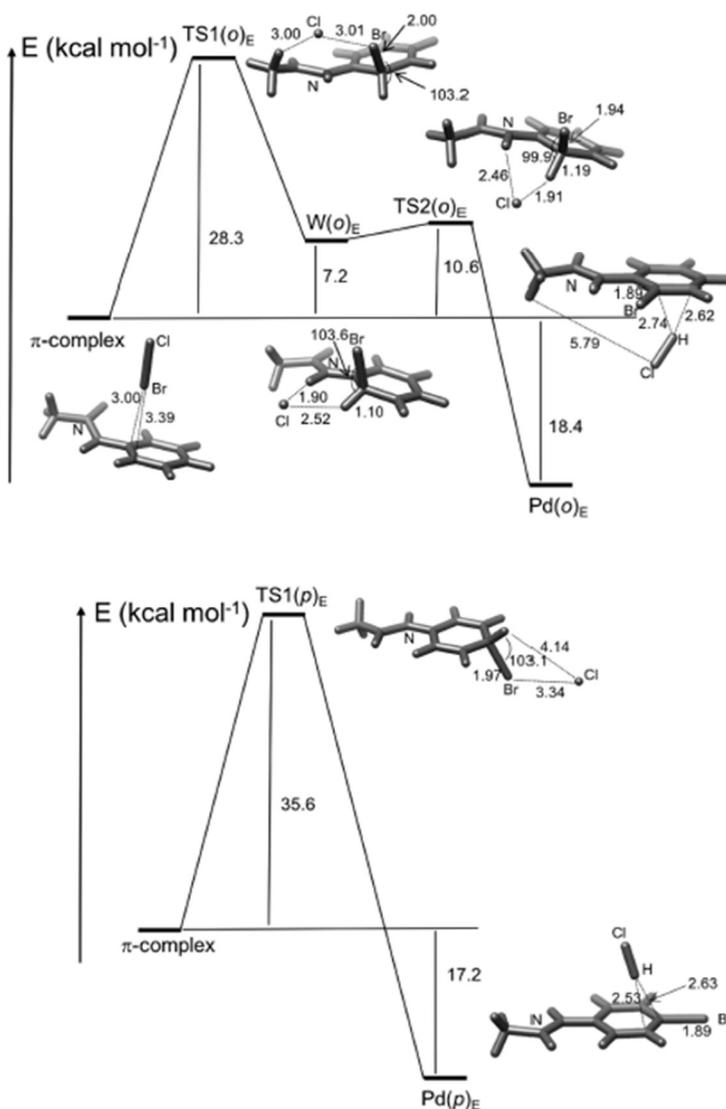
Fig.3.1 Reaction profile computed at the M06-2X/6-31+G* level for the *ortho* (upper side) and *para* (lower side) aromatic bromination of *N*-phenylacetamide occurring in ethylbenzene. A schematic representation of the various critical points is given. Energy values are in kcal mol^{-1}. Bond distances are in Ångstroms and angles in degrees.

Along the *para* pathway (bottom side of **Fig.3.1**) no Wheland intermediate was located. The reaction proceeds *via* a concerted direct substitution mechanism involving transition state **TS1(p)$_E$**, higher than **TS1(o)$_E$** (35.6 kcal mol^{-1} above the starting π complex). This transition structure shows an almost complete C-Br bond (C-Br = 1.97 Å) with the chloride weakly interacting with the bromine atom

(Cl··Br distance = 3.34 Å). Clearly, for the *para* attack, in the absence of a polar solvent, the ion pair (positive arenium ion and chloride ion that originates after release of the positive bromine) is not sufficiently stabilized to allow the existence of the Wheland intermediate. Also, the NH(CO)CH$_3$ substituent is too far away to form a stabilizing bridge structure where chloride simultaneously interacts with the amidic hydrogen and the ring hydrogen. The two transition states **TS1(o)$_E$** and **TS1(p)$_E$** are rather similar (The Cl··Br distance is 3.01 and 3.34 Å in the former and latter case, respectively) and almost equally advanced towards product formation (C··Br bond length is 2.00 and 1.97 Å). However, the absence of the bridge structure with the NH(CO)CH$_3$ substituent makes **TS1(p)$_E$** significantly less stable than **TS1(o)$_E$** and determines a regioselectivity between the *ortho* and *para* products, which is in contrast to the usual assumption based on steric effects. The existence of a concerted reaction path was first demonstrated by Galabov and co-workers[48-52] who carried out a computational investigation in vacuum and non-polar solvent (CCl$_4$) of the reaction of benzene and other arenes with Cl$_2$ and Br$_2$.[48,50-52] Neither for chlorination nor for bromination these authors were able to locate a Wheland-type intermediate. However, they pointed out the important autocatalytic effect of the HCl reaction by-product[51] in agreement with the results obtained many years ago by Andrews and Keefer[53,54] who found that HCl reduces the activation barrier in the chlorination of arenes.

To identify possible catalytic effect of acid chloride in the reaction between Br-Cl and N-phenylacetamide, we re-investigated the reaction in ethylbenzene including a HCl molecule in our model-system. We found again that the *ortho* reaction involves the formation of a σ-complex while the *para* reaction can proceed only in a concerted manner *i.e. via* direct substitution (see **Fig.3.2**).
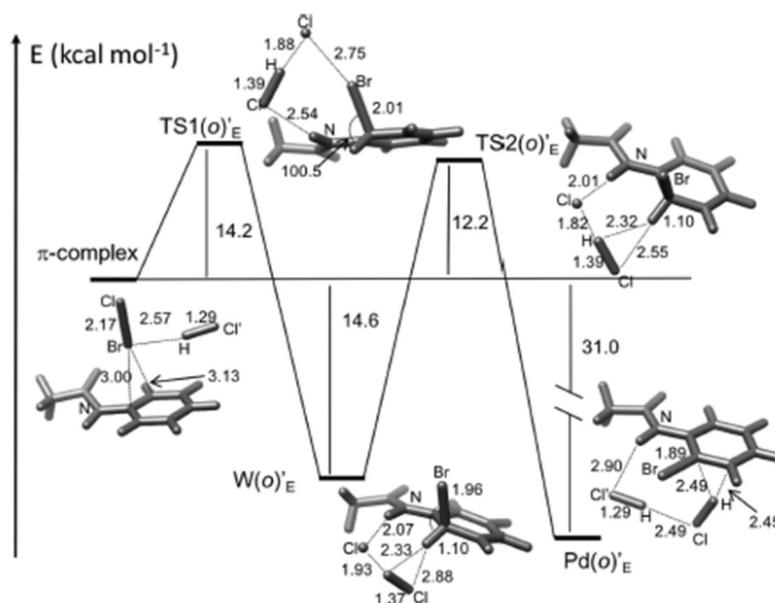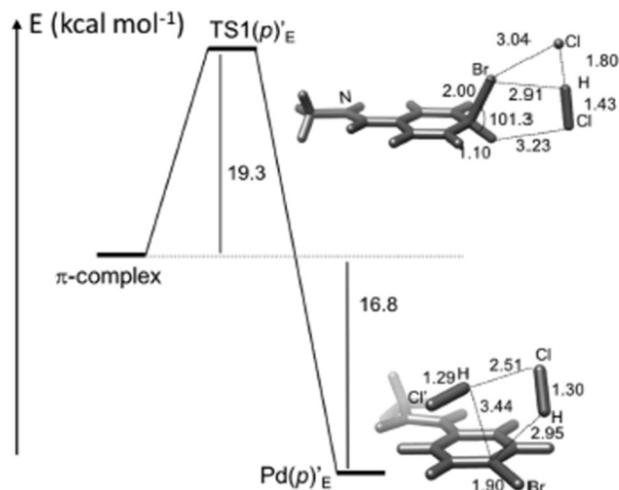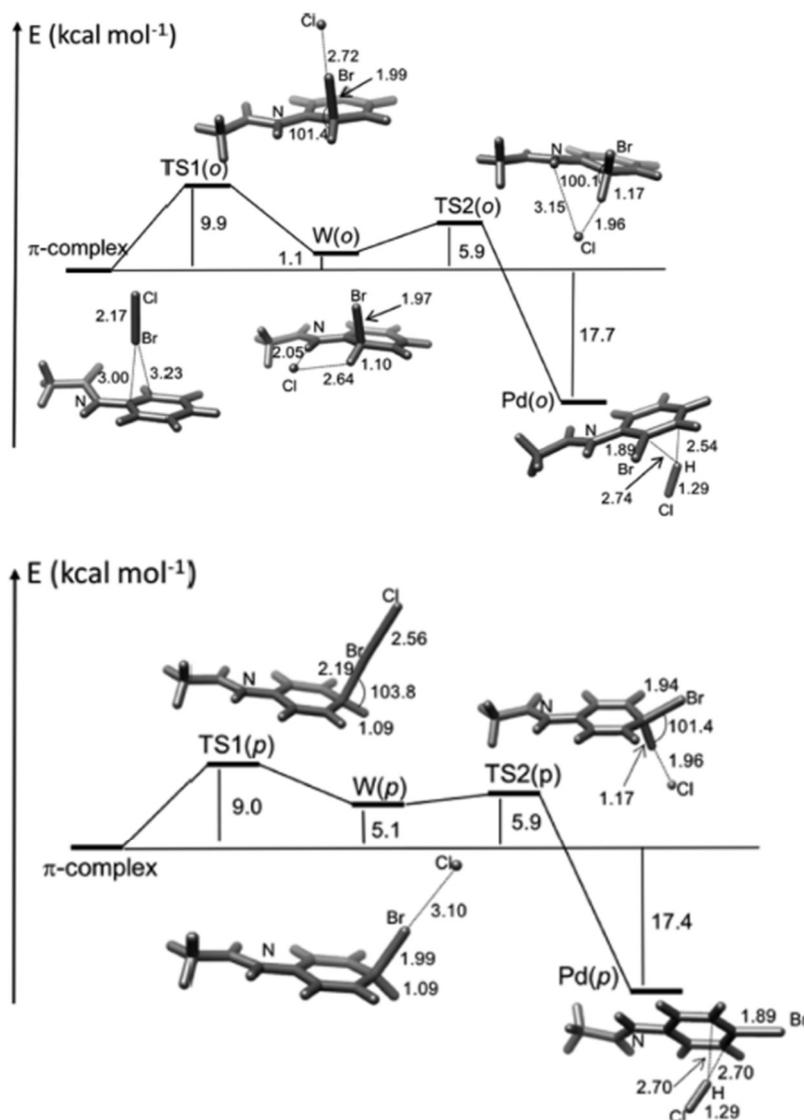
Fig.3.2 Reaction profile computed at the M06-2X/6-31+G* level for the HCl-catalyzed *ortho* (upper side) and *para* (lower side) aromatic bromination of *N*-phenylacetamide occurring in ethylbenzene. A schematic representation of the various critical points is given. Energy values are in kcal mol⁻¹. Bond distances are in Ångstroms and angles in degrees.

In the presence of HCl the *ortho* attack remains significantly favored with respect to the *para* attack (against the most common belief) and both *ortho* and *para* activation barriers considerably decrease, demonstrating the catalytic effect exerted by the reaction by-product HCl. The two reaction profiles (**Fig.3.2**) show that the rate-determining step for the *ortho* path (**TS1(o)′ᴇ**) has an intrinsic barrier of 14.2 kcal mol⁻¹, while for the direct *para* substitution (**TS1(p)′ᴇ**) the barrier is 19.3 kcal mol⁻¹. In the figure we give a schematic representation of all critical points located along the two reaction paths. The stabilization of both *ortho* and *para* reaction paths can be ascribed to a network of hydrogen contacts involving hydrogen chloride. This network is more effective in the *ortho* case since HCl interacts with the polar amidic hydrogen (Cl···HN distance is 2.54 and 2.07 Å in **TS1(o)′ᴇ** and in the intermediate **W(o)′ᴇ′** respectively). The stronger Cl···HN interaction and the additional interactions of HCl with the ring hydrogen in the *ipso* position make possible the existence of the Wheland intermediate. These interactions are also responsible of the larger stabilization of the product complex along the *ortho* pathway (31.0 kcal mol⁻¹) with respect to the *para* pathway (16.8 kcal mol⁻¹). The mechanistic scenario obtained in the simulated water solution is rather different. In these conditions we could locate for both *ortho* and *para* attack a Wheland intermediate (**W(o)ᴡ** and **W(p)ᴡ**, respectively) and the corresponding transition states **(TS1(o)ᴡ, TS2(o)ᴡ, TS1(p)ᴡ** and **TS2(p)ᴡ)**. We did not include the H-Cl molecule in the model-system since in aqueous solution H-Cl is mostly dissociated. The *ortho* and *para* reaction profiles reported in **Fig.3.3** (with a schematic representation of the various critical points) show that in both cases the formation of the intermediate is the rate-

determining step and the *para* reaction is now favoured with respect to the *ortho* reaction, the two activation barriers being 9.0 and 9.9 kcal mol⁻¹, respectively.

Fig.3.3 Reaction profile computed at the M06-2X/6-31+G* level for the *ortho* (upper side) and *para* (lower side) aromatic bromination of *N*-phenylacetamide occurring in water. A schematic representation of the various critical points is given. Energy values are in kcal mol⁻¹. Bond distances are in Ångstroms and angles in degrees.

The steric effects (reasonably due to the interactions between the chlorine lone-pairs and the nitrogen lone-pair) seem now to prevail in favouring the *para* regioselectivity. To check the accuracy of our computational approach we reinvestigated the reaction in water using the more accurate 6-311++G** basis set and the same functional M06-2X. The mechanistic scenario does not change significantly at this level of accuracy. The *para* reaction remains favoured with respect to the *ortho* reaction, the two activation barriers (transition states **TS1(p)w** and **TS1(o)w**) being now 9.1 and 7.7 kcal mol⁻¹, respectively. In Table 1 we report the energy values (relative to the π-complex) of all critical points

involved along the two reaction pathways in water. The above results are consistent with those recently obtained by Galabov and co-workers[48-52] who pointed out that charged arenium ions (corresponding to the Wheland intermediate) can form under favourable conditions, such as, for instance, a polar solvent or suitable substituents on the benzene ring. Here, the polar water environment stabilizes the charged Wheland intermediate formed by arenium ion and chloride along both *ortho* and *para* reaction paths. In the absence of water, the NH(CO)CH₃ substituent can stabilize the Wheland intermediate and allow its existence only along thee *ortho* pathway.

3.2.2 The CNT Confined Reaction. We examined both the non-catalyzed and HCl-catalyzed CNT-confined bromination reaction of *N*-phenylacetamide. The computed reaction profiles are shown in **Fig.3.4**.
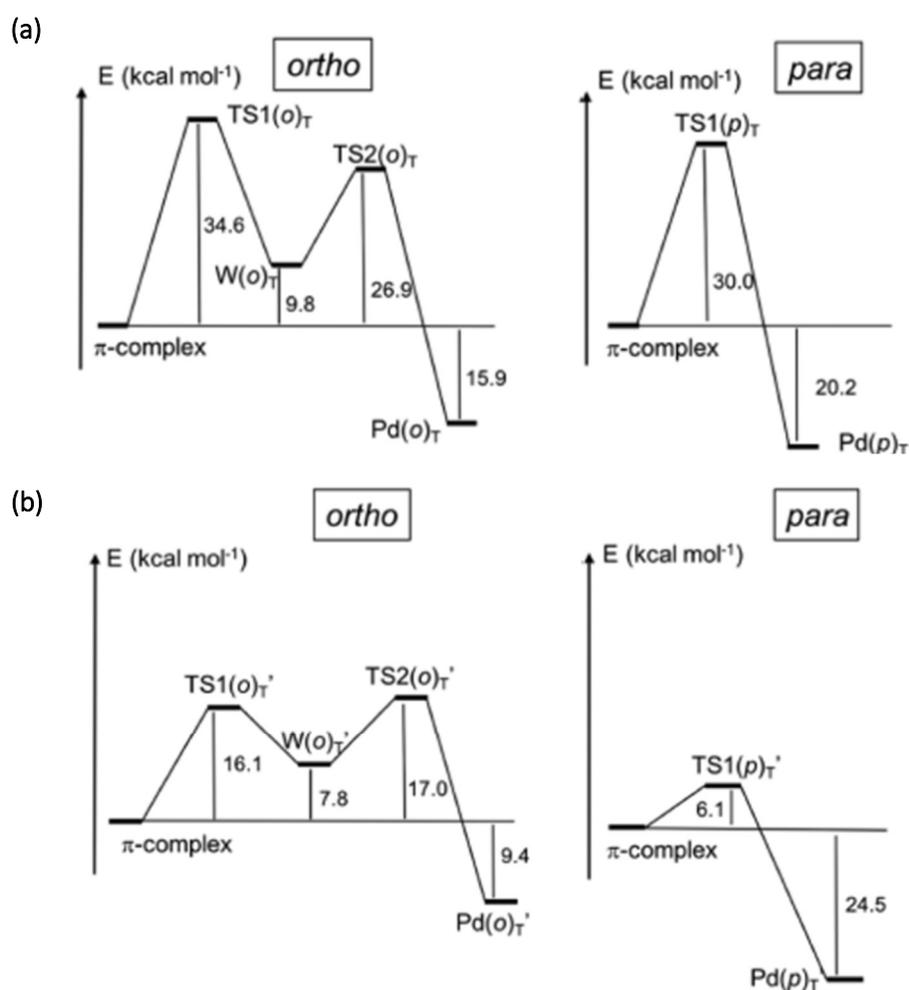
Fig.3.4 Reaction profiles computed at the QM/MM level for the ortho and para aromatic bromination of *N*-phenylacetamide occurring inside the CNT: (a) Non-catalyzed reaction, (b) HCl catalyzed reaction. Energy values are in kcal mol⁻¹.

A schematic representation of the most important critical points (π complex and **TS1** transition states: **TS1(o)**$_T$, **TS1(p)**$_T$, **TS1(o)'**$_T$ and **TS1(o)'**$_T$) is given in **Fig.3.5-3.8**. The remaining critical points located along the four reaction profiles of **Fig.3.4** are depicted in **Fig.3.9** and **3.10**. The reaction mechanism, in the absence and presence of HCl, remains approximately unchanged with respect to the non-confined reaction in ethylbenzene and the structure of the various critical points is very similar to that of the corresponding critical points computed for the non-confined process. The hydrogen contacts involving chlorine, the amidic hydrogen and the HCl molecule (see **Fig.3.5** and **3.7**) play again a key-role in stabilizing the structure and allow the existence of the Wheland intermediate only along the *ortho* pathway. In the absence of HCl the formation of **W(o)**$_T$ occurs by overcoming an activation barrier of 34.6 kcal mol^{-1} (transition state **TS1(o)**$_T$). The concerted *para* transition state **TS1(p)**$_T$ is favoured, the corresponding activation barrier being 30.0 kcal mol^{-1}. Even if the effect of the tube is that of favouring the *para* substitution with respect to the *ortho* one, the energy barrier is rather large for a reaction, which proceeds easily at room temperature.
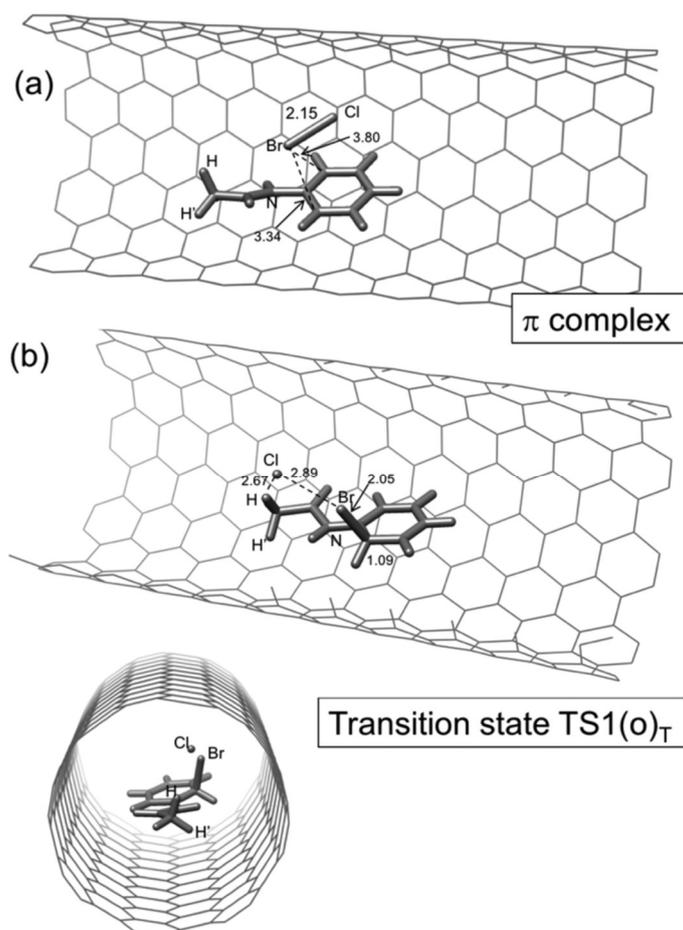
Fig.3.4 Reaction profiles computed at the QM/MM level for the ortho and para aromatic bromination of *N*-phenylacetamide occurring inside the CNT: (a) Non-catalyzed reaction, (b) HCl catalyzed reaction. Energy values are in kcal mol^{-1}.
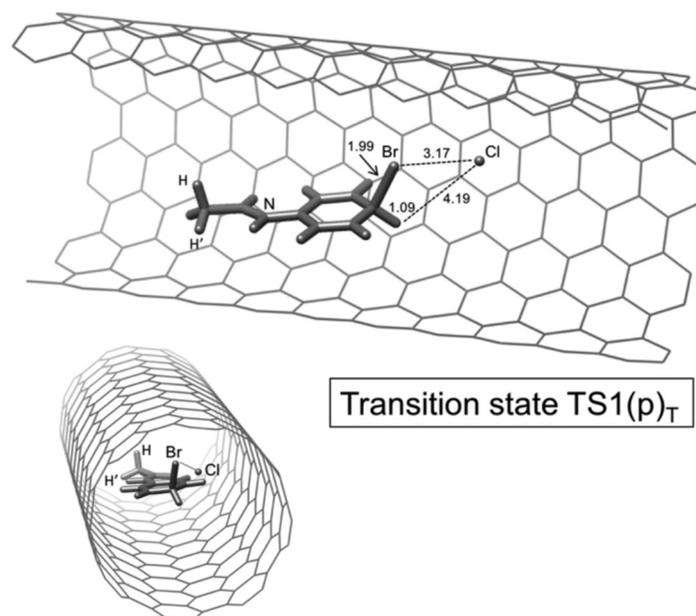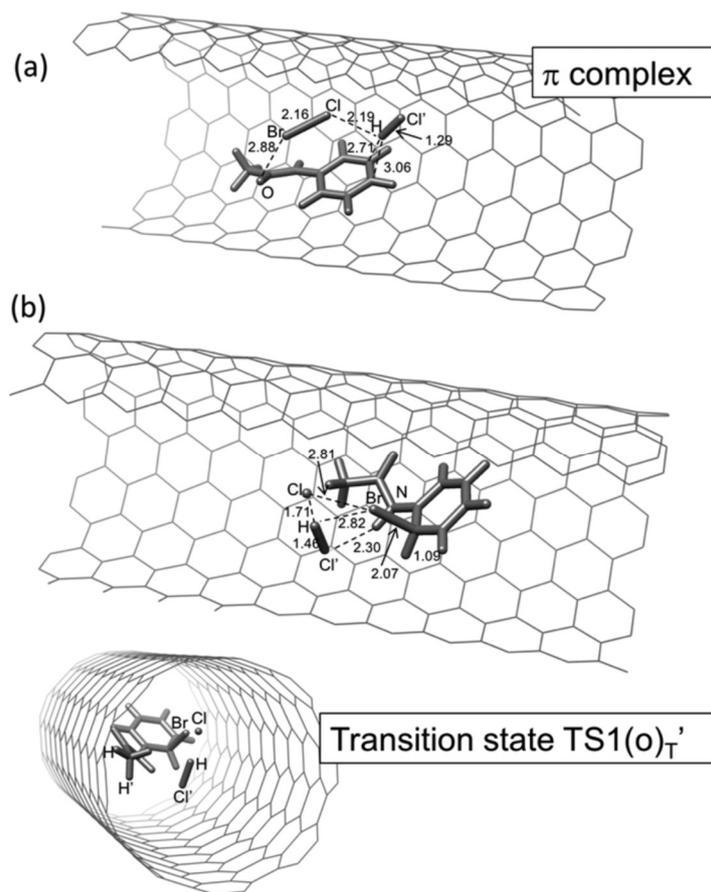
Fig.3.6 A schematic representation of the transition state **TS1(p)**ᴛ for the non-catalyzed bromination of *N*-phenylacetamide occurring inside the CNT (QM/MM computations). Two different perspectives are given for the transition state. Bond lengths are in Ångstroms.

The inclusion of the HCl catalyst makes **TS1(o)**ᴛ′ less stable by 0.9 kcal mol⁻¹ than **TS2(o)**ᴛ′ that becomes the rate-determining step and determines two important effects: (i) a strong decrease of both *ortho* and *para* activation barriers (they become 17.0 and 6.1 kcal mol⁻¹, respectively) and (ii) a much larger *ortho-para* difference, this being now 10.9 kcal mol⁻¹ (it was 4.6 kcal mol⁻¹ in the absence of HCl).

Thus, the results obtained for the HCl-catalyzed process are in much better agreement with the observed easy reaction (acceleration with respect to the reaction in water) and the extremely high regioselectivity towards the *para* product. The preference for the *para* product inside the tube, already evident in the non-catalyzed reaction and emphasized in the presence of HCl, can be explained in terms of non-covalent interactions (van der Waals)[55-59] between the endohedral system and the electron cloud of the surrounding CNT. This was initially suggested by a rough analysis (in terms of interatomic distances) of the change of non-covalent interactions between the various groups of the endohedral system and the tube wall on going from the starting π complex to the *ortho* and *para* transition states.

We assumed that in the π complex (where no bonds are breaking) the spatial arrangement of these groups is optimal and corresponds, to a good approximation, to the minimum of the potential energy curves describing the various non-covalent interactions. Thus, a significant deviation

(positive or negative) in the transition states of the distances of these groups from the tube wall suggested that a given interaction was becoming more repulsive (or less stabilizing).

The groups showing the most significant variations in their positions, are the Cl atoms, the methyl group, the carbonyl group and the amidic group. To bypass this qualitative picture and obtain a quantitative estimate of the change of overall non-covalent interactions between the various groups of the endohedral system and the tube wall on going from the π-complex to the *ortho* and *para* transition states, we considered the van der Waals (vdW) contributions estimated within the UFF scheme used in our QM/MM computations.

The vdW contributions are collected in **Table 3.1** for the non-catalyzed and the HCl-catalyzed reaction occurring within the tube. The overall effect of these interactions is stabilizing in both cases.

Table 3.1. UFF van der Waals contributions (kcal mol^{-1}) computed for the non-catalyzed and HCl-catalyzed reaction within CNT. These contributions refer to the interactions between the various groups of the endohedral system and the tube wall.

Non-catalyzed reaction		HCl-catalyzed reaction	
π-**complex**	-51.2	π-**complex**	-54.8
TS1(o)т	-49.2	**TS2(o)**т′	-45.0
TS1(p)т	-49.7	**TS1(p)**т′	-55.6

Importantly, it is more stabilizing for the *para* reaction path (**TS1(p)**т and **TS1(p)**т′) with respect to the *ortho* reaction path (**TS1(o)**т and **TS2(o)**т′). The difference in stabilization between *ortho* and *para* is small for the non-catalyzed reaction (-49.2 and -49.7 kcal mol^{-1}, respectively) and becomes much more significant for the catalyzed process (-45.0 and -55.6 kcal mol^{-1}, respectively), thus following the trend of the activation barriers. Interestingly, the vdW interactions become highly stabilizing for **TS1(p)**т′, which corresponds to the lowest activation barrier (only 6.1 kcal mol^{-1}). These results enforce our hypothesis that the differential in the vdW interactions between the *ortho* and *para* reaction paths is the key-factor determining the regioselectivity and the reaction speed.

Fig.3.7 A schematic representation of the (a) **π-complex** and (b) transition state **TS1(o)**$_T$' for the non-catalyzed bromination of *N*-phenylacetamide 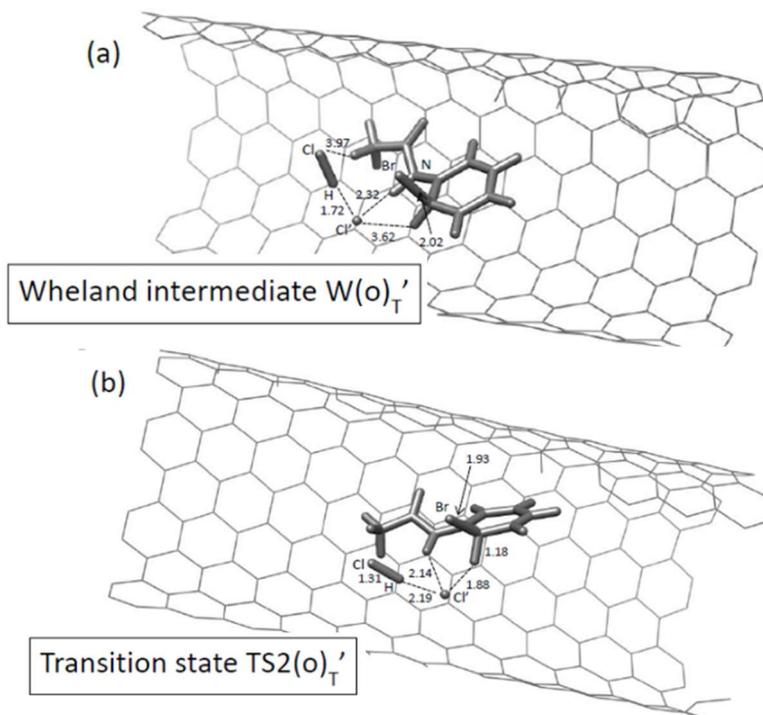occurring inside the CNT (QM/MM computations). Two different perspectives are given for the transition state. Bond lengths are in Ångstroms.
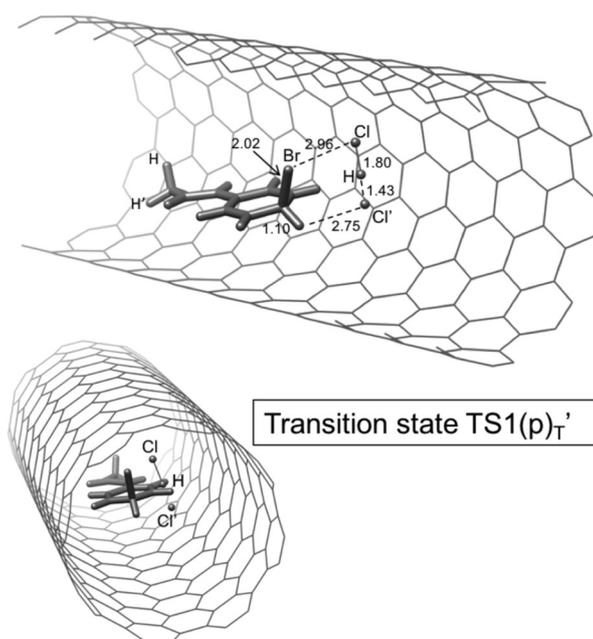
Fig.3.8 A schematic representation of the transition state **TS1(p)**$_T$' for the non-catalyzed bromination of *N*-phenylacetamide occurring inside the CNT (QM/MM computations). Two different perspectives are given for the transition state. Bond lengths are in Ångstroms.
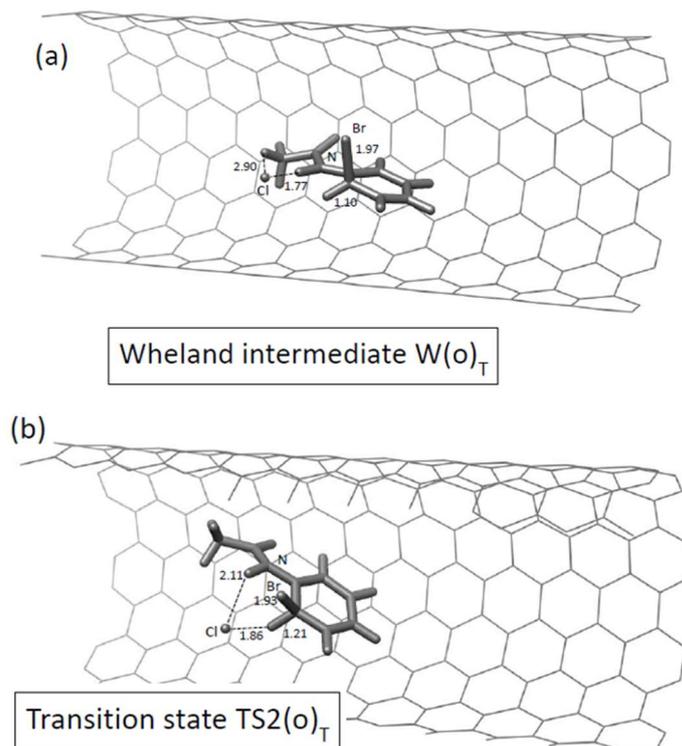
Fig.3.9 A schematic representation of (a) the Wheland intermediate **W(o)ᴛ** and (b) the transition state **TS2(o)ᴛ** for the non-catalyzed bromination of *N*-phenylacetamide occurring inside the CNT (QM/MM computations). Bond lengths are in Ångstroms.

Fig.3.10 A schematic representation of (a) the Wheland intermediate **W(o)ᴛ′** and (b) the transition state **TS2(o)ᴛ′** for the HCl-catalyzed bromination of *N*-phenylacetamide occurring inside the CNT (QM/MM computations). Bond lengths are in Ångstroms.

The mechanistic scenario, which stems from these computational results, can be described as follows. At the beginning of the process inside the tube, when no HCl molecules are present, rather large activation barriers (34.6 and 30.0 kcal mol^{-1}) must be surmounted. When HCl is gradually produced, the catalyzed mechanism is activated and the reaction proceeds almost exclusively along the *para* pathway, which requires the overcoming of a rather small barrier of 6.1 kcal mol^{-1}. The barrier of 6.1 kcal mol^{-1} is significantly lower than the barrier in water solution (9.0 kcal mol^{-1}). The difference between the two values explains satisfactorily the acceleration of the reaction in CNTs with respect to water.

3.4. Conclusions

We carried out a computational investigation on the mechanism of the CNT-confined bromination reaction of *N*-phenylacetamide. Our model-system was formed by the substrate N-phenylacetamide and bromine chloride BrCl, a complex providing positive bromine atoms. The reference experimental study on this reaction was recently carried out by Khlobystov and co-workers[25] both in aqueous solvent and inside narrow CNTs. Thus, we investigated the reaction also in water and, for comparison purposes, in an aprotic solvent (ethylbenzene).

In ethylbenzene, a Wheland intermediate exists only along the *ortho* pathway, while the *para* attack proceeds in a concerted manner (concerted direct substitution). The existence of a Wheland intermediate (an ion-pair formed by a positive arenium ion and chloride ion) is due to stabilizing interactions between the chloride and both the amidic hydrogen and the *ortho ipso*-hydrogen. In the transition state structure, resulting from the *para* attack, the chloride and the NH(CO)CH$_3$ substituent are too far away to activate stabilizing hydrogen contacts and to allow the existence of a Wheland intermediate. The activation barriers computed for the *ortho* and *para* attacks are 28.3 and 35.6 kcal mol^{-1}), respectively, which indicates a *ortho/para* regioselectivity in contrast to the common belief based on simple steric effects.

When HCl (a reaction by-product) participates in the process, both *ortho* and *para* activation barriers significantly decrease (14.2 and 19.3 kcal mol^{-1}) demonstrating the catalytic effect exerted by hydrogen chloride. However, the regioselectivity still favours the *ortho* product. The catalytic action of HCl is due to a network of hydrogen contacts that involves hydrogen chloride and is more stabilizing in the *ortho* case thanks to additional interactions between HCl and the polar amidic hydrogen.

In water solution the arenium ion and chloride ion pair is sufficiently stabilized to allow the existence of a Wheland intermediate for both *ortho* and *para* attacks. In this case hydrogen chloride was not considered since in aqueous solution HCl is greatly dissociated. The formation of the *para* product is now favoured with respect to the *ortho* product (9.0 and 9.9 kcal mol^{-1} are the corresponding activation barriers): the steric hindrance seems now to be dominant in favouring the *para* regioselectivity. This is reasonable because of the interactions between the chlorine lone-pairs and the nitrogen lone-pair.

When the reaction was examined inside CNTs (with and without HCl) we found a mechanism very similar to that identified in ethylbenzene: again a Wheland-type arenium ion exists only along the *ortho* pathway, while the *para* attack proceeds in a concerted manner. In the absence of HCl the concerted *para* reaction path (activation energy of 30.0 kcal mol^{-1}) is favoured with respect to the *ortho* one that involves a Wheland intermediate (activation energy = 34.6 kcal mol^{-1}). However, the *para* energy barrier is too large since experimentally the reaction is rather fast at room temperature. When HCl was included in the computations we observed a strong decrease of both *ortho* and *para* activation barriers (17.0 and 6.1 kcal mol^{-1}, respectively). These results suggest that, as soon as HCl is produced, the catalyzed mechanism is activated and the reaction proceeds almost exclusively along the *para* pathway. This finding, demonstrating the catalytic action of HCl, are in agreement with a fast reaction (faster than in water) and an almost exclusive *para* product.

The regioselectivity and the low activation barrier for the CNT-confined reaction are due to non-covalent (vdW) stabilizing interactions between the endohedral system and the electron cloud of the surrounding CNT. The effect of vdW interactions was estimated quantitatively within the UFF scheme used in our QM/MM computations. We demonstrated that these interactions are more stabilizing for the *para* reaction path (**TS1(p)**т and **TS1(p)**т′) with respect to the *ortho* reaction path (**TS1(o)**т and **TS2(o)**т′). Also, they become particularly stabilizing for the *para* catalyzed process (**TS2(o)**т′), suggesting that vdW interactions are the key-factor determining the acceleration of the confined reaction (compared to water) and the almost exclusive para regioselectivity.

References

a. Marforio, T.D.; Bottoni, A.; Giacinto, P.; Zerbetto, F.; Calvaresi, M. *J. Phys. Chem.* C **2017**, *121*, 27674-27682.

[1]   Brinker, U. H.; Mieusset, J.-L. Wiley: New York, 2010.

[2]   Santiso, E. E.; George, A. M.; Turner, C. H.; Kostov, M. K.; Gubbins, K. E.; Buongiorno-Nardelli, M.; Sliwinska-Bartkowiak, M.. *Appl. Surf. Science* **2005**, *252*, 766-777.

[3] Zhao, L.; Li, Y.; Yu, P.; Han, X.; He, J. *ACS Catal.* **2012**, *2*, 1118-1126.

[4] Li, Y.-P.; Head-Gordon, M.; Bell, A. T. *ACS Catal.* **2014**, *4*, 1537-1545.

[5] Li, J.; Wei, Y.; Chen, J.; Xu, S.; Tian, P.; Yang, X.; Li, B.; Wang, J.; Liu, Z. *ACS Catal.* **2015**, *5*, 661-665.

[6] Goehry, C.; Besora, M.; Maseras, F. *ACS Catal.* **2015**, *5*, 2445-2451.

[7] Goltl, F.; Michel, C.; Andrikopoulos, P. C.; Love, A. M.; Hafner, J.; Hermans, I.; Sautet, P. *ACS Catal.* **2016**, *6*, 8404–8409.

[8] Salavati-fard, T.; Caratzoulas, S.; Lobo, R. F.; Doren, D. J. *ACS Catal.* **2017**, *7*, 2240-2246.

[9] Yokoyama, H.; Kobayashi, H.; Hasegawa, J.; Fukuoka, A. *ACS Catal.* **2017**, *7*, 4828-4834.

[10] Van der Mynsbrugge, J.; Janda, A.; Sharada, S. M.; Lin, L.-C.; Van Speybroeck, V.; Head-Gordon, M.; Bell, A. T. *ACS Catal.* **2017**, 7, 2685-2697.

[11] Khlobystov, A. N.; Britz, D. A.; Briggs, G. A. D. *Acc. Chem. Res.* **2005,** *38*, 901-909.

[12] Khlobystov, A. N. *ACS Nano*, **2011**, *5*, 9306–9312.

[13] Britz, D. A.; Khlobystov, A. N. *Chem. Soc. Rev.* **2006**, *35*, 637–659.

[14] Troche, K. S.; Coluci, V. R.; Braga, S. F.; Chinellato, D. D.; Sato, F.; Legoas, S. B.; Rurali, R.; Galvao, D. S. *Nano Lett.* **2005**, *5*, 349–355.

[15] Guan, L.; Suenaga, K.; Shi, Z.; Gu, Z.; Iijima, S. *Nano Lett.* **2007**, 7, 1532–1535.

[16] Takenobu, T.; Takano, T.; Shiraishi, M.; Murakami, Y.; Ata, M.; Kataura, H.; Achiba, Y.; Iwasa, Y. *Nat. Mater.* **2003**, *2*, 683-688.

[17] Plank, W.; Pfeiffer, R.; Schaman, C.; Kuzmany H.; Calvaresi, M.; Zerbetto, F.; Meyer, J. *ACS Nano* **2010**, *4*, 4515-4522.

[18] Liu, X.; Kuzmany, H.; Ayala, P.; Calvaresi, M.; Zerbetto, F.; Pichler, T. *Adv. Funct. Mater.* **2012**, *22*, 3202-3208.

[19] Calvaresi, M.; Zerbetto, F. *J. Mater. Chem. A*, **2014**, *2*, 12123-12135.

[20] Miners, S. A.; Rance, G. A.; Khlobystov, A. N. *Chem. Soc. Rev.*, **2016**, *45*, 4727-4746.

[21] Mestl, G.; Maksimova, N. K.; Roddatis, V. V.; Schlogl, R. *Angew. Chem. Int. Ed.* **2001**, *40*, 2066-2068.

[22] Su, D. S.; Perathoner, S.; Centi, G. *Chem. Rev.* **2013**, *113*, 5782–5816.

[23] Schaetz, A.; Zeltner, M.; Stark, W. J. *ACS Catal.* **2012**, *2*, 1267-1284.

[24] Tavakkoli, M.; Holmberg, N.; Kronberg, R.; Jiang, H.; Sainio, J.; Kauppinen, E. I.; Kallio, T.; Laasonen, K. *ACS Catal.* **2017**, *7*, 3121-3130.

[25] Miners, S. A.; Rance, G. A.; Khlobystov, A. N. *Chem. Commun.,* **2013**, *49*, 5586-5588.

[26] Zhang, H.; Pan, X.; Han, X.; Liu, X.; Wang, X.; Shen, W.; Bao, X. *Chem. Sci.* **2013**, *4*, 1075-1078.

[27] Solomosz, W. A.; Ramce, G. A.; Suyetin M.; La Torre, A.; Bichoutskaia, E.; Khlobystov, A. N. *Chem. Eur. J.* **2012**, *18*, 13180-13187.

[28] Xiao, J.; Pan, X.; Guo, S.; Ren, P.; Bao, X. *J. A. Chem. Soc.* **2015**, *137*, 477-482.

[29] Solomonsz, W. A.; Rance, G. A.; Khlobystov, A. N. *Small* **2014**, *10*, 1866-1872.

[30] Chamberlain, T. W.; Biskupek, J.; Skowron, S. T.; Markevich, A. V.; Kurasch, S.; Reimer, O.; Walker, K. E.; Rance, G. A.; Feng, X.; Mullen, K.; Turchanin, A.; Lebedeva, M. A.; Majouga, A. G.; Nenajdenko, V. G.; Kaiser, U.; Besley, E.; Khlobystov, A. N. *ACS Nano* **2017**, *11*, 2509-2520.

[31] Muathen, H.A. *Synthesis* **2002**, 169-171.

[32] Nikolaev, P.; Bronikovsky, M. J.; Bradley, R. K.; Rohmund, F.; Colbert, D. T.; Smith, K. A.; Smalley, R. E. *Chem. Phys. Lett.* **1999**, *313*, 91-97.

[33] Giacinto, P.; Bottoni, A.; Calvaresi, M.; Zerbetto, F. *J. Phys Chem. C* **2014**, *118*, 5032-5040.

[34] Giacinto, P.; Bottoni, A.; Calvaresi, M.; Zerbetto, F. *J. Chem. Theory. Comp.* **2016**, *12*, 4082-4092.

[35] Tavares, I. S.; Figueiredo, C. F. B. R.; Magalhães, A. L. *J. Phys. Chem. C* **2017**, *121*, 2165-2172.

[36] Humphrey, W.; Dalke, A.; Schulten, K.; VMD - Visual Molecular Dynamics. *J. Molec. Graphics* **1996**, *14*, 33-38.

[37] Gaussian 09, Revision **D.01**, Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; et al Gaussian, Inc., Wallingford CT, **2009**.

[38] Zhao, Y.; Truhlar, D. G. *Theor. Chem. Acc.,* **2008**, *120*, 215-241.

[39] Tomasi, J.; Mennucci, B.; Cammi, R. *Chem. Rev.* **2005**, *105*, 2999–3093.

[40] Vreven, T.; Morokuma, K.; Farkas, Ö.; Schlegel, H. B.; Frisch, M. J. *J. Comp. Chem.* **2003**, *24*, 760-769.

[41] Vreven, T.; Byun, K. S.; Komaromi, I.; Dapprich, S.; Montgomery, J. A. Jr.; Morokuma, K.; Frisch, M. J. *J. Chem. Theory. Comp.* **2006**, *2*, 815-826.

[42] Rappé, A. K.; Casewit, C. J.; Colwell, K. S.; Goddard III, W. A.; Skiff, W. M. *J. Am. Chem. Soc.*, **1992**, *114*, 10024-10035.

[43] Rappé, A. K.; Goddard III, W. A. *J. Phys. Chem.*, **1991**, *95*, 3358-3363.

[44] Olah, G. A. *Acc. Chem. Res.* **1971**, *4*, 240-248.

[45] Olah, G. A.; Staral, J. S.; Asencio, G.; Liang, G.; Forsyth, D. A.; Mateescu, G. D. *J. Am. Chem. Soc.* **1978**, *100*, 6299-6308.

[46] Lyerla, J. R.; Yannoni, C. S.; Bruck, D.; Fyfe, C. A. *J. Am. Chem. Soc.* **1979**, *101*, 4770-4772.

[47] Lenoir, D. *Angew. Chem. Int. Ed.* **2003**, *42*, 854-857.

[48] Kong, J.; Galabov, B.; Koleva, G.; Zou, J.-J.; Schaefer III, H. F.; Schleyer, P. *Angew. Chem. Int. Ed.* **2011**, *50*, 6809-6813.

[49] Koleva, G.; Galabov, B.; Kong, J.; Schaefer III, H. F.; Schleyer, P. *J. Am. Chem. Soc.* **2011**, *133*, 19094-19101.

[50] Galabov, B.; Koleva, G.; Kong, J.; Schaefer III, H. F.; Schleyer, P. *Eur. J. Org. Chem.* **2014**, 6918-6924.

[51] Galabov, B.; Koleva, G.; Simova, S., Hadjieva, B.; Kong, J.; Schaefer III, H. F.; Schleyer, P. *Proc. Natl. Acad. Sci. USA* **2014**, *111*, 10067-10072.

[52] Galabov, B.; Nalbantova, D.; Schleyer, P. v. R.; Schaefer H. F. *Acc. Chem. Res.*, **2016**, 49, 1191-1199.

[53] Andrews, L. J.; Keefer, R. M. *J. Am. Chem. Soc.* **1957**, *79*, 5169-5174.

[54] Andrews, L. J.; Keefer, R. M. *J. Am. Chem. Soc.* **1959**, *81*, 1063-1067.

[55] Prasanna, M. D.; Guru Row, T. N. *Crystal. Eng.* **2000**, *3*, 135-154.

[56] Saraogi, I.; Vijay, V. G.; Soma, D.; Sekar, K.; Guru Row, T. N. *Crystal. Eng.* **2003**, *6*, 69-77.

[57] Imai, Y.N.; Inoue, Y.; Nakanishi, I; Kitaura, K. **2008**, *17*, 1129-1137.

[58] Jain, A.; Purohit, S, C.; Verma, S.; Sankararamakrishnan, R. *J. Phys Chem.* **2007**, *111*, 8180-8683.

[59] Gung, B.W.; Zou, Y.; Xu, Z.; Amicangelo, J. C.; Irwin, D. G.; Ma, S.; Zhou, H.-C. *J. Org. Chem.* **2008**, *73*, 689-693.

Part III

Biocatalysis

1. Introduction

Life would not be possible without enzymes since these molecular machines catalyse almost every metabolic process in living cells. The number of reactions they speed up is countless. The complete enzymatic mechanism is a complex process, that requires the adsorption of the substrate in the binding pocket, the formation of the Michaelis complex, the transformation of the substrate into the products and the restoration of the catalytic site, and finally the desorption of the product (**Fig.1.1**).
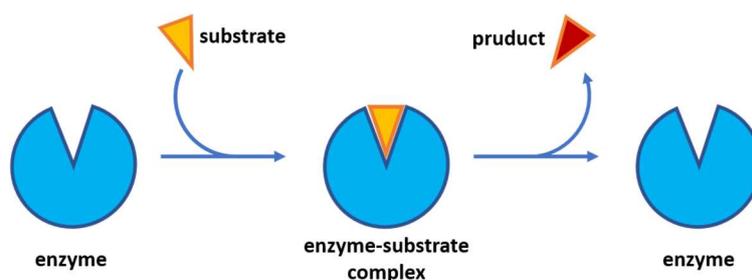
Fig.1.1 General representation of enzymatic mechanism.

Anciently enzymes were unconsciously employed to produce bread, wine and other provisions. Nowadays, thanks to the scientific progress, enzymes are exploited in a wide range of sectors such as in the production of detergents, food, biofuels and are the main targets of pharmaceutical industries.[1] In synthetic area, enzymes are interesting catalysts because of their i) chemo-selectivity (preferential product over possible outcomes), ii) regio-selectivity (preferential bond formation or cleavage among all possibilities), iii) stereo-selectivity (preferential formation of an enantiomer or diastereoisomer) and iv) stereo-specificity (specification of the stereochemical outcome). Because of their ability of generating asymmetric products, industry takes advantage of enzymes to synthetize enantiopure fine chemicals in medium- and large- scale production. Chemoenzymatic approaches, *i.e.* the use of enzymes in synthetic organic chemistry, have the enormous advantage of having their best operating conditions in aqueous solutions and low temperatures. These characteristics make enzymes the ultimate environmental friend catalysts at a chemist's disposal.[2] Despite the

advantages in using enzymes in synthesis, some drawbacks have to be taken in account: most of enzymes are quite sensitive to pH and temperature, and they must be handled with care to avoid the denaturation of the protein which would cause the catalytic activity to be switched off. Also, they are quite selective towards their natural substrates, and in many cases, they hardly recognize modified substrates. This latter problem can be overcome by nano-bio-technology, which modifies genes to obtain mutated enzymes (mutagenesis), where few amino acids are substituted with other residues, to push the enzyme to recognize non-standard substrates. Mutagenesis is also employed to modify substrate specificity and reaction mechanism, by changing the active site residues.

The establishment of enzymatic mechanism may be, in some cases, too complex to be solely resolved by experimental techniques and computer-based models are a tool at chemists' disposal to provide information which are elusive to experiments. In particular, the study of the reaction mechanism by computational approaches provides structures and energetics of transition states and reaction intermediates and supplies insights into the factors that govern the enzymatic stereo- and regioselectivity.[3] The recent development of high-performance computers helps the study of enzymatic reactivity in a high-detailed fashion, and the computational information are nowadays complementary to experimental results. A reliable computational study of an enzymatic reaction is based on experimental evidences. The investigation of the reaction mechanism is generally carried out on a *model-system*, which is extracted from the enzyme three-dimensional crystal structure (obtained by X-ray crystallography or NMR techniques), usually deposited at the protein data bank (PDB)[4], which collects more than 140000 resolved structures, ranging from proteins to nucleic acids. The *model-system* may have different sizes, specifically it can be: i) a small cluster, where the model is cut out from the enzyme active site and contains the substrate and some active residues involved in the catalysis, ii) an extended cluster, where a great portion of the enzyme is comprised into the model, overcoming some limitations of "small cluster" approaches, and iii) a full enzyme model, which can allow to investigate conformational rearrangements of the whole protein during the reaction course.[5] For small clusters, usually first principles QM methods are employed, where all the atoms that comprise the model system are treated quantum-mechanically. When the size increases, hybrid QM/MM approaches are necessary. As described in Part I, Section 5, with this methodology the model system is divided into two (or three) layers, where the higher one comprises the reactive atoms are treated at the QM level, while the atoms of the lower level, which account for the protein environment, are treated at the MM level. A full QM approach has the advantage of using the highest possible accuracy and it has been proven to be particularly efficient in

discriminating between different mechanistic proposals[6] but when the enzymatic environment has a strong effect on the catalyzed reaction, by means of long-range interactions, the enzyme has to be emulated by a dielectric continuum.[7,8] When the dielectric continuum is not sufficient to simulate the protein environment, or when the reaction involves significant conformational changes that are crucial in the catalytic process, hybrid QM/MM approaches are necessary to take into account such modifications and interactions. Increasing the size of the model has an undeniable disadvantage of rising the computational time, and the need of lowering the computational accuracy may have the disadvantage of losing mechanistic details.

In this thesis we report two examples of small cluster approaches, where a full QM approach was employed to study an enzymatic reaction and another example were a QM/MM methodology was adopted.

References

[1]   Reetz, M.T. *J. Am. Chem. Soc.,* **2013**, *135*, 12480-12496.
[2]   Sheldon, R.A.; Woodley, J.M. *Chem. Rev.,* **2018**, *118*, 801-838.
[3]   Tunon, I.; Moliner, V. Simulating Enzyme Reactivity, **2017**, RCS.
[4]   Berman, H.M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T.N.; Weissig, H.; Shindyalov, I.N., Bourne, P.E. *Nucleic Acids Research*, **2000**, *28*, 235-242.
[5]   Sousa, S. F.; Ribeiro, A.; Neves, R.; Bras, N.; Cerqueira, N.; Fernandes, P.; Ramos, M.J. *WIREs Comput. Mol. Sci.,* **2017**, *7*, 1281
[6]   Himo, F. *J. Am. Chem. Soc.*, **2017,** *139*, 6780-6786.
[7]   Marforio, T.D; Giacinto, P.; Bottoni, A.; Calvaresi, M. *Biochemistry*, **2015**, *54*, 4404-4410.
[8]   Campesato, L.; Marforio, T.D.; Giacinto, P.; Calvaresi, M.; Bottoni, A. *ChemPhysChem,* **2018**, *19*, 1514-1521.

2. A promising enzyme for novel chemoenzymatic approaches: Linalool Dehydratase-Isomerase. The catalytic mechanism unveiled by computational investigations

2.1. Introduction

Monoterpenes are highly volatile molecules, formed by two isoprene unities. Vegetal organisms are the principal source of monoterpenes and release more than 100 Ton/year of these molecules in the atmosphere.[1]

Some microorganisms exploit the degradation of monoterpenes for energy reservoir and as carbon source. Aerobic degradation of monoterpenes has been carefully studied for the *Pseudomonas* species,[2,3] where the degradation takes place by hydroxylation of terpenes catalyzed by oxigenases, a step of the metabolic route known as *acyclic terpene utilization (Atu) pathway.* The final products of the *Atu* route are acetyl-CoA and acetoacetate. Anaerobic degradation is an alternative metabolic process identified in denitrificans bacteria, like the gram-negative beta-protobacteria *Castellaniella defragrans* and *Thauera terpenica*. These two microorganisms have been taken as model to study anaerobic metabolic processes of terpenes.[4,5] In particular, *C. defragrans 65phen* exploits β-myrcene, limonene and pinene as carbon atom source.

The complete metabolic pathway of β-myrcene is still unclear, in fact only the first part of this process, which foresees the transformation of β-myrcene to geranic acid is known. Three oxidation stages occur, and three different enzymes catalyse each step. The Linalool dehydratase-isomerase (Ldi EC 4.2.1.127 e EC 5.4.4.4) is the first enzyme encountered and it displays a dual functionality: it accomplishes the hydration of β-myrcene to (*S*)-linalool and the isomerization of (*S*)-linalool to geraniol (**Scheme 2.1**). Then geraniol is oxidized to geranial by the Geraniol dehydrogenase (GeDH), and finally to geranic acid by the Geranial dehydrogenase (GaDH).[6] The fate of geranic acid has not been discovered yet.

Scheme 2.1 Reversible reactions catalyzed by Ldi: hydration of β-myrcene to (*S*)-linalool and isomerization to geraniol.

The synthetic interest for Ldi is due to its capability of catalysing hydration and dehydration reactions to produce enantiomeric pure alcohols from double bonds and obtain alkenes by the inverse reaction. Conventionally, harsh conditions (strong acids or bases, high temperatures) are required to perform dehydration of tertiary alcohols for the preparation of olefins, and the presence of other sensitive functional groups limits the scope of substrates.[7] The reverse reaction, *i.e.* hydration, foresees the nucleophilic attack of a water molecule on a double bond for which acid conditions are usually required and the stereochemical control of the resulting alcohol is challenging.

The reaction conditions required for the enzymatic catalysis carried out by Ldi are milder than conventional catalysis and the biosynthetic exploitation of this enzyme to obtain olefins (or asymmetric alcohols) is undoubtedly attractive.

In 2017 Nestl and co-workers tested the dehydratase and hydratase activity of Ldi on various substrates,[8] reporting excellent conformational selectivity and good conversion percentages. Despite the potential exploitation of Ldi at industrial level, only two crystallographic structures of this enzyme are reported in literature (PDB codes 5G1U[8], 5HSS[9]). In both crystals, **Ldi** exhibits a cyclic homo-pentameric structure, with 5 catalytic sites located at the interface between each monomer (**Fig.2.1**). No cofactors or metal ions were as found within 5G1U or 5HSS catalytic sites.
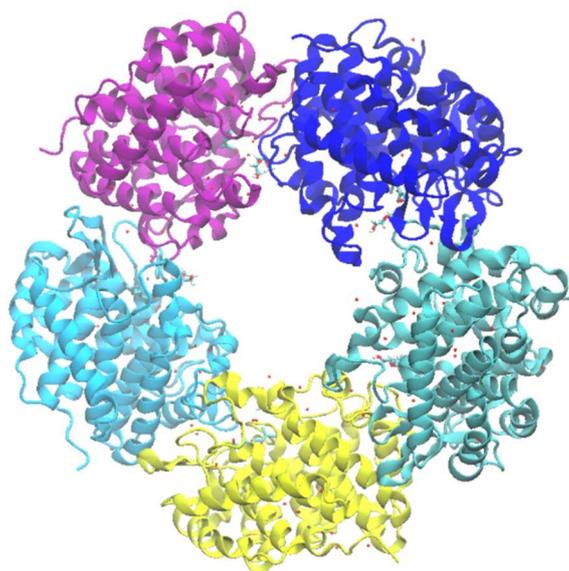
Fig.2.1 Homo-pentameric structure of Ldi from PDB 5HSS.

From mutagenesis experiments[8] two cysteine residues (Cys171, Cys180) are reported to be crucial for the transformation of myrcene to geraniol or for the isomerization of geraniol to linalool. Other residues (His129, Tyr65, Tyr45'and Asp39' - the quoted residues belong to the adjacent monomer)

are also reported to be important for the catalytic activity of Ldi, but their role has not been confirmed yet. Furthermore, studies on enzyme activity demonstrated that Ldi needs a reducing environment and the total absence of O_2 for maintaining the catalytic activity with the two cysteines in the reduced state.[10] The optimal temperature and pH for Ldi's enzymatic activity are 35 °C and pH 9 respectively.

Up to date, the reaction mechanism by which Ldi catalyses the isomerization and hydration of the ligand has never been studied in detail and diverse mechanisms have been suggested.[8,9] In **Scheme 2.2** and **Scheme 2.3**, we report the first two mechanistic hypotheses reported in literature.[8]

The first mechanism hypothesizes a nucleophilic attack by the thiolate group of Cys180 on the terminal double bond of linalool, followed by the allylic rearrangement and the release of the hydroxyl which is being protonated by Cys171. A covalent intermediate is then formed, and isomerization or dehydration steps occur. Following the blue arrows in **Scheme 2.2**, a deprotonated Tyr45 abstracts the methyl proton, engaging a 1,4-elimination that generates the diene moiety while cleaving the covalent bond with Cys180. The green arrows in **Scheme 2.2** describe the isomerization process, by which His129 activates a second water molecule which attacks the electrophile carbon, breaking the C-S bond with the cysteine residue.

The second hypothesis is the formation of a carbocationic intermediate, as schematically depicted in **Scheme 2.3**. Here, Cys171 activates the hydroxyl group of linalool, driving the formation of the carbocation. As for the covalent intermediate, two paths are now available: i) isomerization to geraniol due to the nucleophilic attack of an activated water molecule (green arrows) or ii) dehydration performed by the alcoholate group of Tyr45 (blue arrows).

The third hypothesis[9] suggests an acid-base catalysis, where the protonable aminoacids carry out the catalysis. Under this hypothesis, the dehydration reaction (blue arrows in **Scheme 2.4**) would be acted by Asp39', which abstracts the proton of methyl group of linalool, followed by the β-elimination of the hydroxyl group, protonated by Cys171. The isomerization (green arrows in **Scheme 2.4**) would be activated by the deprotonation of the water molecule by the thiolate group of Cyr180 while the nucleophilic attack promotes the double bond shift and the expulsion of a water molecule.

Even if these mechanisms seem to be plausible, there are many aspects that raise concerns. Up to date, there is no consensus on the effective reaction mechanism and none of the hypotheses reported in literature have been corroborated by theoretical computations.
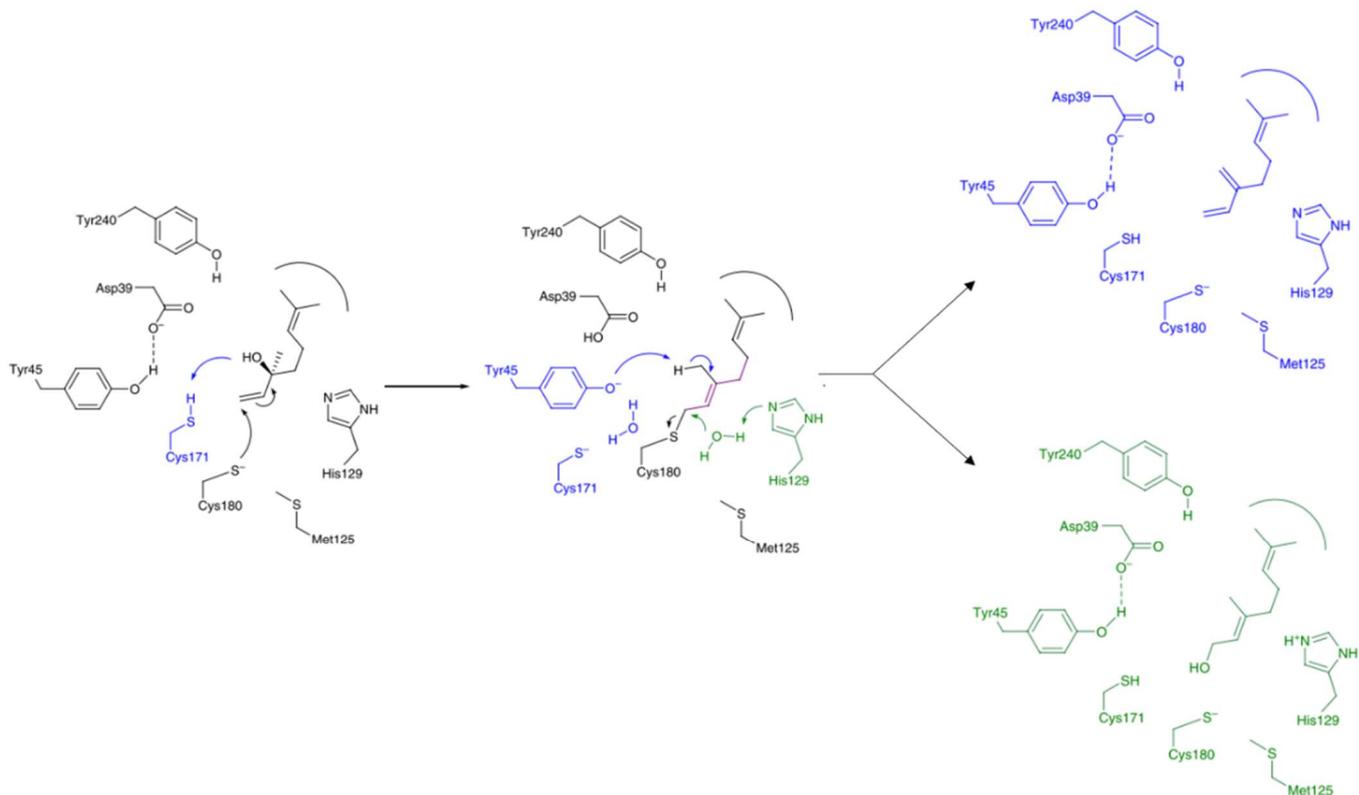
Scheme 2.2 Schematic representation of the hypothetical mechanism of dehydration of linalool to β-myrcene (blue arrows) and isomerization of linalool to geraniol (green arrows), with the formation of a covalent intermediate.
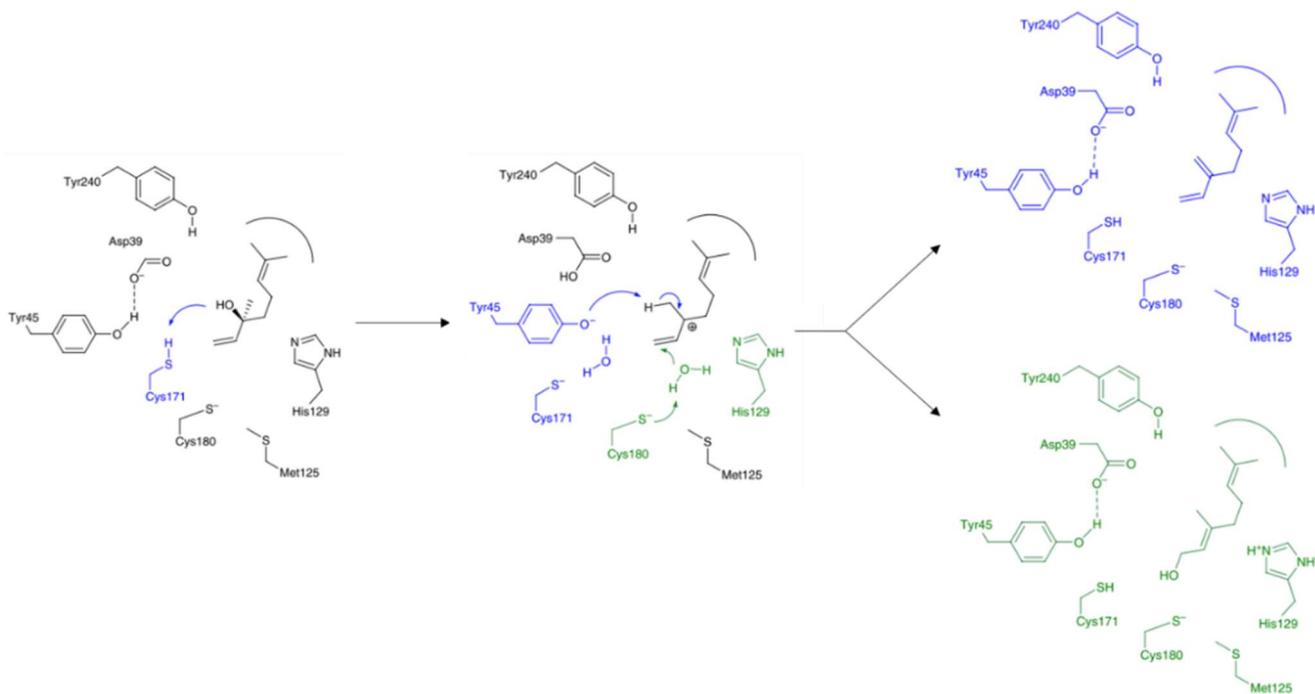
Scheme 2.3 Schematic representation of the hypothetical mechanism of dehydration of linalool to β-myrcene (blue arrows) and isomerization of linalool to geraniol (green arrows), with the formation of a carbocationic intermediate.
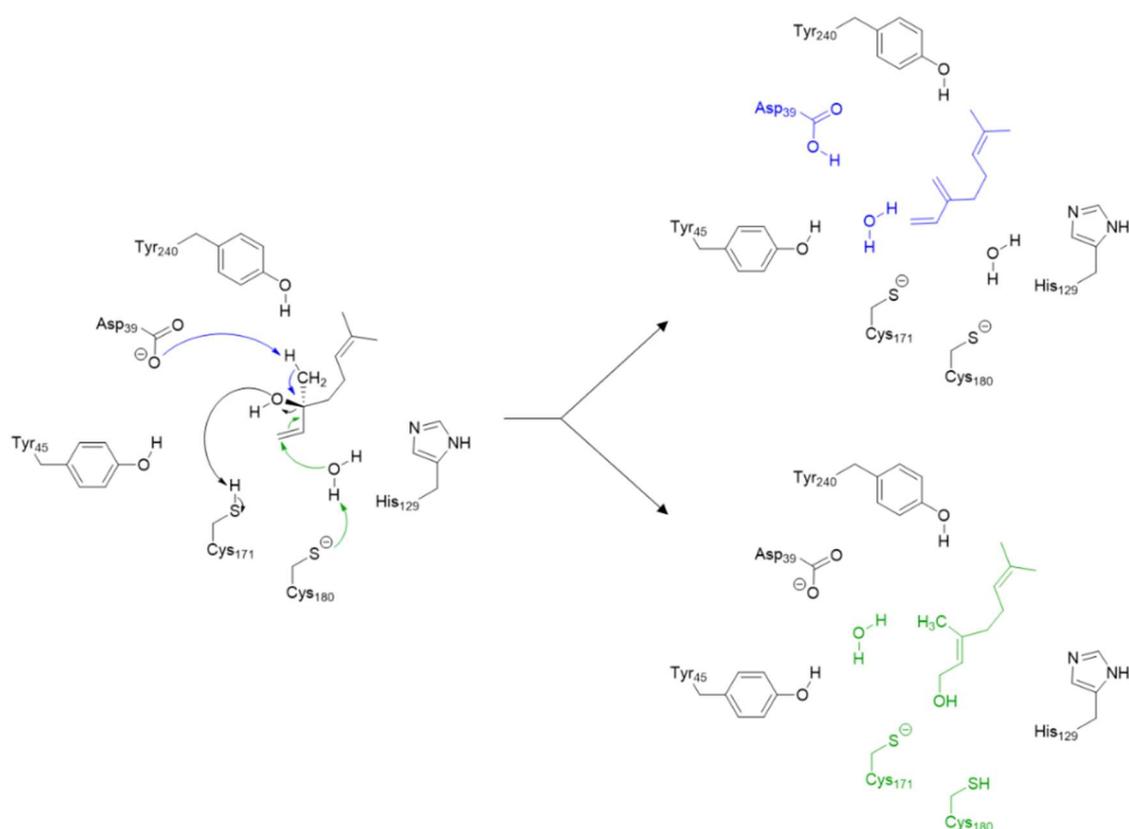
Scheme 2.4 Schematic representation of the hypothetical mechanism of dehydration of linalool to β-myrcene (blue arrows) and isomerization of linalool to geraniol (green arrows), through an acid-base catalysis.

In this work we endeavour to answer to some questions that we believe crucial for the comprehension of the Ldi catalytic activity. In particular: i) which is the reaction mechanism of isomerization and dehydration of geraniol or linalool? ii) which is the real role of the residues that are suggested as crucial for the catalysis? Do they directly participate to the process forming covalent intermediates with the substrates, or do they catalyse the reaction *via* non-covalent interactions? iii) how does the protonation state of the cysteine residues affect the mechanism?

Using detailed and accurate quantum-mechanical computations we seek to give a theoretical response to the above open questions. We believe this mechanistic study of Ldi catalysis will help in the design of efficient approaches to exploit Ldi in chemoenzymatic reactions.

2.2. Computational Details

2.2.1 Choice of the Quantum-Mechanical Model-System. The first two reaction hypothesises **Scheme 2.2** and **2.3**, show two water molecules in the active site, and each water participates in the reaction in a precise way. Visualizing the active site of 5G1U, obtained by the authors of these formulations,[8] there are no water molecules near the ligand molecule, nor in the active site

(**Fig.2.2a**). A notable point is that a disulphide bond is contacted between the two catalytic cysteines, pushing the hydroxyl group of geraniol to contact a plausible hydrogen bond with His129. In **Fig.2.2b** we give a three-dimensional representation of 5HSS active site, on which the authors suggested the acid-base catalysis reported in **Scheme 2.4**. This crystal was obtained after soaking Ldi in a linalool solution,[9] so that the active site can be considered as a picture of the enzyme taken right after it accomplished the dehydration or isomerization of linalool. In fact, the dual activity of the 5HSS-Ldi is also experimental corroborated by the presence of three catalytic sites complexed with geraniol and two with β-myrcene.
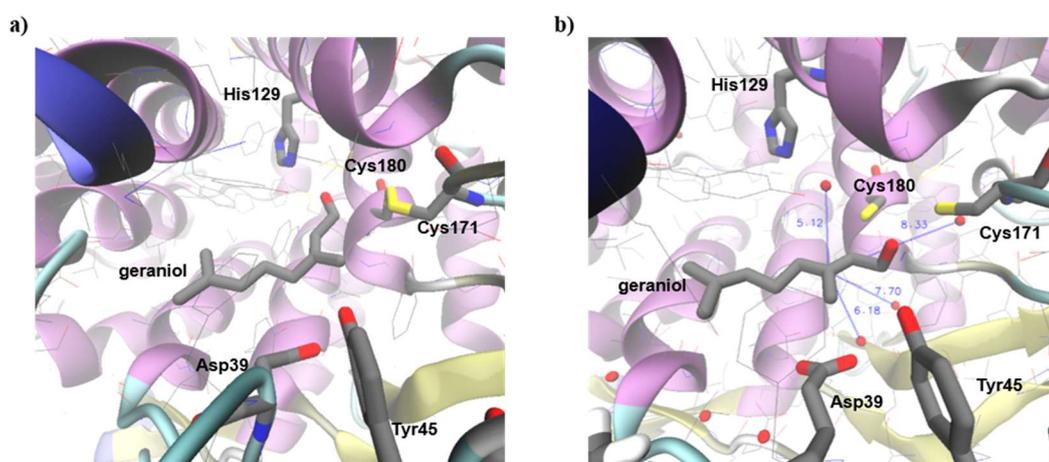
Fig.2.2 Three-dimensional representation of the active site of **a**) 5G1U and **b**) 5HSS.

Comparing the two active sites, the 5HSS active site (**Fig.2.2b**) can be considered the most realistic configuration of the active site, with the hydroxyl group of geraniol pointing towards one of the two cysteines, suggesting the presence of a H-bond interaction, and a water molecule (at 5.12Å) from C3 of geraniol, that may participate in the isomerization/dehydration reactions. The three hypotheses reported above depict a scenario where Cys180 has a thiolate sidechain, rather than a thiol. The pK_a of "unperturbed" cysteine side-chain is 8.5, and at a working pH 9 the residue should be protonated. It is known that the protein environment effects the amino acids pK_a. In particular, hydrogen bond donors and an electropositive local environment tend to lower the pK_a of cysteine sidechain, stabilizing the formation of the thiolate. On the contrary, a hydrophobic and local electronegative surrounding cause the pK_a to raise, enhancing the stabilization of the neutral thiol group.[11]

Since no experimental evidence supports the actual protonation state of the two cysteine residues present in the active site of Ldi, we must take into account the possibility that one of the two cysteines may be deprotonated. We enquired the 5HSS PDB to H++[12] and web PDB2PQR[13] web servers to protonate the 5HSS crystal structure at pH 9: in both cases the two catalytic cysteines were in the

thiol form. Previous to the modelling of the QM model system, we did not carry out any MD sampling since the crystallographic structure of 5HSS can be considered as an optimum starting point for the mechanistic investigation.

We chose to cut out, from 5HSS, the active site complexed with geraniol and we designed the QM model system by comprising all the residues within a distance less than 7 Å from the geraniol. The selected sphere of amino acids was resized by appropriate cuts of chemical bonds (in particular, those involving the major part of α carbons) and the deleted atoms were replaced with hydrogen atoms. Following this procedure, we generated **Model A**, which consists of Tyr65, Hys129, Cys171, Phe177, Cys180, Tyr240, Trp242, Asp39′ and Tyr45′, the geraniol closest water molecule (5.1 Å) and the geraniol ligand, resulting in a system of 163 atoms with a total charge of -1. In order to evaluate the influence of the protein micro-environment, we manually deprotonated Cys171, creating **Model B**, made of 162 atoms with a total charge of -2. Since **Model B** differs from **Model A** just for a proton, only the schematic two-dimensional and three-dimensional representations of the **Model A** are given (**Fig. 2.3a and 2.3b**).
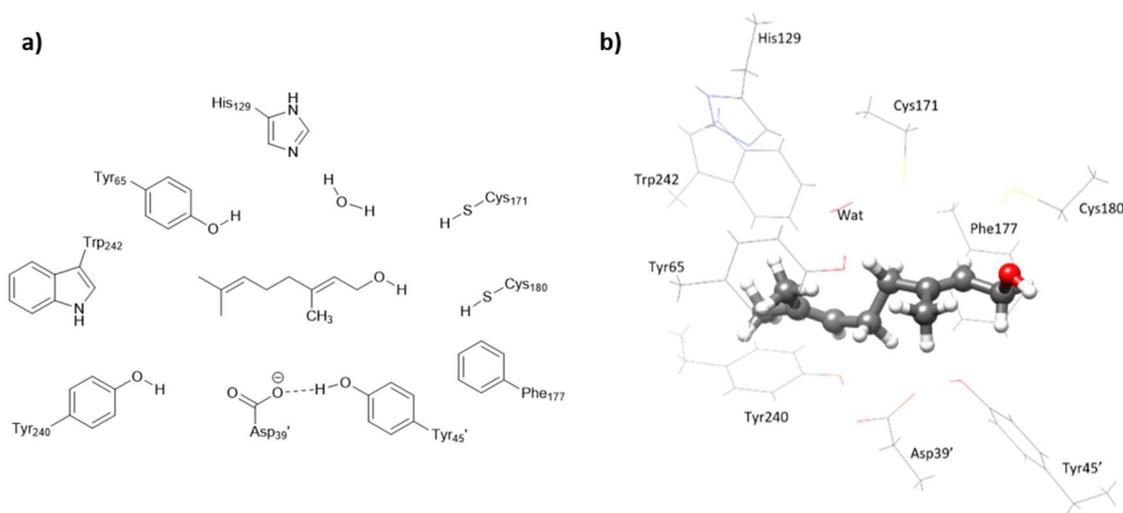
Fig. 2.3 A schematic **a**) two-dimensional and **b**) three-dimensional representation of **Model A**.

2.2.2 Preliminary QM Calculations. All DFT computations were carried out with the Gaussian09[14] series of programs employing the M06-2X functional.[15] The geometry of all critical points was optimized with the gradient method available in Gaussian09 and the nature of all critical points was evaluated by harmonic vibrational frequency calculations. IRC computations[14] starting from transition states in both reactant and product direction were carried out to prove the connection of the various critical points. Two basis sets of different accuracy were used. The atoms directly involved in the reaction were described by the 6-31+G* basis set, [14] while for all the remaining atoms

we adopted the 3-21G* basis set. [14] The double layer basis set (6-31+G*/3-21G* level) corresponded to 1007 basis function for **Model A** and 1003 for **Model B**. This computational approach is denoted here as M06-2X//6-31+G*/3-21G* level. To preserve the geometry of the active site cavity and emulate the constraining effect of the protein environment, during the QM calculations we "froze" the hydrogens replacing the removed atoms along the cut bonds and the atoms at the edges of the model-system at their original crystallographic positions.

2.3. Results and Discussion

2.3.1 Model A. In this section we discuss the preliminary results obtained in the investigation of the isomerization of geraniol to linalool (**Scheme 2.5**), and its dehydration to β-myrcene (**Scheme 2.6a** and **2.6b**).

Our computations suggest that the isomerization occurs in one kinetic step (**Scheme 2.5**). In the reactant complex **Rx(A)**, a network of hydrogen bonds stabilizes the position of water and geraniol molecules. In particular, the hydroxyl group of geraniol contacts a H-bond with Cys180 (O5···H7 2.15 Å), while the water is stabilized by three interactions (N13···H13 2.04 Å, O12···H15 1.80 Å and S10···H11 2.27 Å). The nucleophilic attack of the water molecule in **TS1(A)** takes place in a late stage while the expulsion of the leaving group is already advanced (O5---C3 2.23 Å, O5---H11 1.08 Å). Despite extensive search, we were not able to locate nor the carbocationic or the covalent intermediates, which have been hypothesized previously in literature. Our computational results depict a scenario where the protonated cysteine residues act as electrostatic catalysts in the formation of the leaving group: the hydrogen bonds contacted with Cys180 (O5···H7 1.91 Å) and Cys171 (S10···H11 2.50 Å) accompany the formation of the water molecule. The residue Tyr65 follows the nucleophilic attack of O12 by partially donating H15 to the nucleophile (O12···H15 1.12 Å, O14···H15 1.32 Å). Surmounting **TS1(A)** which lies at 30.6 kcal mol^{-1} above **Rx(A)**, linalool is formed as **Int1(A)**. The isomerization process is complete (O12-C1 1.54 Å) and a water is produced. Strong hydrogen bonds are contacted between the polar residues of the catalytic site and linalool or with the water molecule. From linalool **Int1(A)**, dehydration follows.

We tried to locate the transition state corresponding to the abstraction of the methyl proton by the water present in the active site (**Scheme 2.6a**) or by the activated base of Tyr45' (**Scheme 2.6b**).
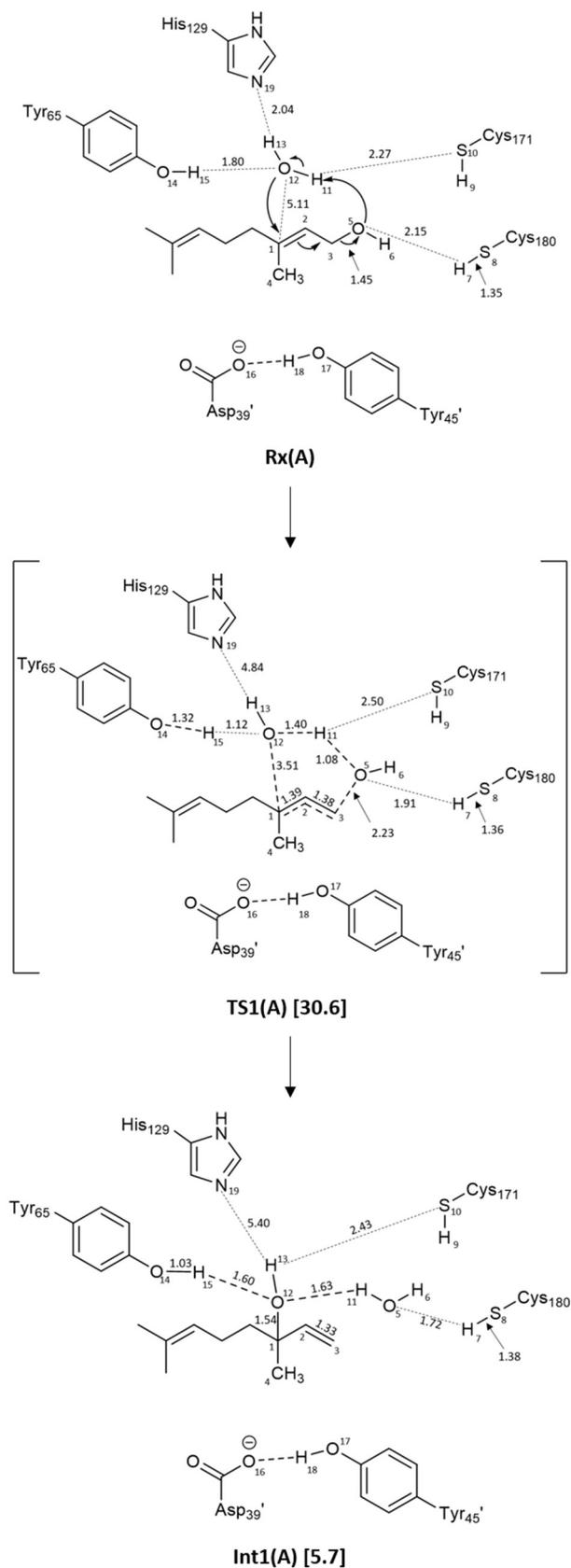
Scheme 2.5 Schematic representation of the **a)** isomerization of geraniol to linalool. Bond distances are in Ångstroms. Energy values (kcal mol^{-1}) are relative to **Rx(A)**.

In the former case (**Scheme 2.6a**), the computations suggest that the water molecule, generated in **TS1(A)**, abstracts the proton of the methyl group of linalool. The β-elimination takes place in **TS2(A)**, characterized by the formation of a 6-member ring which comprises two concomitant proton transfers (O11---H11 1.28, O5---H19 1.27 Å) and the expulsion of water molecule, which takes place in an early stage of the transformation (O12---C1 2.93 Å). For the β-elimination acted by a water molecule, we computed an activation barrier above 40 kcal mol^{-1}, which corresponds to an unfavourable path at the reaction conditions. The mechanism reported in **Scheme 2.6b** corresponds to the β-elimination exerted by Tyr45', which is activated by Asp39'deprotonation. In transition state **TS2(A)'**, the phenate group of tyrosine abstracts the methyl proton of CH$_3$ while the incipient water molecule is expelled (O17---H19 1.15, H19---C4 1.71, C1---O12 1.55 Å). In our preliminary results, this transformation results to be highly unreliable, since the barrier required to surmount **TS2(A)'** is above 40 kcal mol^{-1} as well as **TS2(A)**.
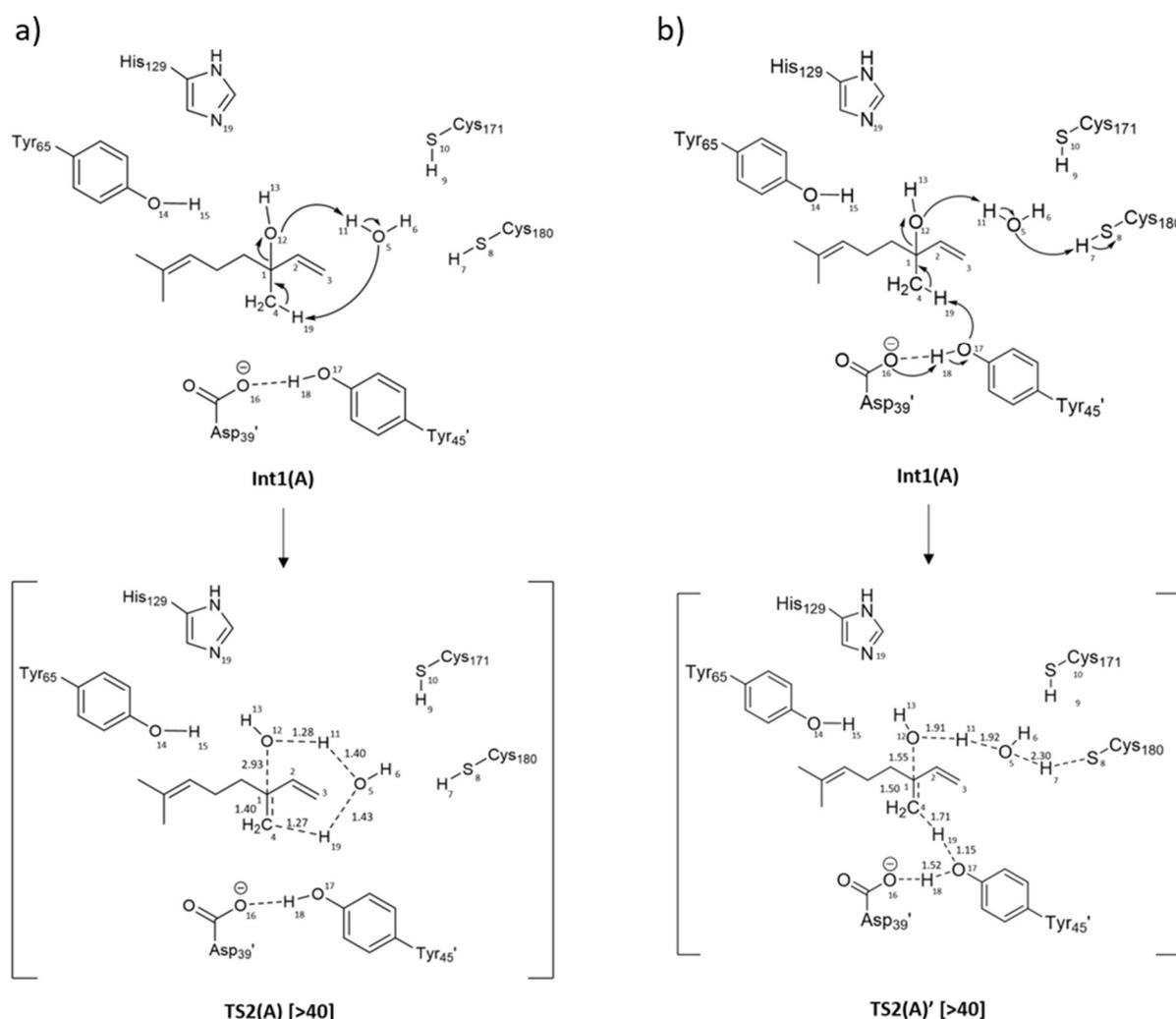
Scheme 2.6 Schematic representation of the linalool dehydration to β-myrcene acted by **a**) the water molecule, **b**) Tyr45'. Bond distances are in Ångstroms. Energy values (kcal mol^{-1}) are relative to **Rx(A)**.

It is evident that the activation barriers for the two possible dehydration paths of linalool to β-myrcene are highly unrealistic. We computed the possibility that the dehydration process takes place on the geraniol ligand, rather than linalool. Our preliminary results suggest that this transformation (**TS3(A)** in **Scheme 2.7**) is more favourable that the dehydration of linalool. In **TS3(A)** we found that Tyr65 acts as the catalytic base in the methyl proton abstraction (O14---H19 1.95 Å) while the water molecule activates the tyrosine (O12---H15 1.32 Å). This transformation has an energy cost of 30.8 kcal mol^{-1}, more than 10 kcal mol^{-1} less than the dehydration of linalool.
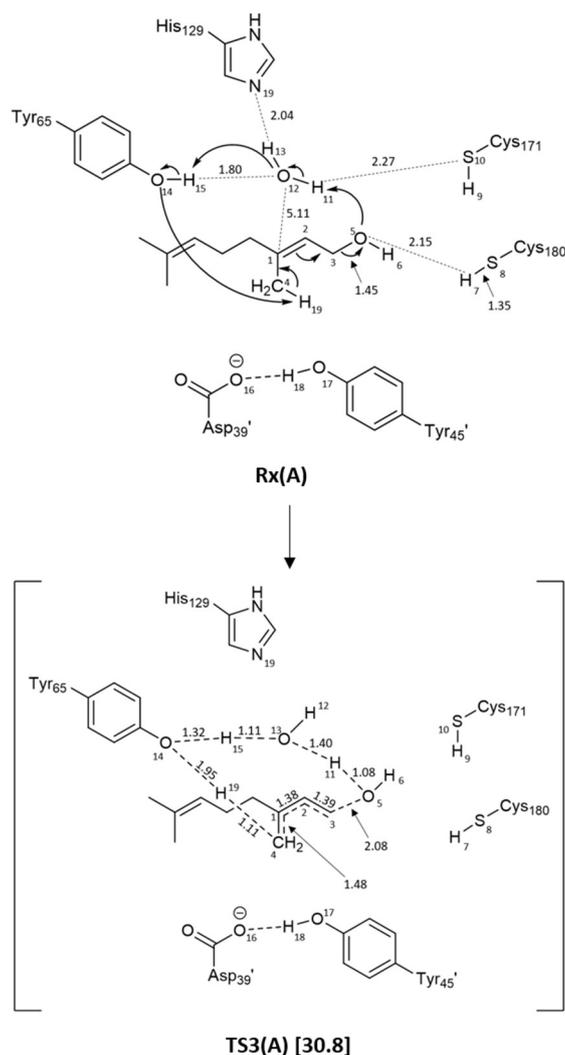
Scheme 2.7 Schematic representation of the geraniol dehydration to β-myrcene. Bond distances are in Ångstroms. Energy values (kcal mol^{-1}) are relative to **Rx(A)**.

Upon these results, it is reasonable to hypothesize that β-myrcene may be obtained by dehydration of geraniol, rather than linalool. Nevertheless, other paths leading to β-myrcene may take place. In fact, we also computed the critical points that corresponds to the formation of *trans*-β-myrcene, which is less stable than the *cis* stereoisomer of ~16 kcal mol^{-1}. Interestingly, the reaction path that

leads to the formation of the *trans* product, passes through the formation of a covalent intermediate with Asp39′, by overcoming an energy barrier of 20.1 kcal mol⁻¹. Once the covalent intermediate is generated (-6.5 kcal mol⁻¹), the 1,4-elimination follows (ΔE=21.1 kcal mol⁻¹) where, while the water abstracts H19, the π electrons shifts from C1C2 to C2C3 breaking the covalent bond, releasing the *trans*-β-myrcene which lies 16 kcal mol⁻¹ above the *cis* stereoisomer. The barriers computed for the formation of the *trans* stereoisomer are significantly lower than those required for the formation of the *cis* product. It is possible that the path towards the *cis*-β-myrcene passes through the *trans* form, so we are carrying out further investigations to evaluate the activation energy barrier for the *trans-cis* isomerization.

2.3.2 Model B. In order to evaluate the effect of the microenvironment of the active site upon the protonable sidechains, we studied the reaction mechanism for the isomerization of geraniol in **Model B**, in which Cys171 is deprotonated.

We located the critical point corresponding to the initial reactant complex **Rx(B)**, where the hydrogen bond network does not differ significantly to **Rx(A)**. Despite the similarities in the H-bond connections, our computational results suggest that a hydrogen bond network and an electron-rich environment may affect the nature of the reaction path (**Scheme 2.8**).

Interestingly we were able to locate the carbocationic intermediate **Int1(B)**. The transition state that generates **Int1(B)** is represented in **TS1(B)**, where the protonation of the hydroxyl group of geraniol is carried out by Cys180 (O5---H7 1.05 Å, S8---H7 1.94 Å) rather than Cy171 as suggested by Nestl.[16] The water molecule present in the model contacts a strong hydrogen bond with the incipient leaving group (H11··O5 1.84 Å), assisting the expulsion of the water molecule. Then the intermediate **Int1(B)** formation follows and a stable allylic carbocation is formed. The carbocationic intermediate lies at 30.4 kcal mol⁻¹ above the reactant complex, and only 1.5 kcal mol⁻¹ under **TS1(B)**, suggesting that this region of the PES is rather flat.

From **Int1(B)** the reaction may follow two different paths, as suggested in **Scheme 2.5**: i) the formation of linalool, by the nucleophilic attack of the water molecule on the positive carbon C1 or ii) the H-elimination to β-myrcene by the abstraction of the methyl proton of C4. Additional computations will be carried out to elucidate the geometries and energetics of these two transformations.
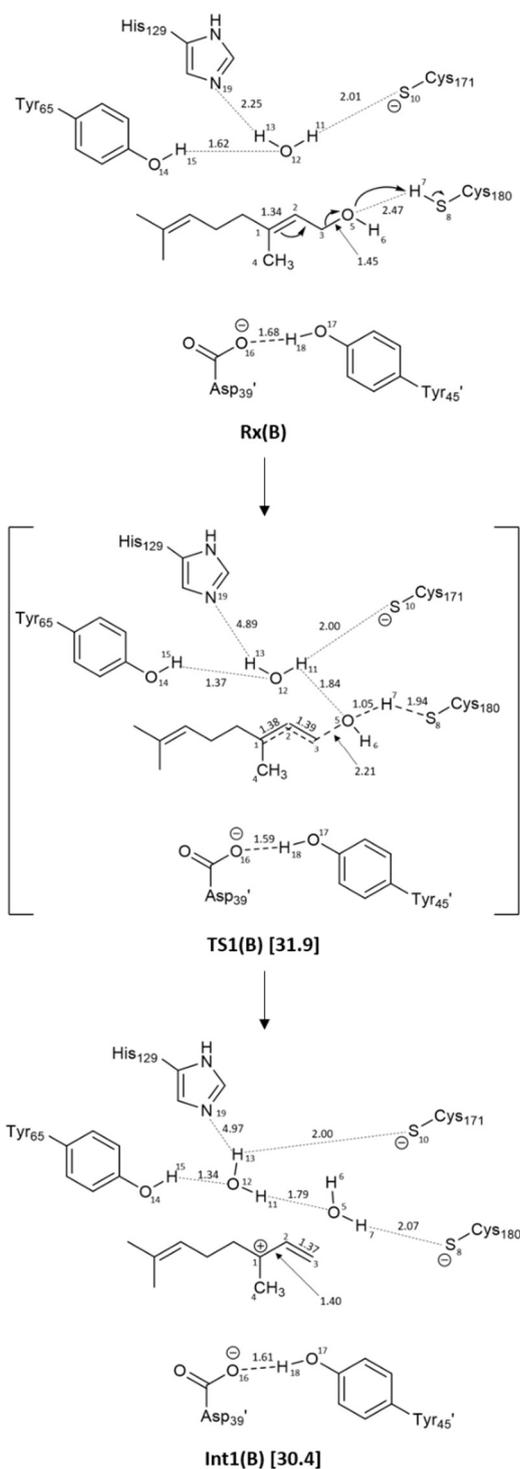
Scheme 2.8 Schematic representation of the carbocationic intermediate formation for Model B. Bond distances are in Ångstroms. Energy values (kcal mol^{-1}) are relative to **Rx(B)**.

2.4. Conclusions

In this work we investigated the catalytic mechanism of Linalool dehydrase-isomerase from *Castellaniella defragrans* using a full QM approach. The preliminary results of our computations can be summarized as follows:

i) In **Model A**, where both the cysteines are protonated, the isomerization of geraniol to linalool follows a concerted path. Two possible concerted paths were computed for the dehydration of linalool, but the activation barriers were too high. An alternative may be that β-myrcene is generated by dehydration of geraniol or isomerization of the *trans*-β-myrcene. The latter transformation follows a step-wise mechanism, generating a covalent intermediate with Asp39′. The energetics for the dehydration of linalool to *cis*-β-myrcene are quite implausible for an enzymatic reaction, and it is possible that the most favourable path to obtain *cis*-β-myrcene is that Ldi firstly generates the *trans*-β-myrcene and then isomerizes it to the *cis* stereoisomer.

ii) In **Model B**, where one of the two cysteines is deprotonated, we computed the formation of a carbocationic intermediate, from which linalool or β-myrcene can be obtained.

iii) In **Model A**, Cys171 and Cys180 act as hydrogen-bond acceptor and donor respectively, suggesting that their catalytic role is not to form covalent bonds with the substrate but rather to stabilize the transition state by non-covalent interactions. From our computations, His129 is confirmed to have an electrostatic role by interacting with the water molecule that is present in the active site. The hydroxyl group of Tyr65 seems to play a crucial role in the isomerization of linalool to geraniol, by assisting the expulsion of the leaving group by partially donating its proton. The residues Asp39′ and Tyr45′ have a critical role in the formation of *trans*-β-myrcene, where Asp39′ acts as a nucleophile, creating a covalent bond with the substrate. In **Model B**, Cys180 acts as an acid catalyst, protonating the leaving group during the carbocationic intermediate formation. The other cysteine, Tyr65 and His129 stabilize the transformation by engaging hydrogen bonds with the second water molecule.

In both models, at least one water molecule is necessary to aid the catalytic steps of isomerization and dehydration, assisting the proton transfers and/or directly participating in the reaction as a nucleophile.

iv) The protonation state of the cysteine residues present in the active site is decisive for the nature of the reaction mechanism that bring the substrate to undergo the isomerization and dehydration transformations. The carbocationic intermediate was located only when one of the two cysteines were deprotonated. This can suggest that a negative charged environment allows the formation of a positively charged intermediate, that can be stabilized by electrostatic interactions found in the surroundings.

2.5. Perspectives

At this stage of our investigation, conclusive and definitive considerations cannot be made. In fact, we are aware that further investigations are necessary, and we are carrying on our computations to elucidate some critical aspects, in particular we would like to: i) validate the hypothesis that Ldi firstly produces the *trans*-β-myrcene and then, through an isomerization process, generates the *cis*-product; ii) complete the computations to understand the evolution process of the carbocation intermediate in **Model B**; ii) consider the effect of the protein environment to evaluate if other short- or long-range interactions have a role in the catalytic processes since the activation barriers found are slightly high, and are not consistent with an enzymatic reaction (even if Ldi works at 35-40 °C); iii) take into account the possibility that more than one water molecule participate to the process, enhancing the catalytic activity of Ldi.

References

[1]   Zimmerman, P. R.; Chatfield, R. B.; Fishman, J.; Crutzen, P. J.; Hanst, P.L. *Geophys. Res. Lett.* **1978**, *5*, 679–682.
[2]   Förster-Fromme, K.; Jendrossek, D. *FEMS Microbiol.*, **2008**, *286*, 78–84.
[3]   Trudgill, P. W. Academic Press, New York, **1986**, 483–528.
[4]   Harder, J.; Probian, C. *Appl. Environ. Microbiol.* **1995**, *61*, 3804–3808.
[5]   Foss, S.; Heyden, U.; Harder, J. *System. Appl. Microbiol.* **1998**, *21*, 237–244.
[6]   Lüddeke, F. *et al. Appl. Environ. Microbiol.* **2012**, *78*, 2128–2136.
[7]   Hoffman, R. V.; Bishop, R. D.; Fitch, P. M.; Hardenstein, R. *J. Org. Chem.* **1980**, *45*, 917.
[8]   Nestl, B.M.; Geinitz, C.; Popa, S.; Rizek, S.; Haselbeck, R.J.; Stephen, R.; Noble, M.A. *et al. Nat Chem Biol.* **2017**, *13(3)*, 275-281.
[9]   Weidenweber, S.; Marmulla, R.; Ermler, U.; Harder, S. *FEBS Lett.*, **2016**, *590*, 1375–1383.
[10] Brodkorb, D.; Gottshall, M.; Marmulla, R.; Lüddeke, F.; Harder, J. *J. Biol. Chem.* **2010**, *40*, 30436-30442.
[11] Poole, L.B. *Free Rad. Biol. Med.* **2015**, *80*, 148-157.
[12] Anandakrishnan, R.; Aguilar B.; Onufriev, A.V. *Nucleic Acids Res.*, **2012**, *40*, W537-W541.
[13] Dolinsky, T.J.; Nielsen, J.E.; McCammon, J.A.; Baker, N.A. *Nucleic Acids Research*, **2004**, *32*, W665-W667.
[14] Gaussian 09, Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; et al Gaussian, Inc., Wallingford CT, **2009**.
[15] Zhao, Y. ; Truhlar, D.G. *Theor. Chem. Acc.* **2008**, *120*, 215-241.

3. A Full QM Computational Study of the Catalytic Mechanism of α-1,4-Glucan Lyases[a]

3.1. Introduction

Carbohydrates are amongst the most abundant and important bio-molecules on earth. They are commonly found as oligosaccharides and polysaccharides (*i.e.* carbohydrates polymers) that play numerous and crucial roles in living organisms: structural components and store of vital energy are the most common functions of these molecular species. The action of carbohydrates implies, in general, the rupture of the stable glycosidic bond. For instance, α-1,4-glycosidic bonds bind together D-glucose units to form α-1,4-glucans. Glycogen and starch are examples of α-glucans that provide plants, animals and fungi with energy which is obtained by their degradation (*via* the break of the 1,4-glycosidic bonds) releasing carbohydrate monomers in the cell.

Nature has developed various enzymatic systems capable of cleaving the glycosidic bond.[1] The most common are glycoside hydrolases (GHs), which undergo the hydrolytic degradation of oligo- and polysaccharides. GHs form a large class of enzymes that catalyse crucial biological reactions in almost all realms of life and are classified in more than 130 families on the basis of the conserved aminoacidic primary structure. However, the cleavage of the glycosidic bond can be carried out also by non-hydrolytic enzymes known as lyases *via* elimination reactions.[2]

For instance, polysaccharide lyases (PLs) cleave the O-C4' glycosidic bond through a β-elimination involving the abstraction of a proton in position C5' followed by the formation of a C4'-C5' double bond. α-1,4-glucan lyases (GLases)[3] form another group of lyases that catalyse the rupture of the glycosidic bond along with the formation of keto 1,5-anhydro-D-fructose (keto-AF).

The production of keto-AF by Glases is the first fundamental step in the anhydrofructose pathway[4] that is a secondary route of glucan metabolism. AF and its metabolites found within this metabolic pathway, are known to occur in microorganisms[5], red algae[6,7] and rat liver tissues[8]; furthermore, these molecules show antimicrobial[9-11] and antitumoral[12-14] properties, together with antioxidant[14] and anti-inflammatory[15] activities. A 1,4-glucan lyase was firstly described in 1993 by Yu *et al.*[6]

Although being classified as a PL, GLases diverge from the other members of this family: while PLs demand the presence of a uronic moiety in the ligand and the mechanism relies on the enhanced acidity of the hydrogen in position 5' resulting from the adjacent C6' carboxylic group, GLases ligands are α-1,4-glucans with no carboxylic functions. This determines for GLases a completely

different regioselectivity of the β-elimination. These enzymes cleave the C1-O glycosidic bond and generates a double bond between C1 and C2 by abstraction of the C2 proton, thus leading to the enol-AF that undergoes tautomerization to keto-AF.[2]

Beyond these inconsistencies between PLs and GLases, it was demonstrated that GLases share a 23-28% of primary sequence identity with α-glycosidases[16] of the glycoside hydrolase family 31 (GH31).[17] Because of these similarities, α-1,4-glucan lyases were grouped as special members of GH31 where they form the subgroup 2.[16] The similar aminoacidic sequence suggested a similar catalytic mechanisms of α-1,4-glucan lyases and α-glucosidases. This hypothesis was enforced by mechanistic studies showing that inhibitors of α-glucosidases also inhibit Glases.[16,18,19]

Other experimental studies (based on kinetic isotope effects) suggested for GLases a two-step mechanism similar to that of retaining α-glucosidases.[18-20] The commonly accepted mechanism of retaining glycosidases is a double displacement mechanism involving a glycosylation and a deglycosylation step.[21] As depicted in **Scheme 3.1**, in the first step (glycosylation) a carboxylate residue acts as a nucleophile at the anomeric carbon forming a glycosyl-enzyme covalent intermediate. The leaving group –OR (cleavage of the glycosidic bond) is activated by another catalytic residue that behaves as a general acid catalyst. In the second step (deglycosylation; path a in **Scheme 3.1**) this residue acts as a general base by removing a proton from a water molecule that attacks the anomeric carbon, thus cleaving the covalent glycosyl-enzyme bond. The overall effect is a retention of stereochemistry at the anomeric centre of the monosaccharide unit and the reactivation of the enzyme in its pre-catalytic state. A similar two-step mechanism was proposed for GLases in 2003 by Lee *et al.*[19] Within this mechanistic hypothesis two residues behave as a basic and acid catalysts. An intermediate with an oxocarbenium ion character that enhances the acidity of the C2 proton, was suggested to be involved in the process.[19] The cleavage of the covalent glycosyl-enzyme bond proceeds through the abstraction of this proton by a basic residue rather than by a water molecule (path b in **Scheme 3.1**). However, even if a detailed discussion of the mechanism was given, the identity of the basic residue was not clear at that stage.[19,20]

A good candidate proposed for this role was the Asp[553] residue, which seemed to be appropriately placed with respect to the β-face of the carbohydrate unit. This residue ambiguity was solved in 2013, by Rozeboom *et at.*[22] who published the crystallographic structure of α-1,4-glucan lyase from the red seaweed *Gracilariopsis lemaneiformis*.
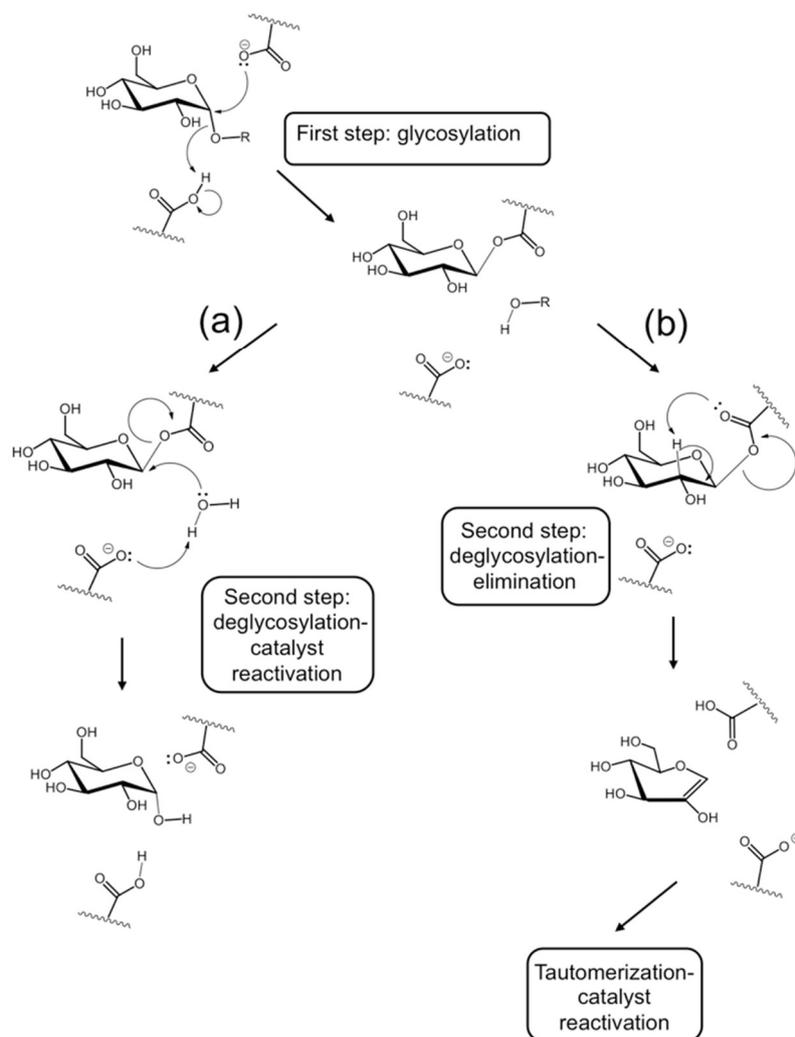
Scheme 3.1 The two reaction pathways that differentiate (a) retaining α -glucosidases and (b) α-1,4-glucan lyases.

They identified the catalytic dyad with the two residues Asp[553] and Asp[665]. They clearly demonstrated that Asp[553] is correctly positioned to abstract the sugar C2 proton and that this task cannot be performed by Asp[665] which is on the opposite side of the ring. Also, a separate base or a water molecule able to abstract a proton were not observed in the active site.

Recently Su *et al.*[23] carried out a computational study on the reaction mechanism of α-1,4-glucan lyases up to the formation of enol-AF. They used a hybrid QM/MM approach and the X-ray crystal structure obtained by Rozeboom[22] to build their model-system. They explored the glycosylation and deglycosylation steps and confirmed the importance of the two catalytic residues Asp[553] and Asp[665] in agreement with the hypothesis of Rozeboom.[22] However, they did not examine how the initial state of the catalyst can be restored and the final tautomerization process where the enol-AF species is transformed into keto-AF. These authors found that glycosylation occurs in two kinetic

steps. The first step corresponds to the protonation of the glycosidic oxygen and the cleavage of the glycosidic bond (barrier of 15.8 kcal mol^{-1}). In the second step the nucleophilic attack of the carbonyl oxygen of Asp[553] leads to the formation of the glycosyl-enzyme intermediate. These intermediate forms very easily since the corresponding barrier is almost negligible (0.5 kcal mol^{-1}). Furthermore, they demonstrated that deglycosylation-elimination is the bottleneck of the entire enzymatic process. The deglycosylation-elimination occurs in three kinetic steps and the abstraction of the C2 proton is rate limiting. However, the corresponding transition state is rather high, being 28.9 kcal mol^{-1} above the starting complex. These authors pointed out the important dual role of the Asp[553] residue: it behaves as a nucleophile and as a base in the glycosylation and elimination step, respectively.

In the present paper, we re-examine some aspects of this enzymatic reaction that, in our opinion, are still unclear. In particular, we try to answer the following questions: (i) Does a large QM region significantly modify the mechanistic picture or the reaction energetics? In short, what is the effect of the inclusion, with respect to previous computations,[23] of additional residues that could be responsible of important stabilizing effects? The importance of the size of the QM region in determining the reaction profile has been a subject of significant interest during the last decade. In particular, it has been noted that for enzymatic reactions the size of the QM region can significantly affect the computed energy profile.[24,25] (ii) Does a reaction pathway exist characterized by a low energy barrier more consistent with an enzymatic reaction? (iii) What is the effect on the reaction energetics of nowadays available functionals explicitly designed to take into account dispersion forces? (iv) What is the mechanism that restores the initial state of the enzyme? (v) Finally, how the final enol-keto tautomerization occurs? How its energetics affects the entire enzymatic process? In particular, points (iv) and (v) are examined here for the first time.

To answer these questions in the present paper we carry out a full QM investigation of the mechanism of α-1,4-glucan lyases including the final tautomerization process. This type of approach was already demonstrated to provide reliable results in the study of enzymatic mechanisms or biological/pharmaceutical processes.[26-35] To this purpose, we use the M06-2X functional and a QM model-system built from the X-ray crystal structure obtained by Rozeboom.[22] This QM model-system (218 atoms) is rather large and includes all the aminoacidic residues that in principle can be responsible for important interactions.

3.2. Computational Details

3.2.1 Choice of the Quantum-Mechanical Model-System. To design a reliable model for the enzyme-substrate complex, we superimposed the X-ray crystal structures deposited by Roozebom PDB = 2X2I (α-1,4-glucan lyase and acarbose), PDB = 2X2J (α-1,4-glucan lyase and deoxynojirimicin), PDB = 4AMW (α-1,4-glucan lyase covalent intermediate complex with 5-fluoro-idosyl-fluoride) and PDB = 4AMX (α-1,4-glucan lyase covalent intermediate complex with 5-fluoro-glucosyl-fluoride) in order to verify possible conformational modifications on the active site residues (see **Fig.3.1**).
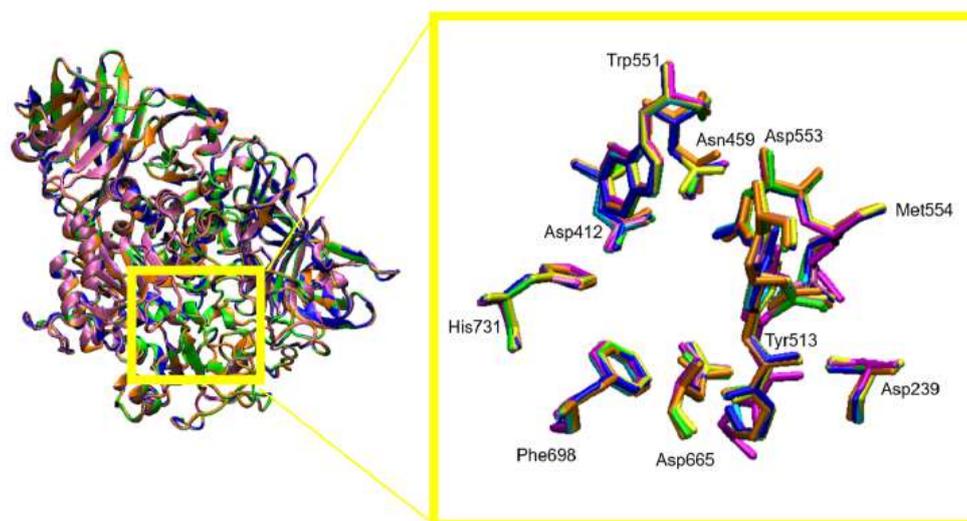
Fig.3.1 Structural superposition of: 2X2H-Chain A apo structure (blue); 2X2J-Chain A with substrated analogue acarbose (orange); 2X2J-Chain A with substrate analogue deoxymojirimycin (green); 4AMW-Chain A with intermediate analogue 5-fluoro-idosyl-fluoride (pink); 4AMW-Chain D with product analogue (2-oxo-1,2-dideoxy-5F-D-idopyranose (cyan); 4AMX-Chain A with intermediate analogue 5-fluoro-beta-D-glucopyranose (yellow); 4AMX-Chain D with product analogue 2-oxo-1,2-dideoxy-5F-D-glucopyranose (magenta).

Since we did not find significant conformational discrepancies in the comparison of the four crystals, we did not carry out any MD sampling[37] and we constructed the QM system-model upon the PDB = 2X2I structure. In PDB=2X2I, the enzyme is crystallized with acarbose, a potent GLase inhibitor. We kept the acarbose skeleton in its crystallized position and modified its atoms in order to obtain maltose, one among the shortest α-glucans. The nitrogen atom connecting the two monosaccharide units was replaced by an oxygen atom to obtain the glycosidic bond. Then we removed the third unit of acarbose and we replaced this unit by a hydrogen atom. The hydrogen atoms were then added to the enzyme structure at pH 7.4 using the H++ online program. [38] In addition to maltose and the two residues Asp[553] and Asp[665] (postulated as the catalytic dyad), in the construction of our

model-system we considered all residues surrounding the substrate within a distance less than 8 Å from the glycosidic oxygen. In particular we included residues involved in important hydrogen contacts with maltose and the catalytic dyad, *i.e.*: (a) Asn[459] and Arg[649] as hydrogen bond donors to Asp[553]; (b) Asp[412], Asn[459], Trp[551], Arg[649], Phe[698] and Hys[731], which are responsible for strong interactions with the non-reducing end of maltose; (c) Asp[239], Asp[512], Tyr[513], Gly[514] and Met[554] interacting with the maltose reducing-end. Then, taking into account mechanistic experimental evidences,[18,19] a hydrogen was manually added to one of the oxygen of Asp[665] carboxylic group. The residues of the model-system were resized by appropriate cuts of chemical bonds (in particular, those involving the major part of α carbons) and the deleted atoms were replaced with hydrogen atoms. The resulting enzyme-substrate model counted 218 atoms, with a total charge of -1. Schematic two-dimensional and three-dimensional representations of the model-system are given in **Fig.3.2a** and **2b**. A point concerning **Fig.3.2a** should be stressed: since we adopted a two-dimensional representation, several atomic distances are not realistic and appear much longer (or shorter) than in the real protein. **Fig.3.2b** provides a much more realistic picture (three-dimensional) of the model-system.
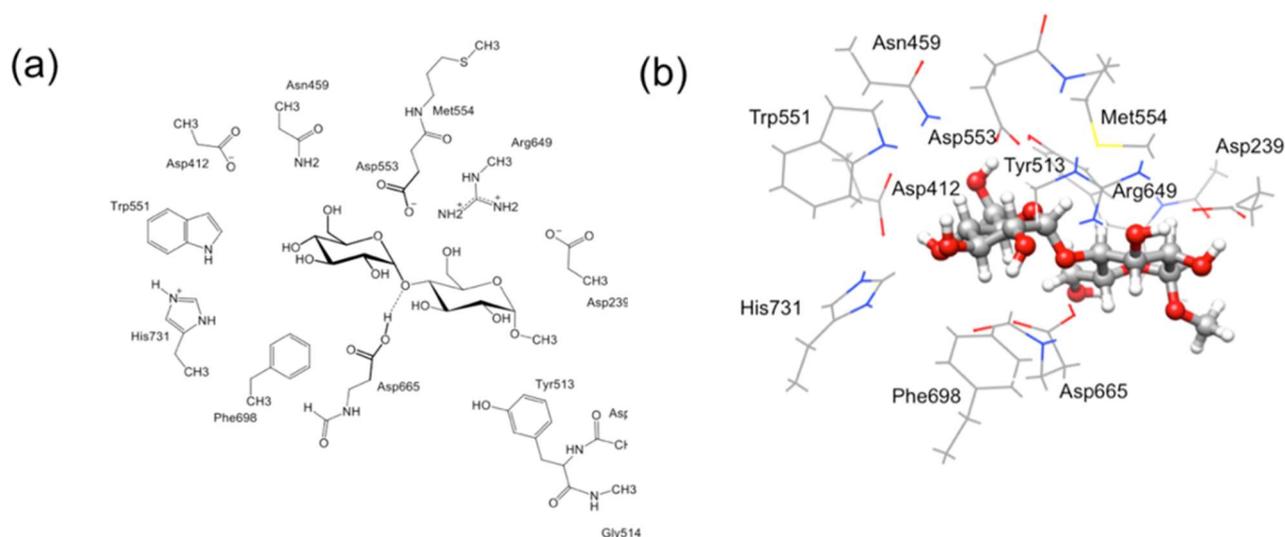
Fig.3.2. A schematic **a**) two-dimensional and **b**) and three-dimensional representation of the model-system.

3.2.2 QM Calculations. All DFT computations were carried out with the Gaussian09[39] series of programs employing the M06-2X functional.[40] The geometry of all critical points was optimized with the gradient method available in Gaussian09 and the nature of all critical points was evaluated by harmonic vibrational frequency calculations. IRC computations starting from transition states in both reactant and product direction were carried out to prove the connection of the various critical

points. Two basis sets of different accuracy were used. In our experience this approach can provide an accurate description of enzymatic reactions at a reasonable computational cost.[41-43] The atoms directly involved in the reaction (*i.e.* the non-reducing end of maltose, Asp^{553} and Asp^{665}) were described by the 6-31+G* basis set, while for all the remaining atoms we adopted the 3-21G* basis set. The double layer basis set (6-31+G*/3-21G* level) corresponded to 1351 basis function. This computational level is denoted here as M06-2X//6-31+G*/3-21G* level. To preserve the geometry of the active site cavity and emulate the constraining effect of the protein environment, during the QM calculations we "froze" the hydrogens replacing the removed atoms along the cut bonds and the atoms at the edges of the model-system at their original crystallographic positions (the "frozen" atoms are evidenced in **Fig.3.3**).
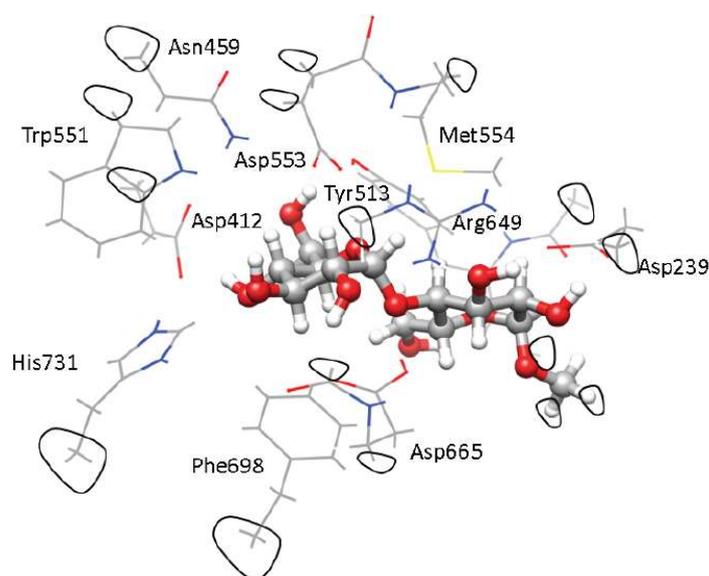
Fig.3.3 A schematic three-dimensional representation of the model-system. Atoms which are kept frozen during optimization are enclosed by contour lines.

To clarify the effect of the functional on the reaction profile we carried out single-point computations with the B3LYP functional and corrections for dispersion forces (B3LYP+D level). Also, we used this functional to re-optimize the most important critical points *i.e.* the transition states and intermediates for glycosylation and deglycosylation/elimination. To obtain a more accurate reaction energetics we carried out single-point computations on the optimized geometries by replacing the 6-31+G* and 3-21G* basis with the 6-311++G** and 6-31+G*, basis, respectively (M06-2X//6-311++G**/6-31+G* computational level). We adopted the same strategy in the benchmark computations with the B3LYP+D method (B3LYP+D//6-311++G**/6-31+G* computational level). Furthermore, in all computations the effect of the whole protein environment was taken into account using the

polarizable continuum model (PCM)[44] approach (single-point computations on the gas-phase geometries) as implemented by Gaussian09. We used a dielectric constant ε of 4.24 (diethyl ether). As suggested elsewhere[45-47] this value is suitable to account for the presence of hydrophilic and hydrophobic groups around the enzyme active site. To elucidate the effect of solvent on geometry, we re-optimized the structures of the most important critical points (in particular, the rate-determining step) in the presence of solvent at the M06-2X//6-31+G*/3-21G* level, then we carried out again single point computations with the more extended basis 6-311++G**/6-31+G*.

3.3. Results and Discussion

In this section we discuss the results obtained in the investigation of the singlet potential energy surface for the whole enzymatic process reported in **Scheme 3.1** and consisting of three fundamental phases: glycosylation, deglycosylation/elimination and keto-enol tautomerization. The computed energy profile is displayed in **Fig.3.4**.
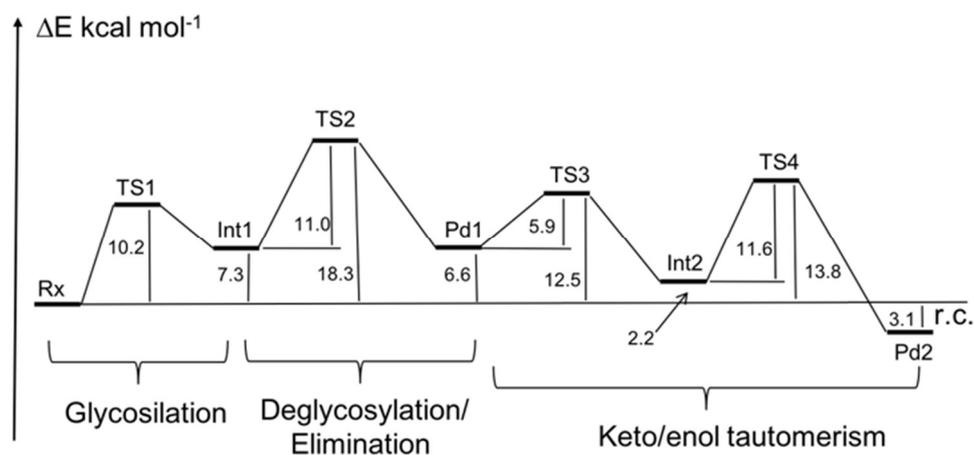
Fig.3.4 Energy profile for the enzymatic process of α-1,4-glucan lyases computed at the M06-2X/6-311++G**/6-31G* computational level (energy values include ZPE corrections).

3.3.1 Glycosylation. Our computations showed that glycosylation occurs in one kinetic step characterized by a highly asynchronous S_N1-like transition state **TS1** (see **Scheme 3.2**). An activation energy of 10.2 kcal mol^{-1} must be overcome to form the glycosyl-enzyme intermediate **Int1**, which is 7.3 kcal mol^{-1} higher than reactants **Rx** (the Michaelis complex).
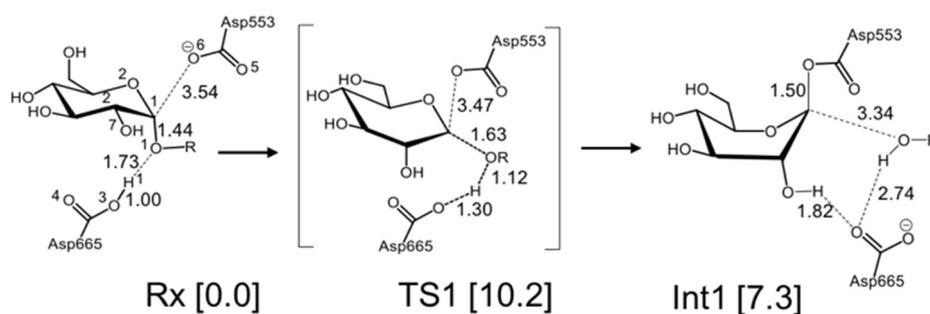
Scheme 3.2 Schematic representation of the glycosylation step (bond lengths are in Å). Energy values (kcal mol⁻¹) are relative to reactants.

In reactants the Asp[665] hydrogen H1 forms a rather strong hydrogen bond with the glycosidic oxygen O1, the O1···H1 distance being 1.73 Å. This contact anticipates the proton transfer occurring in the subsequent transition state **TS1**. The Asp[553] residue is positioned on the opposite side of the ring (the β face of glucose) and one of its carboxylate oxygens (O6) points in the direction of the anomeric carbon C1 (O6···C1 distance = 3.54 Å). In **TS1** the proton H1 is moving from Asp[665] to the glycosidic oxygen O1 and simultaneously Asp[553] carries out a nucleophilic attack at the anomeric carbon C1. However, while the proton is approximately half-way between O1 and O3 (O3···H1 and H1···O1 distances are 1.30 and 1.12 Å, respectively) with a consequent weakening of the glycosidic bond (C1···O1 distance = 1.63 Å), the nucleophile is still rather far from the anomeric carbon, the O6···C1 distance being 3.47 Å (only slightly shorter with respect to reactants). In spite of extensive search, we were not able to locate in this case an oxocarbenium intermediate. In this region the surface is rather flat and **TS1** leads directly to the glycosyl intermediate **Int1**. Actually, these results are similar to those found by Su and coworkers:[46] the oxocarbenium ion intermediate located by these authors (only 0.47 kcal mol⁻¹ lower than **TS1**) is probably only a computational shortcoming with no real physical meaning, whose existence strongly depends on the level of theory and accuracy. Its arguable nature simply indicates that the formation of the glycosyl-enzyme intermediate is a very asynchronous concerted process (S_N1-like transition state) consisting of two phases: in the first phase a proton moves from Asp[665] to the glycosyl oxygen causing a significant weakening of the glycosidic bond; in the second phase the nucleophile Asp[553] attacks the anomeric carbon without overcoming any barrier. In the glycosyl intermediate **Int1** the O6-C1 bond is completed (1.50 Å) and the leaving group is far away from the anomeric carbon C1, the C1···O1 distance being 3.34 Å. Asp[665] establishes a contact with the O7-bonded hydrogen (O7H···O4 distance = 1.82 Å) and maintains a weak hydrogen interaction with the leaving sugar (O4···H(sugar) distance = 2.74 Å).

3.3.2 Deglycosylation/Elimination Step. An intrinsic barrier of 11.0 kcal mol⁻¹ must be overcome to break the glycosyl-enzyme bond. In the corresponding transition state **TS2** (see **Scheme 3.3**) the C1··O6 distance is 2.36 Å. The cleavage of this bond causes a planarization of the anomeric carbon (the dihedral angle C2-C1-O2-C5 is 1.9°) and a shortening of the endocyclic C1-O2 bond (1.27 Å), which takes a significant double bond character.
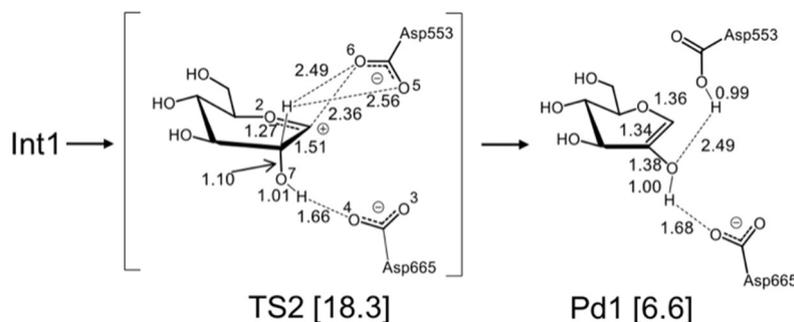
Scheme 3.3 Schematic representation of the deglycosylation/elimination step (bond lengths are in Å). Energy values (kcal mol⁻¹) are relative to reactants.

The two Asp[553] carboxylate oxygens O5 and O6 form two hydrogen bonds with the C2-bonded hydrogen: the O5··H and O6··H distances are 2.49 and 2.56 Å, respectively, and the CH bond is longer than normal (1.10 Å). These hydrogen contacts anticipate the proton transfer to the Asp[553] carboxylate oxygen, which will be completed in the resulting species **Pd1**. Thus, in this phase of the catalytic process, the Asp[553] residue behaves as a general base. Furthermore, along the transformation **Int1 → TS2 → Pd1**, the Asp[665] carboxylate oxygen O4 maintains a rather strong contact with the O7-bonded hydrogen: the O7··H and H··O4 distances are 1.01 and 1.66 Å in **TS2**, respectively and 1.00 and 1.68 Å in **Pd1**. In **Pd1** the C1-C2 bond is unquestionably an olefin bond (1.34 Å). These results agree with the experimental evidence[47] suggesting that the second phase of the catalytic process is an E1-like elimination where the proton abstraction and the expulsion of the leaving group Asp[553] (acting simultaneously as a base) occur asynchronously in the same kinetic step. This *syn* elimination is favored by the enhanced acidity of the C2-bonded hydrogen, which is due to the adjacent positive charge on C1 in **TS2**, which has a significant oxocarbenium ion character (the Mulliken charge on C1 being 0.41), as suggested by Lee *et al.*[42] Importantly, **TS2**, corresponding to deglycosylation-elimination, is the rate determining step of the catalytic process, as pointed out by Su and coworkers.[23] Since it is 11.0 kcal mol⁻¹ above **Int1** (7.3 kcal mol⁻¹ above reactants) the overall barrier that must be surmounted to form **Pd1** is 18.3 kcal mol⁻¹. Two points of our computations are noteworthy. The first point concerns the whole reaction energetics. The first phase of the reaction is endothermic, the enol product **Pd1** (enol-AF) being 6.6 kcal mol⁻¹ less stable than

the starting Michaelis complex **Rx**). The second and more important point, concerns the energy of the rate-determining transition state **TS2**, which is 18.3 kcal mol[-1] above reactants, a value which is consistent with an enzymatic reaction. Two factors can be, in principle, responsible, for these results that differ from previous computations[23][63] *i.e.* the rather large QM region together with the type of functional used in our computations. The M06-2X functional used here, is specifically intended for non-covalent long-range interactions, which become more important with the increasing number of residues included in the QM region. The same reasoning holds for **Int1** reported in the previous reaction step, which is very close in energy to **Pd1**.

3.3.3 Tautomerism and catalyst reactivation. We found that it is possible to restore the initial state of the catalyst *via* a double proton transfer requiring a barrier of 5.9 kcal mol[-1] (transition state **TS3** in **Scheme 3.4**). A proton moves from the carboxylate oxygen O5 of Asp[553] to oxygen O7 (in **TS3** O5···H and H···O7 distances are 1.17 and 1.29 Å, respectively) and simultaneously a proton abandons O7 to bind the carboxylate oxygen O4 of Asp[665], thus restoring the initial protonation state of the enzyme (intermediate **Int2**, 2.2 kcal mol[-1] above reactants).
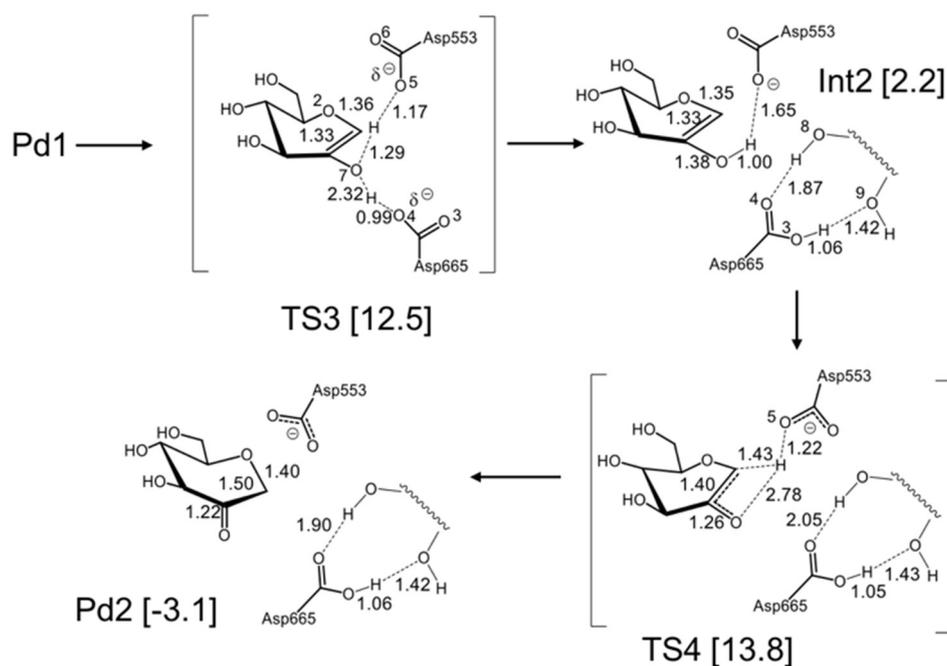
Scheme 3.4 Schematic representation of the catalyst reactivation and subsequent tautomerization (bond lengths are in Å). Energy values (kcal mol[-1]) are relative to reactants.

The second proton transfer is highly asynchronous, the two O7···H and H···O4 distances being 2.32 and 0.99 Å, respectively. Thus, one proton moves from Asp[553] when the other proton transfer is almost complete. These results simply indicate that, once the enol-AF product is formed (**Pd1**), the system restores easily the initial protonation state of the catalyst by surmounting a rather small

energy barrier (5.9 kcal mol^{-1}). The restoring of the catalyst initial state makes possible to carry out the enol-keto tautomerization by overcoming a reasonable barrier of 11.6 kcal mol^{-1}. In this passage Asp[553] behaves as a proton shuttle (see transition state **TS4** in **Scheme 3.4**). A proton moves from O7 to C1 and its motion is "assisted" by the Asp[553] carboxylate oxygen O5. This oxygen atom strongly interacts with the proton (O5$\cdots$H distance is 1.22 Å) and accompanies its "journey" along a parabolic trajectory where the proton is hung over the C1-C2-O7 plane (O7$\cdots$H and H$\cdots$C1 distances are 2.78 and 1.43 Å, respectively). Simultaneously the C2-O7 bond becomes shorter (1.26 Å, *i.e.* its double bond character increases) with a consequent lengthening of the C1-C2 bond (1.40 Å). The resulting product **Pd2** is keto-AF, which is 3.1 kcal mol^{-1} more stable than the starting Michaelis complex and 9.7 kcal mol^{-1} more stable than the enol form **Pd1**. Interestingly, the **Pd1 → TS3 → Int2** passage plays a double role: (i) it reactivates the initial catalytic state of the enzyme and (ii) allows Asp[553] to behave as a proton shuttle, which makes possible to complete the enol-keto tautomerism process by overcoming a moderate energy barrier. Furthermore, a structural modification of the active site in the last phase of the process contributes to stabilize the final **Int2 → TS4 → Pd2** transformation. The newly protonated Asp[665] carboxylate group changes its orientation and forms two strong stabilizing hydrogen bonds with two adjacent OH groups (H-O8 and H-O9) of the leaving sugar unit, which is still present in the catalytic site (see **Scheme 3.4**). In **Int2** the O4$\cdots$HO8 and O9$\cdots$HO3 distances are 1.87 and 1.42 Å, respectively, and remains almost constant along the **Int2 → TS4 → Pd2** transformation (1.90 and 1.42 Å in **Pd2**). Alternatively, tautomerization could occur in solution once the enol-AF has left the enzyme pocket. To verify this hypothesis, we investigated at the same computational level the tautomerization process enol-AF → keto-AF outside of the protein environment. In our model the proton transfer was "assisted" by a water molecule (explicitly considered) in aqueous solvent (PCM computations).[44] We found an activation energy barrier of 35.8 kcal mol^{-1} (see **Scheme 3.5**). Thus, it seems more reasonable to hypothesize that enol-AF undergoes the tautomerization process inside the enzyme before leaving the active site rather than in the aqueous solution.
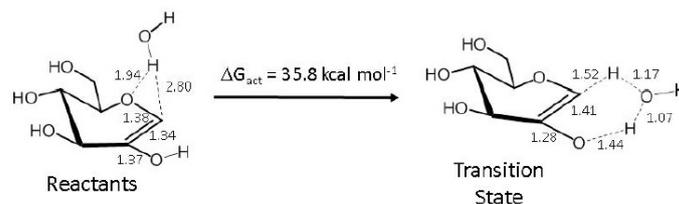
Scheme 3.5 A schematic representation of the tautomerization process (reactant and transition state) occurring in water outside the enzyme. Bond lengths are in Å.

3.3.4 The solvent effect on molecular structures. To elucidate the effect of solvent on geometry optimization we re-optimized the geometry of the most important critical points in the presence of solvent effects at the M06-2X//6-31+G*/3-21G* level. We considered **Rx**, **TS1**, **Int1**, **TS2** and **TS3**. Even in this case after geometry optimization we carried out single-point computations with solvent effects at the more accurate 6-311++G**/6-31+G* level. We found that the solvent does not significantly affect the geometries and relative energies of the critical points we have considered. The re-optimized geometries are reported in **Scheme 3.6**. The energies of **TS1**, **Int1**, **TS2** and **TS3** (relative to reactants **Rx**) are 9.3, 7.4, 20.0 and 14.5 kcal mol^{-1}, respectively. These values do not affect the mechanistic picture which stems from the diagram of **Fig.3.4** and compare very well with those reported there. The rate determining step remains the glycosylation/elimination (**TS2**). The overall energy that must be surmounted to carry out the reaction is 20.0 kcal mol^{-1}, a value rather close to that previously discussed (18.3 kcal mol^{-1}). All these results indicate that single point computations including solvent effects on the gas-phase optimized geometries provide a reliable description of the energy profile.
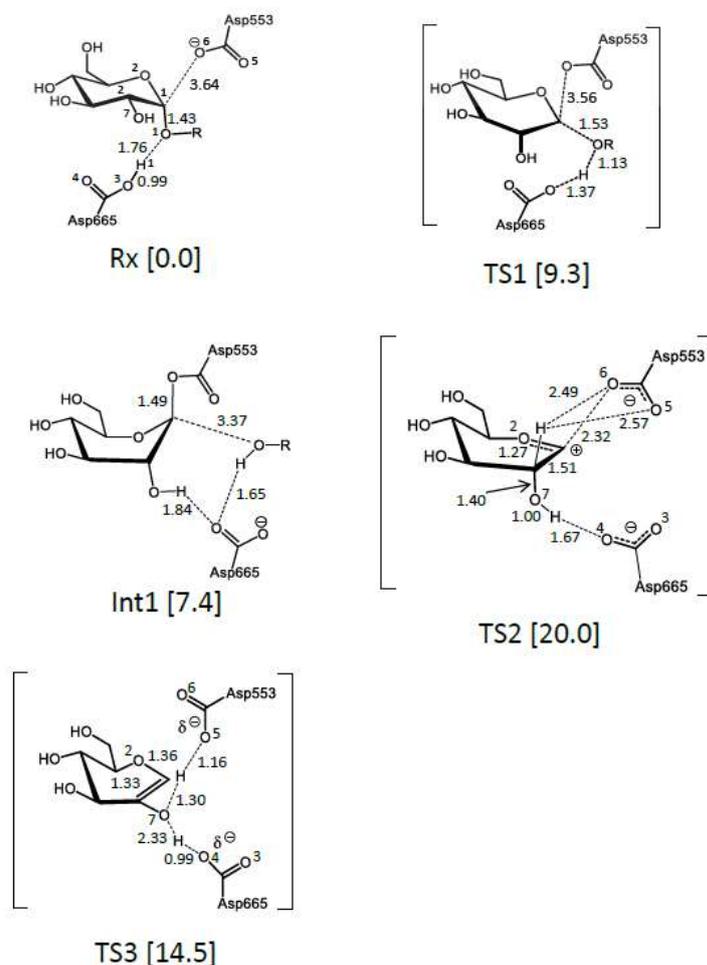
Scheme 3.6 Schematic representation of **Rx**, **TS1**, **Int1**, **TS2** and **TS3** computed at M06-2X//6-31+G*/3-21G* level including solvent effect in the geometry optimization process.

3.3.5 The effect of the functional on the energy profile. We used the B3LYP functional with Grimm corrections (B3LYP + D) to carry out single-point computations on the structures obtained with the M06-2X functional. Computations were carried out at the 6-311++G**/6-31+G* level of accuracy including solvent effect. The corresponding energy values are reported in **Table 1**.

Table 1. Energy values (kcal mol^{-1}) relative to reactants obtained with single point computations at the (B3LYP + D level)//6-311++G**/6-31+G* level including solvent effects. On the right column the geometries were first re-optimized at the (B3LYP + D)//6-31+G*/3-21G* level.

	Rx	TS1	Int1	TS2	Pd1	TS3	Int2	TS4	Pd2
Single-point computations (B3LYP+D)	0.00	11.8	10.4	16.9	8.7	19.4	3.9	18.2	-3.1
Geometry optimization (B3LYP+D)	0.00	17.6		22.7		19.8			

Interestingly, while the energy of **TS1** (11.8 kcal mol^{-1}) is rather close to the value computed with the M06-2X functional (10.2 kcal mol^{-1}), more significant changes concern **TS2**, **TS3** and **TS4**: the energy of **TS2** decreases (16.9 kcal mol^{-1}), while those of **TS3** and **TS4** significantly increase (19.4 and 18.2 kcal mol^{-1}) with respect to the M06-2X values (18.3, 12.5 and 13.8 kcal mol^{-1}, respectively). The picture which stems from single-point computations is that **TS3** (restoring of the catalytic active site) represents the bottle-neck of the entire process.

To establish even more clearly the effect of the functional, we re-optimized the geometry of **Rx**, **TS1**, **TS2** and **TS3** in gas-phase at the (B3LYP + D)//6-31+G*/3-21G* level and we carried out again, for the sake of consistency, single point computations with the 6-311++G**/6-31+G* basis including solvent effect. The re-optimized geometries are reported in **Scheme 3.7**.

The new energy values are reported on the bottom row of **Table 1**. Interestingly, the geometry optimization restores the energetic sequence of **TS2** and **TS3**, suggesting that the geometry optimization is crucial: **TS2** becomes again the rate-determining step of the process (22.7 kcal mol^{-1}), but both **TS1** and **TS3** are rather high in energy (17.6 and 19.8 kcal mol^{-1}, respectively) compared to the M06-2X values. These results indicate that M06-2X and B3LYP+D provide a similar mechanistic picture (**TS2** rate determining), even if the energetics of glycosylation (**TS1**) and catalyst regeneration (**TS3**) change significantly.
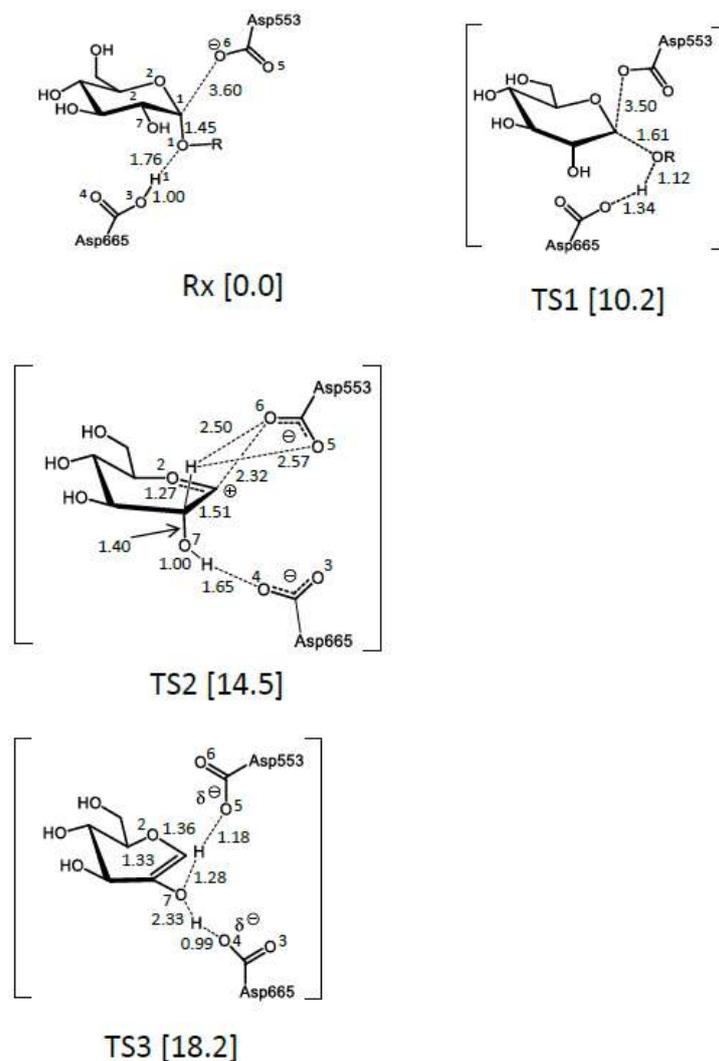
Scheme 3.7 Schematic representation of **Rx**, **TS1**, **Int1**, **TS2** and **TS3** computed at the (B3LYP+D3//6-31+G*/3-21G* level.

3.4. Conclusions

In this work we examined the catalytic mechanism of α-1,4-glucan lyases from the red seaweed *Gracilariopsis lemaneiformis* using a full QM approach. The results of our computations can be summarized as follows:

(i)   The glycosylation occurs in one kinetic step characterized by a highly asynchronous S_N1-like transition state. A barrier of 10.2 kcal mol^{-1} must be overcome to form the glycosyl-enzyme intermediate. A proton transfer from Asp^{665} (general acid) to the glycosidic oxygen activates the rupture of the glycosidic bond. A nucleophilic attack of Asp^{553} at the anomeric carbon follows in the same kinetic step. Even if these results are similar to those obtained by Su and coworkers,[23] our computations show that the glycosyl intermediate is significantly less stable than the starting Michaelis complex (7.3 kcal mol^{-1} above it) and not

almost degenerate to reactants (which would be an anomalous situation for a reactive intermediate species).

(ii) The deglycosylation-elimination is the bottleneck of the entire process (rate-determining step) as pointed out by Su and coworkers.[23] The corresponding transition state **TS2** has a strong oxocarbenium ion character and describes an E_1-like elimination where the proton abstraction and the expulsion of the leaving group Asp[553] occur asynchronously. Importantly, the overall barrier computed with respect to reactants (18.3 kcal mol^{-1}) is consistent with an enzymatic reaction.

(iii) The high energy of **Pd1** (enol-AF representing the initial product of the catalytic process) makes very easy to restore the initial state of the catalyst (protonated Asp[665]). Our computations indicate that this process occur easily by overcoming a small barrier of 5.9 kcal mol^{-1}.

(iv) A very asynchronous double proton transfer restores the catalyst in its initial state.[50] The restored catalyst makes possible to obtain the final keto form of the product (keto-AF significantly more stable than enol-AF) by overcoming a moderate barrier (11.6 kcal mol^{-1}) since the Asp[553] carboxylate group can act as a proton shuttle and assist the proton transfer in the tautomerization process.

(v) Our computations clearly indicate that the tautomerization must occur inside the enzyme before leaving the active site rather than in the aqueous solution. Outside of the protein environment the enol-AF → keto-AF process "assisted" by a water molecule (in aqueous solvent emulated by PCM computations) has an energy barrier of 35.8 kcal mol^{-1}.

(vi) The solvent effect used to emulate the protein environment does not modify significantly the geometry of the various critical points with respect to gas-phase. Thus, single-point computations including solvent effect on the gas-phase optimized geometries provide reliable reaction energetics.

(vii) On the whole the M06-2X and the B3LYP+D functionals provide a similar general mechanistic picture, even if the activation barriers of glycosylation (first step) and catalyst regeneration (third step) change significantly. Importantly, however, **TS2** remains the rate-determining step at the two levels of theory and its energy does not change dramatically (18.3 and 22.7 kcal mol^{-1} in the former and latter case, respectively).

(viii) All above results suggest that the large QM region used in our computations is crucial in determining the difference with respect to previous work.[23] An additional effect (related

to the previous one) can be ascribed to the type of functional used for estimating long-range interactions, whose importance increase with the increasing number of residues included in the QM region. Even if the contribution of every single interaction is very small, the overall effect can become important and modify significantly the structure and energy of the various critical points along the reaction profile. As a result of the complex interplay of these two factors (QM size and functional), we found that deglycosylation-elimination can occur in only one kinetic step (requiring a barrier of 18.3 kcal mol^{-1}) where the breaking of the glycosidic bond and the abstraction of the C2 proton are highly asynchronous.

References

a. Campesato, L; Marforio, T.D; Giacinto, P.; Calvaresi, M.; Bottoni, A. ChemPhysChem, **2018**, *19(12)*, 1514-1521.

[1] Davies, T.; Gloster, M.; B. Henrissat. *Current Opinion in Structural Biology*, **2005**, *15*, 637-645.

[2] Yip, V. L. Y.; Withers, S. G. *Current Opinion in Structural Biology*, **2006** ,*10*, 147-155.

[3] Yu, S.; Kenne, L.; Pedersén, M. *Biochim. Biophys. Acta.*, **1993**, *1156*(3), 313-320.

[4] Yu, S. *IUBMB Life*, **2008**, *60*, 798-809.

[5] Yu, S. *Zuckerindustrie*, **2004**, *129*, 26-30.

[6] Yu, S.; Pedersén, M. *Planta*, **1993**, *191*, 137-142.

[7] Hirano, K.; Ziak, M.; Kamoshita, K.; Sukenaga, Y.; Kametani, S.; Shiga, Y.; Roth, J.; Akanuma, H. *Glycobiology*, **2000**, *10*, 1283-1289.

[8] Kametani, S.; Shiga, Y.; Akanuma, H. *Eur. J. Biochem.*, **1996**, *242*, 832-838.

[9] Maier, P.; Anderse, S. M.; Lundt, I. *Synthesis*, **2006**, 827-830.

[10] Baute, M.A.; Deffieux, G.; Baute, R.; Badoc, A.; Vercauteren, J.; Léger, J.M.; Neveu, A. *Phytochemistry*, **1991**, *30*, 1419-1423.

[11] Hisaku, S.; Takeda, Y.; Abe, J. I.; Muroya, T.; Yoshinaga, K.; Fujisue, M. *Jap. Patent Appl.*, **1999**, JP11-266336.

[12] Daisuke, U.; Yoshiaki, T.; Ichiro, W.; Yoshimasa, H. *Chem. Lett.*, **1979**, *12*, 1481-1482.

[13] Brehm, M.; Göckel, V. H.; Jarglis, P.; Lichtenthaler, F. W. *Tetrahedron Asym.*, **2008**, *19*(3), 358-373.

[14] Cardellina, J. H.; Hendrickson, R. L.; Manfredi, K. 0. P.; Strobel, S. A.; Clardy, J. *Tetrahedron Lett.*, **1987**, *28*, 727-730.

[15] Abeyama, K.; Maruyama,Y.; Yoshimoto, Y.; Yoshinaga, K. *Jap. Patent Appl.*, **2006**, JP2006306814.

[16] Yu, S.; Bojsen, K.; Svensson, B.; Marcussem, J. *Biochimica et Biophysica Acta*, **1999**, *1433*, 1-15.

[17] Hernissat, B.; Davies, G. *Curr. Opin. Struct. Biol.*, **1997**, *7*, 637-644.

[18] Lee, S.; Yu, S.; Withers, S. G. *J. Am. Chem. Soc.*, **2002**, *124*, 4948-4949.

[19] Lee, S.; Yu, S.; Withers, S. G. *Biochemistry*, **2003**, *42*, 13081-13090.

[20] Lee, S.; Yu, S.; Withers, S. G. *Biologia, Bratislava 60/Suppl.*, **2005**, *16*, 137-148.

[21] Zechel, D.; Withers, S.G. *Acc. Chem. Res.*, **2000**, *33*, 11-18.

[22] Rozeboom, H. J.; Yu, S.; Madrid, S.; Kalk, K. H.; Zhang, R.; Dijkstra, B. W. *J. Biol. Chem.*, **2013**, *288*, 26764-26774.

[23] Su, H.; Dong, L.; Liu, Y. *RSC Adv.*, **2014**, *4*, 54398-54408.

[24] a) Jindal, G.; Warshel, A. *J. Phys. Chem. B*, **2016**, *120*, 9913-9921; b) Siegbahn, P. E. M.; Borowski, T. *Faraday Discuss.*, **2011**, *148*, 109-117.

[25] Cerqueira, N. M. F. S. A.; Fernandes, P. A.; Ramos, M. J. *ChemPhysChem*, **2017**, 10.1002/cphc.201700339.

[26] Williams, R. T.; Wang, Y. *Biochemistry*, **2012**, *51*, 6458-6462.

[27] Wòjcik, A.; Broclawik, E.; Siegbahn, P.E.M.; Borowski, T. *Biochemistry*, **2012**, *51*, 9570-9580.

[28] Faraji, S.; Wirz, L.; Dreuw, A. *ChemPhysChem*, **2013**, *14*, 2817-2824.

[29] Chen, S.L.; Liao, R.Z. *ChemPhysChem*, **2014**, *15*, 2321-2330.

[30] Aranda, J.; Cerqueira, N. M. F. S. A.; Fernandes, P. A.; Roca, M.; Tunon, I.; Ramos, M.J. *Biochemistry*, **2014**, *53*, 5820-5822.

[31] Cassimjee, K. E.; Manta, B.; Himo, F. *Org. Biomol. Chem*, **2015**, *3*, 5466-5470.

[32] Marforio, T.D.; Giacinto, P.; Bottoni, A.; Calvaresi, M. *Biochemistry*, **2015**, *54*, 4404-4410.

[33] Li, J.; Li, J.; Zhang, D.; Liu, C. *ChemPhysChem*, **2015**, *16*, 3044-3048.

[34] Makurat, S.; Chomicz-Manka, L.; Rak, J. *ChemPhysChem*, **2016**, *17*, 2572-2578.

[35] Larranaga, O.; de Cjzar, A.; Cossio, F. P. *ChemPhysChem*, **2017**, *18*, 3390-3401.

[36] Lintuluoto, M.; Lintuluoto, J. M. *Biochemistry*, **2016**, *55*, 210-223.

[37] Olàh, J.; Mulholland, A.J.; Harvey, J. N. *PNAS*, **2011**, *108*, 6050-6055.

[38] Anandakrishnan, R.; Aguilar, B.; Onufriev, A.V. *Nucleic Acids Res.*, **2012**, *40*, W537-W541.

[39] Gaussian 09, Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; et al Gaussian, Inc., Wallingford CT, **2009**.

[40] Zhao, Y.; Truhlar, D. G. *Theor. Chem. Acc.*, **2008**, *120*, 215-241.

[41] Calvaresi, M.; Bottoni, A.; Garavelli, M. *J. Phys. Chem. B*, **2007**, *111*, 6557-6570.

[42] Calvaresi, M.; Garavelli, M.; Bottoni, A. *Proteins: Struct., Funct., Genet.*, **2008**, *73*, 527-538.

[43] Bottoni, A.; Miscione, G.P.; Calvaresi, M. *Phys. Chem. Chem. Phys.*, **2011**, *20*, 9568-9577.

[44] Tomasi, J.; Mennucci, B.; Cammi, R. *Chem. Rev.*, **2005**, *105*, 2999-3093.

[45] Blomberg, M. R. A.; Siegbahn, P. E. M.; Babcok, G. T. *J. Am. Chem. Soc.*, **1998**, *120*, 8812-8824.

[46] Borowski, T.; Siegbahn, P. E. M. *J. Am. Chem. Soc.*, **2006**, *128*, 12941-12953.

[47] Liao, R.Z.; Ding, W.J.; Yu, J.G.; Fang, W.H.; Liu, R.Z. *J. Comput. Chem.*, **2008**, *29*, 1919-1929.

[48] Yip, V. L. Y.; Withers, S. G. *Org. Biomol. Chem.*, **2004**, *2*, 2707-2713.

[49] Trujillo, C.; Sanchez-Sanz, G.; Alkorta, I.; Elguero, J. *ChemPhysChem*, **2015**, *16*, 2140-2150.

[50] Lopata, A.; Jambrina, P. G.; Sharma, P. K.; Brooks, B. R.; Toth, J.; Vertessy, B. G.; Rosta, E. *ACS Catal.* **2015**, *5*, 3225-3237.

4. Unveiling the reaction mechanism of Sulfotransferases in Heparan Sulfate biosynthesis by means of hybrid QM/MM computational investigations

4.1. Introduction
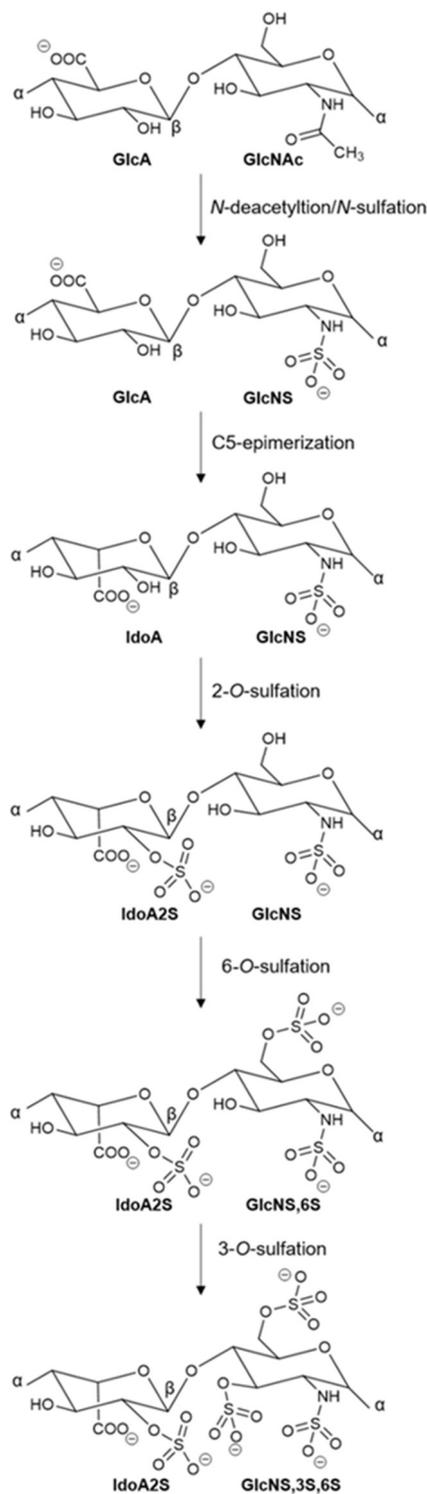
Fig.4.1 Heparan sulfate biosynthesis.

Heparan sulfate (HS) is a glycosaminoglycan expressed on cell surfaces and in the extracellular matrix. This complex polymer is involved in a wide range of physiological and pathophysiological functions, from embryonic development to respond to viral and bacterial infection.[1] The HS chain is a polysaccharide and consists of repeating disaccharide units of glucoronic acid (GlcA) or iduronic (IdoA) and *N*-acetylglucosamine (GlcNAc). The disaccharide units (-GlcA-GlcNAc-) can carry sulfuryl (also referred to as sulfo) SO_3^- groups in different positions. The biosynthesis of HS takes place in the Golgi apparatus and involves glycosyltransferases, an epimerase, and several sulfotransferases, which are the enzymes responsible for the sulfonation of the saccharides. The universal sulfate donor is 3'-phosphoadenosine-5'-phosphosulfate (PAPS). Once the backbone of the disaccharide unit (-GlcA-GlcNAc-) is generated by HS polymerase, it undergoes modifications by *N*-deacetylase/*N*-sulfotransferase (NDST), *D*-Glucoronyl-C5-epimerase (Glce), 2-*O*-sulfotransferase (2OST), 6-*O*-sulfotranferase (6OST) and 3-*O*-sulfotransferases (3OST). The order of the modifications is essentially sequential, so that the products of a step is the substrate for the next. Even if the order of the enzymes in the biosynthetic path of HS is preserved, the sulfuryl groups can end to be located in slightly different positions, and this modulations in the pattern of SO_3^- plays crucial role in determining the functions of HS.[2] In fact, HS executes a wide range of functions by interacting with other

proteins, such protease inhibitors, proteases, cytokines, chemokines and morphogens[3-5] and these interactions depend on the position of the sulfuryl groups on HS. NDST is the first enzyme found in the biosynthetic path of HS and heparin (see **Fig.4.1**) and displays a dual function: the *N*-deacetylase domain converts GlcNAc into *N*-glucosamine (GlcNH$_2$) and the *N*-sulfotransferase domain transfers a sulfuryl group to GlcNH$_2$, generating *N*-sulfo-glucosamine GlcNS. This dual enzyme is crucial, since all the other enzymes found in the HS path formation require the presence of GlcNS residues to complete their modification on the polysaccharide chain. Despite great efforts, only one crystallographic structure extracted from *Homo sapiens* is available, and it dates back to 1999.[6] Recent experimental insights have been published in 2015 by Liu[1] indicating that the ND and the NS domain are capable of working separately, but the highest efficiency (in terms of deacetylation-sulfonation levels) is obtained when the two enzymes are coupled. Then, the second enzyme we encounter is Glce, which is responsible for the conversion of GlcA into IdoA by the C5 epimerization. The most recent crystal structure of Glce has been deposited in the Protein Data Bank in 2015 under the accession code 4PW2 and few insights into the mechanism of C5 epimerization have been provided.[7] Successively, 2OST performs the transfer of the sulfo-group preferentially on IdoA, and to a minor extent on GlcA, generating 2-*O*-sulfated iduronic acid (IdoA2S) or 2-*O*-sulfated glucuronic acid (GlcA2S) respectively. The IdoA2S was revealed to have physiological functions in growth factor-mediated signal transduction pathways.[8,9] In 2014 Pedersen and coworkers[10] deposited the crystallographic structure of 2OST (from *Gallus gallus*, 92% identity to *Homo sapiens*) with 3'-phosphoadenosine-5'-phosphate (PAP) an a heptasaccharide acceptor substrate (PDB 4NDZ). The second-last enzyme is 6OST, which transfers the sulfuryl group to the 6-OH position of GlcNH$_2$. The crystallographic structure of 6OST of human (or mouse) has never been resolved, but in 2017 the Zebra fish 6OST (>70% identity to human) has been obtained and deposited in the PDB under the accession code 5T05. 3OST is responsible for the catalysis of the last sulfurylation and it catalyzes the formation of GlcNS,3S,6S.

Because of the physiological implication of HS and its analogues, it is evident that many experimental and computational efforts should be made to understand and explain the mechanisms underlying the enzymes that play crucial roles in HS biosynthesis. Up to date, only the reaction mechanism of 3OST has been studied in detail by means of hybrid QM/MM methodologies, underlying that the SO$_3^-$ transfer takes place in a S$_N$2-like reaction, where the triad Asp189-His186-Glu184 exert the catalytic action.[12] Beyond the biological importance of HS, heparin, which differs from HS for being produced only by mast cells and containing a higher sulfurylation and IdoA

levels, is the oldest anticoagulant drug used in medicine. Heparin used to be isolated from porcine intestine or bovine lung and in 2008 contaminated lots of heparin were withdrawn from the US market.[13] Alternative ways to obtain pure heparin were desirable. The chemoenzymatic approach resulted to be a promising method to accomplish this goal.[14] The disclosure of the catalytic mechanism of the sulfotransferases that participate in the HS biosynthesis will help the development of more efficient chemoenzymatic approaches to synthesize heparin and HS-based therapeutics.

Our computational work is being carried out in collaboration with Tarsis F. Gesteria (Huston University, Texas, USA) and attempts to give mechanistic insights into the sulfotransferase catalysis and binding modes exerted by NDST, 2OST and 6OST enzymes by means of MM and hybrid QM/MM computations. We believe that our results will help in the comprehension of the reactions that generate HS and will furnish clues for the development more efficient chemoenzymatic approaches to obtain heparin derivatives. Moreover, the computational insights into the critical points that describe the sulfotransferase catalyzed reactions will help to design analogues capable of inhibit/enhance some steps of the HS biosynthesis. The preliminary results resumed in this thesis focus on the QM/MM investigation of the role of 2OST in transferring the SO_3^- group to IdoA or GlcA units. We seek to determine the reactant, transition state and product structures, as well as the role of the catalytic base, which is presumed to be His142, responsible for the activation of the nucleophile.

4.2. Computational Details

The model systems of 2OST were obtained from the Protein Data Bank (PDB codes 4NDZ). The original crystallized ligand (NPO-GlcA-GlcNS-IdoA-GlcNS-GlcA-GlcNAc-GlcA) is reported in **Fig.4.2**.
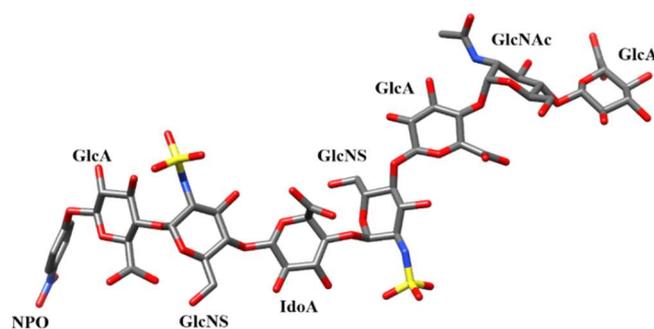
Fig.4.2 Three-dimensional representation of the crystallized ligand of 2OST (PDB 4NDZ).

In order to study the sulfurylation of IdoA and GlcA, we manually removed the NPO (*para*-nitro phenol) group from the ligand and then i) replaced all GlcA with IdoA units to create ligand A **Fig.4.3a**), and ii) changed the only IdoA unit with GlcA, to obtain ligand B (from the reducing end: **Fig.4.3b**).
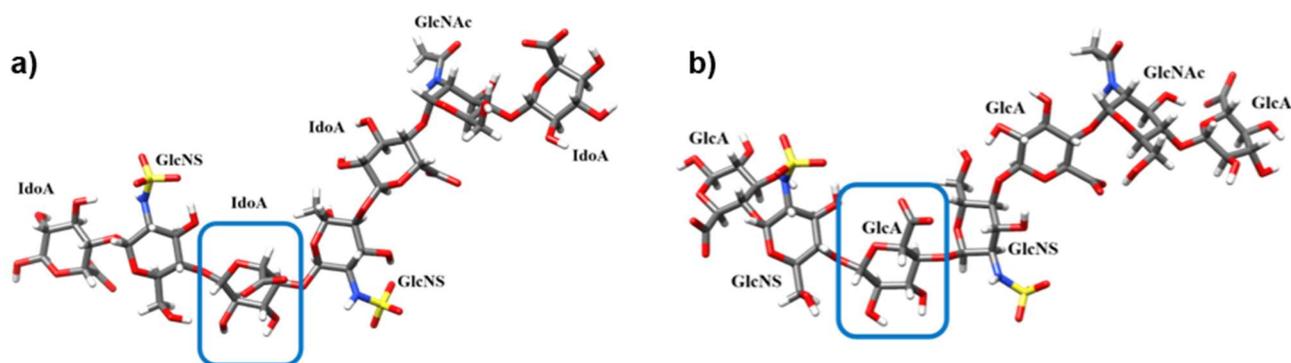
Fig.4.3 Three-dimensional representation of the modified **a)** ligand A from the reducing end: GlcA-GlcNS-IdoA-GlcNS-IdoA-GlcNAc-IdoA **b)** ligand B from the reducing end: GlcA-GlcNS-GlcA-GlcNS-GlcA-GlcNAc-GlcA.

Once the MM parameterization of PAPS and ligands A and B was completed, we simulated the dynamical behavior in water of 2OST – PAPS – ligand A and 2OST – PAPS – ligand B systems by means of two Molecular Dynamic simulations of 500ns each. Parameterization and trajectory analysis were completely produced by our external collaborator.

From the two trajectories, a selected snapshot for each simulation was extracted and two different QM/MM model systems were built upon it. We created **Model A** and **B** (4949 atoms) to study the sulfonation of IdoA and GlcA units respectively. **Model A** comprises the 2OST monomer, PAPS and ligand A, and **Model B** the 2OST monomer, PAPS and ligand 2. All calculations were carried out using the combined quantum mechanical and molecular mechanical (QM/MM) method in the ONIOM formalism, as implemented within the Guassian09 program suite.[15] The coupled QM/MM method adopted here is a two-layer ONIOM scheme,[16] where the reactive region is treated at DFT level using the density functional M06-2X[17] and 6-31+G* basis set.[15] The QM layer comprises the target sugar unit, the lateral chain of PAPS, Lys83, Thr90, Arg80 and His142, which are the residues that are suggested to be crucial in the catalytic activity of 2OST. The remaining region is treated using the UFF force field.[15] The global potential can be referred as M06-2X/6-31+G*:ff99SB.

ONIOM calculations were performed considering both mechanical and electrostatic embedding,[16] in order to give an accurate description of the electrostatic interaction between both layers and

polarization of the QM layer by the MM region. Residues more than 10Å away for the active site were frozen during the calculations.

The potential energy surface (PES) was obtained by selecting the principal component of the linear combination that describes the reaction coordinate: in our case the nucleophile-electrophile distance. Following a scanning approach, the $2OH - SO_3^-$ distance was incrementally diminished, while the rest of the system was allowed to relax. When needed, a second component (*e.g.* nucleophile activation by the catalytic base) was scanned, creating a multidimensional energy surface.

In this thesis, instead of transition state structures we report the geometry and energetics of the maximum points obtained by the scanning approach described: these points correspond, in a first approximation, to the region where the transition states are more likely to be found. Actual transition state optimizations and frequency calculations will be run to confirm the nature of these points located at the maximum of the PES.

4.3. Results and Discussion

4.3.1. Model A: 2-OH sulfonation of IdoA. The full optimization of our initial point generated the Michaelis complex **Rx(A)** depicted in **Scheme 4.1**.

The two positively charges residues considered at QM level contact strong hydrogen bond with the $SO_3^- - PO_4^{2-}$ chain of the donor PAPS (H13$\cdots$O5 2.58 Å and H10$\cdots$O6 1.76 Å respectively) and Thr90 engages a H-bond of 1.79 Å with the 2-OH of the IdoA unit (H2$\cdots$N3 1.86 Å). The transfer of the sulfuryl group from PAPS to the iduronic unit takes place in one stage, in the transition state region identified as **Max1(A)**, where the nucleophile is activated by His142, which abstracts the proton from 2-OH group (H2 ---N3 1.57 Å). Simultaneously, the activated nucleophile O1 attack the electrophilic sulfur atoms (O1---S4 2.85 Å) while the $SO_3^- - PO_4^{2-}$ bond is cleaved (S4---O6 2.25 Å). The incipient negative charge on O6 is mitigated by a proton transfer acted by Lys83, which donated H10 to O6 (N11---H10 1.87, H10---O6 0.99 Å).

It is interesting to note that the guanidinium portion of Arg80 assists the transfers of SO_3^- by acting as a shuttle that follows the sulfuryl "travel" from PAPS the iduronic acid. The product of the reaction, that foresees the formation IdoA2S, is described by the critical point **Pd1(A)**, where the bond O1-S4 (1.74 Å) is formed, and the phosphate moiety is protonated. The residue Lys83 stabilizes the structure by engaging a H-bond with the HPO_4^- moiety (N11$\cdots$H10 2.70 Å).
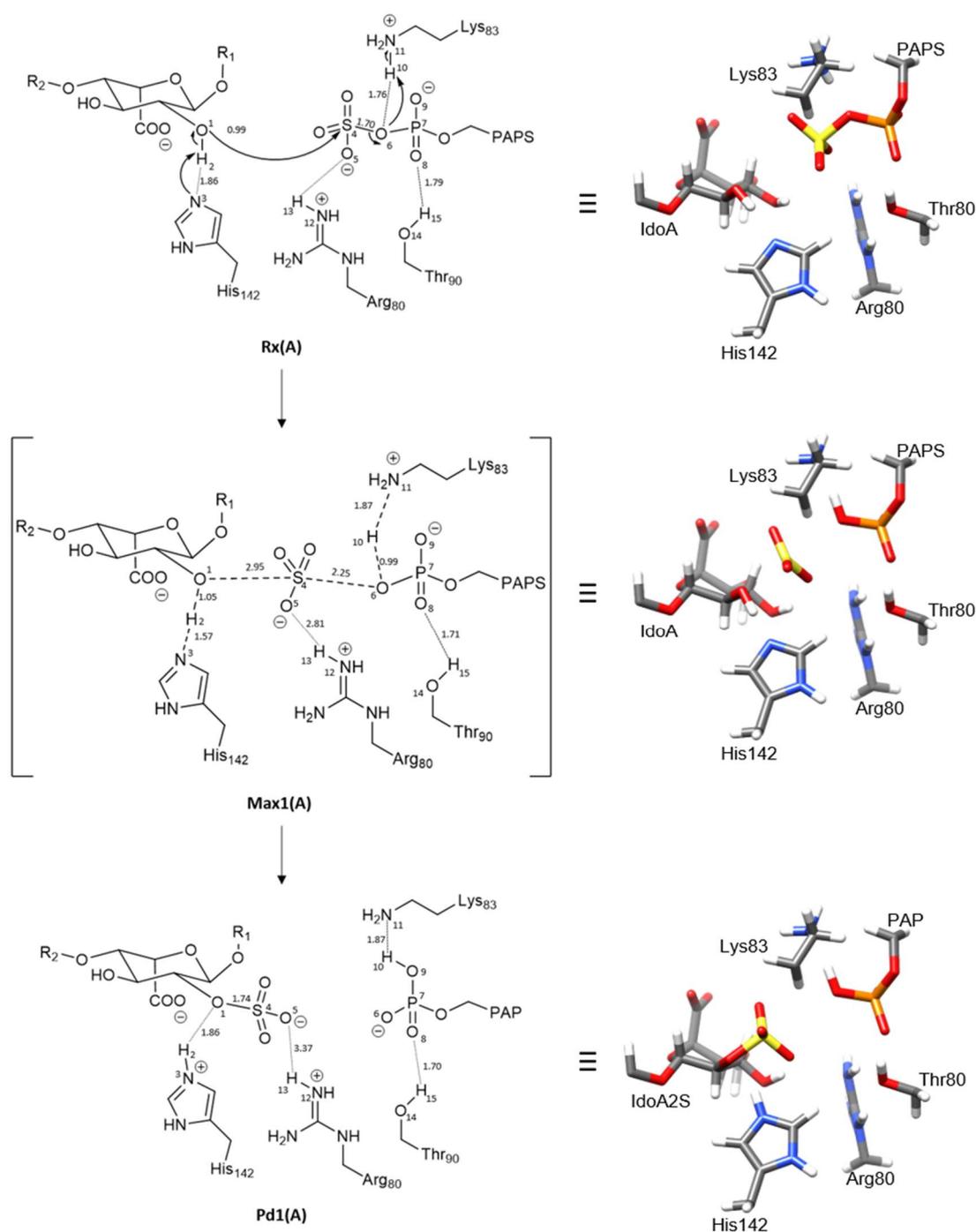
Scheme 4.1. On the left side, schematic representation of the sulfurylation of IdoA unit (bond lengths are in Å). R_1 =-GlcNs-IdoA and R_2=-GlcNS-IdoA-GlcNAc-IdoA. On the right side, the corresponding three-dimensional representation of the computed points.

4.3.2. Model B: 2-OH sulfonation of GlcA. The reaction mechanism computed for the transfer of the sulfuryl group of the glucuronic acid unit is schematically given in **Scheme 4.2**. The Michaelis complex **Rx(B)** computed for **Model B** describes a situation quite analogue to **Rx(A)**, even if the sulfuryl moiety is protonated (O5-H10 1.05 Å) while Lys83 is in its neutral form. Arg80 and Thr90 engage H-bonds with SO_3^- (O5′···H13 2.20 Å) and PO_4^{2-} (O8···H18 2.80 Å). From **Rx(B)**, the reaction

path computed for the sulfonation of the glucuronic acid is a concerted mechanism, where the nucleophile is activated a proton abstraction acted by His142 (N3---H2 1.29 Å). The 2-OH of GlcA attacks the sulfuryl group (O1---S4 2.00 Å), while the SO_3^- — PO_4^{2-} bond is cleaving (S4---O6 1.83 Å). Interestingly, the proton H10 shifts from the sulfuryl to the phosphate group in the region corresponding to the transition state (**Max1(B)**).
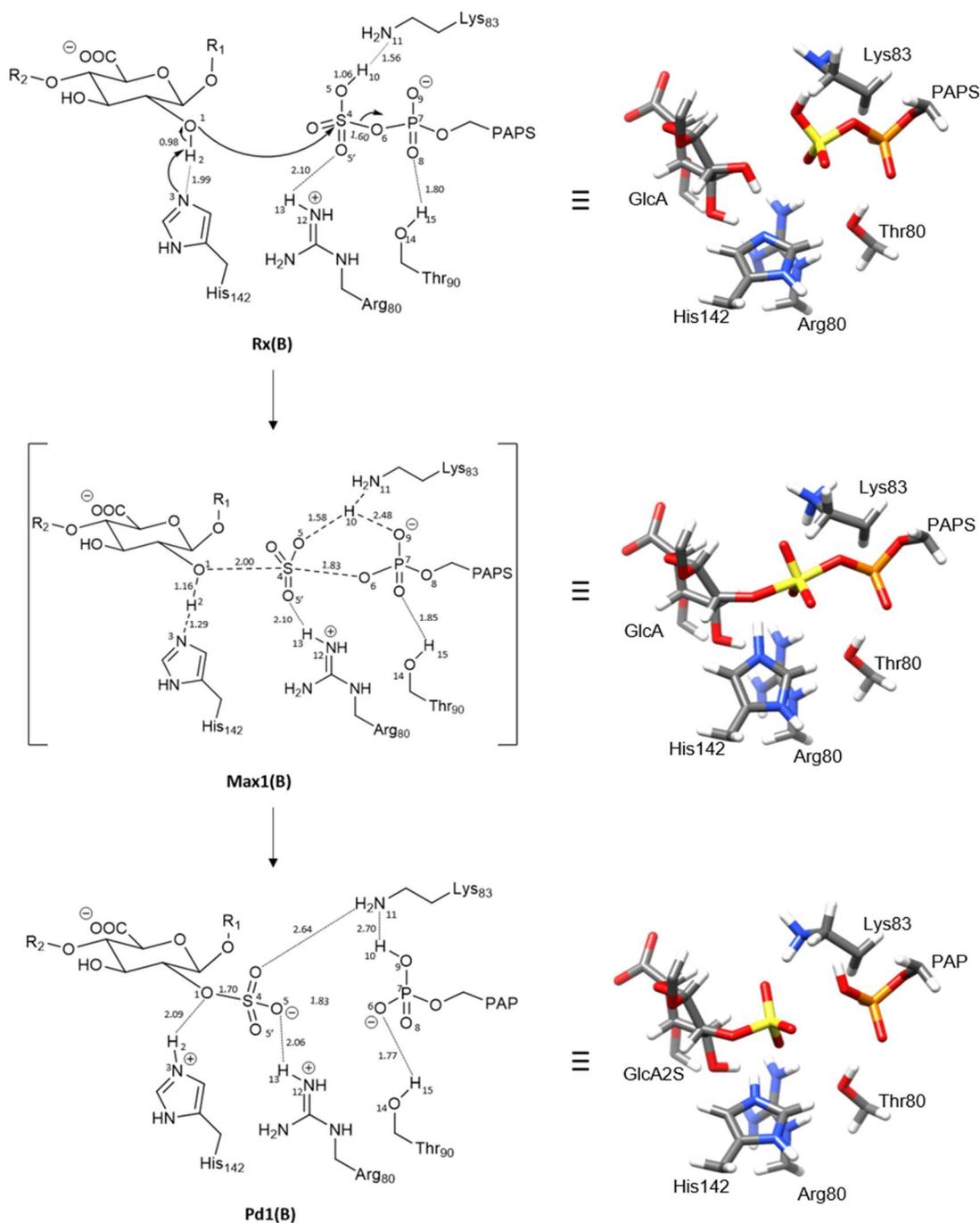
Scheme 4.2. On the left side, schematic representation of the sulfurylation of GlcA unit (bond lengths are in Å). R_1 = -GlcNs-GlcA and R_2= -GlcNS-Glca-GlcNAc-GlcA. On the right side, the corresponding three-dimensional representation of the computed points.

113

In the transition state region, both Arg80 and Lys83 can be ascribed to behave as "shuttles" of the sulfuryl group: they follow the SO_3^- transfer by preserving strong hydrogen bond with the partially negative oxygens of the sulfuryl group (H10⋅⋅O5 1.58 Å and H13⋅⋅O5' 2.10 Å). In the computed product complex **Pd1(B)**, the sulfurylation of the glucuronic acid is complete (O1-S4 1.70 Å), and the phosphate moiety of PAP is protonated by Lys83, with which it contacts a weak H-bond (N11⋅⋅⋅H10 2.70 Å). The residue Thr90, in **Pd1(B)**, stabilizes the negative charge of the HPO_4^- (H15⋅⋅O6 1.77 Å) and Arg80 engages a strong H-bond with the GlcA2S sulfuryl group (H13⋅⋅O5 2.06 Å).

Our preliminary results for **Model A** and **B** are promising in terms of energetics (see **Fig.4.4a** and **4.4b**). Concerning **Model A**, **Max1(A)** is found to lie around ~ 20 kcal mol^{-1} (**Fig.4.4a**); we believe that constraint-free transition state optimization upon this point will help the structure to relax, obtaining an activation energy barrier below 20 kcal mol^{-1}. The formation of **Pd1(A)** is a slightly endothermic process, since the product complex lies at 8.8 kcal mol^{-1} above **Rx1(A)**. This may not be alarming, since we can consider that the successive steps in the HS biosynthesis may overcome this energy lost, by pushing the overall equilibrium towards the formation of the final HS product (Le Chatelier's principle). In **Model B**, we identified maximum **Max1(B)** of the energy surface around ~36 kcal mol^{-1}: even after transition state optimizations, this value is less likely to decrease under 30 kcal mol^{-1}. This energetic value is coherent with the experimental evidence that 2OST prefers to accomplish the sulfuryl transfer on IdoA rather than on GlcA units.
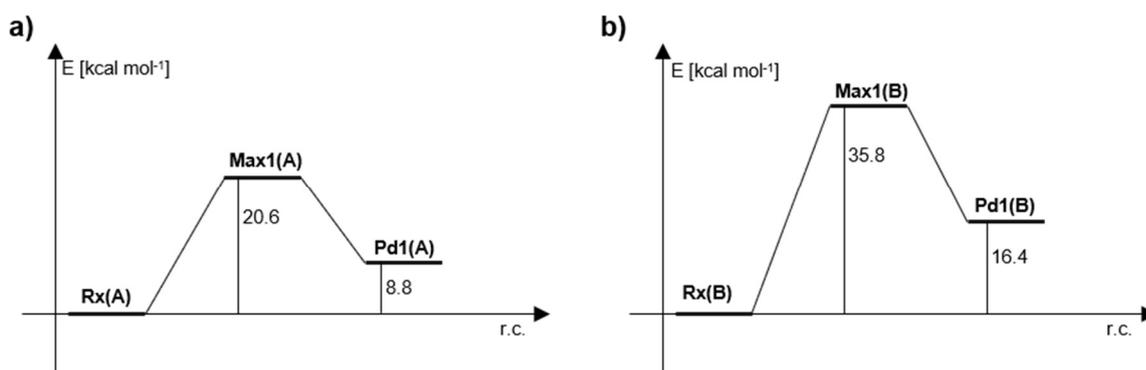
Fig.4.4 Energy profile for the enzymatic process of sulfurylation of **a)** IdoA and **b)** GlcA computed at M06-2X/6-31+G*:ff99SB.

4.4 Conclusion

The preliminary results reported in this thesis for the investigation of sulfurylation of IdoA and GlcA units in HS biosynthesis suggest that the catalytic mechanism occurs through one kinetic step, characterized by a S_N2-like character. In both cases examined here, His142 is the catalytic base, which activates the nucleophile by abstracting the proton of the 2-OH group. The region where we found

the maximum of the energy surface, corresponds to the transfer of the SO_3^- group from PAPS to the saccharide substrate and it involves the formation of a bi-pyramidal trigonal sulfate, stabilized by electrostatic interactions with the active site residues. The guanidium head of Arg80 facilitates the detachment of the SO_3^- by shuttling this incipient negative group from PAPS to the nucleophile, engaging a strong H-bond with one of the oxygens of the sulfuryl group.

The energies computed for the two mechanisms suggest that 2OST tends to catalyze the sulfurylation of iduronic acid substrates preferentially to glucuronic acids.

4.5. Perspectives

We seek to locate the critical point for the transition states starting from **Max1(A)** and **Max1(B)** structures, which seem to be very promising starting point in terms of geometry and energetics. Frequency computations will be carried out to confirm the nature of the computed points and to describe an accurate free energy surface of the reactive process. Also, we are endeavoring to study the reaction mechanisms of NDST and 6OST, in order to give the first detailed computational insights into the HS biosynthetic path.

References

[1] Dou, W.; Xu, Y.; Pagadala, V.; Pedersen, L.; Liu, J. *J. Biol. Chem.*, **2015**, *290*, 20427-20437.

[2] Gama, C. I.; Tully, S. E.; Sotogaku, N.; Clark, P. M.; Rawat, M.; Vaidehi, N.; Goddard, W. A.; Nishi, A.; Hsieh-Wilson, L. C. *Nat. Chem. Biol.*, **2006**, *2*, 467–473.

[3] Lindahl, U.; Li, J. P. *Int. Rev. Cell Mol. Biol.*, **2009**, *276*, 105−159.

[4] Sarrazin, S.; Lamanna, W. C.; Esko, J.D. *Cold Spring Harbor Perspect. Biol.*, **2011**, *3*, a004952.

[5] Bishop, J.; Schuksz, M.; Esko, J. D. *Nature*, **2007**, *446*, 1030− 1037.

[6] Kakuta, Y.; Sueyoshi, T.; Negishi, M.; Pedersen, L. *J. Biol. Chem.*, **1999**, *274*, 10673-10676.

[7] Qin, Y.; Ke, J.; Gu, X.; Fang, J.; Wang, W.; Cong, Q.; Li, J.; Tan, J.; Brunzelle, J.; Zhang, C.; Jiang, Y.; Melcher, K.; Li, J.; Xu, E.; Ding, K. *J. Biol. Chem.*, **2015**, *290*, 4620-4630.

[8] Kreuger, J.; Salmivirta, M.; Sturiale, L.; Giménez-Gallego, G.; Lindahl, U. *J. Biol. Chem.*, **2011**, *276*, 30744–30752.

[9] van Kuppevelt, T. H.; Fuller, M.; Chan, J. A.; Hopwood, J. J.; Seeberger, P. H.; Segal, R. A. *J. Biol. Chem.*, **2013**, *288*, 26275–2628.

[10] Liu. C.; Sheng, J.; Krahn, J.; Perera, L.; Xu, Y.; Hsieh, P.; Dou, W.; Liu, J.; Pedersen, L. *J. Biol. Chem.*, **2014**, *289*, 13407-13418.

[11] Xu, Y.; Moon., A.; Xu, S.; Krahn, J.; Liu, J.; Pedersen, L. *ACS Chem. Biol.*, **2017**, *12*, 73-82.

[12] Sousa, R.; Fernandes, P.; Ramos, M.; Bras, N. *Phys. Chem. Chem. Phys.*, **2016**, *18*, 11488-11496.

[13] Liu, H.; Zhang, Z.; Linhardt, R. *Nat. Prod. Rep.*, **2009**, *26*, 313-321.

[14] Liu. J.; Linhardt, R. *Nat. Prod, Rep.*, **2014**, *31*, 1676-1685.

[15] Gaussian 09, Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; *et al.* Gaussian, Inc., Wallingford CT, **2009**.

[16] a) Vreven, T.; Morokuma, K.; Farkas, Ö.; Schlegel, H. B.; Frisch, M. J. *J. Comp. Chem.* **2003**, *24*, 760-769. b) Vreven, T.; Byun, K. S.; Komaromi, I.; Dapprich, S.; Montgomery, J. A. Jr.; Morokuma, K.; Frisch, M. J. *J. Chem. Theory. Comp.* **2006**, *2*, 815-826.

[17] Zhao, Y.; Truhlar, D. G. *Theor. Chem. Acc.*, **2008**, *120,* 215-241.

Part V

Metal and Organo Catalysis

1. Introduction

1.1. Metal catalysis

Transition metal catalyzed reactions are a powerful approach in organic synthesis, both within academia and industry. The scope of organic reactions catalyzed by metal-catalysts is vast: amination, cyclization, dehydrogenation, hydrogenation, decarboxylation, decarbonylation, metathesis, oxidation reactions are just few examples of the possible catalyzed reactions.

Among all the possibilities, the catalytic processes that generate new C-C or C-heteroatom (C-X) bonds are very attractive for the synthesis of fine chemicals and building blocks. Generally, the formation of new C-C and C-X bonds is accomplished by cross-coupling reactions, which bring together two starting molecules that react with the aid of the metal centre, promoting the formation of a new covalent bond between the fragments. These reactions are usually carried out by using precious metals (*e.g.* Ir, Ru, Pd) which are a limited resource, and result in expensive catalysts. Their low abundance requires more invasive mining and refinement processes, creating economic and environmental concerns. A recent trend in catalysis is to develop new catalysts based on earth-abundant non-precious metals (*e.g.* Cu, Fe, Co).[1-3]

Some synthetic approaches, such as decarboxylative cross couplings, also attempt to achieve the sustainability goal by producing non-toxic wastes. The combination of eco-friendly and cheap metals, low-impact ligands, toxic waste-free synthetic approach and a sustainable solvent is the best choice to reduce the environmental impact of transition metal catalysis.

1.2. Organo catalysis

In organo-catalysis the species responsible for the enhancement of the reaction rate are organic molecules, that carry out the catalytic action by formation of transient covalent bonds or by creating strong non-covalent interactions as in the hydrogen bond (H-bond) catalysis.

In the case of covalent organo-catalysis, intramolecular bonds are created in intermediate species. Iminium, enamine, phosphine, *N*-heterocyclic carbene, sulphur ylide catalysis are the most common processes exploited in organic chemistry to synthetize new C-C or C-heteroatom bonds.

The main advantages of this type of catalysts are the absence of metals, their low cost compared to rare-earth metals, and their low toxicity, proclaiming organo-catalysis as one of the greener approaches in synthetic chemistry. Organo-catalysis is generally employed for the enantioselective construction of C-C and C-heteroatom (*e.g.* O, N, P, S) bonds by translating the chiral information from the catalysts to the reaction intermediates/products. The asymmetric catalysis carried out by these organic molecules enhances the possibility of governing the chiral outcomes, raising the synthetic interest towards these eco-friendly catalysts. Because of its abundance in both enantiomeric forms, and the presence of a carboxylic and an amine portion with unique nucleophilic reactivity, proline and its derivatives have been largely studied as catalysts in aldol, Mannich and Michael reactions for C-C bond formation reactions.[4]

The most common example of organo-non-covalent catalysis relies on the formation of hydrogen-bonds, which promote the reaction by stabilizing charged intermediates and transition states. In particular, the proton transfers that occur during the reaction are assisted by the formation of H-bond, where the donor can be the catalyst itself (*i.e.* proline) or water molecules at the interface between solvents. Covalent and H-bond catalyses, depending on the reaction conditions, can exert a concomitant catalytic effect, taking place simultaneously during a catalytic step.

The underlying reaction mechanisms in metal- and organo-catalyzed processes are non-trivial and for new catalyst development and rational design a profound understanding of the mechanistic aspects is essential. Modern computational chemistry, once again, comes in our aid to understand the fundaments of electronic structures and reaction mechanism in organic catalysis.

References

[1] Ludwig, J.R; Schindler, C. *Chem.*,**2017**, *2*, 313-316.
[2] Enthaler, S.; Junge, K.; Beller, M. *Angew. Chem. Int. Ed.*, **2008**, *47*, 3317-3321.
[3] Chemler, S.; Beilstein J. *Org. Chem.*, **2015**, *11*, 2252-2253.
[4] List, B. *Tetrahedron*, **2002**, *58*, 5573-5590.

2. A Mechanistic insight into the Cu(II)-catalyzed C-N and C-O Coupling reaction of Arylglyoxylic acids with Isatins. A DFT investigation

2.1. Introduction

Transition-metal catalyzed cross-coupling reactions represent a powerful and effective strategy for the formation of C-C and C-heteroatom bonds.[1,2] It allows the construction of complex organic molecules needed in the preparation of natural products and functional materials. However, the traditional approach to these reactions requires the use of highly toxic and expensive transition metals, the involved organometallic reagents are often sensitive to air and moisture and unwanted side-products are produced in stoichiometric amounts. An interesting alternative, which has attracted considerable attention during the last decade, is the decarboxylative cross-coupling reaction where stable carboxylic acids are used as a key-substrate. Carboxylic acids generate reactive organometallic intermediates by removal of CO_2 that represents the only waste material. Various examples of the employment of decarboxylative cross-coupling for C-C bond formation can be found in literature.[3] In general, decarboxylative metal catalyzed cross-coupling has been employed in the synthesis of C-heteroatom bonds.[4,5] In particular, inexpensive and widely available aryl carboxylic acids with various transition metal salts[6] have raised interest as coupling partners in decarboxylative cross coupling for C-N bond formation, a procedure widely used to obtain aromatic amines.[7]

Very recently Gogoi and coworkers[8] reported that α-oxocarboxylic acids, in the presence of Cu(OAc)2, undergo decarboxylative cross-coupling with isatin (1*H*-indole-2,3-dione), used as N-nucleophile. In particular, they employed phenylglyoxylic acid as acid reaction partner (see **Scheme 2.1**) and found that the best reaction conditions were 30 mol% of copper salt, t-amyl alcohol (t-AmOH) as solvent and a temperature of 95°C for 24 hours. They considered various isatin molecules substituted with electron-donating and electron-withdrawing groups on the aromatic ring. Interestingly, because of the nature of the two substrates (isatin and phenylglyoxylic acid), the reaction leads at the same time to the formation of a C-N and C-O bond in the coupling product 4*H*-[3,1]benzoxazin-4-one, which is representative of a series of pharmaceutically important aromatic amines.[9,10] In particular, benzoxazin-4-one derivatives are the keys scaffolds in the synthesis of quinazolinone analogues[11] and 2-substituted benzoxazin-4-ones show interesting biological activities.[12-15]
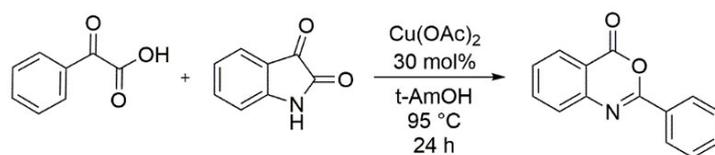
Scheme 2.1. Gogoi's reported Copper(II) catalyzed decarboxylative C-N, C-O cross coupling.
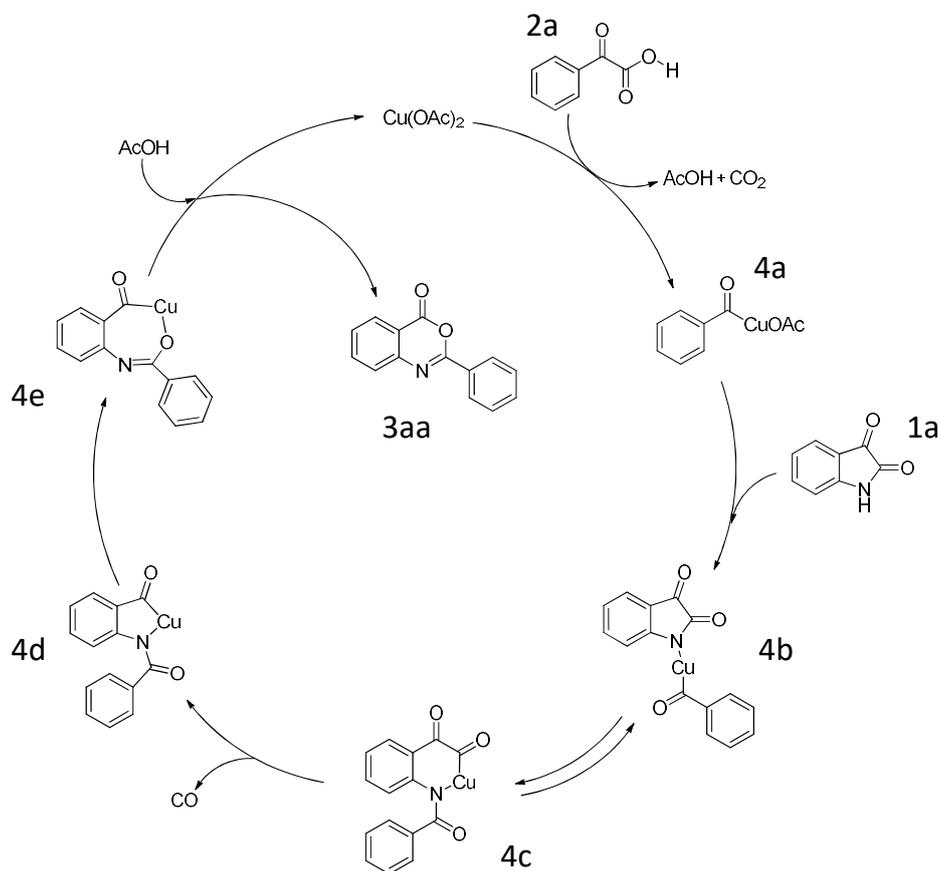
Scheme 2.2 The catalytic cycle proposed by Gogoi for the Cu(II)-catalyzed decarboxylative C-N, C-O cross coupling reaction.

Following the author guidelines, the reaction should start with the decarboxylative process where $Cu(OAc)_2$ reacts with the phenylglyoxylic acid (**2a**), releasing carbon dioxide. This leads to the benzoyl copper complex intermediate **4a** and triggers the catalytic process. After decarboxylation the catalytically active organometallic intermediate (**4a**) reacts with isatin (**1a**) forming species **4b** where copper is coordinated to the isatin nitrogen. The subsequent step is a reversible equilibrium where Cu inserts into the amide bond and forms a new species **4c** with the copper atom coordinated to nitrogen and one carbonyl atom in a six-membered cycle. A decarbonylation and a structural rearrangement form intermediate **4d**. A further structural rearrangement, corresponding to an amide-iminol tautomerism, leads to the seven-membered Cu intermediate **4e**. Finally, a reductive

elimination in the presence of air and acetic acid affords product **3aa** and the Cu(OAc)$_2$ catalyst is recovered. Importantly, the proposed catalytic cycle does not involve any change of the copper(II) oxidation state.

Important results that must be taken into account and must be consistent with this mechanistic hypothesis are the following: (a) The reaction between phenylglycoxylic acid and isatin molecule where the N-H bond is replaced by a N-Me bond is not observed and benzoic acid is the only product. This suggests that a free N-H bond in the isatin ring is essential; (b) the reaction of isatin alone or the reaction of isatin with benzoic acid does not lead to product **3aa**, suggesting that decarboxylation involving a α-oxocarboxylic acid is an essential step in the catalytic cycle; (c) insertion of different substituents at different positions of the benzene ring does not affect significantly the yield of the reaction. This supports the idea that the aromatic ring is not directly involved in the reaction.

However, even if the mechanistic hypothesis suggested by Gogoi is plausible and is not at odds with the experimental evidences collected by this author, many aspects are still unclear and need to be elucidated in detail. In particular: (i) What is the real nature of the various intermediates hypothesized in the Gogoi's catalytic cycle? Do they actually exist along the reaction pathway? In particular, what is the mechanism leading to the formation of the first complex (**4b** adduct) when isatin enters the catalytic cycle? (ii) Is the transformation following the formation of the first encounter complex (*i.e.* **4b** → **4c** in Gogoi's scheme) a reversible step? (iii) Why is the presence of a free N-H isatin bond an essential feature for the reaction to occur? (iv) What is the fate of carbon monoxide? In the **4c** → **4d** transformation does carbon monoxide immediately abandon the reaction mixture or maintain a copper coordination? (v) What is the rate determining step of the reaction? (vi) Does the initial copper oxidation state (II) change during the process? To elucidate the above points, in the present work we carry out a quantum-mechanical investigation at the DFT level of the Cu(OAc)$_2$ catalyzed decarboxylative C-N, C-O cross coupling cycle between isatin and phenylglyoxylic acid.

2.2. Computational Details

All reported DFT computations were carried out with the Gaussian09 series of programs.[16] Since the investigated model-system involves aromatic groups and metal ligands, a reliable estimate of non-covalent interactions involving π-systems and metal coordination is essential in the computation of the potential surface. Thus, we chose the M06 functional[17-20] (one of the recent

Minnesota functionals) which was demonstrated to be capable of treating transition metals, non-covalent interactions and medium-range correlation effects. The 6-31+G(d) basis set[16] was used for all atoms, except for the copper atoms, which was described with the LAN2DZ basis set.[21-23] The molecular structures of all critical points on the reaction surface (PES) were fully optimized with the gradient method available in Gaussian09 in the presence of solvent effects using the Polarizable Continuum model (PCM) method.[24] We used the dielectric constant of chlorobenzene (ε = 5.70) implemented in the Gaussian program as a standard solvent. This value is very close to the dielectric constant of 2-methyl-2-butanol (t-amyl alcohol, ε = 5.82)[25] which was the solvent actually used in the experiment and corresponding to the best reaction conditions found by Gogoi.[8]

For all critical points harmonic frequencies and Gibbs free energies were computed. Since the oxidation state of copper in the $Cu(AcO)_2$ catalyst is +2, the electron configuration of the metal is $[Ar]d^9$ and the resulting electronic state of the complex and the entire model-system is a doublet. To check the quality of the wavefunction computed at the DFT level, we examined the spin contamination after Lodwin corrections for all computed critical points. These values are reported in **Table 2.1**. In all cases the $<S^2>$ values are closed to the theoretical total spin expectation value *i.e.* 0.75 for multiplicity 2.

Table 2.1. Total spin expectation values $<S^2>$ after Lowdin corrections computed for all critical points located on the reaction surface.

	Spin Multiplicity	Computed $<S^2>$ values after Lowdin Corrections
I2	2	0.7534
TS2	2	0.7533
I3	2	0.7534
TS3	2	0.7548
I4	2	0.7536
TS4	2	0.7550
I5	2	0.7542
TS5	2	0.7538
I6	2	0.7535
TS6	2	0.7544
Pd	2	0.7548

2.3. Results and Discussion

The free energy reaction profile is depicted in **Fig.2.1**. The process can be divided into four fundamental stages: (i) decarboxylation, (ii) C-N bond formation, (iii) decarbonylation and (iv) C-O bond formation. Schematic representations of the various critical points located on the potential surface are reported in **Scheme 2.3-2.5**. A point concerning **Scheme 2.3-2.5** should be stressed: since we adopted a two-dimensional representation, several atomic distances are not realistic and appear much longer (or shorter) than in the real system.
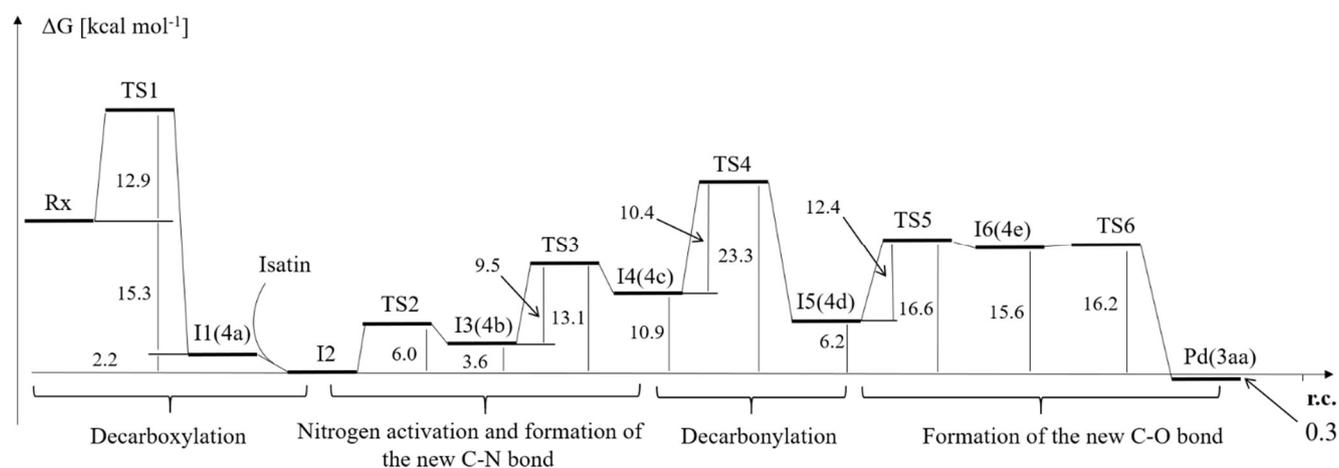
Fig.2.1 Free energy profile computed for the Cu(OAc)$_2$ catalyzed cross-coupling reaction between phenylglyoxylic acid (**2a**) and isatin (**1a**).

2.3.1 Decarboxylation. The initial decarboxylation involves the phenylglyoxylic acid (point **2a** in **Scheme 2.2**) and the catalytic complex Cu(OAc)$_2$. In the corresponding transition state **TS1** (see **Scheme 2.3**) a proton transfer occurs from the phenylglyoxylic oxygen O$_2$ to the acetate ligand oxygen O$_A$. Simultaneously the C$_3$C$_5$ bond breaks and the C$_5$ carbon enters the copper coordination sphere. The whole process is very asynchronous since, while the H$_1$-O$_A$ bond is almost completed (1.01 Å), the C$_3$C$_5$ bond is breaking (1.88 Å) and the C$_5$ carbon is still rather far from the metal (C$_5$Cu distance = 3.07 Å). The computed activation free energy for this process is 12.9 kcal mol^{-1}.

In the resulting intermediate **I1** a benzoyl moiety forms a new bond with the metal (C5-Cu = 2.00 Å). The acetic acid fragment AcOH acts as a bidentate ligand: the carbonyl group engages one copper coordination site (O$_B$-Cu 2.06 Å), and the hydroxyl hydrogen forms a rather strong H-bond with the benzoyl oxygen (H$_1$···O$_6$ distance = 1.72 Å). The process is significantly exothermic since a free energy stabilization of 15.3 kcal mol^{-1} characterizes the formation of **I1** from **Rx** (phenylglyoxylic acid + Cu(AcO)$_2$). Thus, decarboxylation should occur rather easily under the reaction conditions used in the experiment (95°C for 24 hours) triggering the catalytic cycle via the formation of **I1**, which

represents the initial active organometallic complex (**4a** in Gogoi's scheme) needed to start the cross-coupling reaction. When isatin enters the catalytic cycle and interacts with **I1**, a new adduct **I2** is formed. A complex structural rearrangement, mainly concerning the coordination of copper, characterizes the formation of **I2**. The metal now simultaneously interacts with three oxygen atoms ($Cu-O_B = 2.06$ Å, $Cu-O_E = 1.97$ Å, $Cu-O_{10} = 2.30$ Å) and the benzoyl carbon ($Cu-C_5 = 2.03$ Å). Furthermore, two strong hydrogen bonds ($N_7-H_8 \cdots O_D$ and $O_A-H_1 \cdots O_{10}$) connect the isatin molecule to the acetate and acetic acid ligands ($H_8 \cdots O_D$ and $H_1 \cdots O_{10}$ distances are 1.71 and 1.73 Å, respectively). On the whole the formation of **I2** leads to a stabilization of 2.2 kcal mol^{-1}. These results indicate that, after decarboxylation, the resulting acetic acid remains anchored to the metal and isatin (*via* hydrogen contact) rather than dissolving in solution. These interactions contribute to stabilizes the encounter complex **I2** and determine the relative arrangement of the coupling partners needed for the subsequent catalytic steps.
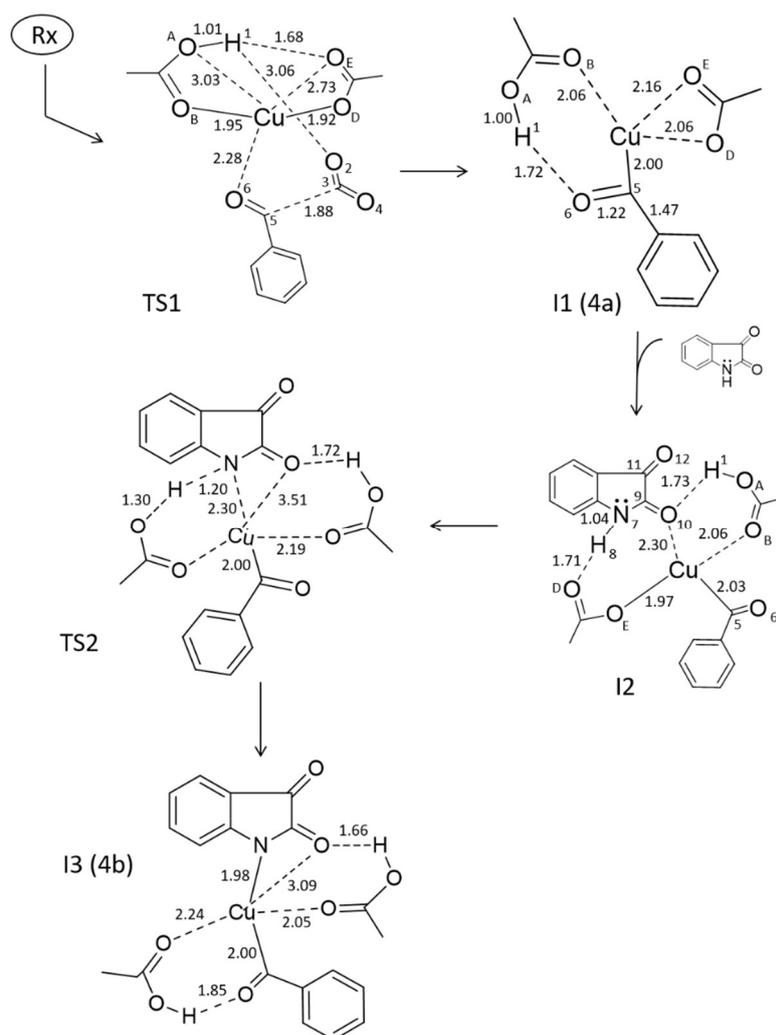
Scheme 2.3 A schematic representation of the structure of critical points **TS1, I1, I2, TS2** and **I3** (bond lengths are in Ångstroms).

2.3.2 Nitrogen activation and formation of the new C-N bond. The nucleophilic character of the isatin ring nitrogen atom in **I2** is enhanced by a proton transfer from nitrogen to oxygen O_D. Simultaneously the nitrogen attacks the metal and replaces the isatin oxygen (O_{10}) in the metal coordination sphere (see **Scheme 2.4**). In the corresponding transition state **TS2** the hydrogen is approximately half way between N_7 and O_D ($O_D\cdots H_8$ and $N_7\cdots H_8$ distances are 1.30 and 1.20 Å, respectively) and the $N_7\cdots Cu$ distance is 2.30 Å. The result is the formation of a strong coordinate bond between nitrogen and the metal (N_7-Cu = 1.98 Å) in the subsequent intermediate complex **I3**, 3.6 kcal mol^{-1} above **I2**. **I3** corresponds to the intermediate **4b** hypothesized in the Gogoi's catalytic cycle (**Scheme 2.2**). The proton transfer causes a rotation of the resulting acetic acid that forms a strong H-bond with the benzoyl fragment ($H_8\cdots O_6$ =1.85 Å) and a weakening of the O_E-Cu bond that becomes 2.24 Å. The transformation **I2 → I3** is a rather easy process characterized by a free energy barrier (**TS2**) of 6.0 kcal mol^{-1}. The formation of **I3** *via* deprotonation of the nitrogen N_7 explains why the reaction is not observed in the absence of a free N-H bond on the isatin ring. If the nitrogen cannot be deprotonated, its nucleophilic character is not strong enough to coordinate the copper by replacing the isatin oxygen O_{10}.
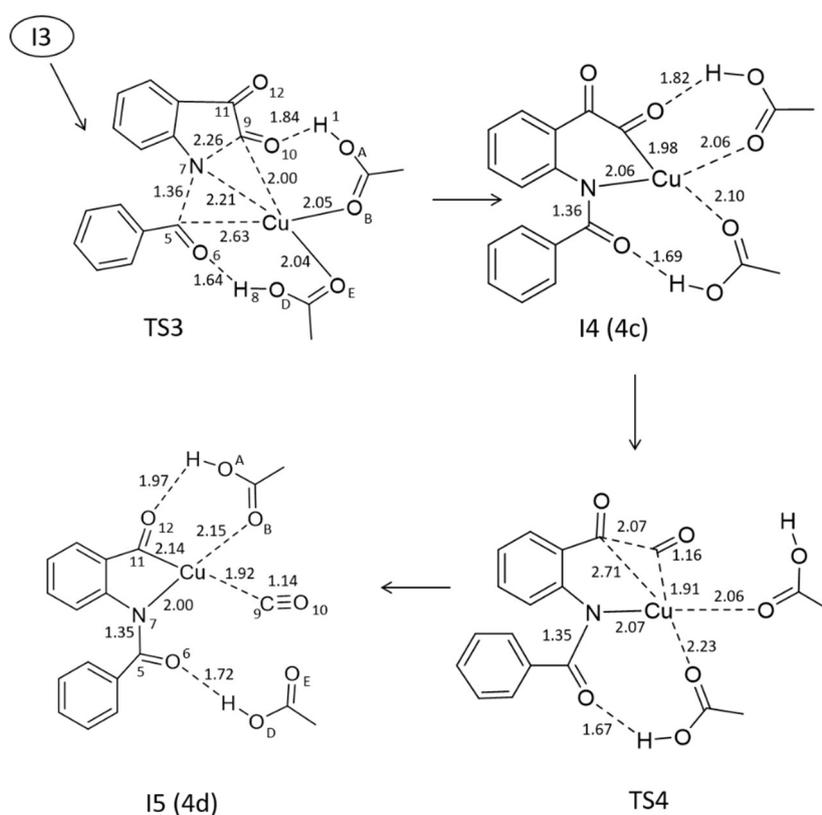
Scheme 2.4 A schematic representation of the structure of critical points **TS3**, **I4**, **TS4** and **I5** (bond lengths are in Ångstroms).

As suggested by Gogoi,[4] in the subsequent step the metal atom inserts into the amide bond and a new bond forms between nitrogen and the carbonyl carbon of the benzoyl fragment. The migration-insertion of Cu requires the overcoming of a free energy activation barrier of 9.5 kcal mol^{-1} (transition state **TS3** in **Scheme 2.4**). Thus, also this step occurs easily under the experimental conditions used by Gogoi. It is a rather late process where **TS3** has a pronounced product-like character with the new N_7C_5 and CuC_9 bonds are almost completed (1.36 and 2.00 Å, respectively).

The resulting product intermediate (**I4**) is characterized by a six-membered ring that includes the copper atom in the same initial oxidation state +2. In **I4** we can recognize the intermediate **4c** proposed by Gogoi. The copper atom has a tetrahedral structures with a complete coordination sphere (Cu-C_9 = 1.98 Å, Cu-N_7 = 2.06 Å, Cu-O_B and Cu-O_D = 2.00 Å).

Additionally, two strong H-bonds (O_{10}···H_1 = 1.82 Å and O_6···H_8 = 1.69 Å) tight intermediate **I4** in the proper conformation needed for the next catalytic step. The low barriers computed along the transformation **I2 → I3(4b) → I4(4c)** indicate that it corresponds to a reversible process as suggested by Gogoi.

2.3.3 Decarbonylation. A migration-insertion of copper into the C-C bond connecting the two adjacent carbonyl groups of isatin leads to the formation of carbon monoxide from the C_9-O_{10} carbonyl and a five-membered cycle in the resulting complex intermediate **I5**. In the corresponding transition state **TS4** the C_9C_{11} bond is breaking (C_9-C_{11} distance = 2.07 Å) and the carbonyl C_{11}-O_{12} is entering the metal coordination sphere (C_{11}-Cu distance = 2.71 Å).

During the transformation the new forming monoxide molecule (the original C_9-O_{10} carbonyl group) remains bonded to the copper atom and in intermediate **I5** it occupies one of the metal coordination sites: the Cu-C_9 distance remains almost constant on passing from **TS4** to **I5**, the corresponding bond distances being 1.91 and 1.92 Å, respectively. **TS4** is much higher in energy than transition states **TS3** and **TS2**, its free energy barrier being 23.3 kcal mol^{-1} above **I2**. However, it is lower in energy than **TS1** (30.4 kcal mol^{-1} above **I2)**, which represents the rate-determining step of the catalytic cycle. The transformation **I4 → TS4 → I5** has an intrinsic free energy barrier of 12.4 kcal mol^{-1}, very close to the barrier for decarboxylation (12.9 kcal mol^{-1}).
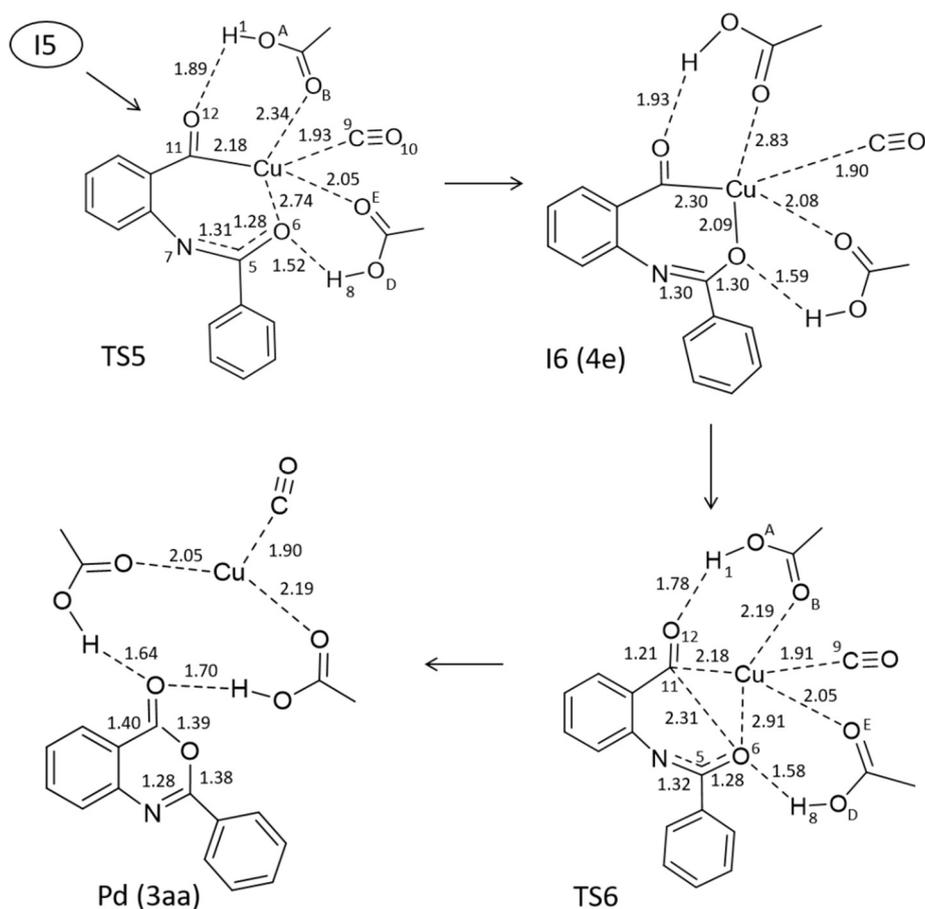
Scheme 2.5 A schematic representation of the structure of critical points **TS5**, **I6**, **TS6** and **Pd** (bond lengths are in Ångstroms).

2.3.4 Formation of new C-O bond. A further structural rearrangement occurs in the two final steps. First a cyclic electronic reorganization involving the benzoyl carbonyl group, the nitrogen and the copper atom causes the rupture of the five-membered cycle and the formation of a seven-membered cycle that includes the benzoyl moiety. In the corresponding transition state **TS5** (free energy barrier of 10.4 kcal mol^{-1}) the benzoyl oxygen enters the metal coordination sphere (O_6-Cu distance is 2.74 Å). A rotation around the N_7-C_5 bond (N_7-C_5 = 1.31 Å), the breaking of the N_7-Cu bond (N_7-Cu = 3.06 Å) and the subsequent formation of a N_7C_5 double bond increases the electron density on O_6 that intercepts the copper centre forming the intermediate **I6** (see **Scheme 2.5**), 15.6 kcal mol^{-1} above the starting complex **I2**. **I6** can be associated to species **4e** in **Scheme 2.2**.

The final product 4*H*-[3,1]benzoxazin-4-one (**Pd**) forms immediately by overcoming a very small barrier of 0.6 kcal mol^{-1}. In the corresponding transition state **TS6** oxygen O_6 attacks the electrophilic carbon C_{11} (O_6-C_{11} distance = 2.31 Å). This causes the breaking of the two coordinated bonds O_6-Cu and C_{11}-Cu (O_6···Cu and C_{11}···Cu distances are 2.91 and 2.18 Å, respectively) and the expulsion of the

copper atom from the seven-membered cycle. The two strong hydrogen contacts involving the hydrogens of the acetic acid fragments (H_1 and H_8) remain almost untouched on passing from **I6** to **TS6** (O_A-H_1···O_{12} = and O_D-H_8···O_6 distances = 1.78 and 1.58 Å, respectively) and contribute to stabilize the transition state. The negligible barrier found for the transformation **I6 → TS6 → Pd** (only 0.6 kcal mol^{-1}) has not a real physical meaning and can be considered a computational shortcoming. This finding simply indicates a very flat region of the free energy surface for the final passage from the intermediate **I6** (**4e**) to the final product **Pd** (**3aa**). Product **3aa**, which is almost degenerate to **I2** (only 0.3 kcal mol^{-1} lower in energy), shows two strong H-bonds (O_{12}···H_1 = 1.64 Å and O_{12}···H_8 = 1.70 Å) involving the carbonyl oxygen and the two acetic acid molecules that simultaneously coordinate the copper atom. The expelled monoxide molecule remains anchored to the metal. Reasonably, it will be definitely eliminated in the recovery process (not examined here) of the initial catalyst Cu(AcO)$_2$ by the oxygen of air as suggested in the last step of the catalytic cycle depicted in **Scheme 2.6.**
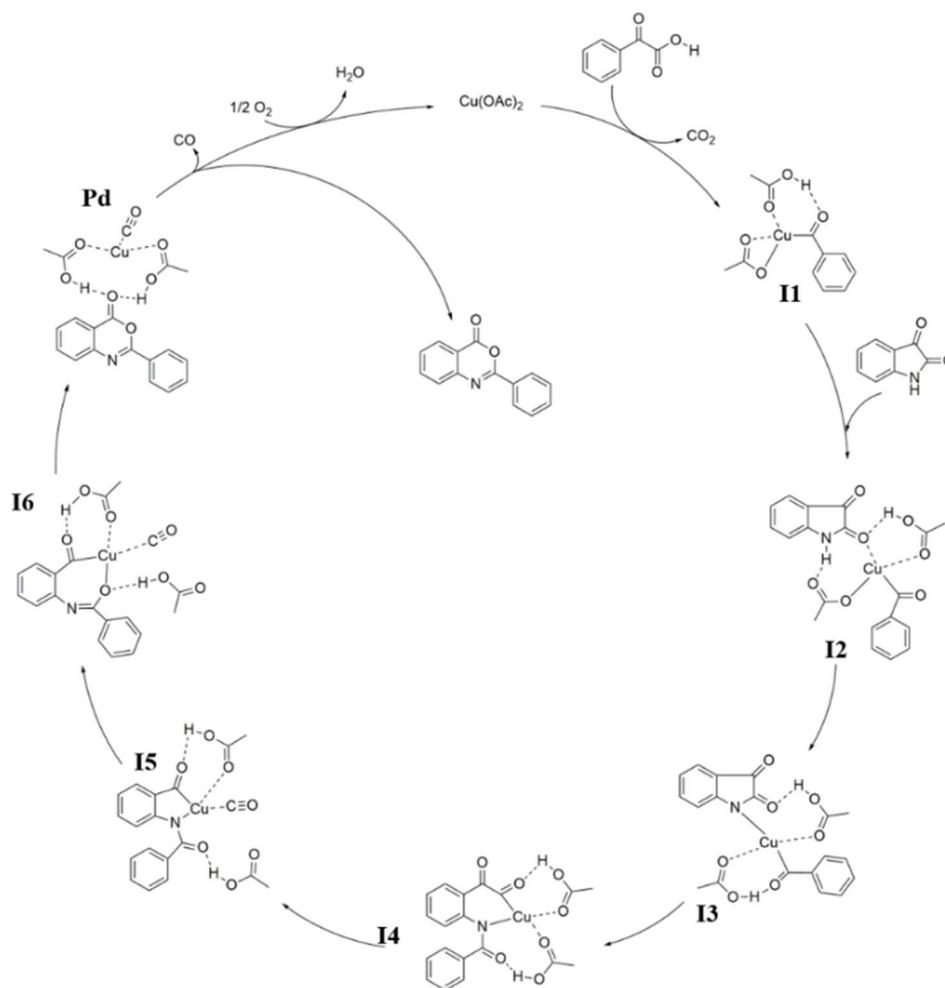
Scheme 2.6 The catalytic cycle computed in this work for the Cu(II)-catalyzed decarboxylative C-N, C-O cross coupling reaction.

2.4 Conclusions

In this work we carried out a computational investigation at the DFT level on the mechanism of the copper(II) catalyzed C-N, C-O cross coupling reaction recently reported by Gogoi. The mechanistic hypothesis proposed by this author is discussed in detail and is confirmed on the whole by our computations. The most important results of our study can be summarized as follows:

i) The initial decarboxylation is characterized by the highest in energy transition state along the catalytic cycle and represents the rate-determining of the entire process. It is an exothermic reaction (reaction free energy is -15.3 kcal mol^{-1}) with a free energy of activation of 12.9 kcal mol^{-1} and should occur rather easily under the experimental conditions used by Gogoi. It triggers the catalytic cycle leading to the formation of the initial active organometallic complex (**I1**) that reacts with isatin. **I1** corresponds to species **4a** of **Scheme 2.2**.

ii) After formation of the active organometallic species (**I1/4a**) a first encounter complex (**I2**) is formed when isatin enters the catalytic cycle. One acetate ligand plays a key-role in deprotonating the isatin nitrogen leading to the intermediate species **I3** (**4b** proposed by Gogoi) where the nitrogen is tightly bound to metal.

iii) The presence of a free N-H bond on the isatin ring is essential to allow the reaction to occur. The deprotonation of the isatin nitrogen enhances its nucleophilic character and makes possible the formation of a nitrogen-Cu bond and the formation of **I3 (4b)**. This finding explains why, when N-H is replaced by N-Me, the reaction is not observed.

iv) The insertion of the metal atom into the amide bond requires the overcoming of a moderate barrier of 9.5 kcal mol^{-1}. Thus, the two subsequent steps **I2 → I3(4b) → I4(4c)** represent on the whole a reversible transformation connecting the initial intermediate with isatin (encounter complex) and intermediate **4c**. This agrees with Gogoi's hypothesis.

v) The expelled carbon monoxide remains anchored to the copper atom in the final steps of the reaction, suggesting that it will be definitely eliminated in the recovery of the initial catalyst Cu(AcO)$_2$ during which the molecular oxygen contained in air restores the catalytic species.

vi) Copper does not change its oxidation state (II) in the course of the catalytic reaction.

vii) The mechanism proposed by Gogoi and confirmed by our computations never involves directly the benzene ring. Thus, it is consistent with the evidence showing that only slight

changes in the reaction yield are observed when substituents are inserted at different positions of the aromatic ring.

References

[1]   Meijere, A.; Dederich, F. Metal-catalyzed Cross Coupling reactions, II Edition, Wiley, 2008.

[2]   Bariwal, J.; Eycken, E. *Chem. Soc. Rev.,* **2013**, *42*, 9283-9303.

[3]   Rodriguez, N.; Goossen L. J. *Chem. Soc. Rev.,* **2011**, *40*, 5030-5048.

[4]   Shang, R.; Liu, L. *Sci. China Chem.,* **2011**, *54*, 1670–1687.

[5]   Wei, Y.; Hu, P.; Zhang, M.; Su, W. *Chem. Rev.,* **2017**, *117*, 8864–8907.

[6]   Zhang, Y.; Patel, S.; Mainolfi, N. *Chem. Sci.,* **2012**, *3*, 3196-3199.

[7]   Wang, Q.; Su, Y.; Li, L.; Huang, H. *Chem. Soc. Rev.,* **2016**, *45*, 1257-1272.

[8]   Prakash, R.; Gogoi, S. *Adv. Synth. Catal.,* **2016**, *358*, 3046–3049.

[9]   Jarvest, L.R.; Parratt, M.J.; Debouck, C.M.: Gorniak, J.; Jennings, J.; Serafinowska, H.; Strickler, J. *Bio. & Med. Chem. Letters,* **1996**, *20*, 2463-2466.

[10]   Yang, J.; Barniol-Xicota, M.; Nguyen, M.T.N.; Ticha, A.; Strisovsky, K.; Verhelst S. *Bio. & Med. Chem. Letters,* **2018**, *28*, 1423-1427.

[11]   Noolvi, M.; Patel, H.; Bhardawaj, V.; Chauhan, A. *Eur. J. Med. Chem.,* **2011**, *46*, 2327-2346.

[12]   Hedstrom, L.; Moorman, A.R.; Bodds, J.; Abeles, R.H. *Biochemistry,* **1984**, *23(8)*, 1753-1759.

[13]   Hays, S. J.; Caprathe, B.W.; Gilmore, J. L.; Amin, N.; Emmerling, M. R.; Michael, W.; Nadimpalli, R.; Nath, R.; Raser, K. J.; Stafford, D.; Watson, D.; Wang, K.; Jaen, J.C. *J. Med. Chem.,* **1998**, *41*, 1060–1067.

[14]   Kopelman, P.; Bryson, A.; Hickling, R.; Rissanen, A.; Rossner, S.; Toubro, S.; Valensi, P. *Int. J. Obes.,* **2007**, *31*, 494–499.

[15]   Neumann, U.; Schechter, N. M.; Gutschow, M. *Bioorg. Med. Chem.,* **2001**, *9*, 947–954.

[16]   Gaussian 09, Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; et al Gaussian, Inc., Wallingford CT, **2009**.

[17]   Zhao, Y.; Truhlar, D.G. *J. Phys. Chem. A,* **2004**, *108*, 6908–6918.

[18]   Zhao, Y.; Truhlar, D.G. *J. Phys. Chem. A,* **2005**, *109*, 5656–5667.

[19]   Zhao, Y.; Tishchenko, O.; Truhlar, D. G. *J. Phys. Chem. B,* **2005**, *109*, 19046–19051.

[20]   Zhao, Y.; Truhlar, D.G. *Acc. Chem. Res.,* **2008**, *41*, 157–167.

[21]   Hay, P. J.; Wadt, W. R. *J. Chem. Phys.,* **1985**, *82*, 270-283.

[22]   Wadt, W. R.; Hay, P. J. *J. Chem. Phys.,* **1984**, *82*, 284-298;

[23]   Hay, P. J.; Wadt, W. R. *J. Chem. Phys.,* **1985**, *82*, 299-310.

[24]   Tomasi, J.; Mennucci, B.; Cammi, R. *Chem. Rev.,* **2005**, *105*, 2999–3093.

[25]   Bertolini, D.; Cassettari, M.; Salvetti, G. *Mol. Phys.,* **1985**, *55*, 475–488.

3. The Hajos-Parrish-Eder-Sauer-Wiechert Reaction: New Mechanistic Details unveiled by a DFT computational investigation

3.1. Introduction

The Hajos–Parrish–Eder–Sauer–Wiechert (HPESW) reaction was the first proline-catalyzed intramolecular asymmetric aldol reaction to be discovered in the early seventies.[1,2,3] The reaction takes the name of the chemists that have developed it and represented the pioneer exploitation of the asymmetric (S)-proline aminoacid to obtain highly enantiopure products in excellent yields.

The HPESW reaction foresees the conversion of substrate **1**, known as the Hajos-Parrish diketone, into the aldol addition intermediate **(S,S)-2**, then, by the loss of a water molecule, the addition intermediate turns into the aldol condensation product **(S)-3** through the dehydration process, as depicted in **Scheme. 3.1.**
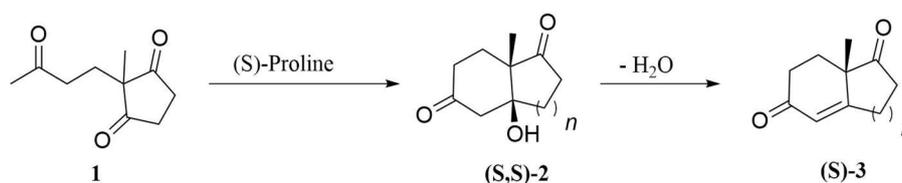
Scheme. 3.1 Schematic representation of the HPESW reaction, from the Hajos-Parrish diketone **1** to the dehydrated product **(S)-3**.

Since its description, the HPESW reaction has been exploited in medicinally relevant steroids and natural products synthesis[3], but although its importance, several and alternative mechanistic models (in particular Hajos[1], Agami[4] and Swaminathan[5] hypotheses) have been proposed during the years. The debate dealt with the number of proline molecules involved and with the mechanism by which the amino acid participates in the C-C bond formation during the aldol cyclization. A recent computational study published by Houk *et al.*[6] and other experimental works[7] established that the HPESW reaction proceeds through the covalent catalysis of one proline molecule. To prove the mono-molecular catalysis, Houk presented the computed pathway[6] for the formation of the aldol addition intermediate **(S,S)-2**, depicted in **Scheme 3.2**. From the computational results, it emerges that, starting from the separated reagents (**1a** + **(S)-4**), the condensation between the two molecules takes place in step **5**, followed by the minimum that corresponds to the zwitterionic adduct **6**. The successive minimum is the neutral adduct **7**, but the transition state between point **6** and **7** is missing. Then, the adduct **7** undergoes the elimination of a water molecule in transition state **8**, forming the iminium ion in step **9**.
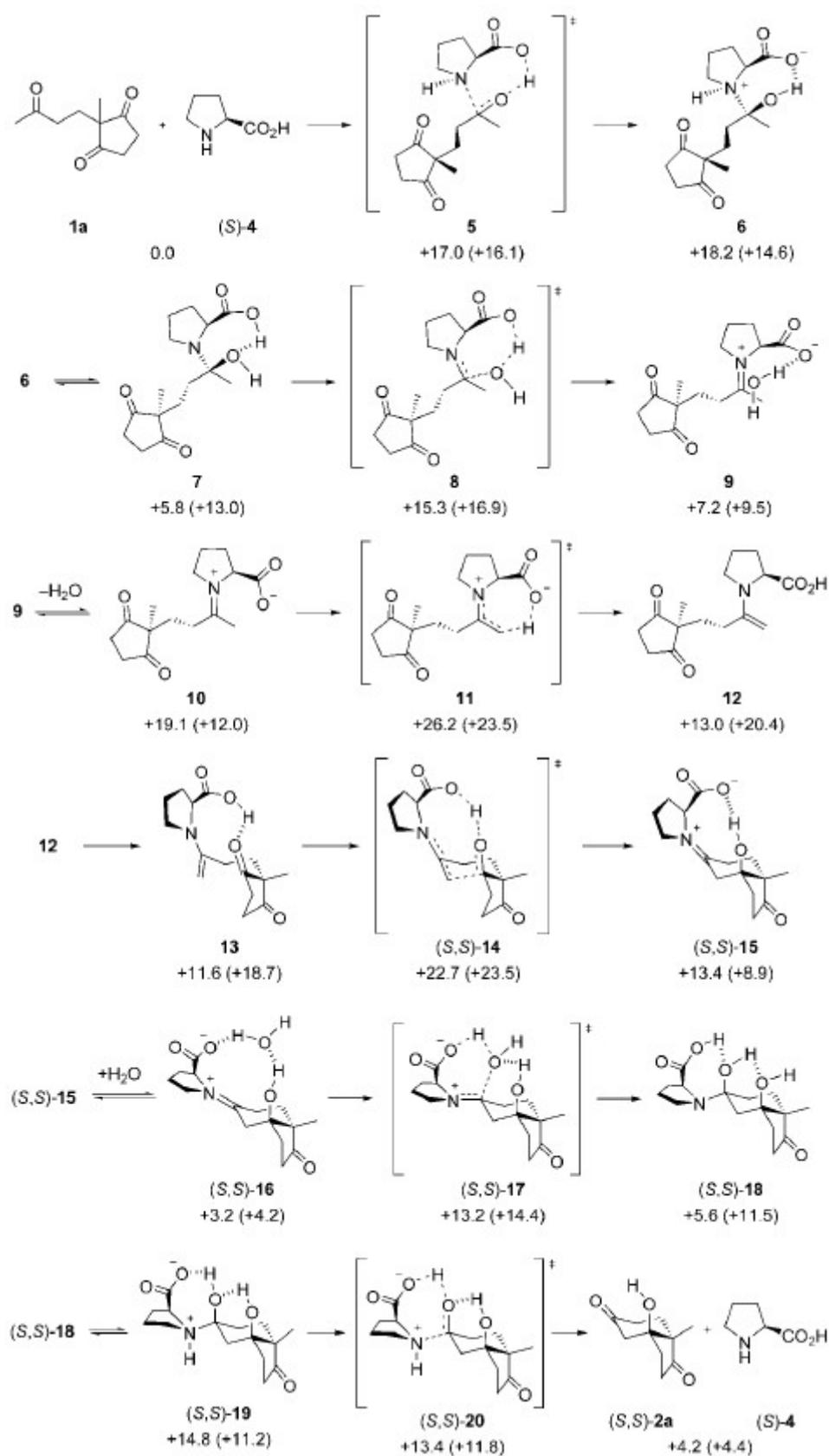
Scheme 3.2. The computed HPESW reaction, image taken from Houks's paper [6].

Proceeding, step **10** corresponds to point **9** where the water molecule was deleted. At this point, the iminium is transformed into the enamine passing through the transition state **11**, which is one of the key transformations in the HPEWS reaction. The enamine **12** undergoes a conformational conversion into **13**, but the transition state between **12** and **13** was not calculated. Then, through the action of the acid carboxylic moiety of proline, in **(S,S)-14** the C-C bond forming step takes place and then collapses into the iminium ion **(S,S)-15**. The next point **(S,S)-16** corresponds to **(S,S)-15** where the water molecule was re-inserted into the calculations. This water molecule attacks the iminium ion **(S,S)-16** in **(S,S)-17** transition state which generates the intermediate **(S,S)-18**. A proton transfer from the carboxylic acid of proline **(S,S)-18** follows, and forms **(S,S)-19**, but the connecting transition state is missing. At this point, the zwitterionic intermediate **(S,S)-19** undergoes the separation of the product **(S,S)-2a** from the proline **(S)-4**. The last point, which represents the separated products, corresponds to the aldol addition intermediate and the regenerated catalyst proline. Surprisingly, even if it has been demonstrated that the aldol condensation product **(S)-3** occurs as the major product of the HPESW reaction[7], the computational study performed by Houk neglects to compute the dehydration process, focusing the mechanistic investigation only on the first part of the reaction. A part from the lack of the computational investigation of the final dehydration step, analysing in detail Houk's work[6] we identified some aspects that are still unclear, in particular the formation of the enamine intermediate (**11**) and the C-C forming step (**(S,S)-14**) have equal energies (23.5 kcal mol^{-1}) in solvent and both represent the rate-determining steps of the reaction. Furthermore, a water molecule, which is explicitly computed in the initial and final points, has been removed from the key transformations (enamine intermediate and C-C bond formation), and its energetic contribute was added only algebraically, not considering nor its stabilizing effect or its direct participation in these transition states. Also, the activation energies ΔEs (kcal mol^{-1}) in presence of solvent (DMSO) were calculated by adding up values obtained from different levels of theory, in particular summing the energy obtained by optimization procedures in gas phase at DFT level (B3LYP/6-31G* corrected with the ZPE factor) plus the results from single point calculations in DMSO at Hartree-Fock level (HF/6-31+G**). This sum was then multiplied by a scale factor of 0.9806 (as suggested by Foresman[8]). The convoluted procedure in calculating the activation barriers in solvent and the choice of neglecting the explicit water molecule from the enamine and C-C bond transition states may have affected the energy profile, causing the two key transformations **11** and **(S,S)-14** to result into equi-energetic critical points. In support of our consideration, in 2009 an experimental paper published by Houk's group[9] suggested that the RDS is more likely to be previous to the C-C bond

formation, refuting the theoretical results obtained in 2004 by the same author. We are in the opinion that an accurate theoretical revision of the HPESW reaction would be useful in elucidating some unresolved issues and, in the present work, we aim to clarify through a meticulous computational analysis: i) which is the structure and energetics of the missing critical points corresponding to the transformation of **6→7**, **12→13** and **(S,S)-18→(S,S)-19**, and how does the dehydration process take place, ii) which is the RDS of the process, iii) which is the role of water in the HPESW reaction. To answer these questions, we carried out a detailed and accurate computational investigation.

3.2. Computational details

All reported DFT calculations were carried out with the Gaussian09[10] series of programs. We used the DFT M06-2X[11] functional and the 6-311+G** [10] basis sets for geometry optimization. As starting points, we used the geometries supplied in Houk's work.[6] The structure of the various critical points (minima and saddle points) was fully optimized using the gradient method available in the Gaussian package. Frequency calculations were carried out at the same level of theory to check the nature of critical points. We found only real frequencies for reactants, products and intermediates (minima), while all transition states (saddle-points) were characterized by one imaginary frequency. IRC computations were carried out to establish the connection between reactants and products (or intermediates). The effect of the solvent (DMSO) was taken into account using the polarizable continuum model (PCM)[12] approach as implemented in Gaussian09.[10]

3.3. Results and discussion

In this section, we discuss the results obtained by re-computing the entire pathway of the HPESW reaction from the separated reactant to the formation of the aldol condensation product **(S)-3**. The computed energy profile is reported in **Fig.3.1**.
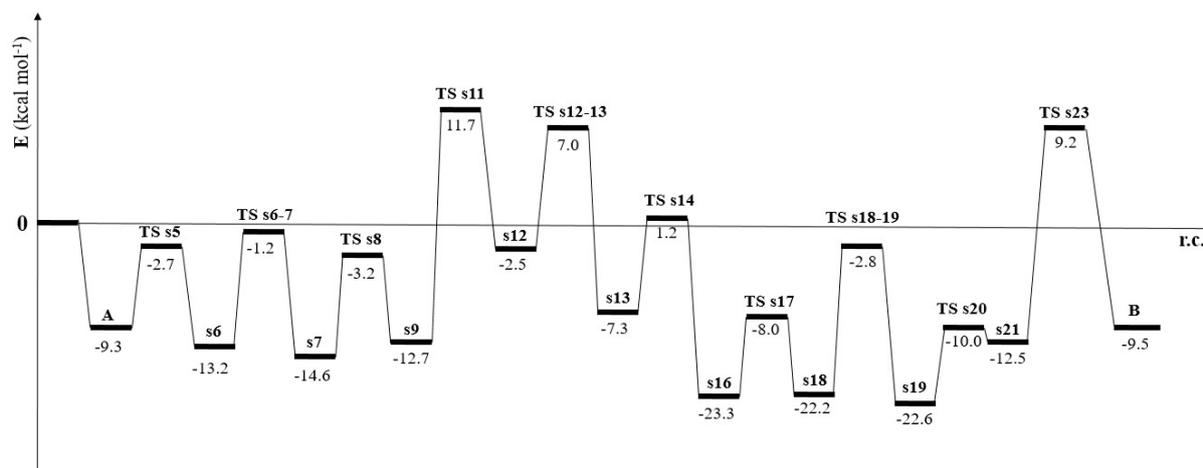
Fig.3.1 Reaction profile computed at M06-2X/6-311+G** level for the HPESW reaction, taking into account ZPE corrections and solvent effects. Energy values are in kcal mol^{-1}.

The HPESW can be divided into five fundamental stages: (i) iminium ion formation, (ii) enamine formation, (iii) C-C bond formation, (iv) aldol addition intermediate formation and (v) aldol intermediate dehydration.

3.2.1 Iminium ion formation. Starting from non-interacting proline and substrate, we computed the reactant complex **A** (**Scheme 3.3**). This critical point was missing in Houk's profile,[6] but is undeniable that, in order to react, substrate and proline need to approach each other. During this process two H-bonds are contacted between the proline and the diketone.

The H1 of the carboxylate group of proline is engaged in a strong H-bond with the carbonyl oxygen O14 of the lateral chain of substrate (1.73 Å), while the hydrogen of N-proline contacts a slightly less marked interaction with one of the two ketonic groups in the cyclopentanedione moiety of the substrate (H7··O8 2.18 Å). The interaction between the reaction partners has a substantial stabilizing effect respect to the separated molecules, which represents the asymptotic limit. The following critical point **TS s5** is the condensation between the proline and the Hajos-Parrish ketone.

This transition state corresponds to the formation of the N6---C13 bond formation and the synchronous proton transfer from the carboxyl group of proline to the carbonyl oxygen of the lateral chain of the substrate (O14 --- H1 1.49 Å). The subsequent minimum is the zwitterionic species **s6**.
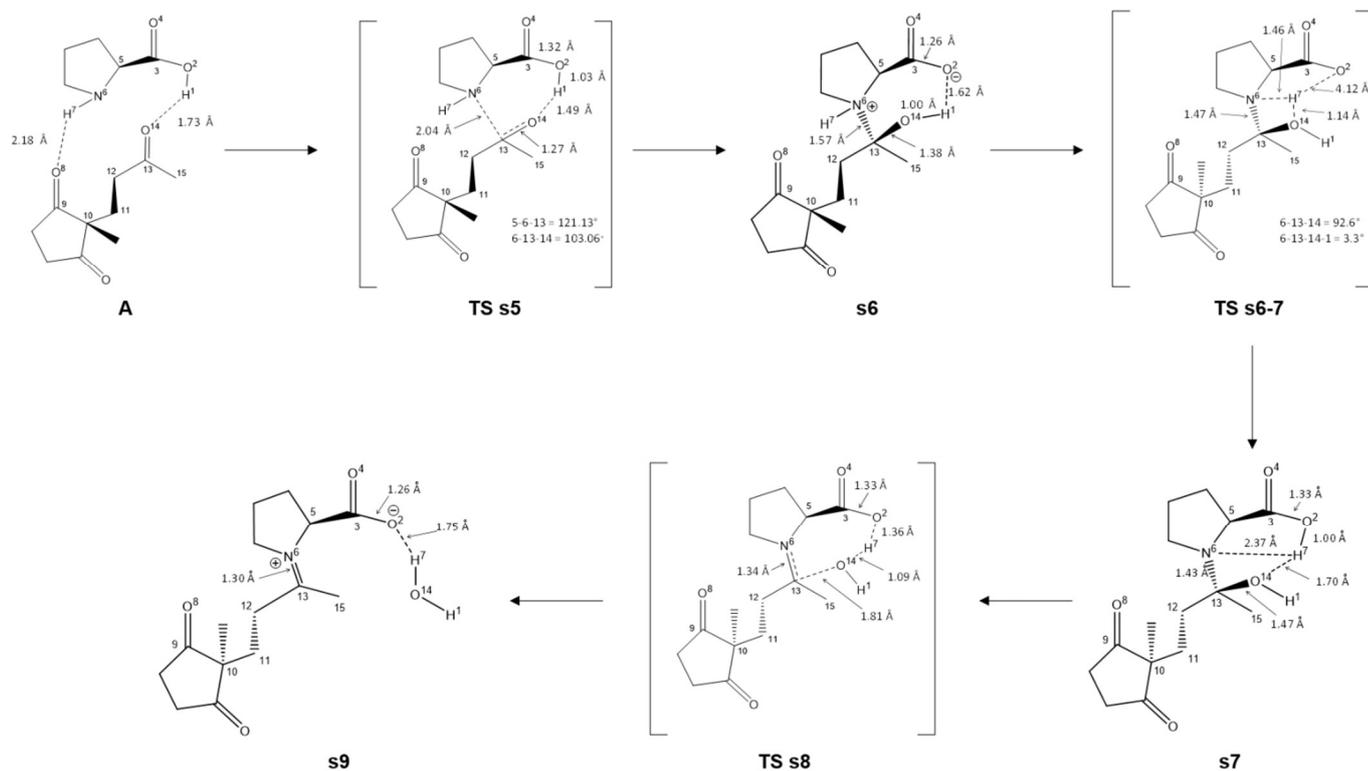
Scheme 3.3 A schematic representation of the iminium ion formation computed. Bond lengths are in Ångstroms.

Employing the geometry of **s6**, we analysed the proton transfer of H[7] from the proline nitrogen (N6) to the O2 of the carboxylic group. During this transition state (**TS s6-7**), the proton migrates from the positively charge nitrogen to the negatively charge carboxyl group of proline. This proton transfer is assisted by the hydroxyl group (O14···H7 1.14Å). We computed an energy barrier of 39.2 kcal mol[-1], suggesting that this step is not likely to occur in the experimental conditions. Nevertheless, the presence of this bottleneck is supported by Houk's hypothesis[9] and would corroborate the hypothesis of a slow stage previous to the C-C bond formation. Nevertheless, an energy barrier of almost 40 kcal mol[-1] is basically unsurmountable at room temperature.

We examined the possibility of alternative routes, where the process described by **TS s6-7** is assisted by a specie that can be easily found in the reactive environment. We analysed a plausible catalysis acted by: i) an additional zwitterionic proline (**Scheme 3.4a**) which lowered the barrier from 39.2 to 20.9 kcal mol[-1] (**TS s6-7'**) and ii) two additional water molecules (**Scheme 3.4b**).
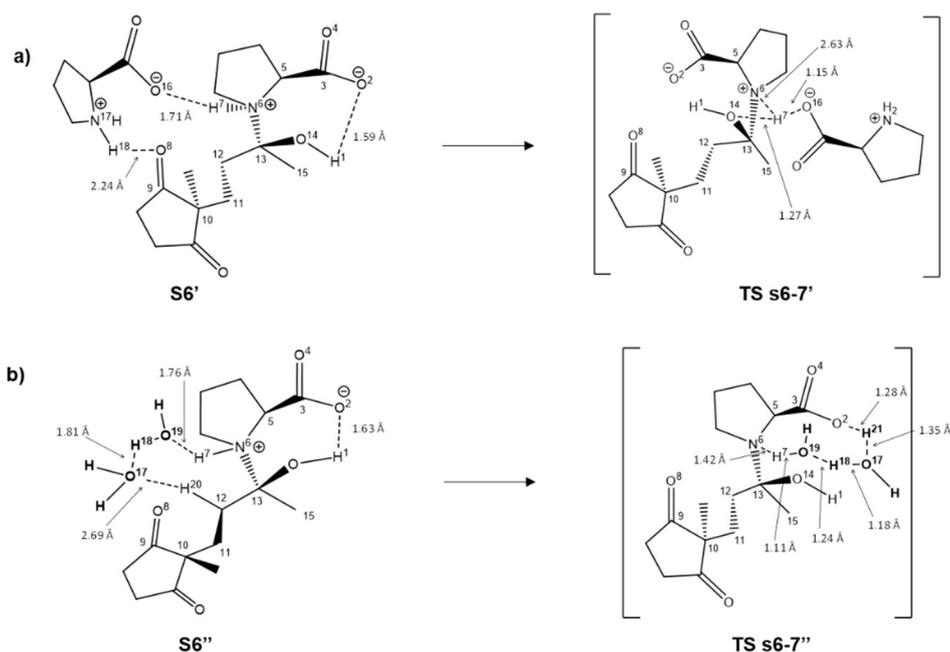
Scheme 3.4 A schematic representation of the alternative catalysis for the transition state **TS s6-7** acted by **a**) an additional zwitterionic proline, and **b**) two water molecules. Bond lengths are in Ångstroms.

Both the water and the zwitterionic proline may be present in the reaction medium, since it is possible that a percentage of proline is not involved in the catalysis and water molecules are present in the reactive solution, since the experimental solvent is not anhydrous, and water is also the sub-product of the reaction. We computed that one water molecule is not able to catalyse the reaction, since the distance between the two groups is wider than the water dimension. We computed that two water molecules can act as proton shuttles (**TS s6-7''**), facilitating the proton transfer and lowering the activation barrier to 12.0 kcal mol^{-1}. Since the presence of free water molecules is reliable, we can suggest that the **s6**→**s7** transformation is unlikely to be the reaction bottleneck and that the activation barrier of this step is 12.0 kcal mol^{-1} rather than 39.2 kcal mol^{-1}.

Moving on, **s7** point, which corresponds to the non-zwitterionic form of **s6**, connects to **TS s7-8** which corresponds to formation of the iminium ion (N6---C13 1.34 Å) and the expulsion of the water molecule (C13---O14 1.81 Å). In the computed iminium ion intermediate **s9**, the expelled H$_2$O interacts with the carboxyl group of proline through a pronounced H-bond (O2$^{...}$ H7 1.75 Å).

3.3.2. Enamine formation. From the iminium ion **s9**, **TS s11** corresponds to the iminium-enamine tautomerism which is governed by the proton transfer from the terminal methyl group to the carboxyl group of proline (O2 --- H16 1.23Å), with the incipient formation of the C-C double bond (C13 --- C15 1.41Å). The explicit water molecule does not participate in the process but interacts with the transient complex (see **Scheme 3.5**).
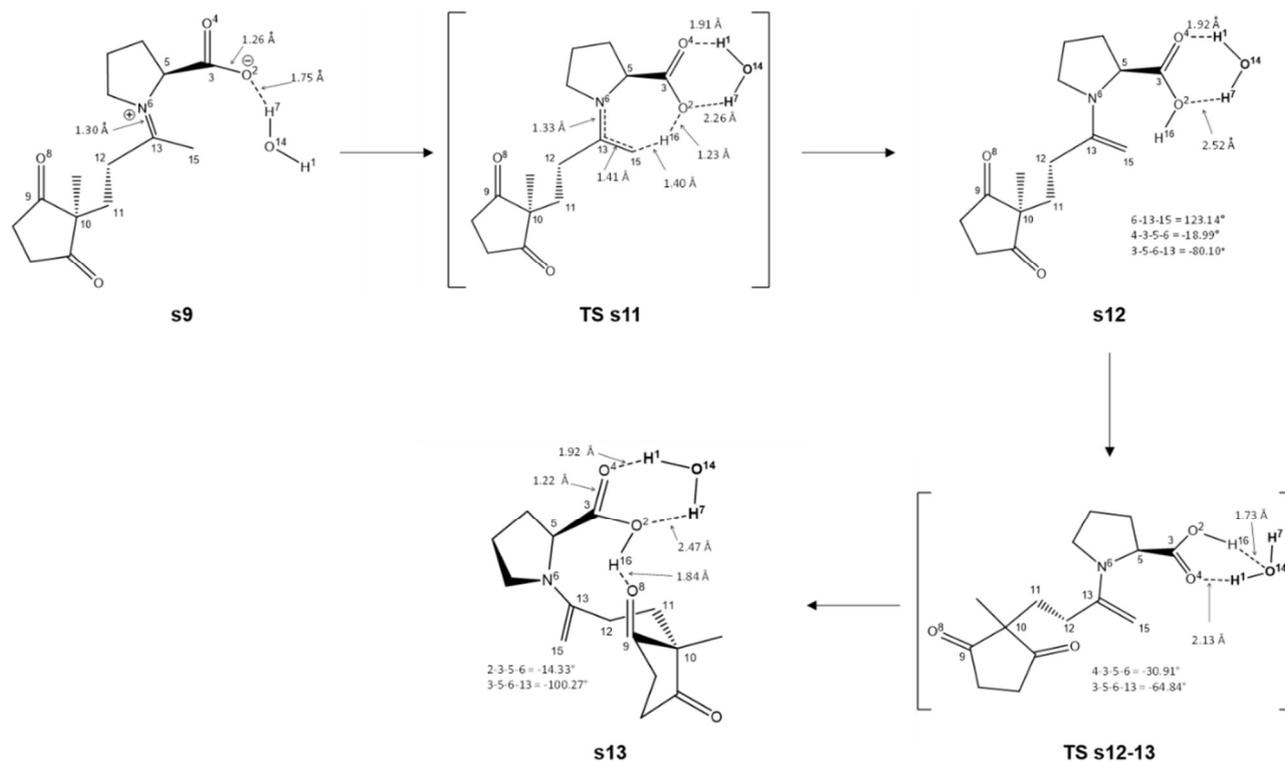
Scheme 3.5 A schematic representation of the enamine formation computed. Bond lengths are in Ångstroms.

We analysed the water participation in the proton transfer, but the result (intrinsic activation barrier 24.4 kcal mol^{-1} and a distorted geometry) suggests that, if the water molecule participates in the process, the carboxyl group of proline would undergo a sturdy deformation, due to the steric hindrance of the water molecule located halfway between the methyl group and the carboxyl group of the proline. So, we support the hypothesis that the water molecule only stabilizes this transformation by acting as a H-bond catalyst. The protonated carboxyl group of the enamine **s12** interacts with the water molecule through two hydrogen bonds (O4$\cdots$H1 1.92Å and O2$\cdots$H7 2.52 Å). The species **s12** and **s13** are conformers, and **TS s12-13** connects the two points by a conformational transition state which is described by the rotation along the N6-C13-C12-C11 dihedral angle. The water molecule, again, only stabilizes the structure, and does not participate in the catalytic process. The enamine double bond in **s13** is now located in the pro-reactive position for the next nucleophilic attack on the carbonyl group of the ketone.

3.3.3. C-C bond formation. In **TS s14** the C-C bond formation process is taking place (C15---C9 2.11 Å), triggered by the shortening of the N-C bond of the enamine (N6---C13 1.34 Å), as depicted in **Scheme 3.6**. Together with the formation of a new C-C bond, the iminium ion is restored, and the bicycle are also generated (**s16**). The activation barrier computed for this transformation is 8.5 kcal

mol^{-1} above **s13** and 1.2 kcal mol-1 above the asymptotic limit, suggesting that the C-C bond formation is not the rate determining step of the HPESW reaction.
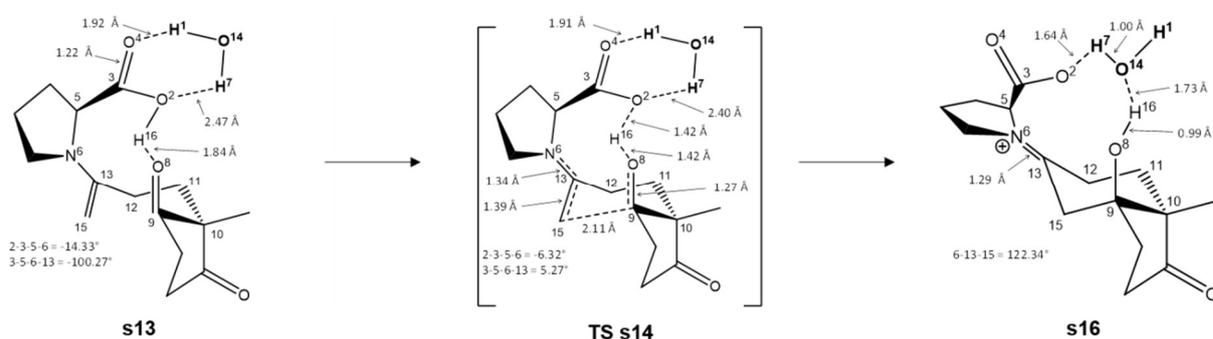
Scheme 3.6 A schematic representation of the C-C bond formation computed.

3.3.4 Aldol intermediate formation. The highly electrophilic character of C13 summons the nucleophilic attack of the water molecule, computed as **TS s17**. The water binds the carbon of the iminium ion (O14---C13 1.92 Å), while the carboxylate group of proline deprotonates the nucleophile (O2---H7 1.21 Å).
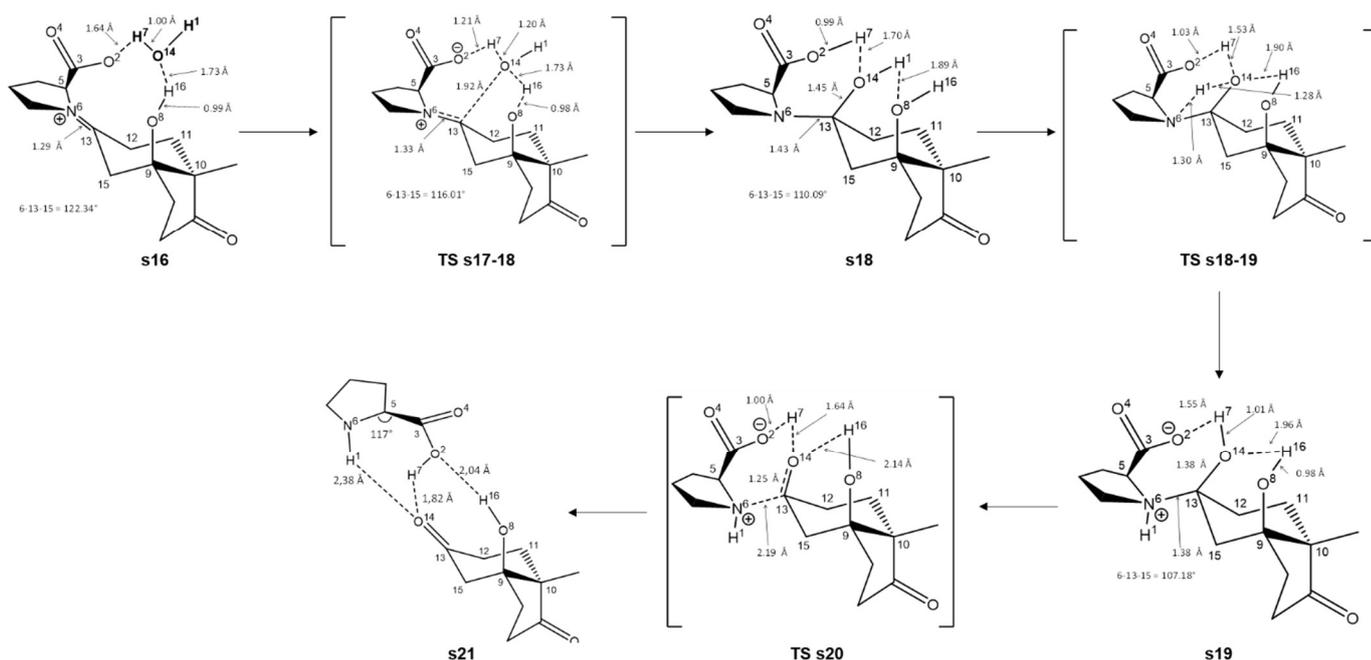
Scheme 3.7 A schematic representation of the aldol intermediate formation computed. Bond lengths are in Ångstroms.

Collapsing from **TS s17**, we computed specie **s18**, where the carboxylic group of proline is protonated, and the neutral amine is restored as schematically reported in **Scheme 3.7**. The next step corresponds to an intramolecular protonation of the nitrogen by the carboxylic acid of proline (**TS**

s18-19). This double proton transfer is located at 16.1 kcal mol-1, which correspond to an overall barrier of 31.1 kcal mol^{-1}.

This barrier results to be too high and we examined the possibility of a catalysis acted by two water molecules (**TS s18-19'**), finding an appreciable stabilizing effect that lowers the barrier from 31.1 to 19.4 kcal mol^{-1} (see **Scheme 3.8**).
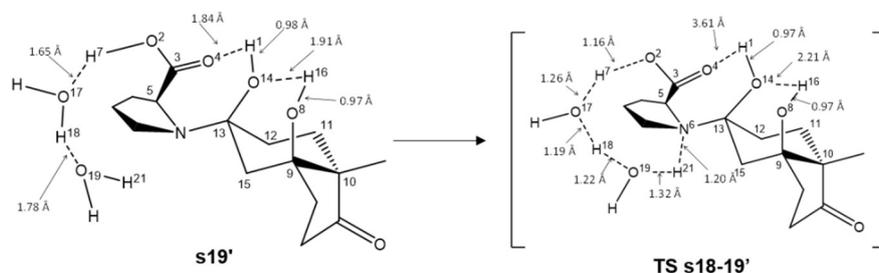
Scheme 3.8 A schematic representation of the two water molecules catalysis in the nitrogen protonation. Bond lengths are in Ångstroms.

Once the proton transfers are completed, the specie **s19** is generated. The aldol addition intermediate of is generated in **TS s20**, with the breaking of the N6---C13 bond (2.19 Å) followed by the expulsion of a proline molecule. The complex (**s21**) comprises the protonated proline that interacts with the bicycle through three H-bonds between O2···H16 (2.04 Å), O14 ··· H7 (1.82 Å), O14 ··· H1 (2.38 Å).

3.3.5 Aldol intermediate dehydration. Once the aldol intermediate is generated, the dehydration takes place (**Scheme 3.9**). We computed the dehydration process (**TS s23**) where the expulsion of one water molecule (C9 ---O8 1.81 Å) occurs rather early respect to the proton abstraction by the nitrogen of the proline (N6---H22 1.03 Å, C15---H22 3.07). In **s24**, which corresponds to the final dehydrated complex product of the HPESW reaction, we obtained the α-β unsaturated bicycle that interacts with the zwitterionic proline interact, by a strong hydrogen bond (O14 ··· H1 1.71 Å). The last transformation has an activation energy of 21.7 kcal mol^{-1}, which can be easily overcome at the experimental conditions.
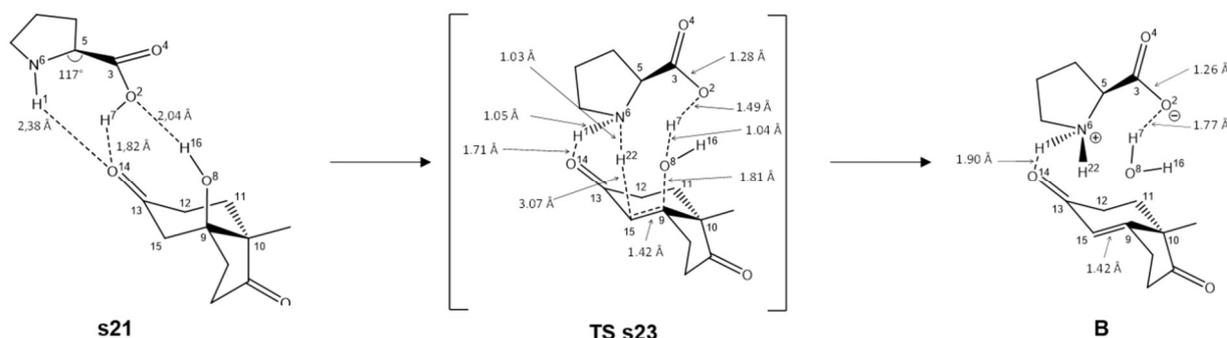
Scheme 3.9 A schematic representation of the dehydration process computed. Bond lengths are in Ångstroms

3.4. Conclusions

In this work we employed a computational approach based on the DFT theory to examine in detail the complete profile of the HPESW reactions and to disclose some issues that, in our opinion, were still open.

In particular, we determined:

i)   the structure and energetics of the critical points that were missing from Houk's investigations, *i.e.* the reactant complex (**A**), the carbinolamine formation (**TS s6-7**), the conformational TS of the enamine specie (**TS s12-13**), a crucial proton transfer in the aldol addition intermediate (**TS s18-19**), the product complex (**s21**) and the dehydration process (**TS s23**) that generates the α-β unsaturated bicycle;

ii)   That the enamine formation can be considered the rate determining step of the overall HPESW and that the C-C bond formation has an easily surmountable activation energy;

iii) the water molecules that are present in the reaction environment can exert a hydrogen bond catalysis or participate directly to the many proton transfers that characterize this reaction, lowering the activation energies of few transformations.

Also, our computational approach strongly stabilizes the reaction profile compared to the one computed by Houk. Even if the intrinsic barriers do not change dramatically, the barrier of the slow transformation are strongly decreased, suggesting that the reaction takes place easily at room temperature. We are in the opinion that these results give an interesting and detailed insight into the reaction, confirming the role of proline and water molecules in the asymmetric catalysis of α-β-unsaturated diketones.

References

[1]   Hajos, Z. G.; Parrish, D. R. *J. Org. Chem.*, **1974**, *39*, 1612–1615.
[2]   Eder, U.; Sauer, G.; Wiechert, R. *Angew. Chemie Int. Ed. English,* **1971**, *10*, 496–497.
[3]   Danishefsky, S. J.; Masters, J.J.; Young, W.; Link, J.; Snyder, L.; Magee, T.; Jung, D. *et al. J. Am. Chem. Soc.,* **1996**, *118*, 2843–2859.
[4]   Puchot, C.; Samuel, O.; Dunach, E.; Zhao, S.; Agami, C.; Kagan, H.B. *J. Am. Chem. Soc.,* **1986**, *108*, 2353–2357.
[5]   Rajagopal, D.; Moni, M.; Subramanian, S.; Swaminathan, S. *Tetrahedron: Asymmetry*, **1999**, *10*, 1631–1634.
[6]   Clemente, F. R.; Houk, K. N. *Angew. Chemie Int. Ed.,* **2004**, *43*, 5766–5768.
[7]   List, B.; Hoang, L.; Martin, H. J. *Proc. Natl. Acad. Sci. U. S. A.,* **2004**, *101*, 5839–5842.
[8]   Foresman, J.B.; Frisch, A. *Gaussian, Inc.* **2013,** 335.
[9]   Zhu, H.; Clemente, F. R.; Houk, K. N.; Meyer, M. P. *J. Am. Chem. Soc.,* **2009**, *131*, 1632–1633.

[10] Gaussian 09, Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; et al Gaussian, Inc., Wallingford CT, **2009**.

[11] Zhao, Y.; Truhlar, D. G. *Theor. Chem. Acc.*, **2008**, *120*, 215-241.

[12] Tomasi, J.; Mennucci, B.; Cammi, R. *Chem. Rev.*, **2005**, *105*, 2999-3093.

4. The reaction pathway of cellulose pyrolysis to a multifunctional chiral building block. The role of water unveiled by a DFT computational investigation[a]

4.1 Introduction

Cellulose is the most abundant biopolymer on Earth occurring ubiquitously in all vascular plants, but also found in some algae, bacteria and even in animals (Urochordates).[1] Due to its abundance in inedible biomasses, agro-industrial wastes and forestry residues, cellulose is an ideal candidate in the route towards a carbohydrate bio-based economy.[2] However, the exploitation of cellulose in bio-refinery is facing challenging obstacles caused by its structure and the way it is combined with lignin and hemicellulose in the extracellular matrix of plant cells. Cellulose is a polysaccharide made up of d-glucose units connected by b-1-4 glycosidic bonds that allow the formation of crystalline microfibrils. The crystalline features of cellulose and the shielding action of lignin complicate the biological and chemical transformation of cellulose into simple sugars and derived biofuels and chemicals (*e.g.* biomethane, bioethanol).[3-5] As an alternative to bio/chemical conversions, pyrolysis breaks apart plant biopolymers into lower molecular weight compounds by the action of thermal energy without using chemical reagents. This phenomenon places the process in the green chemistry context. However, the apparent simplicity of the process is counteracted by the complexity of the reactions involved in the thermal degradation of biomass that lead to the formation of hundreds of compounds.[6,7] Interestingly, in the case of cellulose, pyrolysis produces unique chiral compounds (collectively termed as anhydroglucoses), which are not directly obtainable by any other means. These saccharides possess a defined stereochemistry and can be effectively used as chiral building blocks in organic synthesis. Among the various anhydrosugars, levoglucosan is one of the 25 most preferred chemicals from carbohydrate biomass in terms of atom economy and high functionality.[8] Levoglucosan, which forms as a primary pyrolysis product of cellulose, received most of the attention because of its potentially high yields[9,10] and is suitable as a starting material for a variety of synthetic compounds as well as a fermentation substrate for biofuel.[11] The hydroxylactone species (1*R*,5*S*)-1-hydroxy-3,6-dioxabicyclo[3.2.1]octan-2-one, usually denoted as LAC, is also raising particular interest. This is due to the presence of a lactone and hydroxyl functionality, a tetrahydrofuran ring and two chiral centres (**4, Scheme 4.1**) and the easy conversion of the lactone unit into useful functional groups without affecting the stereochemistry.[12] The importance of LAC

as a versatile scaffold towards novel substances was confirmed by recent studies. LAC was successfully employed as a monomer in the synthesis of biodegradable co-polyesters, specifically with the aim of preparing a poly(l-lactide) functionalised with hydroxy groups.[13] Also, LAC was used as starting material for the efficient synthesis of a novel branched tetrahydrofurane amino acid. This, being an isoster of the glycine-alanine dipeptide, could, in turn, be a useful intermediate to peptidomimetics.[14] Notably, the structure of this LAC-derived amino acid resembles closely that of (+)-muscarine, a fungi toxin widely investigated in pharmacology because of its activity in profound effect on parasympathetic nervous system. On this ground, a novel class of muscarine analogues was synthesised starting from LAC and their biological activity resulted similar to that of muscarine in terms of the affinity towards human cloned muscarinic receptors.[15]

The identification of LAC was firstly reported by Furneaux and co-workers after pyrolysis of cellulose in the presence of chlorides of Fe, Al, Zn, Ni, Co Mn, Sn.[16] This study remained silent for almost twenty years, when the enhanced production of LAC from the catalytic pyrolysis of cellulose was confirmed in a variety of conditions.[17-22] A few years ago it was reported that zeolite catalysts reduce the yields of anhydrosugars, while all nanopowder oxides, with the exception of silicon oxide, provide higher yields of LAC.[17]
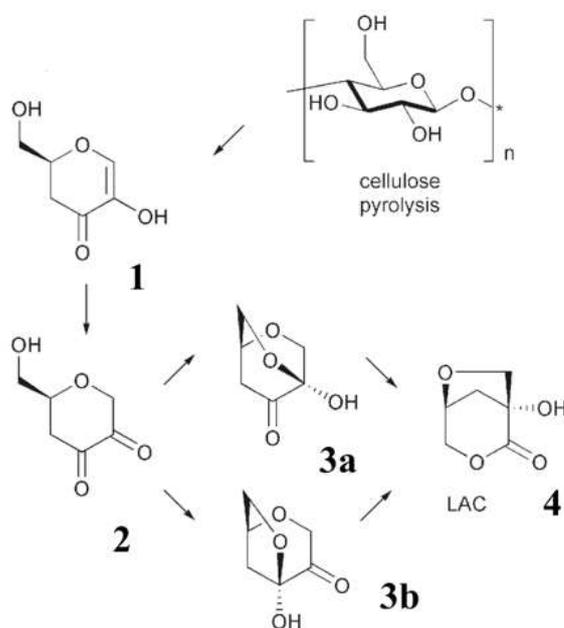
Scheme 4.1. Mechanism of LAC formation during cellulose pyrolysis.

The importance of nanopowder characteristics was confirmed by Eibner *et al.* who observed that metal-based nanoparticles, especially Zn and Co, promoted the formation of LAC over levoglucosan in the pyrolysis of eucalyptus at 500 °C.[18] Mesoporous phases of the MCM-41 type doped with Sn

or other metals were found to enhance the yields of LAC.[19-21] The clay montmorillonite K10 was found to catalyze the production of LAC. Even if the yields are lower with respect to Sn-MCM-41 and nanopowder aluminium titanate, montomorillonite has the advantage of being more economical and environmentally sustainable.[21,22] However, the most substantial yields of LAC in the pyrolysis of cellulose did not reach 10% (8.6 ± 0.7 with aluminium titanate nanoparticles at 350 °C).[12] Yields of 7.6 ± 0.1 with Sn-MCM-41 and 4.8 ± 0.1 with montmorillonite were found at 500 °C,[21] but only small amounts of LAC were obtained in the absence of catalyst (0.4% ± 0.2 and 1.2% ± 0.2 at 350 and 500 °C, respectively).[12] Since the pioneering article by Furneaux and coworkers,[16] no fundamental studies have been published on the mechanism of LAC formation. These authors proposed that LAC forms after rearrangement of the pyrolysis product 1 following the transformation shown in **Scheme 4.1**.[16] **1** is the enol tautomer of the 2,3-diketone **2**, which is hypothesized to undergo an internal rearrangement to form intermediates **3a** and **3b**. A further rearrangement of these two intermediates would lead to LAC (**4**). However, the energetics behind this process remains unknown and, as far as we know, no theoretical investigations on its mechanism are available in literature. Very recently, only computational studies at the Density Functional Theory (DFT) level on the mechanism of formation of the anhydrosugars levoglucosan and levoglucosenone have appeared.[23-29] Since higher yields of LAC are desirable to place conveniently this chemical platform into biorefinery streams, it becomes crucial to understand in detail the mechanism of this reaction. In the present work, we start our mechanistic investigation by examining the mechanism of the non-catalyzed reaction yielding LAC and we use a DFT approach to investigate the corresponding potential surface. Even if in the absence of catalyst only small amounts of LAC are obtained, we believe it is crucial to identify the key species involved in the process, to estimate the energy barriers along the reaction pathway and elucidate the nature and role of the forces that drive the reaction. Recently several examples in literature indicate DFT as a suitable approach to investigate a variety of reactivity problems relevant to sustainability in chemistry.[23-37] Our model-system is the starting species **1** (see **Scheme 5.1**), *i.e.* a molecule formed by two eliminations from a d-glucopyranosyl residue of the cellulose chain, namely (2*S*)-5-Hydroxy-2-(hydroxymethyl)-2,3-dihydro-4H-pyran-4-one (ascopyrone P).[16] We examine in detail the mechanism leading to the formation of the 2,3-diketone **2** and that of the two hemiketal intermediates, **3a** and **3b**, that eventually form LAC (**4**).

4.2. Computational Details

All DFT and MP4 calculations were carried out with the Gaussian09 B[38] series of programs. We used the DFT M06-2X[39] functional and the 6-31+G** [38] basis sets. The structure of the various critical points (minima and saddle points) was fully optimized using the gradient method available in the Gaussian package. Frequency calculations were carried out at the same level of theory to check the nature of critical points.

We found only real frequencies for reactants, products and intermediates (minima), while all transition states (saddle-points) were characterized by one imaginary frequency. IRC computations were carried out to establish the connection between reactants and products (or intermediates). Rate constants, k, were calculated using Eyring theory as

$$ k = \frac{k_B T}{h} \frac{Q_{TS}}{Q_{React}} exp\left(-\frac{\Delta E}{k_B T}\right) $$

where k_B is Boltzmann constant, T is the temperature in K, Q_{TS} is the partition function of the transition state, Q_{React} is the partition function of the reactant, and ΔE is the energy difference between transition state and reactants at 0 K.

5.3. Results and Discussion

In this section we discuss the results obtained in the investigation of the singlet potential energy surface for the transformation reported in **Scheme 4.1**.
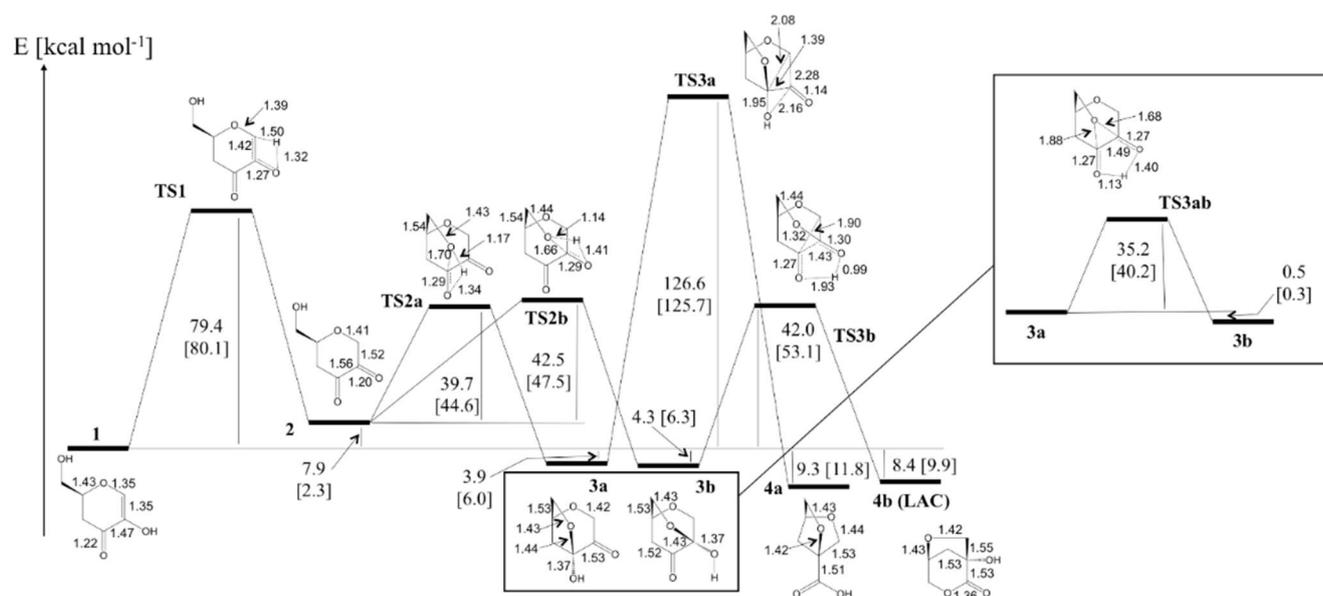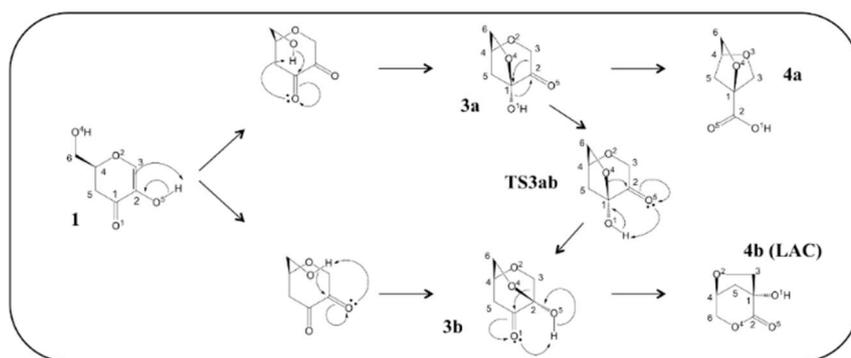
Fig.4.1. Energy profile computed for the reaction reported in **Scheme 4.1**. The model-system is the enol **1**. Values in brackets were obtained at the MP4(SDQ) computational level. Energy values (kcal mol⁻¹) are relative to **1**.

The computed energy profile is displayed in **Fig.4.1** with a schematic representation of the structure of the various critical points. At the bottom of the figure the entire process is represented by means of canonical Lewis structures in which, for the sake of clarity, we numbered the various atoms involved in the internal rearrangements.

The first step of the process is a keto-enol tautomerization, which requires a rather large activation barrier (79.4 kcal mol⁻¹). The starting enol form **1** is slightly more stable (7.9 kcal mol⁻¹) than the keto form **2**. Reasonably, this is due to the presence of a stabilizing hydrogen bond between the carbonyl oxygen O1 and the adjacent hydroxyl group (O1··H-O5 distance is 2.16 Å). In the keto form this stabilizing interaction is replaced by the repulsion between the oxygens of the two vicinal carbonyl groups. Two reaction channels (**path A** and **path B**) stem from the diketone **2**: they correspond to the attack of the hydroxyl group O4H on the carbonyl carbons C1 and C2, respectively. The related activation barriers are rather close (39.7 and 42.5 kcal mol⁻¹, respectively) and the resulting intermediate **3a** and **3b** are almost degenerate, the difference between **3a** and **3b** being only 0.4 kcal mol⁻¹. Along path B an internal rearrangement leads to the hydroxylactone 4b (LAC) by overcoming a barrier of 46.3 kcal mol⁻¹. Another product 4a can be obtained through a different internal rearrangement (path A). However, in this case a rather large barrier (130.5 kcal mol⁻¹) must be overcome, which makes this transformation very unlikely. Since within the hypothesis of **Scheme 5.1**, both **3a** and **3b** would lead to the formation of LAC, we investigated the possibility of 3a-3b interconversion. We located a transition state (**TS3ab**) corresponding to the rearrangement of the oxetane bridge, which moves from carbon C1 to carbon C2. A schematic representation of **TS3ab** is given in **Fig.5.1** (top right): in the picture we have evidenced the breaking of the O-C1 bond (1.88 Å) and the formation of the new O-C2 bond (1.68 Å) together with the simultaneous transfer of the

hydrogen from O1 to O5. A barrier of 35.2 kcal mol^{-1} must be overcome for the **3a → 3b** transformation.

As a partial conclusion, the computed energy surface indicates that it is possible to obtain LAC either following path B or path A by crossing the "gate" (**TS3ab**) which connects the two reaction channels. This agrees with the hypothesis proposed by Furneaux[16] and depicted in **Scheme 5.1**. In both cases the rate-determining step of the process is represented by the starting keto-enol tautomerization. Interestingly, the keto-enol tautomerization was recognized to be the rate-determining step also for the pyrolytic formation of levoglucosenone from cellobiose due to the high energy of the distorted four-memberd transition state.[26] The product **4a** can be ruled out because of the large activation barrier found for **TS3a**. This result is in agreement with the experiment where no traces of **4a** are found.[12] Furthermore, to check the reliability of the DFT results we carried out benchmark single-point computations at the MP4(SDQ)/6-31+G** level. The MP4 results (reported in brackets in **Fig.4.1**) are very close to the DFT results and do not affect the mechanistic scenario. However, the barrier for the rate-determining step (79.4 kcal mol^{-1} at the DFT level and 80.1 kcal mol^{-1} at the MP4(SDQ) level) is very large and the computed rate constants for LAC formation are 1.97×10^{-16} s^{-1} and 4.52×10^{-9} s^{-1} at 300°C and 500°C, respectively. These values are excessively low and inconsistent with the presence of LAC, even in small amounts, in the cellulose pyrolysates in the absence of catalyst.[12,40] Interestingly, a rather large activation energy of about 55.0 kcal mol^{-1} (230 kJ mol^{-1}) was determined for the rate-determining step of primary pyrolysis of cellulose[41] but lower values were also reported.[42,43]

Therefore, we examined alternative reaction paths for the keto-enol tautomerization. A possibility could be the involvement in the transformation of a second molecule assisting the hydrogen transfer from O5 to C3. This could be another carbon-containing molecule, for instance the enol form **1**, or a water molecule. Water can be present as moisture in the original sample: it can be physically desorbed at about 100-220 °C and abundantly generated by chemical reactions at higher temperatures.[44] Although water is a key molecule in the thermal degradation of cellulose, its formation and reactivity has not been so widely investigated as other pyrolysis products. An interesting review on the evolution of water during the heating of cellulose is given by Scheirs *et al.* in ref. 44. These authors found that in the temperature region 25-400°C a significant amount of water is produced. Only a small fraction (about 9%) of the total water is physically desorbed while the remaining 91% is chemically eliminated. Also, it was shown that water can participate as catalyst enhancing the thermal degradation of cellulose and the formation of char.[45] The influence of water

on the fate of anhydrosugars under pyrolytic conditions was highlighted in the case of levoglucosan.[9] Water was supposed to favour the evaporation of levoglucosan by hydrogen bonding, thus preventing the polymerization and charring of levoglucosan. We recomputed the whole reaction profile using a model-system formed by species 1 and one water molecule (see **Fig.4.2**).

This molecule can participate, in principle, to the various reaction steps involving a proton transfer. The added water molecule forms initially (starting species **1**) a rather strong hydrogen bond with the hydroxyl fragment O5H (O5H··O(water) = 1.87 Å). The proton transfer from O5 to C3 is effectively "assisted" by the water molecule acting as a proton shuttle: it accepts a proton from O5 and transfers a proton to C3. This process requires a much lower activation barrier (40.9 kcal mol^{-1}) than that computed in the absence of water (79.4 kcal mol^{-1}). The role played by water is similar to that recently evidenced by Zhang and coworkers who examined at the DFT level the mechanism of glucose conversion in supercritical water.[46] They demonstrated that water behaves as a "magic catalyst" that transfers hydrogen atoms and lowers activation barriers.
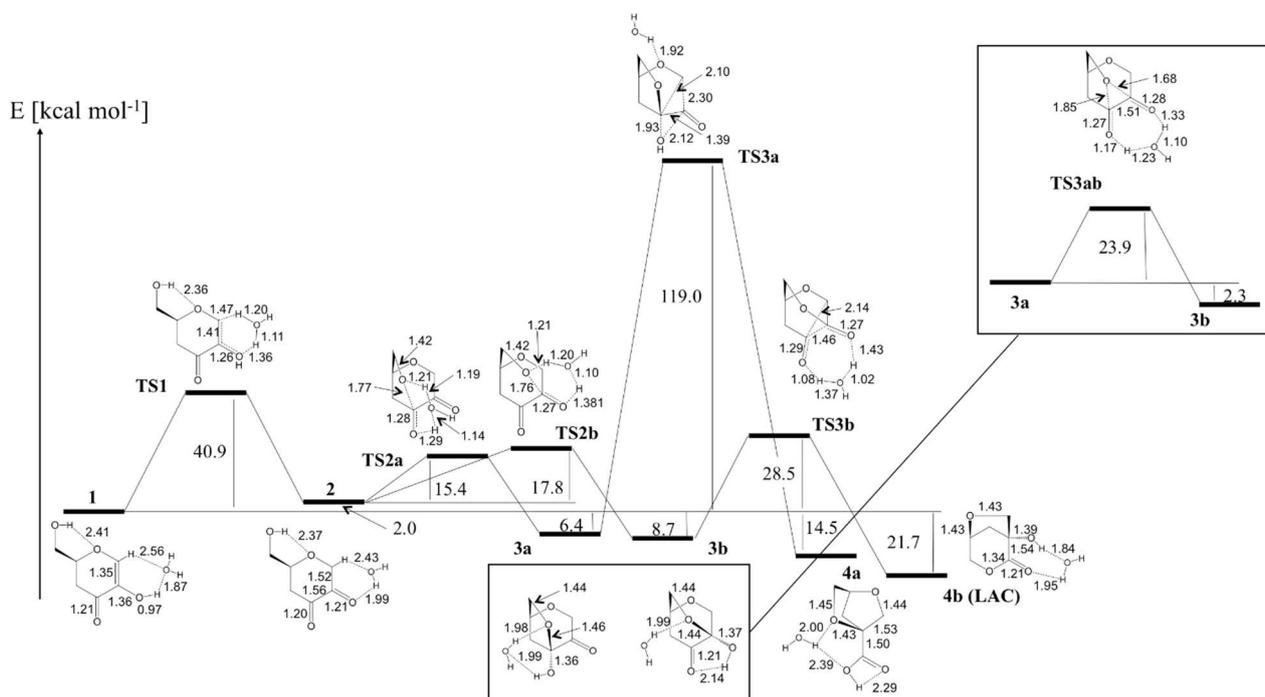
Fig.4.2. Energy profile computed for the reaction reported in **Scheme 4.1**. The model-system is formed by species **1** and one water molecule (reactants). Energy values (kcal mol^{-1}) are relative to reactants.

The resulting keto form **2** is approximately degenerate to **1**, being only 2.0 kcal mol^{-1} higher than **1**. In **2** the water molecule remains stacked to the substrate molecule *via* two hydrogen contacts, one involving the carbonyl oxygen O5 and the other the hydrogen transferred to C3, as shown in **Fig.4.2**.

These two interactions are probably responsible for the stabilization of 2 (in the absence of water, **2** was 7.9 kcal mol^{-1} higher than **1**). The two transition states **TS2a** and **TS2b** (corresponding to the two reaction channels **path A** and **path B** that originate from **2**) are both stabilized by the participation of the water molecule in the proton transfer required by the internal rearrangement leading to intermediates **3a** and **3b**. In both cases, water behaves again as a proton shuttle with a significant lowering of the corresponding activation barriers that become 15.4 kcal mol^{-1} (**TS2a**) and 17.8 kcal mol^{-1} (**TS2b**). In 3a and 3b (6.4 and 8.7 kcal mol^{-1} lower than **1**, respectively) the water molecule is still interacting with the substrate through hydrogen contacts as evidenced in **Fig.4.2**. Since the transformation **3a → 4a** does not involve any proton transfer, the water molecule cannot stabilize the corresponding transition state **TS3a**, whose structure is almost identical to that computed in the absence of water. The corresponding activation barrier (125.4 kcal mol^{-1}) is rather close to that previously obtained and formation of the product **4a** can again be discarded on energy ground. The effect of water is much more significant along **path B** that leads directly to LAC. The transition state **TS3b** is significantly stabilized by the participation of water in the proton transfer from O1 to O5. The mechanism is the usual one: water bridges the two oxygens O1 and O5 and the proton migrates from one oxygen to the other using water as a shuttle. Simultaneously an internal rearrangement of the CC bonds occurs to form LAC, which is 21.7 kcal mol^{-1} lower than the starting reactant. The "gate" between the two reaction channels (represented by transition state TS3ab) is still active. Since the key-transformation is again a proton transfer, which accompanies the breaking of the O4C1 bond and the formation of the new O4C2 bond, the presence of water stabilizes the transition state and the "gate" is now easier to cross, the activation barrier for the **3a → 3b** passage being now 23.9 kcal mol^{-1}.

Qualitatively, the mechanistic scenario obtained after inclusion of a water molecule in the reacting model-system is identical to that previously depicted in **Fig.4.1**. The rate-determining step is still the initial keto-enol isomerization and it is possible to obtain LAC either following **path B** or starting the process along **path A** and then cross the easily accessible gate **TS3ab**. The major difference outlined by the computations on the two model-systems is the entity of the rate-determining barrier, which changes from 79.4 kcal mol^{-1} in the absence of water to 40.9 kcal mol^{-1} in the "water-catalyzed" process. The rate-constant computed at 500 °C becomes 4.34 x10 s^{-1}, a value which is consistent with the experimental evidence showing that the life times of intermediate pyrolysis products of cellulose can be shorter than 1 sec.[47] Interestingly, similar energy values were obtained in a mechanistic study on the formation of 5-hydroxy-2-methylfuraldehyde (HMF), a well-recognized carbohydrate

building block.[48] This study showed that the tautomerization of glucose required a large activation barrier of 73 kcal mol^{-1} that was reduced to 40 kcal mol^{-1} when a water molecule participates in the transition state.

It is worth to mention that our discussion is based on total energy values for a hypothetical gas-phase reaction and entropy contributions were not included. In gas-phase these contributions should disfavour the assistance of water (in particular in the formation of the substrate-water complex). However, they should become much more favourable in the solid/liquid phase temporarily formed in the course of the degradation process.

Furthermore, we examined the possibility that another molecule of **1** can assist the initial keto-enol tautomerization. We found a transition state (see **Fig.5.3**) showing the existence of a cooperative effect involving two molecules in enol form (**1**), one of them acting as a proton shuttle. The hydroxyl fragment O5H of one molecule (**A**) accepts a proton from the hydroxyl fragment of another molecule (**B**) and simultaneously transfers a proton to C3 of the same molecule **B**.
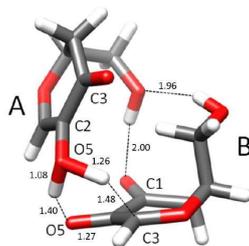
Fig.4.3. A three-dimensional representation of the transition state for the initial keto-enol tautomerization in the cooperative mechanism (bond lengths are in Ångstroms).

However, even if the activation barrier significantly decreases with respect to that computed in the absence of water, its value is still rather large (54.5 kcal mol^{-1}) and cannot compete with the "water-catalyzed" process.

4.4. Conclusions

One of the most interesting products of the pyrolysis of cellulose is the hydroxylactone species (1*R*,5*S*)-1-hydroxy-3,6-dioxabicyclo[3.2.1]octan-2-one, usually denoted as LAC. This species, characterized by a lactone unit and two chiral centres with unambiguous configuration, represents a useful chiral building block in organic synthesis. In this paper we carried out a computational investigation at the DFT level of the mechanism of formation of LAC (in the absence of catalysts) starting from the known pyrolysis product **1** shown in **Scheme 5.1**. To this purpose we used two

model-systems: one formed just by species **1** and the other including **1** and one water molecule, which was proved to be abundantly present during the heating of cellulose.

Our computations depicted a mechanistic scenario that is in qualitative agreement with the hypothesis shown in **Scheme 5.1**. LAC can be obtained following two reaction paths: **path A** and **path B**. **Path B** leads directly to LAC through a series of internal rearrangement involving in all cases a proton transfer. Alternatively, it is possible to follow initially **path A** and then "jump" on to **path B** by crossing a "gate" connecting the two channels. The gate is represented by an internal rearrangement transition state and is characterized by a low activation barrier. In both cases the rate-determining step of the process is the initial keto-enol isomerization. The final product of **path A** is not experimentally observed in agreement with the large activation barrier that should be surmounted in the final phase of the process.

Because of the large barrier for the rate-determining step (79.4 kcal mol^{-1}) the corresponding rate-constant at 300°C and 500°C are excessively low and cannot explain the presence of LAC, even in small amounts, in the product mixture when the process is carried out in the absence of catalyst. To explain this experimental evidence we examined the possibility that water, which is present in the reaction mixture, can "catalyse" the reaction by assisting the proton transfers present in all the steps of the process. We found that in all cases water acts as a proton shuttle by stabilizing all the transition states that must be overcome to reach LAC. In particular, the barrier of the rate-determining step is significantly reduced, *i.e.* 40.9 kcal mol^{-1}. The corresponding computed rate constant is 4.34x10 s^{-1} at 500°C, a value which is consistent with the presence of LAC in the absence of metal catalysts.

The present study highlights the importance of water in catalysing secondary pyrolysis reactions of cellulose leading to potential chemical platforms. In particular, the role of water resulted important in reducing the energy barrier of keto-enol tautomerization, a bottle-neck not only to the formation of LAC, but also for levoglucosenone.

References

a. Marforio, T.D; Bottoni, A.; Calvaresi, M.; Fabbri, D.; Giacinto, P.; Zerbetto, F. ChemPhysChem, **2016**, 17(23) 3948-395.

[1]   Nakashima, K.; Yamada, L.; Satou, Y.; Azuma, J.I.; Satoh, N. *Dev. Genes Evol.* **2004**, *214*, 81-88.

[2]   a) Bozell, J. J.; Petersen, G. R. *Green Chem.* **2010**, *12*, 539-554; b) Chatterjee, C.; Pong, F.; Sen, A. *Green Chem.* **2015**, *17*, 40-71.

[3]   Luterbacher, J. S.; Martin Alonso, D.; Dumesic. J. A. *Green Chem*. 2014, *3* (16), 4816-4838.

[4]   Schroyen, M.; Vervaeren, H.; Vandepitte, H.; Van Hulle, S. W. H.; Raes, K. *Bioresour. Technol.* **2015**, *192*, 696-702.

[5] Fabbri, D.; Torri, C. *Curren. Opin. Biotechnol.* **2016**, *38*, 167–173.

[1] Oasmaa, B.; van de Beld, P.; Saari, D. C.; Elliott, Y.; Solantausta, Y. *Energ. Fuel* **2015**, *29*, 2471-2484.

[2] Bridgwater, A. V. Taylor & Francis Group, **2015**, 473-513.

[3] Dusselier, M.; Mascal, M.; Sels, B.F. *Top. Curr. Chem*. **2014**, *353*, 1-40.

[4] Bai, X.; Johnston, P.; Brown, R.C. *J. Anal. Appl. Pyrol.* **2013**, *99*, 130-136.

[5] Hoekstra, E.; Van Swaaij, W. P. M.; Kersten, S. R. A.; Hogendoorn, K. J. A. *Chem. Eng. J.* **2012**, *187*, 172-184.

[6] Radlein D. Bridgwater Ed. CPL Press 2002.

[7] Fabbri, D.; Torri, C.; Mancini, I. *Green Chem.* **2007**, *9*, 1374-1379.

[8] Dobrzynski, P.; Fabbri, D.; Torri, C.; Kasperczyk, J.; Kaczmarczyk, B.; Pastusiak, M. *J. Polym. Sci. A,* **2009**, *47*, 247-257.

[9] Defant, A.; Mancini, I.; Torri, C.; Malferrari, D.; Fabbri, D. *Amino Acids,* **2011**, *40*, 633-640.

[10] Defant, A.; Mancini, I.; Matucci, R.; Bellucci Dosi, F.; Malferrari, D.; Fabbri, D. *Org. Biomol. Chem*., **2015**, *13*, 6291-6298.

[11] Furneaux, R. H.; Mason, J. M.; Miller, I. J. *J. Chem. Soc. Perkin Trans. I,* **1988**, 49-51.

[12] Fabbri, D.; Torri, C.; Bavarelli, V. *J. Anal. Appl. Pyrol*., **2007**, *80*, 24-29.

[13] Eibner, S.; Broust, F.; Blin, J.; Julbe, A. *J. Anal. Appl. Pyrol.,* **2015**, *113*, 143-152.

[14] Torri, C.; Lesci, I.; Fabbri, D. *J. Anal. Appl. Pyrol.,* **2009**, *85*, 2192-2196.

[15] Casoni, A. I.; Nievas, M. L.; Moyano, E. L.; Álvarez, M.; Diez, A.; Dennehy, M.; Volpe, M. A. *App. Catal. A,* **2016**, *514*, 235-240.

[16] Mancini, I.; Dosi, F.; Defant, A.; Crea, F.; Miotello, A. *J. Anal. Appl. Pyrol.,* **2014**, *110*, 285-290.

[17] Rutkowski, P. *J. Anal. Appl. Pyrol.* **2012**, *98*, 115-122.

[18] Seshadri, V.; Westmoreland, P. R. *Catal. Today,* **2016**, *269*, 110-121.

[19] Zhang, Y.; Liu, C.; Chen, X. *J. Anal. Appl. Pyrol.* **2015**, *113*, 621-629.

[20] Zhang, Y.; Liu, C. *J. Anal. Appl. Pyrol.,* **2014**, *110*, 297-304.

[21] Lu, Q.; Zhang, Y.; Dong, C-Q.; Yang, Y-P.; Yu, H-Z. *J. Anal. Appl. Pyrol.,* **2014**, *110*, 34-43.

[22] Zhang, X.; Yang, W.; Blasiak, W. *Fuel*, **2013**, *109*, 476-483.

[23] Zhang, X.; Yang, W.; Blasiak, W. *J. Anal. Appl. Pyrol.,* **2012**, *96*, 110-119.

[24] Wang, S.; Gua, X.; Liang, T.; Zhou, Y.; Luo, Z. *Biores. Tech.,* **2012**, *104*, 722-728.

[25] Ren, H.; Yu, W.; Salciccioli, M.; Chen, Y.; Huang, Y.; Xiong, K.; Vlachos, D. G.; Chen, J. G. *ChemSusChem,* **2013**, *6*, 798-801.

[26] Yang, G., Pidko, E. A.; Hensen, E. J. M. *ChemSusChem,* **2013**, *6*, 1688-1696.

[27] Hosoya, T.; Sakaki, S. *ChemSusChem,* **2013**, *6*, 2356-2368.

[28] Yabushita, M.; Kobayashi, H.; Hasegawa, J.-Y.; Hara, K.; Fukuoka, A. *ChemSusChem,* **2014**, *7*, 1443-1450.

[29] Yun, D.; Kim, T. Y.; Park, D. S.; Yun, Y. S.; Han, J. W.; Yi, J. *ChemSusChem,* **2014**, *7*, 2193-2201.

[30] Li, K.; Yang, Z.; Zhao, J.; Lei, J.; Jia, X.; Mushrif, S. H.; Yang, Y. *Green Chem.,* **2015**, *17*, 4271-4280.

[31] Cao, X.; Teong, S. P.; Wu, D.; Yi, G.; Su, H.; Zhang, Y. *Green Chem.,* **2015**, *17*, 2348-2352.

[32] Yang, L.; Tsilomelekis, G.; Caratzoulas, S.; Vlachos, D. G. *ChemSusChem,* **2015**, *8*, 1334-1341.

[33] Gaussian 09, M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, et. al. Gaussian, Inc., Wallingford CT, 2009.

[34] Zhao, Y.; Truhlar, D. G. *Theor. Chem. Acc.,* **2008**, *120*, 215-224.

[35] Lu, Q.; Yang, X. C.; Dong, C. Q.; Zhang, Z. F.; Zhang, X. M.; Zhu. X. F. *J. Anal. Appl. Pyrol.,* **2011**, *92*, 430-438.

[36] Antal, M. J., Varhegyi, G.; Jakab, E. *Ind. Eng. Chem. Res.,* **1998**, *37*, 1267-1275.

[37] Lédé, J. *J. Anal. Appl. Pyrol.,* **2012**, *94*, 17-32.

[38] Li, Y. ; Cho, J. M.; Tompsett, G. A.; Westmoreland, P. R.; Huber, G. W. *J. Phys. Chem. C,* **2009**, *113*, 20097-20107.

[39] Sheirs, J.; Camino, G.; Tumiatti, W. *Europ. Polym. J.,* **2001**, *37*, 933-942.

[40] Varhegyi, G. ; Szabo, P.; Mok, W.S. L.; Antal, Jr, M. J. *J. Anal. Appl. Pyrol.,* **1993**, *26*, 159-174.

[41] Zhang, Y.; Liu, C.; Chen, X. *J. Anal. Appl. Pyrol.,* **2016**, *119*, 199-207.

[42] Boutin, O.; Ferrer, M.; Lédé, J. *J. Anal. Appl. Pyrol.,* **1998**, *47*, 13-31.

[43] Assary, R. S.; Curtiss, L. A. *Energy Fuels,* **2012**, *94*, 1344-1352.

Part V

Side Works

1. Engeneering the Fullerene-protein Interface by Computational Design: The Sum is more than its Parts[a]

1.1. Introduction

Understanding the interactions at the nano–bio interface is the first step for the safe use of nanotechnology and for the design of nanomaterial for biological applications.[1] Engineering specific interactions between proteins and nanoparticles will allow development of new applications and design innovative nanomachines, sensors, and theranostic platforms.[2]

Computational design offers opportunities for engineering protein structure and function.[3] Binding pocket optimization allows the design of specific interactions between proteins and ligands. [3] In these designs, an existing protein is generally used as a scaffold, and its binding pocket is modified to tune the interaction with the target ligand. [3] Generally, the optimization of the protein binding pocket is carried out considering small molecules as ligands.

The ability of carbon nanoparticles (CNP) and fullerenes to interact with proteins was demonstrated for the first time by pioneering work that reported the C_{60} inhibiting activity on HIV-proteases.[4] Protein interactions with fullerene-based compounds were later identified in many other systems both computationally[5-19] and experimentally.[20-41] In the study of protein-nanoparticles interactions, it is often hard to ascertain if the measured interactions are due to (i) the formation of a well-defined 1:1 adduct, (ii) binding of the protein with aggregates, or (iii) average effects deriving by the binding of the nanoparticle(s) to multiple protein binding sites.

For binding pocket optimization it is necessary to work with a well-defined system where the interaction between the nanoparticle and the protein binding pocket is highly specific and localized. For this reason, we chose C_{60}@lysozyme that we recently characterized.[24] NMR chemical shift perturbation analysis unambiguously identified a fullerene-protein binding pocket in solution. [24] The NMR and spectroscopic data showed that lysozyme forms a truly stoichiometric 1:1 adduct with C_{60} where lysozyme maintains its tridimensional structure with only a few well-identified residues that are structurally perturbed.[24] Molecular dynamics (MD) simulations have already provided

information regarding interactions of carbon nanoparticles with proteins showing the dynamics at the molecular level and addressing the effects of surface chemistry on the adsorption of proteins. [42,43]

Using the MM-PB(GB)SA protocol to analyze the MD trajectories, we recently described a computational procedure that provides a detailed analysis of the various components of the binding energy and quantifies the interactions between the protein and the π electron conjugated surface at the level of individual residues. [6.8] This approach allowed us to identify "hot" and "cold" spots for the interaction. Different interactions such as π-π stacking interactions, hydrophobic interactions, surfactant-like interactions, electrostatic interactions govern the wealth of structures that appear when proteins and CNPs interact. The per-residue decomposition of $\Delta G_{binding}$ of C_{60}@lysozyme helped us in understanding the roles of the various aminoacids of the protein.[6] The calculations emphasized the role of Trp residues (Trp62 and Trp63) in the binding process.[6]

It is well-known that protein adsorption onto CNPs improves with the increase of the content of aromatic residues in the protein sequence. [44] Among the aromatic amino acids, tryptophan possesses the highest affinity for CNPs, followed by tyrosine, phenylalanine, and histidine.[44]

The π-stacking contacts between the indolic group of Trp residues and the carbon cage, which are identified in the C_{60}@Lysozyme complex and govern the interactions between lysozyme and C_{60}.[6] This π-stacking interaction may be sandwich-like, as for Trp62, or T-shape-like, as for Trp63. [6] In this work, we endeavor to engineer the fullerene-lysozyme interface by computational design, in order to provide guidelines for the optimal binding of proteins to CNPs.

1.2. Results and Discussion

1.2.1 Lysozyme mutants. In C_{60}@lysozyme, the per-residue decomposition analysis of the $\Delta G_{binding}$[6] showed that there are 9 more residues that have an interaction energy with C_{60} that is higher than 1 kcal mol^{-1}. Apart from the two Trp residues (Trp 62 and Trp 63), these residues make up the fullerene binding pocket of wild-type lysozyme (See **Fig.1.1**).[6] Since Trp is the aminoacid that shows the highest values of interaction with C_{60}, we mutate *in silico*, one at a time, the other 7 "binding" residues of the wild-type lysozyme binding pocket (**Fig.1.1**) to Trp residues. This procedure generates 7 lysozyme mutants (see **Table 1.1**). We then study the effect of the mutation on the structure of the complex and on the $\Delta G_{binding}$, with the aim to optimize the interaction energy of the protein with the C_{60}.
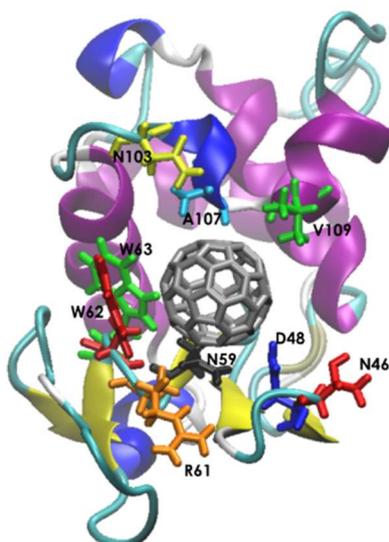
Fig.1.1. Fullerene binding pocket of wild-type lysozyme. The more interacting residues ($E_{interact}$ > 1 kcal mol^{-1}) of wild-type lisozyme with C_{60} are in licorice.

Table 1.1. Lysozyme mutants and interaction energies of the newly introduced tryptophan (kcal/mol).

Wild-Type Lysozyme	Lysozyme Mutant	Interaction energy with C_{60}
WT-N46	N46W	-2.0
WT-D48	D48W	-2.8
WT-N59	N59W	-4.1
WT-R61	R61W	-3.0
WT-N103	N103W	-1.4
WT-A107	A107W	-2.9
WT-V109	V109W	-1.3

1.2.2 MD and MM-GBSA analysis of the C_{60}@Lysozyme-mutants complexes. Starting from the NMR data[24] and from the previous calculation about C_{60}@lysozyme,[6] all the different complexes between C_{60} and the lysozyme mutants were built. Subsequently, a scoring of the interaction energy (see Computational Details) for all the complexes was carried out. The estimation of the binding energy between the lysozyme mutants and C_{60} was performed using a molecular mechanics/Generalized-Born Surface Area (MMGBSA) analysis of the MD individual trajectories. The calculated interaction energies were compared with the results obtained for the wild-type lysozyme ($\Delta\Delta G_{binding}$) in order to estimate the effect of the mutation on the $\Delta G_{binding}$ (**Fig.1.2a**). In principle, it could be expected that the mutation with Trp would increase $\Delta G_{binding}$ of C_{60} to the protein. There are, however, four cases where the mutation worsens the interaction of the protein with the fullerene cage, namely N46W, N59W, N61W, N107W. In three cases, D48W, N103W, V109W, $\Delta G_{binding}$ increases, as expected. This

158

result confirms that the simple presence of amino acids with a strong tendency to bind carbon nanoparticles does not guarantee strong binding of proteins to a fullerene.
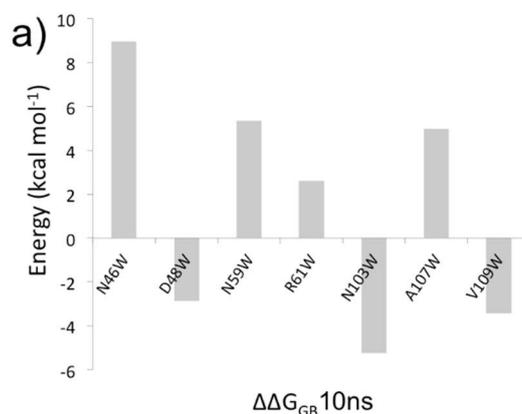
Fig.1.2. Overall variation of the interaction energy $\Delta\Delta G_{binding}$ of C_{60} with lysozyme mutants compared to Wild-Type Lysozyme.

The three-dimensional arrangement of the amino acid side chains around C_{60} is crucial. Interestingly, the mutations that show the larger variation of $\Delta G_{binding}$ involve an arginine (N46 and N103). The two variations have a different sign. In order to analyze in detail, the structural and energetic effects caused by these two cases, we carried out longer MD simulations of 100 ns.

1.2.3 Binding of C_{60} with wild-type lysozyme and the N103W, N46W mutants. The role of surface complementarity in the binding of C_{60} with proteins is known to be crucial,[6,8] in fact the most important energetic terms, vdW interactions and energy de-solvation are roughly proportional to the variation of the Solvent Accessible Surface Area, ΔSASA, that is a measure of the surface complementarity. [6,8] This phenomenon is similar to the well-known encapsulation of C_{60} by macrocyclic receptors, [45] where concave-convex complementarity is the driving force to fullerene binding. [45]

To provide an initial estimation of the effect of the mutations on $\Delta G_{binding}$, we can compare ΔSASA in the mutants and in the wild type lysozyme, upon binding with the C_{60}. Pictorially in **Fig.1.3** and quantitatively in **Table 1.2**, it emerges that when the mutation determines a gain in the $\Delta G_{binding}$ there is an increase in the surface complementarity (compare **Fig.1.3b and Fig.1.3a**) and in the ΔSASA value, otherwise when the mutation determines a loss in the $\Delta G_{binding}$ there is a decrease in the surface complementarity (compare **Fig.1.3c** and **Fig.1.3a**) and in the ΔSASA value. In N103W and wt-lysozyme, the position of C_{60} is the same. In the N46W, the mutation triggers a variation of the fullerene location in the lysozyme binding pocket, which is ultimately responsible for the decrease of the binding energy.
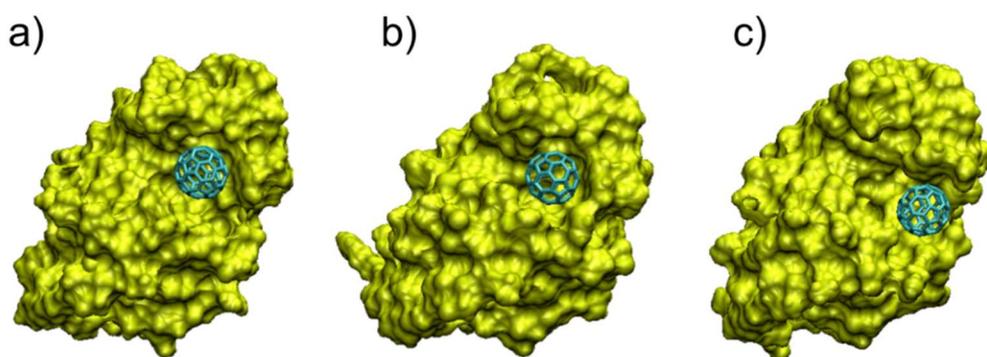
Fig.1.3. Surface complementarity between the C_{60} cage and the surface (in yellow) of a) wild-type lysozyme, b) N103W mutant, c) N46W mutant.

Table 1.2. Estimate of ΔSASA for the wt-lysozyme and mutants upon C_{60} binding.

Protein	ΔSASA (Å^2)	$\Delta\text{Å}^2$ vs wt-lysozyme (Å^2)
WT	302	//
N103W	315	+13
N46W	282	-20

1.2.4 Optimization of the protein-C_{60} interaction: the case of the N103W mutant. In the case of N103W, $\Delta G_{binding}$ between C_{60} and N103W is -22.9 kcal mol^{-1}, which is 4.4 kcal mol^{-1} of increase with respect to wild-type lysozyme. Analysis of the binding components of the energy (**Fig.2.4a**) shows that van der Waals interactions are the driving force of the binding (-48.4 kcal mol^{-1}). Hydrophobic interactions, *i.e.*, nonpolar solvation, assist the binding, even if the corresponding value (-4.5 kcal mol^{-1}) is far smaller than that of the vdW interactions. Polar solvation is detrimental to the binding and the contribution is positive (+12.0 kcal mol^{-1}). C_{60} occupies part of the active site of the lysozyme that is made up by many hydrophilic residues, upon formation of the complex with C_{60}, they are forcedly desolvated, destabilizing the system.
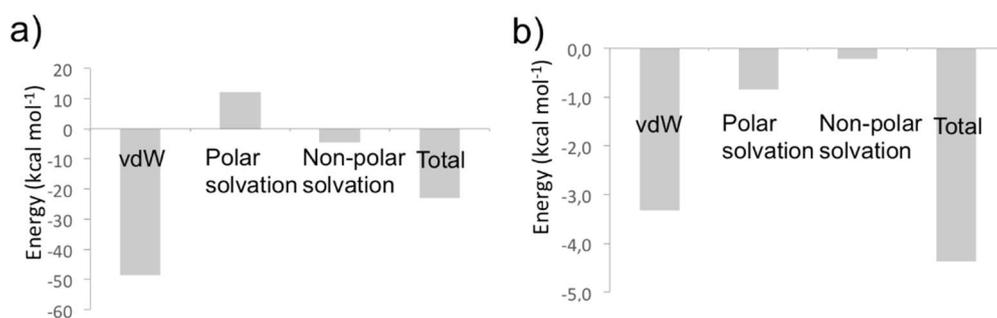
Fig.1.4. Energy components of $\Delta G_{binding}$ for C_{60} binding with a) N103W, b) $\Delta\Delta G_{binding}$ with respect to the complex of wild-type lysozyme. The data are obtained with a simulation of 100 ns.

The trend of the energy components of the $\Delta G_{binding}$ of C_{60} for N103W is similar to that of the complex of wild-type lysozyme/C_{60}. Figure 4b shows that there is an energy gain for all three contributes. The increase of vdW and non-polar solvation terms can be explained with an increase of the shape complementarity between the protein and C_{60}. The Trp 103 is closer to the fullerene cage than Arg 103 (3.55 Å for the Trp vs 4.04 Å for Arg) and interacts strongly (**Fig.2.5**) with additional π-π interactions (sandwich-like as is clear in **Fig.2.5b**).
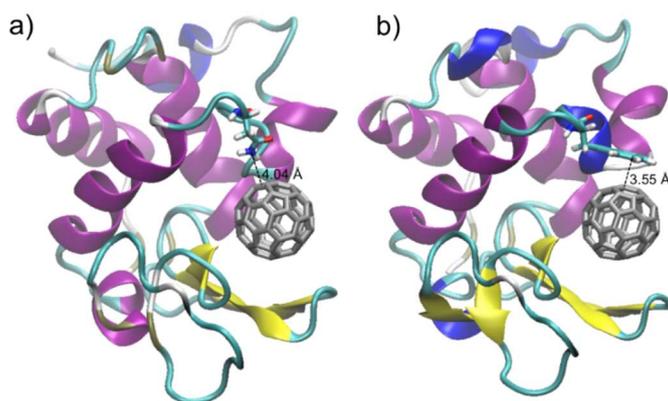
Fig.1.5. Interaction between Arg103 (**a**) and Trp103 (**b**) with C_{60}.

On the other side, the gain in the polar solvation energy is due to the reduction of the polar term, which remains positive. In fact in the wild-type lysozyme during the formation of the complex with C_{60} the hydrophilic part of asparagine (N103), faces the fullerene cage (see **Fig.2.5a**) and is desolvated, causing a destabilization of the system, because this residue, in the presence of the hydrophobic C_{60} molecule, is no longer able to interact with water molecules, usually present in the binding pocket, reducing its solvation energy. In the N103W mutant, this energy penalty is reduced, due to the more hydrophobic character of the Trp mutant that faces C_{60} with its hydrophobic indolic ring, gaining energy and eliminating the desolvation energy penalty due to the C_{60} binding.

1.2.5 Decrease of the protein-C_{60} interaction: the case of N46W mutant. **Fig.1.3c** shows that in N46W the Trp mutation triggers a variation of the fullerene location in the lysozyme binding pocket. The difference is more evident by superimposing the optimized complex of C_{60} with wt-lysozyme and N46W mutant (**Fig.1.6a**). In the N46W mutant, the bulkier Trp occupies partly the original fullerene-binding pocket and as a consequence C_{60} changes position. The physiological substrate of lysozyme is a polysaccharide. The recognition pocket consists of six subsites, designated as "A" to "F", which can accommodate six sugar moieties. In practice, the recognition pocket is a crevice. In wt-lysozyme, C_{60} binds in this crevice.[24] Upon the Trp mutation in position 46, C_{60} molecule slides along the trench. Trp 62 works as a gate and change its conformation during the sliding of the C_{60}, as

underlined in Figure 6a. In **Fig.1.6b** and **1.6c** the two possible sub-pockets that can host the C_{60} molecule and the role of Trp 62 as gate are highlighted.
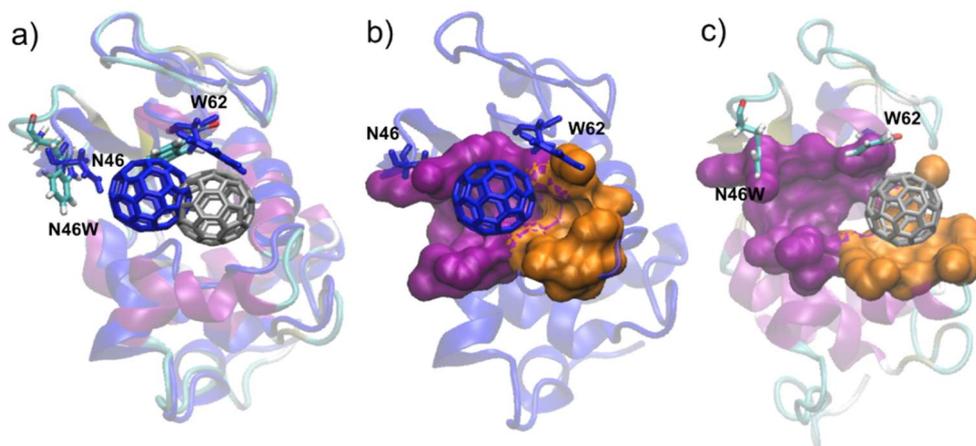
Fig.1.6. Superimposition of the complexes between C_{60} and wild-type lysozyme (in blue) and N46W mutant (in grey). Representation of the two sub-pockets (in purple and orange), able to host C_{60} b) wild-type lysozyme, c) N46W mutant. The surface of the sub-pockets are calculated using DoGSiteScorer,[46] a tool for automated pocket detection and analysis.

The analysis of the root-mean square displacement of the center of mass of the C_{60} molecule (**Fig.1.7**) during the MD trajectories, allows quantification of the mobility of the C_{60} in complex with N46W, especially if compared with wt-lysozyme and N103W.
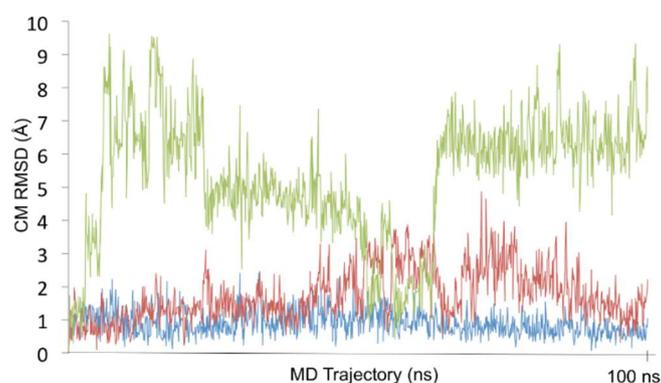
Fig.1.7. Root-mean square displacement of the center of mass of the C_{60} during the MD trajectory of the C_{60} complex with wild-type lysozyme (blue line), N103W (red line) and N46W (green line).

In wt and in N103W, C_{60} remains in its binding pocket during the entire trajectory, apart for a small local re-arrangement in the mutant. In the N46W, C_{60} hops between two sub-pockets.

Compared to the shorter simulation of 10 ns, the energy difference/loss upon binding of C_{60} is smaller since the longer simulation allows the protein to re-organize locally (**Fig.1.8a**). With respect to the wt complex, the energy loss is due to vdW and non-polar solvation terms (**Fig.1.8b**).
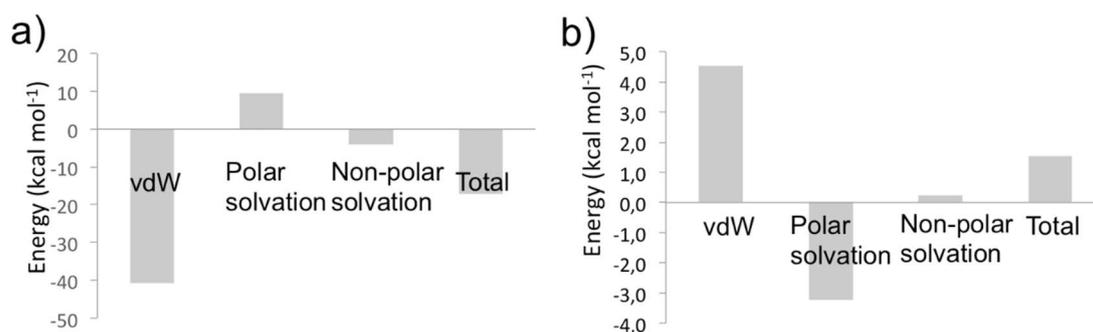
Fig.1.8. a) Energy components of the $\Delta G_{binding}$ for C_{60} binding to N46W, **b**) $\Delta\Delta G_{binding}$ with respect to the complex of wild-type lysozyme. The data are obtained with a simulation of 100 ns.

It can be explained with a decrease of shape complementarity between the protein and C_{60}, due to the continuous movement of C_{60}. As explained previously in the N46W mutant the bulkier Trp occupies partly the wt-lysozyme fullerene binding pocket, generating a steric clash that decreases the vdW interaction energy and the solvent accessible surface area. In a few worlds the Trp 46 continuously "kicks" the C_{60} ball to a different sub-pocket.

Also the positive effect of the mutation on the polar solvation term is due to the position changing of C_{60}. In fact the wt-lysozyme binding pocket is made up by many hydrophilic residues, when C_{60} moves from the wt-pocket to the other sub-pocket these residues became, as before the C_{60} binding, solvent-exposed and increase the solvation term. However, even if important (-3.2 kcal mol^{-1}), this term energetically do not balance the detrimental variations.

1.3. Conclusions

In this work, we have showed that it is possible to use computational protein design to understand and optimize protein-binding pockets that bind to the π-electron rich surface of a CNP. In particular, the approach was used to engineer the binding pocket of lysozyme to host a fullerene molecule. The presence of amino acids with a strong tendency to bind carbon nanoparticles, as Trp, does not suffice *per se* to guarantee the improvement of the binding of proteins to fullerene. In fact to increase $\Delta G_{binding}$ it is crucial the location where Trp residues are located since mutation of residues in wt-proteins may trigger small structural variations that lead to destabilization of the complex between CNP and the protein.

1.4. Computational Details

1.4.1 In silico generation of the mutants. Experimental data from NMR[24] and a docking protocol[6,8,14,16,24,47] recently validated for the study of interaction between proteins and nanoobjects

were used to generate the initial coordinates of the adduct between protein and C_{60} as previously described.[6] The lateral chain of the residue to be mutated is deleted in the PDB file and the Trp residue is added using the *tleap* module in antechamber.[49]Only to remove severe sterical clashes, the position of the added Trp residues in all the mutants are optimized with SANDER, with the C_{60} and all the protein atoms (except the mutated Trp) frozen.

1.4.2 Setting the Simulation. Chloride counterions were included to exactly neutralize the positively charged lysozyme mutants. All simulations were performed with explicit solvent by using the TIP3P water model (7605 water molecules).[48] The same water box was used for all the simulation. The ff10 force field was used to model lysozyme mutants.[49] The C_{60} atoms were modeled as uncharged Lennard−Jones particles by using sp2 carbon parameters from the ff10 force field.[49]

1.4.3 Minimization and Equilibration. About 10,000 steps of steepest descent minimization were performed for the whole system with PMEMD.[49] The minimized structure was considered for a 3 step equilibration protocol. Particle Mesh Ewald summation[49] was used throughout (cut off radius of 10 Å for the direct space sum). H atoms were considered by the SHAKE algorithm[49] and a time step of 2 fs was applied in all MD runs. Individual equilibration steps included (i) 50 ps of heating to 298 K within an NVT ensemble and temperature coupling according to Berendsen. (ii) 50 ps of equilibration MD at 298 K to switch from NVT to NPT and adjust the simulation box. Isotropic position scaling was used at default conditions. (iii) 900 ps of continued equilibration MD at 298 K for an NPT ensemble switching to temperature coupling according to Andersen.

1.4.4 Production MD. MD simulation was carried out for the equilibrated system using PMEMD.[49] Simulation conditions were identical to the final equilibration step (iii). Overall sampling time was 10 ns for all the mutants. Two longer trajectories of 100 ns were carried out for the N46W and N103W mutants.

1.4.5 Post Processing of Trajectories, MM-PBSA. MM-GBSA[50,51] analysis was carried out to estimate the binding free energy of the C_{60} when complexed to lysozyme mutants. MM/GBSA analysis is a postprocessing method in which representative snapshots from an ensemble of conformations are used to calculate the free energy change between two states (typically a bound and free state of a receptor and a ligand).[50] Free energy differences are calculated by combining the gas phase energy contributions that are independent of the solvent model as well as solvation free energy components (both polar and nonpolar) calculated from an implicit solvent model for each species.[51] The molecular mechanics energies are determined with the SANDER program from Amber[49] and represent the internal energy (bond, angle and dihedral), and van der Waals and

electrostatic interactions. An infinite cutoff for all interactions is used. The electrostatic contribution to the solvation free energy is calculated by generalized Born (GB) methods implemented in SANDER. The nonpolar contribution to the solvation free energy has been determined with solvent-accessible surface-area dependent terms. Individual snapshot structures of all trajectories were analyzed with the program PTRAJ.[49]

References

a. Trozzi, F.; Marforio, T.D.; Bottoni, A.; Zerbetto, F.; Calvaresi, M. Israel J. Chem., **2017**, *57(6)*, 547-552.

[1]   Nel, A. E.; Mädler, L.; Velegol, D.; Xia, T. E.; Hoek, M. V.; Somasundaran, P.; Klaessig, F.; Castranova, V.; Thompson, M. *Nat. Mater.,* **2009**, *8*, 543-557; b) Rana, S., Yeh, Y.-C.; Rotello, V. M. *Curr. Opin. Chem. Biol.* **2010**, 14, 828-834; c) Gagner, J. E.; Shrivastava, S.; Qian, X.; Dordick, J. S.; Siegel R. W. *J. Phys. Chem. Lett.* **2012**, *3*, 3149-3158

[2]   Calvaresi, M.: Zerbetto, F. *Acc. Chem. Res.* **2013**, *46*, 2454-2463.

[3]   a) Looger, L. L.; Dwyer, M.A.; Smith, J. J.; Hellinga, H. W. *Nature* **2003**, *423*, 185-190; b) Lippow, S. M.; Tidor, B. *Curr. Opin. Biotechnol.* **2007**, 18, 1-7; c) Malisi, C.; Schumann, M.; Toussaint, N. C.; Kageyama, J.; Kohlbacher, B. Hocker, *Plos One* **2012**, *7*, e52505, d) C. E. Tinberg, S. D. Khare, J. Dou, L. Doyle, J. W. Nelson, O.; Schena, A.; Jankowski, W.; Kalodimos, C. G.; Johnsson, K.; Stoddard, B. L.; Baker, D. *Nature* **2013**, 501, 212-216; e) Pakulska, M. M.; Miersch, S.; Shoichet, M. S. *Science* **2016**, 351, aac4750.

[4]   Friedman, S. H.; DeCamp, D. L.; Sijbesma, R. P.; Srdanov, G., Wudl, F.; Kenyon, G. L. *J. Am. Chem. Soc.* **1993**, *115*, 6506-6509.

[5]   Junaid, M.; Almuqri, E. A.; Liu, J.; Zhang, H. *Plos One* **2016**, *11*, e0147761.

[6]   Calvaresi, M.; Bottoni, A.; Zerbetto, F. *J. Phys. Chem. C* **2015**, *119*, 28077-28082.

[7]   Leonis, G.; Avramopoulos, A.; Papavasileiou, K. D.; Reis, H.; Steinbrecher, T.; Papadopoulos, M. G. *J. Phys. Chem. B* **2015**, *119*, 14971-14985.

[8]   Calvaresi, M.; Furini, S.; Domene, C.; Bottoni, A.; Zerbetto, F. *ACS Nano* **2015**, *9*, 4827-4834.

[9]   Turabekova, M.; Rasulev, B.; Theodore, M.; Jackman, J.; Leszczynsk, D.; Leszczynski, J. *Nanoscale* **2014**, *6*, 3488-3495.

[10] Radic, S.; Nedumpully-Govindan, P.; Chen, R.; Salonen, E.; Brown, J. M.; Ke, P. C.; Ding, F. *Nanoscale* **2014**, *6*, 8340–8349.

[11] Govindan, P. N.; Monticelli, L.; Salonen, E. *J. Phys. Chem. B* **2012**, *116*, 10676-10683.

[12] Monticelli, L.; Barnoud, J.; Orlowski, A.; Vattulainen, I. *Phys. Chem. Chem. Phys.* **2012**, *14*, 12526-12533.

[13] Kang, S. G.; Huynh, T.; Zhou, R. H. *Sci. Rep.* **2012**, *2*, 957.

[14] Calvaresi, M.; Zerbetto, F. *Nanoscale* **2011**, *3*, 2873-2881.

[15] Zuo, G.; Zhou, X.; Huang, Q.; Fang, H.; Zhou, R. *J. Phys. Chem. C* **2011**, *115*, 23323-23328.

[16] Calvaresi, M.; Zerbetto, F. *ACS Nano* **2010**, *4*, 2283-2299.

[17] Kraszewski, S.; Tarek, M.; Treptow, W.; Ramseyer, C. *ACS Nano* **2010**, *4*, 4158-4164.

[18] Wu, X.; Yang, S.-T.; Wang, H.; Wang, L.; Hu, W.; Cao, A.; Liu, Y. *Nanosci.Nanotechnol.* **2010**, 10, 6298-6304.

[19] H. Benyamini, A. Shulman-Peleg, H. J. Wolfson, B. Belgorodsky,

[20] L. Fadeev, M. Gozin, *Bioconjugate Chem.* **2006**, *17*, 378-386.

[21] S. J. Vance, V. Desai, B. O. Smith, M. W. Kennedy, A. Cooper, *Biophys. Chem.* **2016**, *214-215*, 27-32.

[22] K.-H. Kim, D.-K. Ko, Y.-T. Kim, N. H. Kim, J. Paul, S.-Q. Zhang, C. B. Murray, R. Acharya, W. F. DeGrado, Y. H. Kim, G. Grigoryan, *Nat. Commun.* **2016**, *7*, 11429.

[23] Y. Pan, L. Wang, S: Kang, Y. Lu, Z. Yang, T. Huynh, C. Chen, R. Zhou, M. Guo, Y. Zhao, *ACS Nano* **2015**, *9*, 6826-6836.

[24] S. Zanzoni, A. Ceccon, M. Assfalg, R. K. Singh, D. Fushman, M. D'Onofrio, *Nanoscale* **2015**, *7*, 7197-7205.

[25] M. Calvaresi, F. Arnesano, S. Bonacchi, A. Bottoni, V. Calo`, S. Conte, G. Falini, S. Fermani, M. Losacco, M. Montalti, G. Natile, L. Prodi, F. Sparla, F. Zerbetto, *ACS Nano*, **2014**, *8*, 1871-1877.

[26] Y. Miao, J. Xu, Y. Shen, L. Chen, Y. Bian, Y. Hu, W. Zhou, F. Zheng, N. Man, Y. Shen *et al. ACS Nano* **2014**, *8*, 6131-6144.

[27] P. Chen, S. A. Seabrook, V. C. Epa, K. Kurabayashi, A. S. Barnard, D. A. Winkler, J. K. Kirby, P. C. Ke, *J. Phys. Chem. C* **2014**, 118, 22069-22078.

[28] E. Mentovich, B. Belgorodsky, M. Gozin, S. Richter, H. Cohen, *J. Am. Chem. Soc.* **2012**, *134*, 8468-8473.

[29] S. G. Kang, G. Zhou, P. Yang, Y. Liu, B. Sun, T. Huynh, H. Meng, L. Zhao, G. Xing, C. Chen, *et al. Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109*, 15431-15436.

[30] M. Zhen, J. Zheng, L. Ye, S. Li, C. Jin, K. Li, D. Qiu, H. Han, C. Shu, Y. Yang, *et al.* ACS Appl. Mater. Interfaces **2012**, *4*, 3724-3729.

[31] H. Wu, L. Lin, P. Wang, S. Jiang, Z. Dai, X. Zou, *Chem. Commun.* **2011**, *47*, 10659-10661.

[32] T. A. Ratnikova, P. N. Govindan, E. Salonen, P. C. Ke, *ACS Nano* **2011**, *5*, 6306-6314.

[33] S. Maoyong, J. Guibin, Y. Junfa, W. Hailin, Chem. Commun. **2010**, *46*, 1404-1406.

[34] A. Innocenti, S. Durdagi, N. Doostdar, A. T. Strom, A. R. Barron, C. T. Supuran, *Bioorg. Med. Chem.* **2010**, *18*, 2822-2828.

[35] S. Durdagi, C. T. Supuran, T. A. Strom, N. Doostdar, M. K. Kumar, A. R. Barron, T. Mavromoustakos, M. G. Papadopoulos, *J. Chem. Inf. Model.* **2009**, *49*, 1139-1143.

[36] S.-T. Yang, H. Wang, L. Guo, Y. Gao, Y. Liu, A. Cao, *Nanotechnology* **2008**, *19*, 395101.

[37] X. Zhang, C. Shu, L. Xie, C. Wang, Y. Zhang, J. Xiang, L. Li, Y. Tang, *J. Phys. Chem. C* **2007**, *111*, 14327-14333.

[38] G. Pastorin, S. Marchesan, J. Hoebeke, T. Da Ros, L. Ehret-Sabatier, J.-P. Briand, M. Prato, A. Bianco, A. *Org. Biomol. Chem.* **2006**, *4*, 2556-2562.

[39] B. Belgorodsky, L. Fadeev, J. Kolsenik, M. Gozin, *ChemBioChem* **2006**, *7*, 1783-1789.

[40] B. Belgorodsky, L. Fadeev, V. Ittah, H. Benyamini, S. Zelner, D. Huppert, A. B. Kotlyar, M. Gozin, *Bioconjugate Chem.* **2005**, *16*, 1058-1062.

[41] T. Mashino, K. Shimotohno, N. Ikegami, D. Nishikawa, K. Okuda, K. Takahashi, S. Nakamura, M. Mochizuki, *Bioorg. Med. Chem. Lett.* **2005**, *15*, 1107-1109.

[42] K. H. Park, M. Chhowalla, Z. Iqbal, F. Sesti, *J. Biol. Chem.* **2003**, 278, 50212-50216.

[43] M. Calvaresi, F. Zerbetto, *J. Mater. Chem. A* **2014**, *2*, 12123-21135.

[44] M. Calvaresi, F. Zerbetto, *Chem.-Eur. J.* **2012**, *18*, 4308-4313.

[45] a) V. Zorbas, A. L. Smith, H. Xie, A. Ortiz-Acevedo, A. Dalton, G. R. Dieckmann, R. K. Draper, R. H. Baughman, I. Musselman, *J. Am. Chem. Soc.* **2005**, *127*, 12323-12328; b) H. Xie, E. J. Becraft, R. H. Baughman, A. B. Dalton, G. R. Dieckmann, *J. Pept. Sci.* **2008**, *14*, 139-151; c) S. M. Tomasio, T. R. Walsh, *J. Phys. Chem. C* **2009**, *113*, 8778-8785; d) L. Zheng, D. Jain, P. Burke, *J. Phys. Chem. C* **2009**, *113*, 3978-3985.

[46] a) D. Canevet, E. M. Perez, N. Martin, *Angew. Chem., Int. Ed.* **2011**, *50*, 9248-9259; b) E. M. Perez, N. Martin, *Chem. Soc. Rev.* **2008**, *37*, 1512-1519.

[47] A. Volkamer, A. Griewel, T. Grombacher, M. Rarey. *J. Chem. Inf. Model.* **2010**, *50*, 2041-2052

[48] M. Calvaresi, F. Zerbetto, *J. Chem. Inf. Model.* **2011**, *51*, 1882-1896.

[49] W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, M. L. Klein *J. Chem. Phys*. **1983**, *79*, 926-935.

[50] D. A. Case, T. E. Cheatham, III; T. Darden, H. Gohlke, R. Luo, K. M. Merz, Jr. A. Onufriev, C. Simmerling, B. Wang, R. Woods, *J. Comput. Chem.* **2005**, *26*, 1668-1688.

[51] P. A. Kollman, I. Massova, C. Reyes, B. Kuhn, S. Huo, L. Chong, M. Lee, T. Lee, Y. Duan, W, Wang, *et al. Acc. Chem. Res.* **2000**, *33*, 889-897.

[52] B. R. Miller, III; T. D. McGee, Jr.; J. M. Swails, N. Homeyer, H.

[53] Gohlke, A. E. Roitberg, *J. Chem. Theory Comput.* **2012**, *8*, 3314-3321.

2. A Computational Analysis of a novel Carborane-based Vitamin D Receptor ligand

2.1. Introduction

The term "carborane" applies to a class of boron clusters made of boron, hydrogen and carbon atoms. They have polyhedral molecular structures very similar to those of the isoelectronic polyhedral boranes. These clusters, however, have non-classical binding interactions, which result in a complex overall electronic structure in which both carbon and boron are hexacoordinate. These clusters, being polyhedral, are classified as closo-, nido-, arachno- etc based on whether they represent a complete polyhedron (closo-), or a polyhedron that is missing one (nido-), two (arachno-) or more vertices. The most interesting examples of carboranes are the icosahedral closo-carboranes (see **Fig.2.1**) which are neutrally charged and extremely stable due to a sort of spherical aromaticity, where the electrons being delocalized on the 3D structure of the compound.[1] Icosahedral closo-carboranes have a rigid, nearly spherical shape which occupies a space almost 50% larger than that of a rotating phenyl moiety.[2] They also exhibit unique organo-mimetic properties: their chemical reactivity matches that of classic organic molecules, even if structurally they are more similar to metal-based inorganic or organometallic species.
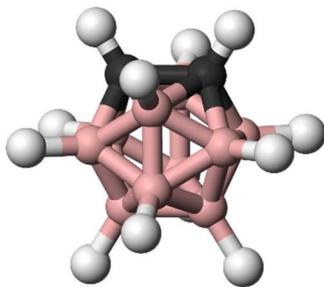
Fig.2.1. Three-dimensional representation of an icosahedral closo-carbornare.

Carboranes have been firstly produced in the '50s, but the results were not published until 1963.[3] Carboranes in medicine are employed for Boron Neutron Capture Therapy, or BNCT,[4] which is an anti-tumoral therapy based on the nuclear capture and fission that occurs when natural non-radioactive ^{10}B nuclei are irradiated with neutrons of the energy needed to form excited $^{11}B^*$, which decay into high energy alpha particles and ^{7}Li nuclei. Both products cause strong ionization in their vicinity, killing the targeted tumoral cell and leaving the surrounding tissue unharmed. Carboranes also exhibit unique characteristics that make them interesting molecules for drug design and

development, since their scaffold is unfamiliar to biological systems. Bacteria and viruses are developing a degree of resistance against the most common drugs in use and this trend can only be reduced with the introduction of new and more efficient drugs. Because of their unique properties, carboranes are arising great attention in this research area. The hydrogen atoms on the cage carry a partial negative charge conferring them a hydric character. The presence of these hydric hydrogens affects carboranes' behaviour in biological environment. First, the partial negative charge prevents them from participating in classical hydrogen bonds, giving the clusters hydrophobic characteristics. Second, it promotes hydrophobic interaction with biological macromolecules.[5] In addition, it allows the clusters to form unconventional dihydrogen bonds - far weaker than classical hydrogen bonds - between their hydric hydrogens and a proton donor. Another useful feature of carborane-based drugs is their increased stability in biological environment. Although no study has been made concerning the metabolism and elimination of carboranes, it's safe to assume that no specific enzymatic system has been evolved to metabolize these boron clusters, them being so unfamiliar to biological life.

Moreover, the rigid 3D arrangement of the carboranes opens new exciting possibilities for drug design. Theoretically, it is possible to arrange substituents in specific spatial configurations, allowing the construction of molecules that can be designed to their biological targets, so that carborane-based drugs may result more versatile scaffolds than the 2D benzene systems. Carboranes can also be chiral or prochiral, so decoration of these clusters with chiral organic components adds an additional chirality factor to the potential drug. A brilliant example is the experimental work carried out by Fujii,[6] who developed a new Vitamin D Receptor (VDR) ligand with a non-secosteroidal skeleton based on the carborane cage (see right molecule in **Fig.2.2**). Naturally, VDR binds to the metabolically activated form of vitamin D, known as calcitriol, and this complex, once bound with DNA, regulates the activity of genes involved in calcium and phosphate homeostasis, bone metabolism, immune regulation, cell proliferation and differentiation.[7,8] So, by turning this gene off and on, VDR and its ligands play an important role in the pathogenesis and therapy of diseases such as osteoporosis, arthritis, psoriasis and even cancer.

The natural VDR ligand calcitriol's main function, when bound with VDR, is to increase calcium level in blood by promoting its adsorption and re-adsorption and by stimulating its release from bones. Calcitriol is largely used as a drug to cure vitamin D deficiency, hypocalcaemia, osteoporosis and other calcium-related diseases. The pharmacophore units on calcitriol to effectively bind to VDR

are three appropriately positioned hydroxyl groups and a hydrophobic core with the right bulkiness, which, in a secosteroidal structure, corresponds to the CD ring (see **Fig. 2.2**).
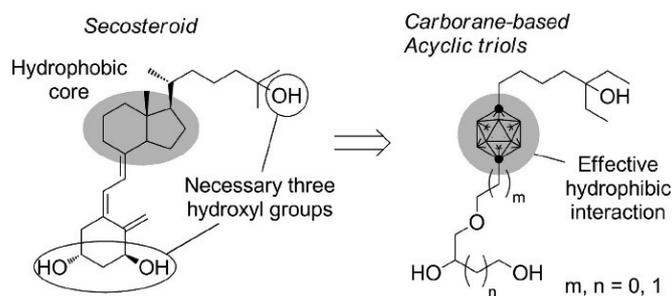
Fig.2.2. The calcitriol and the non-secosteroidal analogue (S)10.

To overcome the problems related to the synthesis of calcitriol and its derivative due to the presence of the secosteroidal core, Fujii synthetized the carborane-based analogue of calcitriol (S)10 depicted on the right side of **Fig.2.2**. The secosteroidal nucleus of calcitriol has been substituted with a para-carborane cage. The groups necessary for the binding with VDR were added to the carborane moiety, to maintain the pharmacophore units on the alternative drug. X-Ray experiments determined the crystal structure of the complexes calcitriol-VDR (PDB 1RK3) and carborane analogue (S)10-VDR (PDB 3VJS).[6]

Experimental evidences reported that the binding affinity of (S)10 toward VDR resembles that of the natural ligand calcitriol,[6] but the driving interactions that govern the binding with the biological target are still to be elucidated. In general, the development of new carborane-based drugs faces many obstacles, among which the difficulty in rationalizing the interactions that govern the binding of these unconventional scaffolds with biological receptors. *In silico* evaluations of binding affinities are valid approaches to explain and rationalize the forces that drive the binding process between a ligand and its receptor. In this work we endeavour to generate an *ad hoc* force field for the carborane moiety and then, by studying the binding energy of calcitriol and (S)10 towards VDR, we seek to give insights into the role of the carborane cage during the binding process.

2.1. Carborane Force Field Parameterization

Up to date, no suitable force field to describe carboranes is available in literature. The starting point for our parametrization work is the study made by Oda in occasion of the 37th Symposium on Chemical Information and Computer Sciences at Toyohashi (JP).[9] Unfortunately, due to translation issues, we couldn't recreate the experimental conditions that lead Oda's team to their results. We chose, however, to use the set of parameters they found as a mean of comparison with our own

results. We conducted two type of parametrizations, one based on the harmonic theory and one exploiting the VFFDT software.[10] The resulting parameters were then tested via MM minimizations, employing the AMBER suite.[11]

2.1.1 Harmonic parameterization. Harmonic parametrization is conducted to find the force constant of a chosen parameter. The parameters we are interested in are bond lengths and angle amplitudes. The equilibrium values (**Table 2.1**) are found by running quantum-mechanical geometry optimizations at a DFT level, employing the Minnesota functional M06-2X and an accurate 6-311++G** basis set, as implemented in Gaussian 09 program.[12]

Table 2.1. Bond and angle equilibrium values obtained by geometry optimizations.
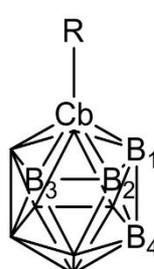
Bond	Equilibrium Bond lenght [Å]
B1-B2	1.76
B1-B4	1.76
B1-Hb	1.18
Cb-Hc	1.08
Cb-B1	1.70

Angle	Equilibrium Angle amplitude [degrees]
B1-B2-B3	108.0
B1-B2-B4	59.7
B1-B2-Hb	119.9
B1-Cb-Hc	117.0
B1-Cb-B2	63.2
B1-B2-Cb	58.4
B1-Cb-B3	115.9
B3-B2-B4	108.1
B1-B4-Hb	126.1
B1-B4-Cb	103.8
B1-B4-B2	60.5
Cb-B1-Hb	120.9

For each bond (or angle) we then operated a series of expansions and restrictions adding 0.01, 0.02 and 0.03 Å (1, 2 and 3 degrees for angles) to the equilibrium value, then we froze each new geometry and run an optimization obtaining the energy for each modified geometry. We used the results thus found to construct the parables, with the energy expressed in kcal mol[-1]. For standard AMBER force

fields,[11] angle amplitude is expressed in degrees, but the relative force constant is expressed in *kcal/mol*[*rad*]². So, in the construction of the angle parables, we chose to convert the degrees in radians. The resulting force constants, bond and angle values can be found in **Table 2.2 and 2.3,** where they are also compared with Oda's parameters.[9]

Table 2.2. Complete parameters set for p-carborane moiety bonds.

Bond	Harmonic results		Oda's results	
	Force constant [kcal mol-1Å-2]	Bond length [Å]	Force constant [kcal mol-1Å-2]	Bond length [Å]
B1-B2	153.393	1.78	97.98	1.74
B1-B4	154.728	1.76	97.98	1.77
B1-Hb	295.631	1.18	298.2	1.18
Cb-Hc	295.631	1.08	419.4	1.09
Cb-B1	151.663	1.70	105.8	1.69

Table 2.3. Complete parameters set for p-carborane angles.

Angle	Harmonic results		Oda's results	
	Force constant [kcal mol-1Å-2]	Angle Amplitude	Force constant [kcal mol-1Å-2]	Angle Amplitude
B1-B2-B3	162.929	108.0	32.012	107.5
B1-B2-B4	279.999	59.7	12.523	56.5
B1-B2-Hb	37.4368	119.9	17.334	119.7
B1-Cb-Hc	50.1668	117.0	22.498	117
B1-Cb-B2	260.357	63.2	16.295	66.4
B1-B2-Cb	259.295	58.4	16.295	57.3
B1-Cb-B3	176.527	115.9	38.492	114.5
B3-B2-B4	180.965	108.1	0	107.9
B1-B4-Hb	38.592	126.1	17.334	123.2
B1-B4-Cb	187.335	103.9	0	104.8
B1-B4-B2	283.664	60.5	12.523	66.5
Cb-B1-Hb	42.1627	120.9	20.387	119.4

The most noteworthy feature in the comparison between these two sets of parameters is the fact that, while the equilibrium values are very similar to each other, our force constants are sensibly larger than those found by Oda *et al.*[6] This is particularly relevant in the case of the angle parameters, where the difference between the two set of constants can exceed 200 *kcal/mol*[*rad*]². Because of these divergences, we decided to adopt an alternative route to obtain the force constants.

2.1.2 VFFDT parametrization. The VFFDT software[10] analyses a vibrational frequency output to obtain force constants and equilibrium values for all the angles and bonds of small molecules. We run frequency calculations on the previously optimized structure at the same level of theory and basis set (M06-2x/6-311++G**)[12] used in the harmonic parametrization. The parameters resulting from the VFFDT software are reported in **Table 2.4** and **2.5**.

Table 2.4. Complete parameters set for p-carborane bonds.

Bond	VFFDT results		Oda's results	
	Force constant [kcal mol-1Å-2]	Bond length [Å]	Force constant [kcal mol-1Å-2]	Bond length [Å]
B1-B2	89.38	1.78	97.98	1.74
B1-B4	90.03	1.76	97.98	1.77
B1-Hb	279.92	1.18	298.2	1.18
Cb-Hc	389.35	1.08	419.4	1.09
Cb-B1	90.7	1.70	105.8	1.69

Table 2.5: Complete parameters set for p-carborane angles.

Angle	VFFDT results		Oda's results	
	Force constant [kcal mol-1Å-2]	Angle	Force constant [kcal mol-1Å-2]	Angle
B1-B2-B3	38.943	108.0	32.012	107.5
B1-B2-B4	20.098	59.7	12.523	56.5
B1-B2-Hb	29.039	119.9	17.334	119.7
B1-Cb-Hc	41.254	117.0	22.498	117.0
B1-Cb-B2	44.958	63.2	16.295	66.4
B1-B2-Cb	35.473	58.4	16.295	57.3
B1-Cb-B3	80.915	116.0	38.492	114.5
B3-B2-B4	40.692	108.1	0	107.9
B1-B4-Hb	38.541	126.1	17.334	123.2
B1-B4-Cb	55.311	103.8	0	104.8
B1-B4-B2	20.098	60.5	12.523	66.5
Cb-B1-Hb	39.608	120.9	20.387	119.4

It's worth to notice that, following this method, both the resulting force constants and equilibrium values are very similar to those found by Oda's team. In particular, the angular force constant values are considerably lower than those found with the harmonic parametrization.

2.2.3 MM validation. The parameters test consists in running both a QM geometry optimization[12] and a MM minimization[11] on a modified p-carborane molecule, in which we vary manually bonds and angles (addendum column in **Table 2.6**). During the minimization, the force constant of the varying geometrical parameters will be frozen. Ideally, the difference in energy between the minimum and the modified geometry (ΔE) will be the same for the QM and the MM calculation. In the case of an unfitting force field, however, there would be a noticeable difference between the two ΔE. To better represent the carborane dynamic behaviour, we chose to modify and freeze three parameters simultaneously. We run a series of tests on our harmonic, Oda's and VFFDT parameters (**Table 2.6**)

Table 2.6. Test results for harmonic, Oda's and VFFDT parameterization.

Frozen parameters	Addendum	QM ΔE [kcal mol^{-1}]	MM ΔE [kcal mol^{-1}]		
			Harmonic parameters	Oda's parameters	VFFDT parameters
B1-B3	0.01 Å				
B1-B3-B4	1 degree	0.1	0.1	0.4	0.2
B3-B7	0.01 Å				
B1-B3	0.02 Å				
B1-B7-C2	2 degrees	1.0	4.7	1.7	2.5
B1-B3-B7	2 degrees				

We found that Oda's parameters are far more suitable (smaller ΔE) in describing p-carborane behaviour than the parameters obtained by the harmonic parameterization. Harmonic parameters are fitting only for small modifications of the carborane geometry, but for stronger modifications the ΔE becomes higher and higher. We can ascribe this trend to the fact that our force constants are higher than Oda's, resulting in an increased rigidity of the p-carborane moiety. The harmonic parametrization failed in giving appropriate results because we analysed one parameter at a time, and since the carborane is a cage, where all bonds and angles influence each other, it is not possible to separately modify one parameter without incurring in relevant errors, because every modification has repercussions on the overall structure of the molecule. On the other hand, the parametrization conducted with the VFFDT software is based on a calculation that involves the overall structure of p-carborane. So, unsurprisingly, the results of the MM test validate the parameters found with this

method. Considering both our results and Oda's parameters, we chose to use the force field described by the parameters found with the VFFDT method in order to carry out our MD simulations.

2.3. Computational Details

2.3.1 Setting the Simulation. Two adducts, VDR-(S)10 and VDR-calcitriol, were generated and sodium counterions were included to exactly neutralize the charged complexes (see **Fig.2.3**). All simulations were performed with explicit solvent by using the TIP3P water model.[11] The ff10 force field was used to model standard residue of the protein.[11] The carborane moiety was modeled by the *ad hoc* force field computed in this work. For calcitriol and side chains of (S)10 RESP charges were computed by a standard quantum-mechanical QM HF/6-31-G* calculation.[12]
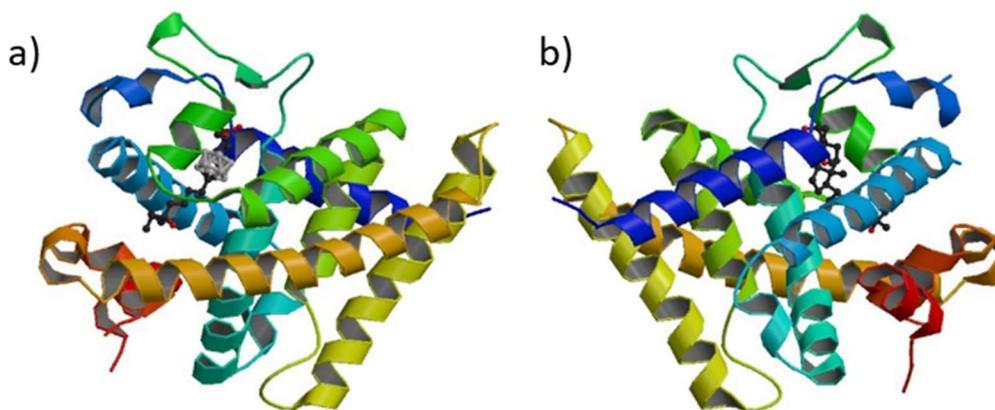
Fig.2.3. Three-dimensional representation of the complexes: **a**) VDR-(S)10 complex (3VJS), b) VDR-calcitriol complex (1RK3).

2.3.2 Minimization and equilibration. About 10,000 steps of steepest descent minimization were performed for the whole system with PMEMD.[11] Particle Mesh Ewald summation[11] was used throughout (cut off radius of 12 Å). Hydrogen atoms were considered by the SHAKE algorithm[11] and a time step of 2 fs was applied in all MD runs. Individual equilibration steps included (i) 50 ps of heating to 298 K within an NVT ensemble and temperature coupling according to Berendsen. (ii) 50 ps of equilibration MD at 298 K to switch from NVT to NPT and adjust the simulation box. Isotropic position scaling was used at default conditions. (iii) 900 ps of continued equilibration MD at 298 K for an NPT ensemble switching to temperature coupling according to Andersen.

2.3.4. Production MD. MD simulation was carried out for the equilibrated system using PMEMD.[11] Simulation conditions were identical to the final equilibration step (iii). Overall sampling time was 100ns for both the complexes.

2.3.3 Post Processing of Trajectories, MM-GBSA.
MM-GBSA[11] analyses were carried out to estimate the binding free energy VDR-(S)10 and VDR-calcitriol complexes. MM/GBSA analysis is a postprocessing method in which representative snapshots from an ensemble of conformations are used to calculate the free energy change between two states (typically a bound and free state of a receptor and a ligand).

2.4. Results and discussion

2.4.1 MD and MM-GBSA results.
The aim of our computational work is to determine if the non-secosteroidal drug (S)10 synthesized by Fujii[6] is a better VDR-ligand than calcitriol. We calculated the binding free energy for both VDR-(S)10 and VDR-calcitriol complexes (called 3VJS and 1RK3 respectively) by post-processing the trajectories, obtained *via* Molecular Dynamics (MD) simulations, with the MM/GBSA method.[11] The computed $\Delta G_{GB,bind}$ for the two complexes are $\Delta G_{GB,3VJS} = -79{,}6$ kcal mol^{-1} and $\Delta G_{GB,1RK3} = -60{,}5$ kcal/mol. The difference between the binding energies of the two complexes is around 20 kcal mol^{-1}, suggesting that VDR prefers to bind (S)10 ligand rather than calcitriol. To better understand the residues that govern the binding process and why there is such a difference in stability between the two complexes, we performed the decomposition of the binding total energy.

2.4.2 Per-residue and Pairwise decomposition.[11]
The per-residue decomposition calculates the energy contribution of a single residue by summing its interaction with the overall system. The comparison between the energy contributes (>1.5 kcal mol^{-1}) of the most interacting residues is reported in **Table 2.7**.

Table 2.7. Per-residue decomposition results. All energies are expressed in kcal mol^{-1}.

Residue	ΔG_{bind} VDR-(S)10	ΔG_{bind} VDR-calcitriol	$\Delta\Delta G$ (absolute value)
LEU 47	-2.7	-1.8	0.9
LEU 50	-1.8	-1.5	0.3
VAL 51	-1.9	-1.1	0.8
SER 54	-1.8	-1.5	0.3
ILE 88	-1.9	-1.9	0.0
SER 92	-2.9	-1.8	1.1

SER 95	-2.2	-0.6	1.6
TRP 103	**-5.3**	**-3.1**	**2.2**
TYR 112	-2.2	-1.4	0.7
HIE 214	-1.2	-1.6	0.4

Except for Ile88 and Hie214 that prefer to interact with calcitriol, the contribute of the selected residues toward the ligand is higher when (S)10 is bound to VDR. The interactions of residue Trp103, with a difference value that surpasses 2 kcal mol^{-1}, are here analysed in detail through a pairwise decomposition which calculates the contributed to the interaction energy between a specific pair of residues (Trp103 and (S)10 and Trp103 and calcitriol).

Table 2.8. Pairwise decomposition of interaction between Trp103-S(10) and Trp103-calcitriol. All energies are expressed in kcal mol^{-1}.

Contribution	Interaction Trp103-S(10) [kcal mol^{-1}]	Interaction Trp103-calcitriol [kcal mol^{-1}]
VDWaals	-4.2	-2.7
Electrostatic	-0.3	-0.2
Polar Solv.	-0.5	-0.1
Non-Polar Solv.	-0.5	-0.6
Total	-5.6	-3.6

We can see from **Table 2.8** that the highest contribution comes from the van der Waals interaction energy. In both complexes, these interactions are contacted between the indole ring of Trp103 and the hydrophobic core of the two ligands respectively. Both ligands interact with the tryptophan in a $\pi - \pi$ stacking fashion, but the hydric hydrogens increase the interaction.

2.5. Conclusions

In this study, we successfully developed an *ad hoc* force field for the p-carborane moiety by using the new VFFDT software developed by Zheng *et al.*[10] We then used the computed force field for running MD simulation on the complex VDR-(S10) and VRD-calcitriol in order to determine whether this new drug could be a better ligand than calcitriol. We can suggest, from our results, that the formation of the VDR-(S)10 complex is energetically more favourable than the formation of the VDR-calcitriol complex. Subsequent *per-residue* and *pairwise* analyses highlighted which residues of

the VDR are most involved in the binding process and demonstrated that the carborane cage, differently from the CD ring of the secosteroidal calcitriol, has an active role in the binding process. In particular, the peculiar hydric hydrogens of the carborane cage favourably interact with residues Leu47, Leu50, Ser92, Trp103 and Tye112 of the VDR receptor. We trust that our results could facilitate the computational studies of carborane and carborane derivatives, by employing the force field generated in this work.

References

[1] King, R. B. *Chem. Rev.*, **2011**, *101(5)*, 1119 - 1152.

[2] Valliant, J. F.; Guenther, K.; King, A. S.; Morel, P.; Schaffer, P.; Sogbein, O. O.; Stephenson, K.A. *Chem. Rev.*, **2002**, *1*, 173-230.

[3] Heying, T. L.; Ager, J. W.; Clark, S. L.; Mangold, D. J.; Goldstein, H. L.; Hillman, M.; Polak, R. J., Szymanski, J. W. Inorganic Chemistry, **1963**, *2(6)*, 1089 - 1092.

[4] Hawthorne, M. F.; Young, D. C.; Wegner, P. A. *J. Am. Chem. Soc.*, **1965**, *87(8)*, 1818 – 1819.

[5] Lesnikowski, Z. J. J. *Med. Chem.*, **2016**, *59*, 7738 - 7758.

[6] Fujii, S.; Masuno, H.; Taoda, Y.; Kano, A.; Wongmayura, A.; Nakabayashi, M.; Ito, N.; Shimizu, M.; Kawachi, E.; Hirano, T.; Endo, Y.; Tanatani, A.; Kagechika, H. *J. Am. Chem. Soc.*, **2011**, *133*, 20933 - 20941.

[7] Thomas, S.L.; Rene, C.; Sandra, R.; Adams, J.; Hewison, M. *J. Bone Miner. Res.*, **2013**, *28*, 1478-1488.

[8] Fleet, J. *Crit. Rev. Clin. Lab. Sci.*, **2010**, *47*, 181-196.

[9] Oda, A.; Ohta, K.; Endo, Y.; Fukuyoshi, S. *37th Symposium on Chemical Information and Computer Sciences, Toyohashi*, **2014**, P02.

[10] Zheng, S.; Tang, Q.; He, J.; Du, S.; Xu, S.; Wang, C.; Xu, Y.; Fu, L. *J. Chem, Inf. Model.*, **2016**, *56*, 811 - 818.

[11] D. A. Case, T. E. Cheatham, III; T. Darden, H. Gohlke, R. Luo, K. M. Merz, Jr. A. Onufriev, C. Simmerling, B. Wang, R. Woods, *J. Comput. Chem.*, **2005**, *26*, 1668-1688.

[12] Gaussian 09, Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; et al Gaussian, Inc., Wallingford CT, **2009**.

Part VI

Conclusions

To a conclusion to this thesis, a brief résumé of the results reported in the previous chapters is given.

(i) Carbocatalysis. We studied the oxidative dehydration of ethylbenzene to styrene catalysed by oxidized carbonanotubes. We considered either epoxy groups on the tube surface or carbonyl groups at the edges of carbon nanotubes. Our full QM DFT computations revealed that the process can occur along three pathways: the first two bifurcate after the initial transformation of an epoxide into a hydroxyl group, which occurs via a biradical transition state and the formation of a benzyl-like intermediate. The epoxide can further react to release styrene and form a water molecule, as observed experimentally, via a highly exothermic process. Alternatively, in the presence of a second adjacent epoxide, styrene is produced without water formation along a less exothermic pathway that leaves two hydroxyl groups on the nanotube surface. Along the third pathway, two adjacent carbonyl groups (quinone functionality) at the edge promote the formation of styrene, with energy barriers similar to those calculated in the presence of epoxy groups.

The second example of carbocatalysed process reported in this thesis is the bromination reaction of *N*-phenylacetamide inside CNTs (used as nanoreactors), either in water or in an aprotic solvent (ethylbenzene). A full QM and a QM/MM approach was used. We endevoured to study the bromination process in different solvents and inside CNTs, to understand the effects of the confinement on the reaction mechanism. In solution (aprotic solvent), a Wheland intermediate exists only for the attack at the *ortho* position, while the *para* attack proceeds in a concerted manner. The reaction is catalyzed by the HCl by-product, which lowers significantly the activation barriers. The *ortho* product is favored, in contrast to the common belief based on simple steric effects. In water solution a Wheland intermediate was located for both *ortho* and *para* attacks. In this case the ion pair is stabilized by the polar protic solvent. Studyin the reaction inside CNTs, as found in aprotic solvent, the Wheland-type arenium ion exists only along the *ortho* pathway. The initial production of the HCl by-product activates rapidly the catalysed mechanism that proceeds almost exclusively along the *para* reaction path. The almost exclusive *para* regioselectivity for the CNT-confined reaction and its acceleration with respect to water are due to van der Waals interactions between the endohedral system and the electron cloud of the surrounding CNT. The effect of these interactions

was estimated quantitatively within the UFF scheme, which is the force-field used in our QM/MM computations. We found that these interactions are particularly stabilizing for the catalyzed *para* process.

(ii) Enzyme catalysis. We used a full QM approach to investigate the catalytic mechanism of *Linalool dehydrase-isomerase* (Ldi). The preliminary results obtained from our computations demonstrate that if the isomerization of geraniol to linalool follows a concerted path, high activation barriers have to be surmounted. Because of the unfeasibility of such an energy-demanding process, we hypothesized that the dehydration of linalool is more favourable if Ldi generates *trans-β*-myrcene first and then isomerizes it to the *cis* stereoisomer. We evaluated the importance of the two cysteine residues present in the active site. We found that Cys171 and Cys180 act as hydrogen-bond acceptor and donor, respectively, suggesting that their catalytic role is to stabilize, *via* non-bonding interactions, the transition state. Also, we confirmed the importance of His129 and Tyr65. Our results suggested that at least one water molecule is crucial in the catalysis, since it assists various proton transfers and/or directly acts as a nucleophile. Then, when a different protonation state is considered for Cys180, this residue protonates the leaving group during the carbocationic intermediate formation. Furthermore, we suggested that a negative charged environment facilitate the formation of a carbocationic intermediate stabilized by electrostatic interactions found in the surroundings. Further investigations are being carried out to confirm our hypotheses.

The second enzymatic process examined in this thesis is the reaction catalyzed by α-glucan-lyase, which was studied using a full QM DFT approach. We found that the entire process can be divided into three phases: glycosylation, deglycosylaton-elimination and tautomerization, and that the rate-determining step corresponds to deglycosylaton-elimination. We also investigated the tautomerization of the substrate in aqueous solution and finding that this step is much more favored inside the enzyme, before leaving the active site.

Preliminary results for the reaction mechanism for the sulfurylation of IdoA or GlcA units catalyzed by 2-*O*-sulfotransferase (2OST) were also reported. A hybrid QM/MM approach was employed, and our computations revealed that the transfer of the SO_3^- group from the donor PAPS to the saccharide units occurs in one kinetic step, characterized by a S_N2-like character. The transition region involves the formation of a bi-pyramidal trigonal sulfate, stabilized by electrostatic interactions with the active site residues. For both sulfurylation reactions, we found that His142 acts as the catalytic base, activating the nucleophile by abstracting the hydroxyl proton. The guanidium group of Arg80 stabilizes the sulfuryl group that is dethatching from the donor, engaging a strong H-bond with SO_3^-.

Our preliminary results also confirmed that 2OST prefers to perform the sulfurylation on iduronic acid unit rather than on glucuronic acid.

(iii) Organometallic catalysis. In Chapter III, we reported the DFT computational investigation on the mechanism of the copper(II) catalyzed C-N, C-O cross coupling reaction involving isatin and phenylglyoxylic acid. Our computations demonstrated that the rate-determining step of the entire process is the initial decarboxylation which generates the initial active organometallic complex. Then, deprotonation of the isatin nitrogen is carried out by a copper ligand, enhancing the isatin nucleophilic character triggering the nucleophilic attack of nitrogen on the Cu atom. Then, a reversible transformation connects the initial complex and an intermediate where the metal atom is inserted into the amide bond. Copper does not change its oxidation state (II) in the course of the catalytic reaction. Our calculations depicted a scenario where the mechanism never involves directly the isatin benzene ring, consistently with the fact that only slight changes in the reaction yield are found when substituents are inserted at different positions of the aromatic ring. Also, we demonstrated that the acetate ligands have a crucial role in the reaction, allowing the formation of the active complex and performing the deprotonation of the nucleophilic nitrogen of isatin.

(iv) Organo-catalysis. We investigated the Hajos-Parrish-Eder-Sauer-Wiechert reaction employing DFT QM methods. Our aim was to solve some open issues that, in our opinion, had not been elucidated by previous investigations. We were able to determine the rate-determining step of the whole process, suggesting that the enamine formation is the slow stage of the reaction. We also confirmed the role of proline in the asymmetric catalysis of unsatured α-β-diketones.

The formation of LAC (hydroxylactone (1*R*,5*S*)-1-hydroxy-3,6-dioxabicyclo [3.2.1] octan-2-one) from cellulose was investigated at the QM DFT level. We demonstrated the existence of two reaction paths starting from ascopyrone *P*. We found that a series of internal rearrangements involving in all cases a proton transfer leads directly to LAC. We also found an alternative path from which, via a "gate" connecting the two reaction channels, it is possible to reach the first path and form LAC. In both cases the rate-determining step of the process is the initial keto-enol isomerization. We found that water, which is present in the reaction mixture, catalyzes the reaction by assisting the proton transfers present in all the steps of the process. The results of this study on the non-catalyzed process demonstrated the important role played by water in the formation of pyrolysis products of cellulose where proton transfer is a key mechanicistic step.

(v) Side-works. In Chapter IV, two side projects are described. In particular, we studied the binding of lysozyme to fullerene, and we endeavored to increase the assembly affinity by engineering the

active site of the protein. We mutated various residues, present in the binding site of the protein, to tryptophan and carried out Molecular Dynamics simulations and post-processing analyses (MM-PBSA) on the assemblies. We found that the single mutation to tryptophan does guarantee the improvement of the binding of proteins to fullerene.

Finally, we have carried out the development of an *ad hoc* force field for the *p*-carborane moiety and the MD simulations on the complexes formed by the vitamin *D* receptor (VRD) and calcitriol or a *p*-carborane based ligand. We computed that VDR preferably binds to the synthetic ligand than calcitriol. Subsequent *per-residue* and *pairwise* analyses highlighted that the carborane cage has an active role in the binding process. In particular, we found that the hydric hydrogens of the carborane cage favourably interact with residues of the protein.

List of Figures

List of Schemes

List of Tables