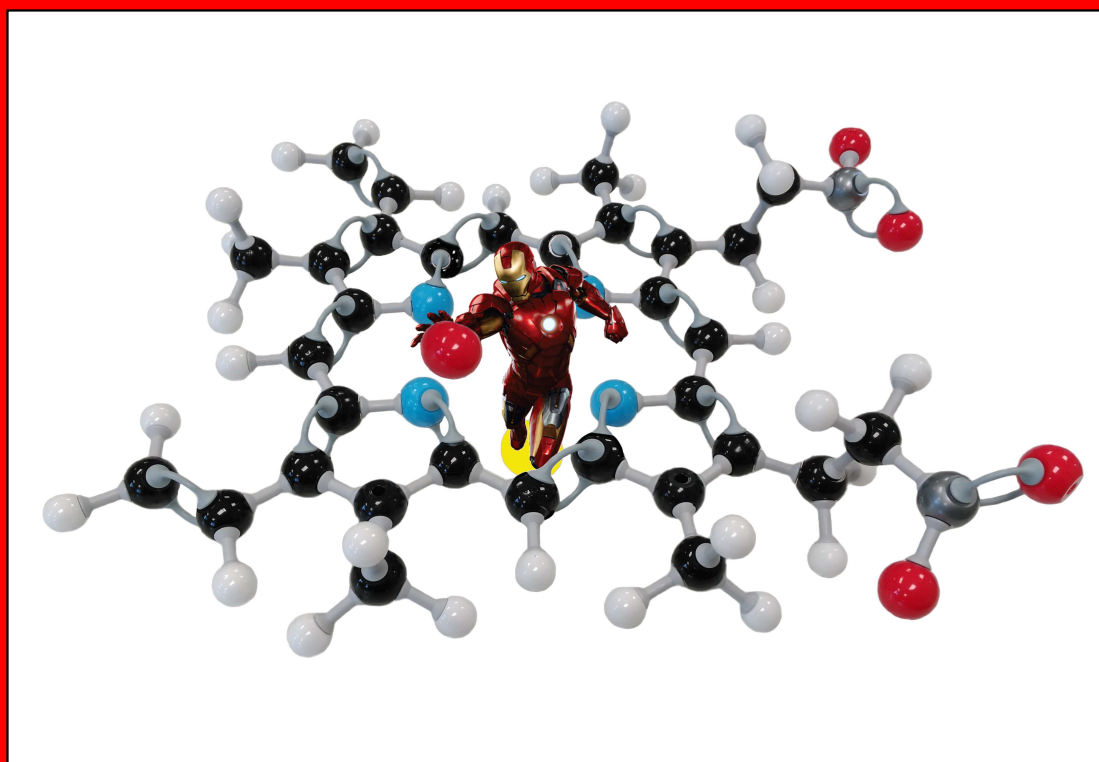


Departament de Química Física i Analítica

Àrea de Química Física



A Theoretical Study on the Mechanism of the Oxidation of Substrates by Human Aromatase Enzyme (CYP19A1)



Ignacio Viciano Gonzalo

Tesis Doctoral – PhD Thesis

Castellón, Julio 2016

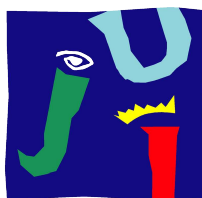
TESIS DOCTORAL

**Programa de Doctorat en Química Teòrica i Modelització
Computacional**

**A Theoretical Study on the Mechanism of the
Oxidation of Substrates by Human Aromatase
Enzyme (CYP19A1)**

Ignacio Viciano Gonzalo

Castellón, Julio 2016



**UNIVERSITAT
JAUME·I**

Escola Superior de Tecnologia i Ciències Experimentals

Departament de Química Física i Analítica

Àrea de Química Física

Sergio Martí Forés, profesor titular del Departamento de Química Física y Analítica de la Universitat Jaume I de Castellón y **Raquel Castillo Solsona**, profesora titular del Departamento de Química Física y Analítica de la Universitat Jaume I de Castellón,

CERTIFICAN:

Que el trabajo con título “*A Theoretical Study on the Mechanism of the Oxidation of Substrates by Human Aromatase Enzyme (CYP19A1)*” ha sido realizada por Ignacio Viciano Gonzalo bajo nuestra dirección, para optar al grado de doctor en Química.

Así, autorizan la presentación de este trabajo a efectos de seguir los trámites correspondientes de la Universitat Jaume I de Castellón.

Castellón de la Plana, a 21 de Junio de 2016.

Dr. Sergio Martí Forés

Dra. Raquel Castillo Solsona

Director de Tesis

Directora de Tesis

A mi familia

A Ana

“A theory can be proved by experiment; but no path leads from experiment to the birth of a theory.”

Albert Einstein

Acknowledgements/Agradecimientos

Se que es muy complicado agradecer a todas las persona que de una u otra manera han colaborado o influido en el desarrollo de esta Tesis Doctoral, por esa razón pido disculpas de antemano si olvido a alguien.

En primer lugar agradecer que esta Tesis Doctoral a mis directores de Tesis, Sergio Martí (Xexo) y Raquel Castillo, por la ayuda y confianza que me han brindado durante todo este tiempo. En especial a Xexo, puesto que ha sido mi mentor científico y porque sus conocimientos en química computacional y en programación han sido cruciales para sacar esta tesis adelante.

Además quiero extender mi agradecimiento a toda la gente que ha “participado” de una u otra forma en el desarrollo de la Tesis:

Al profesor Olivier Parisel, por ser mi tutor durante los 3 meses de estancia en el *Laboratoire de Chimie Theorique* de la UPMC de Paris (Francia). A los estudiantes de Máster y Doctorado de dicho laboratorio: Eléonor, Robin, Gaël, Amélie, Abdou, Mathilde y Lauréline, porque me facilitaron las cosas y porque me hicieron sentir totalmente acogido durante mi estancia allí.

Al profesor Victor Guallar, por ser mi tutor durante los 3 meses de estancia en el *Barcelona Supercomputing Center* (BSC) de Barcelona (España). También a los estudiantes de Doctorado y los Postdocs del departamento de *Life Sciences*: Fátima, Pedro, Mark, Óscar (pinche), Armin, Enmanuele, Didier, Iain, Ryoji, Sandra e Israel, por la ayuda que me brindaron tanto científica como administrativamente, así como por la amistad que me ofrecieron.

Agradezco a Merche, la secretaria del departamento, por la ayuda que tan eficientemente me ha brindado en los trámites administrativos.

A los profesores y personal del departamento de Química Física y Analítica que me han asesorado tanto científica como burocráticamente, así como en mis inicios en la docencia universitaria: Vicente Moliner, Sixte Safont, Mónica Oliva, Rosa LLusar, Armando Beltrán y Juan Andrés. En especial a Vicente Moliner por concederme la beca FPI que he disfrutado durante cuatro años.

Agradezco enormemente a la gente del departamento (grupo de whatsapp “Grupo QTC”) cuyo apoyo ha sido esencial durante la Tesis. Han sido compañeros de congresos, de comidas en la cantina y a veces en el Pinar de Castelló, de cenas en “Las

Tascas” y de farra en “La Antigua”, de voley en la playa y paellas en Moncofa. Los que están en el departamento a día de hoy: Javi, Lourdes, Maite, Silvia, Isabel, Kemel; sin olvidar a aquellos buenos compañeros que estuvieron en el departamento: Natal, Pamela, Conchín, Erica y Pato. Especial mención a Pato y a Silvia por el grandísimo apoyo que me han ofrecido, sin por ello desmerecer a nadie.

Agradezco el compañerismo y el apoyo recibido por el grupo de investigación al que pertenezco actualmente: *Computational Biochemistry at UJI* (BioComp), así como al grupo al cual pertenecí previamente: *Química Teórica y Computacional* (QTC).

Agradecer también a los estudiantes de doctorado que hay (y que ha habido) en el grupo “*Modeling Chemical Processes in Biological Environments*” de Valencia dirigido por el profesor Iñaki Tuñón: Juan, Violeta, Manuel, Rafa, Kiril y María.

Y a los miembros del grupo de “Materiales Moleculares” dirigido por la profesora Rosa Llusar en la UJI: Iván, Eva, Tomás, Elena, Carmina y David.

No se me puede olvidar agradecer a la gente que estuvo de estancia en el departamento: Mauro, Jeronimo, Ederley, Felipe, Natalia, Noelia y André.

Ahora toca agradecer a la gente más cercana.... Agradecer a mi familia y amigos cercanos que siempre estarán a mi lado, por aguantarme tanto en los buenos como en los malos momentos que han aparecido durante este periodo. Y como no a Ana, que apareció en mi vida en los últimos (y más duros) años de la Tesis, sin cuyo apoyo y compañía, estos momentos estresantes hubieran sido todavía más difíciles. ¡¡Gracias Anita!!

Finalmente, quisiera extender mi agradecimiento a las fundaciones y organismos que han financiado mi investigación:

- Fundación BANCAIXA - UJI por la beca de investigación que me fue concedida: Proyecto P1.1B2011-23

- Ministerio de Economía y competitividad (MINECO) (Antiguo Ministerio de Ciencia e Innovación (MICINN)) por la beca FPI que me fué concedida: Proyecto CTQ2009-14541-C02

- HPC-EUROPA2 project (European Community - Research Infrastructure Action of the FP7) por la beca de movilidad que me fue concedida, así como por las horas de computación proporcionadas: project number: 228398

- A los proyectos del MINECO (CTQ2012- 36253-C03-01 y CTQ2012-36253-C03-02) y de la Generalitat Valenciana (Prometeo/2009/053 y Prometeo/2014/022).

Contents

Acknowledgements/Agradecimientos	vii
Contents	ix
List of Figures	xi
List of Tables	xiv
List of Acronyms and Abbreviations	xv
1. Introduction	1
1.1. Mechanisms of Enzyme Action	3
1.1.1. An introduction to Enzymes	3
1.1.2. Enzyme Kinetics	6
1.1.3. Enzyme Inhibition	9
1.1.4. Enzymatic Catalysis	11
1.2. Cytochrome P450 Aromatase	18
1.2.1. Cytochromes P450	18
1.2.2. The Active Species of CYP Enzymes: Compound I	21
1.2.3. Human Aromatase Enzyme (CYP19A1)	23
1.2.4. Aromatase and Breast Cancer	31
Bibliography	34
2. Theoretical Background and Methodology	43
2.1. Theoretical Chemistry and Computational Chemistry	45
2.2. Quantum Mechanics Methods	46
2.2.1. Born-Oppenheimer approximation	46
2.2.2. Hartree-Fock method	49
2.2.3. Basis Sets	60
2.2.4. Post Hartree-Fock methods	62
2.2.5. Density Functional Theory	64
2.3. Molecular Mechanics Methods	72
2.3.1. Force fields	72
2.3.2. Treatment of the Non-Bonding Interactions:	80

2.3.3. Periodic Boundary Conditions	83
2.3.4. Ewald Summation Method	84
2.4. Hybrid Quantum Mechanics / Molecular Mechanics	85
2.4.1. Frontier Bonds	89
2.5. Potential Energy Surface	90
2.5.1. Chemical structure, properties and reactivity	91
2.5.2. Kinetic Isotope Effects	97
2.5.3. Molecular Dynamics	106
2.5.4. Free Energy	114
2.6. Topological Analysis of the Electron Localization Function	118
2.6.1. Electron Localization Function	119
2.6.2. Gradient Dynamical System	122
2.6.3. Topological Analysis of the ELF	124
2.6.4. Properties derived from the ELF	126
2.6.5. Thom's Catastrophe Theory	128
2.6.6. Bonding Evolution Theory	129
Bibliography	132
3. Aim and scope of the thesis	139
4. Results	143
4.1. Study of the Reactive Species of the Enzyme Aromatase: Compound I145	
4.1.1. New Insight into the Electronic Structure of Iron(IV)-oxo Porphyrin Compound I. A Quantum Chemical Topological Analysis	145
4.2. First Catalytic Subcycle of the Enzyme Aromatase	165
4.2.1. QM/MM Modeling of the Hydroxylation of the Androstene- dione Substrate Catalyzed by Cytochrome P450 Aromatase (CYP19A1)	165
4.2.2. Joint Use of Bonding Evolution Theory and QM/MM Hy- brid Method for Understanding the Hydrogen Abstraction Mechanism via Cytochrome P450 Aromatase	189
4.3. Study of the Hydroxylation of an Aromatase Inhibitor: Exemestane .	211
4.3.1. Theoretical Study of the Mechanism of Exemestane Hydrox- ylation Catalyzed by Human Aromatase Enzyme	211
5. Discussion	235
5.1. Study of the Reactive Species of the Enzyme Aromatase: Compound I237	
5.1.1. New Insight into the Electronic Structure of Iron(IV)-oxo Porphyrin Compound I. A Quantum Chemical Topological Analysis	237

5.2. First Catalytic Subcycle of the Enzyme Aromatase	239
5.2.1. QM/MM Modeling of the Hydroxylation of the Androstene- dione Substrate Catalyzed by Cytochrome P450 Aromatase (CYP19A1)	239
5.2.2. Joint Use of Bonding Evolution Theory and QM/MM Hy- brid Method for Understanding the Hydrogen Abstraction Mechanism via Cytochrome P450 Aromatase	241
5.3. Study of the Hydroxylation of an Aromatase Inhibitor: Exemestane	244
5.3.1. Theoretical Study of the Mechanism of Exemestane Hydrox- ylation Catalyzed by Human Aromatase Enzyme	244
6. Conclusions	247
7. Future Perspectives	253
Resumen	261
Appendices	279
A. Software Used	281
B. Scientific Production	289

List of Figures

1.1. Example of reaction profiles for catalyzed and uncatalyzed reactions	4
1.2. Haldane model of enzyme kinetics	7
1.3. Effect of substrate concentration on the rate of an enzyme-catalyzed reaction	8
1.4. Types of enzyme inhibition	10
1.5. Fischer's "Lock and key" and Koshland's Induced fit models	11
1.6. A comparison of free energy profiles for enzyme-catalyzed reaction and uncatalyzed reaction	12
1.7. Energy diagrams of the hydrogen-bonded heteroatoms (X and Y) as a function of R_{A-B}	15
1.8. Variation of charge distribution on going from reactants to the transition state in aqueous solution and in the enzymatic environment	15
1.9. Pictures of Tsuneo Omura and Ryo Sato, and carbon monoxide and ethyl isocyanide difference spectra of liver microsomes	18
1.10. Catalytic cycle of Cytochrome P450	20
1.11. Ball-and-stick model and Lewis structure of Compound I	21
1.12. Molecular orbitals diagram for Compound I	22
1.13. Different representations of the enzyme aromatase	23
1.14. Representation of androgens and estrogens	24
1.15. Biosynthesis of estrogens from androgens (catalytic cycle of aromatase)	26
1.16. Groves mechanism for the hydroxylation of substrates by cytochrome P450	26
1.17. Mechanism for the C–C bond cleavage reaction that occurs in the conversion of androgens to estrogens by CYP19A1	27
1.18. Active site of the enzyme aromatase	30
1.19. Role of estrogens in the proliferation of ER+ breast cancer	31
1.20. Third generation aromatase inhibitors	33
2.1. General scheme of the SCF procedure	57
2.2. MO diagram for a six-electron system: RHF singlet, and ROHF and UHF triplet states	59
2.3. Representation of the main interactions involved in a force field	73
2.4. Representation of the internal coordinates	74

2.5.	Harmonic oscillator potential and Morse potential	75
2.6.	Potential generated by the rotation of the dihedral angle formed by the atoms $ijkl$ (ϕ_{ijkl})	76
2.7.	Illustration of the 1–2, 1–3 and 1–4 non-bonding exclusions for a hydrogen atom in an ethane molecule	78
2.8.	Variation of the Lennard-Jones potential with intermolecular distance	79
2.9.	Plots of different truncation functions, $S(r)$: The direct truncation function, the switch function, and the shift function	82
2.10.	Periodic Boundary Conditions	83
2.11.	Partitioning of the QM/MM system	87
2.12.	Representation of the link atom scheme	90
2.13.	Potential Energy Surface where the main stationary structures have been depicted	92
2.14.	Different free energies of activation for protium, deuterium and tri- tium, resulting from their different zero-point energy at the ground state and at the transition state	103
2.15.	Illustration of the tunneling effect and the recrossings	105
2.16.	Potential Energy Surface in the gas-phase and condensed media, and schematic energy profile of a reaction in a condensed media	107
2.17.	Scheme of the algorithm used to perform a molecular dynamics sim- ulation	113
2.18.	An example of a dynamical system: the Lorenz attractor; Representa- tion of a dynamical system where the attractors and the trajecto- ries belonging to their basins have been depicted	124
2.19.	Representation of the electron localization function (ELF) for the formaldehyde molecule (CH_2O)	126
2.20.	Picture of René Thom and Cusp catastrophe model	129
5.1.	Different SSDs along the hydrogen abstraction from the ASD sub- strate via aromatase enzyme, and the turning points (catastrophes) connecting them.	242
7.1.	Second catalytic subcycle of the enzyme aromatase	255
7.2.	Different mechanisms proposed for the second catalytic subcycle of the enzyme aromatase	256

List of Tables

2.1. Nomenclature of the valence basins according to the synaptic order
(connectivity to core basins) 125

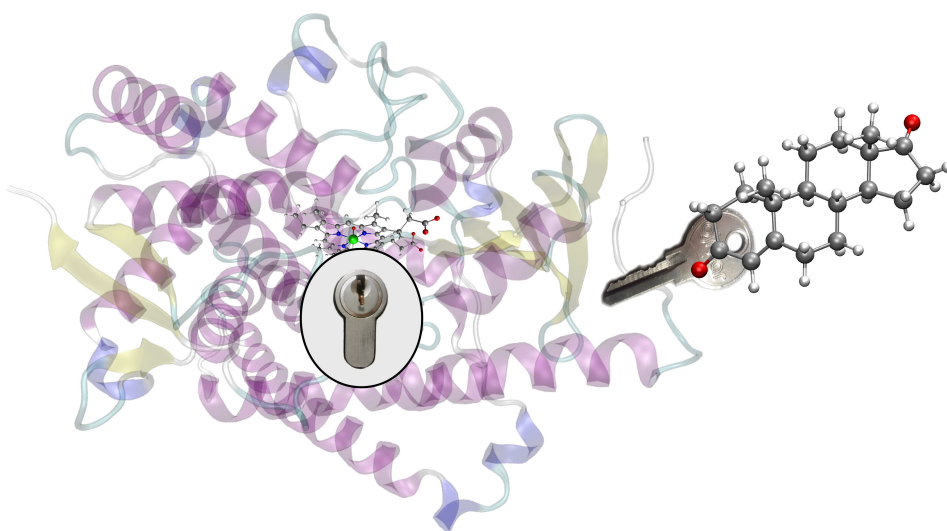
List of Acronyms and Abbreviations

AChE	Acetylcholinesterase	ER+	Estrogen receptor-positive
AI	Aromatase inhibitor	ERα	Estrogen receptor α
AIM	Atoms in molecules	ERβ	Estrogen receptor β
ANZ	Anastrozole	ERE	Estrogen response element
AO	Atomic orbital	EXC	Excited vibrational energy
ASD	Androstenedione	EXE	Exemestane
B	Becke 88 functional	FAD	Flavin adenine dinucleotide
B3LYP	Becke three-parameter Lee-Yang-Parr functional	FEP	Free energy perturbation
BC	Breast cancer	FF	Force field
bend	bending	FFT	Fast Fourier transform
BET	Bonding evolution theory	FR	Fletcher-Reeves algorithm
BFGS	Broyden-Fletcher-Goldfarb- Shanno algorithm	GGA	Generalized gradient approximation
BO	Born-Oppenheimer	GHO	Generalized hybrid orbital
CC	Coupled Cluster	GTO	Gaussian-type orbitals
CG	Conjugate gradient algorithm	GTST	Generalized transition state theory
CGTO	Contracted gaussian-type orbitals	H-GGA	hybrid-generalized gradient approximation
CI	Configuration interaction	HAO	Hydrogenic atomic orbitals
CP	Critical points	HF	Hartree-Fock
Cpd0	Compound 0	HK	Hohenberg-Kohn
CpdI	Compound I	HM-GGA	hybrid-meta-generalized gradient approximation
CPR	Cytochrome P450 reductase	HMM	Henri-Michaelis-Menten
cross	cross-terms	HP	Hartree product
cst	constant	HTST	16 α -hydroxytestosterone
CT	Catastrophe theory	IMOMM	Integrated molecular-orbital/ molecular-mechanics
CYP	Cytochrome P450	impr	impropers
CYP19A1	Cytochrome P450 aromatase	IRC	Intrinsic reaction coordinate
CYP19A1	Cytochrome P450 aromatase	KIE	Kinetic isotope effects
CYP19A1	Cytochrome P450 aromatase	KS	Kohn-Sham
DFF	Davidon-Fletcher-Powell algorithm	L-BFGS	Limited-memory Broyden- Fletcher-Goldfarb-Shanno algorithm
DFT	Density functional theory	LANL	Los Alamos National Laboratory basis set
DIPF	Diisopropylphosphofluoridate	LBHB	Low-barrier hydrogen bond
DNA	Deoxyribonucleic acid	LCAO	Linear combination of atomic orbitals
DS	Dynamical system	LDA	Local density approximation
E1	Estrone	LSCF	Local self-consistent field
E2	17 β -estradiol		
E3	17 β , 16 α -estril		
ECP	Effective core potentials		
elec	Electrostatic		
ELF	Electron localization function		
ER	Estrogen receptor		

LTZ	Letrozole	PP	Pseudopotentials
LYP	Lee–Yang–Parr functional	PR	Polak-Ribiere algorithm
M-GGA	meta-generalized gradient approximation	PW	Perdew–Wang functional
M06	Truhlar-Zhao Minnesota 06 (0.27 HF exchange) functional	PW91	Perdew–Wang 91 functional
M06-2X	Truhlar-Zhao Minnesota 06 (0.54 HF exchange) functional	PW92	Perdew–Wang 92 functional
M06-HF	Truhlar-Zhao Minnesota 06 (HF exchange) functional	QM	Quantum mechanics
MC	Michaelis complex	QM/MM	Quantum mechanics/ Molecular mechanics
MCS	Michaelis complex in solution	QTAIM	Quantum theory of atoms in molecules
MD	Molecular dynamics	R	Reactants
MEP	Minimum energy path	RHF	Restricted Hartree-Fock
MM	Molecular mechanics	RNA	Ribonucleic acid
MMI	Masses and moments of inertia	RNase	Ribonuclease
MO	Molecular orbital	ROHF	Restricted open-shell Hartree-Fock
MP-n	Møller–Plesset perturbation theory of n-order	RP	Reaction path
NAC	Near-attack conformation	RRHO	Rigid-rotor harmonic oscillator
NAD	Nicotinamide adenine dinucleotide	SCF	Self consistent field
NADP	Nicotinamide adenine dinucleotide phosphate	SD	Slater determinant
NADPH	Reduced nicotinamide adenine dinucleotide phosphate	SD	Steepest descent algorithm
nb	non-bonding	SSD	Structural stability domain
NR	Newton-Raphson methods	STO	Slater-type orbitals
ONIOM	Our N-layered integrated molecular orbital and molecular mechanics	stretch	stretching
OPLS	Optimized potentials for liquid simulations	TF	Thomas-Fermi
OPLS-AA	Optimized potentials for liquid simulations - All atoms	TI	Thermodynamic integration
P	Products	TIP3P	Transferable intermolecular potential 3P
P450	Cytochrome P450	TIP4P	Transferable intermolecular potential 4P
P450arom	Cytochrome P450 aromatase	tors	torsional
PBC	Periodic boundary conditions	TP	Turning point
PBE	Perdew–Burke–Ernzerhof functional	TPSS	Tao-Perdew-Staroverov-Scuseria functional
PDB	Protein data bank	TS	Transition state
PEH	Potential energy hyper-surface	TST	Testosterone
PES	Potential energy surface	TST	Transition state theory
PME	Particle mesh Ewald method	TV	Transition vector
Por	Porphyrin	UHF	Unrestricted Hartree-Fock
		VdW	Van der Waals
		VSEPR	Valence shell electron pair repulsion theory
		VSXC	van Voorhis-Scuseria functional
		VWN	Vosko-Wilk-Nusair functional
		WM	Water molecule
		ZPE	Zero-point vibrational energy

Chapter 1

Introduction

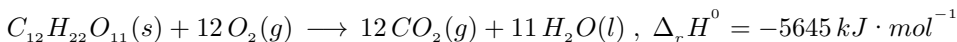


Contents

1.1. Mechanisms of Enzyme Action	3
1.1.1. An introduction to Enzymes	3
1.1.2. Enzyme Kinetics	6
1.1.3. Enzyme Inhibition	9
1.1.4. Enzymatic Catalysis	11
1.2. Cytochrome P450 Aromatase	18
1.2.1. Cytochromes P450	18
1.2.2. The Active Species of CYP Enzymes: Compound I	21
1.2.3. Human Aromatase Enzyme (CYP19A1)	23
1.2.4. Aromatase and Breast Cancer	31
Bibliography	34

1.1. Mechanisms of Enzyme Action

Living beings need energy to perform the physiological processes that support life, such as breathing. One possible source from which to obtain this energy is through the intake of carbohydrates such as sucrose, a molecule found in table sugar. Sucrose is converted into CO_2 and H_2O in the presence of oxygen, according to the following chemical reaction:



Although this reaction is thermodynamically favorable and occurs through an exergonic process, a bag of sugar can remain unchanged for several years without decomposing. However, sucrose is converted into CO_2 and H_2O , and chemical energy immediately after being absorbed by the cells.¹ What is the reason why this process of decomposition occurs in just a few seconds in the human body and is so slow only in the presence of oxygen? The secret lies in enzymes.

1.1.1. An introduction to Enzymes

Enzymes are biological macromolecular entities that are responsible for leading, controlling and accelerating the chemical reactions in biological systems.² These biomolecules are extremely versatile catalysts that catalyze the biochemical reactions very efficiently, transforming one or more compounds (substrates) into one or more different compounds (products), while converting energy from one form into another.^{3,4}

A very representative example of the overall process of catalysis is depicted in Fig. 1.1. As can be observed, an uncatalyzed reaction, $A + B \rightarrow P$, occurs in a single step with high free energy of activation. Conversely, when the same reaction is catalyzed by a catalyst (C), it can occur in three different steps, forming two reaction intermediates (I and J) and the catalyst C is recycled at the end of the process. Each of these reaction steps has a lower activation free energy barrier than the uncatalyzed reaction, which speeds up the overall rate of the reaction.

In fact, enzymes can enhance the rate of the biochemical reaction by factors ranging from 10^6 to 10^{12} with respect to the uncatalyzed reaction,⁶ and sometimes even by a factor of 10^{14} , as in the case of the enzyme urease.⁷ Moreover, since enzymes are catalysts, they are neither consumed nor permanently altered during the biochemical process.¹

In addition, enzymes present a high degree of specificity for their substrates, being able to act on a single substrate or on a small group of them. In fact, enzymes are capable of discriminating between different stereoisomers and even transform nonchiral substrates into chiral products.^{1,3,4} Further features that make

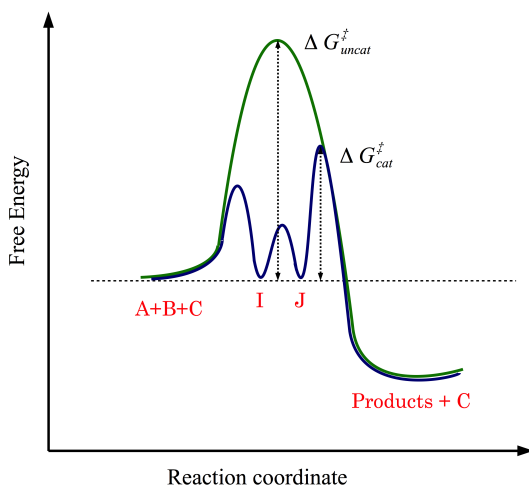


Figure 1.1.: Example of reaction profiles for catalyzed (dark blue line) and uncatalyzed (green line) reactions. Figure adapted from reference [5].

enzymes more versatile than other catalysts are their capacity for regulation and their ability to work in aqueous solutions under conditions that are compatible with life (i.e., mild temperature, neutral pH and atmospheric pressure).⁶ However, it is interesting to highlight that both pH and temperature do have an effect on enzymatic catalysis; the relationship between pH and activity depends on the acid-base behavior of the enzyme. Most enzymes show an optimum pH for which their activity is maximum, usually neutral pH, and above or below this pH the activity decreases sharply.⁷ On the other hand, the reaction rates of the enzyme-catalyzed reactions increase with the temperature, as occurs in other chemical reactions. In an enzymatic reaction, the rate increases until an optimum temperature is reached, usually around human body temperature ($37.5^\circ C$). However, enzymatic activity decreases drastically once this critical temperature is reached, which is related to the thermal denaturation of the enzyme.⁸

Practically all enzymes are proteins, with the exception of a small group of catalytic RNA or ribozymes.³ These are built up of 20 different species of amino acids linked together to form one or more long chains. The molecular weight of these macromolecules ranges from ten thousand to more than one million Daltons.⁷ The success of the catalytic activity of enzymes lies in their protein structure at all levels (primary, secondary, tertiary and quaternary) and therefore, if it is denatured or broken down into its subunits, its efficiency as a biological catalyst is usually lost.¹ An enzyme provides the optimal environment in which the biochemical reaction occurs more rapidly. This environment, called the *active site*, usually takes the form of a three-dimensional pocket or a cleft that consists of a set of amino acid residues where the substrates fit in a very precise way.^{1,7} Catalysis takes place at the

active site, whose surface consists of amino acid residues with substituent groups that bind the substrate and catalyze its biochemical conversion into products or intermediates.⁹

Some enzymes require small non-proteic molecules or metal ions that participate either in substrate binding or catalysis, such as *cofactors*, *coenzymes*, and *prosthetic groups*.⁴ Cofactors are substances of a non-protein nature, which despite being catalytically inactive, are essential for ensuring that certain enzymes are able to perform their catalytic activity. The cofactor binds to a catalytically inactive protein structure called an apoenzyme leading to an apo-cofactor complex called holoenzyme, which is catalytically active.⁶ There are two different kinds of cofactors: 1) inorganic metal ions such as Co^{2+} , Mg^{2+} , Zn^{2+} , Mn^{2+} , Cu^{2+} , Se^{2+} , etc., and 2) coenzymes such as *FAD*, *NAD⁺*, *NADP⁺*, etc.^{1,4,6} Coenzymes are organic/organometallic non-proteic molecules whose function is to transport electrons or specific functional groups from one enzyme to another. In some cases, the coenzymes or metal ions are tightly integrated within an enzyme's structure, while in others they are covalently bound to the enzyme, such as the "Heme group", forming what is known as a prosthetic group.⁴

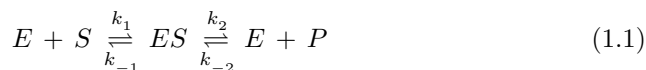
• Historical Perspective

The existence of biological catalysts has been extensively studied since the early nineteenth century, when studies on the digestion of meat by stomach secretions and on the conversion of starch to sugar by saliva were carried out.¹ The first general theory of chemical catalysis was stated in 1835 by the Swedish chemist J.J. Berzelius, who pointed out that the hydrolysis of starch is catalyzed more efficiently by malt diastase (α -amylase) than by sulfuric acid.⁶ Subsequently, in the year 1860, Louis Pasteur concluded that the fermentation of sugar into alcohol by yeast was induced by certain biological catalysts initially called "ferments".¹ Pasteur theorized that these ferments should be attached inseparably to the structure of the living yeast cells. This view, known as "vitalism", prevailed for several decades until, in 1897, E. Buchner demonstrated that some substances extracted from yeast could catalyze the alcoholic fermentation of sugar, even when they were out of the living cell.¹⁰ It was not until 1876 when the term enzyme was first employed by Kühne.¹¹ The word comes from the Greek term meaning "in yeast", since the first studies on these biological catalysts were conducted on this ferment. It was many years later (1926) when J.B. Sumner managed to isolate and crystallize an enzyme, urease.¹² Sumner found that the crystals obtained were formed by proteins and postulated that all enzymes are proteins. In the 1930s, several enzymes such as pepsin,¹³ trypsin¹⁴ and other digestive enzymes were first isolated and crystallized, the proteic nature of these biological catalysts being definitively established. In 1960, Hirs, Moore and Stein published a paper with the first sequence of an

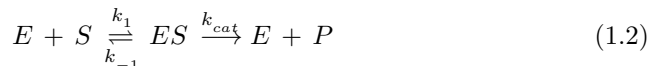
enzyme: ribonuclease (RNase);¹⁵ this enzyme is composed of 124 amino acids and was termed RNase for its ability to cleave molecules of RNA. The first-ever X-ray crystal structure of an enzyme was solved by Phillips et al. in 1965,¹⁶ and in 1969, B. Gutte and R.B. Merrifield announced the first synthesis of an enzyme, ribonuclease A, from amino acid precursors.¹⁷

1.1.2. Enzyme Kinetics

Enzyme kinetics attempts to study enzymes from the standpoint of determining their reaction rates. In 1913, Leonor Michaelis and Maud Menten postulated a general model of enzyme action.¹⁸ They proposed the simplest catalytic scheme based on a single-substrate enzyme-catalyzed reaction. According to this model, the enzyme (E) and the substrate (S) combine reversibly in an initial step, which occurs relatively quickly, leading to the formation of an enzyme-substrate complex (ES).¹ This complex is also known as the Michaelis or Henri-Michaelis coordination complex, after the researchers who first developed it.^{18,19} Subsequently, the ES complex breaks down in a second step, which occurs more slowly than the first step, giving rise to the free enzyme (E) and the reaction product (P):¹



Because the concentration of the product ($[P]$) is negligible at the beginning of the reaction, the reverse reaction $P \rightarrow S$ can be ignored and therefore the expression 1.1 is simplified by excluding k_{-2} .^{1,3} This rapid equilibrium assumption leads to the well-known Henri-Michaelis-Menten (HMM) kinetic scheme:²⁰



This scheme is extremely simple and an oversimplified version of the more general model of Haldane²¹ (1930), which can be seen in Fig. 1.2.²² In fact, the necessary existence of another complex (enzyme-product complex, EP) has also been neglected in the HMM scheme, assuming that the third step ($EP \rightleftharpoons E + P$) is irreversible ($k_{-3} \sim 0$) and relatively fast ($k_3 \gg k_2$).²³

The initial reaction rate, designated v_0 , is determined by the breakdown of ES to form product P and can be expressed as:

$$v_0 = k_{cat}[ES] \quad (1.3)$$

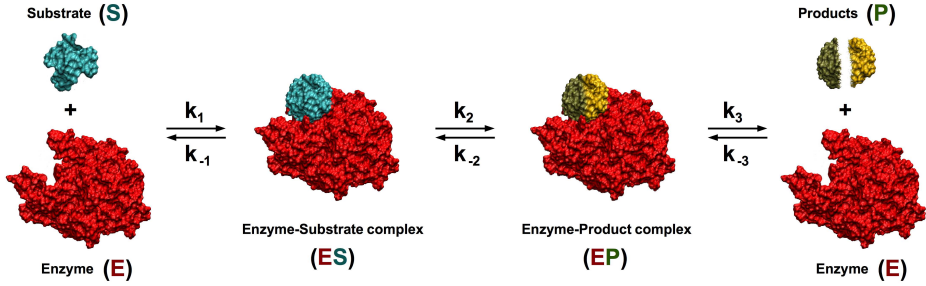


Figure 1.2.: Haldane model of enzyme kinetics.

The steady-state approximation developed by Briggs and Haldane²⁴ assumes that the concentration of the intermediate ES remains constant, because its rate of formation is equal to its rate of breakdown. In this situation, the variation in the concentration of the intermediate is almost zero and, thus, the derivative of the concentration of ES is set to zero:⁶

$$\frac{d[ES]}{dt} = k_1 [E][S] - k_{-1} [ES] - k_{cat} [ES] = 0 \quad (1.4)$$

hence,

$$[ES] = \frac{k_1 [E][S]}{k_{-1} + k_{cat}} \quad (1.5)$$

Knowing that the free enzyme concentration is equal to the total enzyme concentration minus the concentration of the complex $[ES]$ formed, $[E] = [E]_0 - [ES]$:

$$[ES] = \frac{k_1 ([E]_0 - [ES])[S]}{k_{-1} + k_{cat}} = \frac{k_1 [E]_0 [S] - k_1 [ES][S]}{k_{-1} + k_{cat}} \quad (1.6)$$

$$[ES](k_{-1} + k_{cat}) + k_1 [ES][S] = k_1 [E]_0 [S] \quad (1.7)$$

$$[ES] = \frac{[E]_0 [S]}{\frac{k_{-1} + k_{cat}}{k_1} + [S]} = \frac{[E]_0 [S]}{k_M + [S]} \quad (1.8)$$

where,

$$k_M = \frac{k_{-1} + k_{cat}}{k_1} \quad (1.9)$$

is the Michaelis-Menten constant, defined as the concentration of the substrate ($[S]$) at which the reaction rate is one-half the maximum velocity ($v_{max}/2$).³

Therefore, substituting equation 1.8 into 1.3:

$$v = k_{cat} \left(\frac{[E]_0 [S]}{k_M + [S]} \right) = \frac{k_{cat} [E]_0 [S]}{k_M + [S]} \quad (1.10)$$

The maximum reaction rate (v_{max}) occurs when the enzyme is saturated with substrate (that is, when $[ES] = [E]_0$), and therefore the maximum velocity can be defined as:

$$v_{max} = k_{cat} [E]_0 \quad (1.11)$$

Substituting 1.11 in 1.10, the equation of the reaction rate can be rewritten as:

$$v = \frac{v_{max} [S]}{k_M + [S]} \quad (1.12)$$

This is the Michaelis-Menten equation, and expresses the rate for a one-substrate enzyme-catalyzed reaction. The graphical representation of this equation (Fig. 1.3) shows the behavior of the reaction rate (v_0) as a function of the substrate concentration ($[S]$).

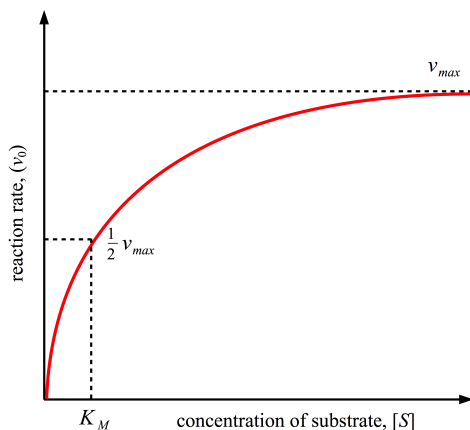


Figure 1.3.: Effect of substrate concentration ($[S]$) on the rate (v_0) of an enzyme-catalyzed reaction. Figure adapted from reference [3].

As can be seen in Fig. 1.3, when $[S] \ll K_M$, the initial reaction rate (v_0) is virtually proportional to the substrate concentration ($[S]$) and thus increases almost linearly with it. In this case, $[S]$ is the limiting factor in the reaction rate.

Conversely, when $[S] \gg K_M$, the initial reaction rate asymptotically approaches a constant velocity ($v_0 = v_{max}$). At this point, it is said that the enzyme has reached its saturation limit for the substrate (that is, $[ES] = [E]_0$), $[E]$ therefore being the limiting factor in the reaction rate.

1.1.3. Enzyme Inhibition

As pointed out in sec. 1.1.1, enzymes are responsible for leading, controlling and accelerating the chemical reactions in biological systems that allow life to occur. However, sometimes we are interested in stopping the catalytic activity of the enzyme, since some enzymes are related to the onset or development of certain diseases. There are a number of substrates that are responsible for slowing or stopping enzymatic reactions, which are known as *enzyme inhibitors*. Many of these inhibitors are substances that structurally resemble their enzyme's natural substrate, and for this reason are commonly used to probe the chemical and conformational nature of a substrate-binding site. In this way, understanding how an enzyme binds its substrate may help us to elucidate the enzyme's catalytic mechanism.⁶ Moreover, these substances are very valuable in the field of the pharmaceutical industry, since many of the drugs circulating on the market throughout the world today, mostly chemotherapeutic agents, are enzyme inhibitors.^{4,6}

Two different classes of enzyme inhibition can be found: reversible and irreversible.

- **Reversible inhibition:**

The reversible inhibitors temporarily combine with the enzyme in a similar manner to the natural substrates, and sometimes they combine with the enzyme-substrate complex rather than with the free enzyme. Three types of reversible inhibition can be found:

1. **Competitive inhibition:** The reversible inhibitor shows a similar geometrical structure to the natural substrate, and thus competes with it for the active site of the enzyme. The inhibitor forms an enzyme-inhibitor complex (EI), in accordance with the scheme $E + I \rightleftharpoons EI$, with similar kinetics to those of the ES complex, but which cannot be decomposed into a free enzyme and products (see Fig. 1.4a).
2. **Uncompetitive inhibition:** In this case, the inhibitor is not combined with the free enzyme nor does it affect its binding with the substrate. The uncompetitive inhibitor is combined with the ES complex leading to an enzyme-substrate-inhibitor complex (ESI), in accordance with the scheme $ES + I \rightleftharpoons ESI$ (see Fig. 1.4b). This complex, which is inactive, is not subsequently decomposed to yield the products. The inhibitor is placed near the active center, so that products are physically prevented from exiting.
3. **Mixed inhibition:** The inhibitor binds to either the free enzyme or the enzyme-substrate complex, interfering with the action of both. In this case, the binding with the inhibitor leads to both complexes, ES and ESI , following the schemes: $E + I \rightleftharpoons EI$ and $ES + I \rightleftharpoons ESI$. None of the above complexes can decompose to lead to products and the free enzyme (see Fig. 1.4c).

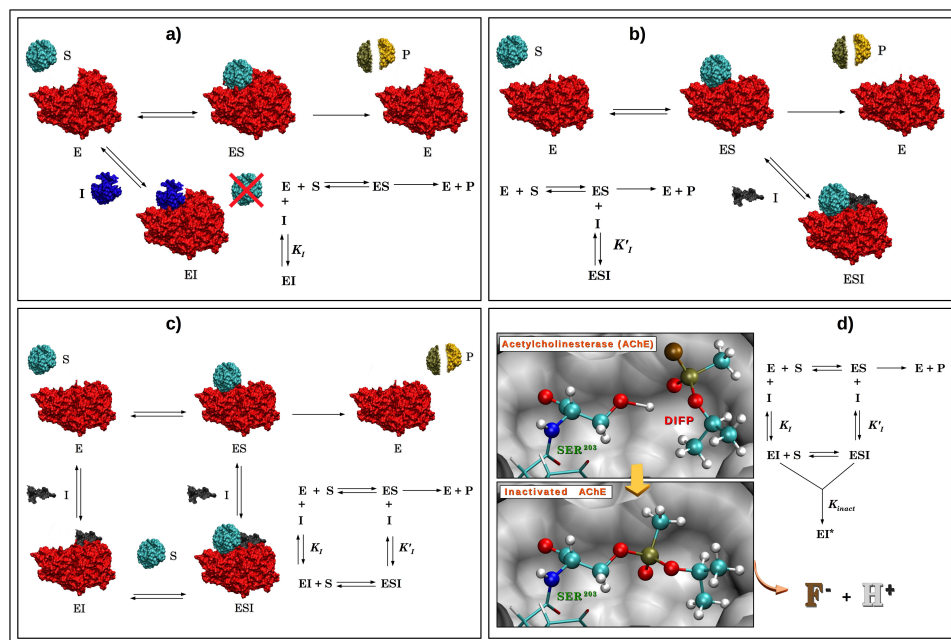


Figure 1.4.: Different types of enzyme inhibition. Reversible competitive inhibition (a), reversible uncompetitive inhibition (b), reversible mixed inhibition (c), and irreversible inhibition (d). All the figures are adapted from reference [1]. Graphical representations have been generated by means of the VMD program [25].

- **Irreversible inhibition:**

The irreversible inhibitors permanently combine with the enzyme, carrying out their catalytic activity in an irreversible way. These inhibitors bind covalently with certain functional groups of the amino acids that are essential for catalysis or are associated with them non-covalently by means of strong interactions. The study of this type of inhibitors is a powerful tool for studying enzymatic mechanisms, since it allows the identification of functional groups which are essential for catalysis in those enzymes inactivated by them.¹ An example of irreversible inhibition can be seen in Fig. 1.4d, where the diisopropylphosphofluoridate (DIPF) molecule binds covalently to the hydroxyl group of a serin residue of the enzyme acetylcholinesterase (AChE), thus inactivating it. Since the enzyme AChE is involved in nervous system activity, the DIPF molecule acts as a nerve poison.³

A special class of irreversible inhibitors is the *mechanism-based inactivators*, also known as *suicide inhibitors*. These inhibitors follow the first steps of an enzyme catalytic route in the same way as the natural substrate. However, at a certain point they become a very reactive intermediate species that is covalently bound to the active site of the enzyme in an irreversible way, thus inactivating it. Suicidal inhibitors play an important role in the *rational drug design*, a novel strategy being

employed in the pharmaceutical industry to develop new drugs.¹

1.1.4. Enzymatic Catalysis

The concept that the enzyme combines temporarily with the substrate to form the enzyme-substrate (or Michaelis) complex (ES) is the cornerstone of enzyme kinetics and of our understanding of enzyme catalysis. One of the most outstanding attributes of enzymes is their high degree of specificity for their substrates, which can be better understood if it is known how the substrate binds to the active site of an enzyme. The first model to explain the specificity of enzymes was proposed by Emil Fischer in 1894^{26,27} (see Fig. 1.5a). Fischer's model proposes that the interaction between enzyme and substrate takes place by the exact fit between the molecular surfaces of both biological entities, which have complementary shapes. This model is known as the "Lock and key" model, where the substrate is the key that fits exactly into the keyhole (active site) of the enzyme (lock). However, this model is very rigid and can be misleading when applied to enzymatic catalysis.¹ For this reason, in 1958, Daniel Koshland proposed a new hypothesis known as the "induced fit" model²⁸ (see Fig. 1.5b). According to this model, since enzymes are flexible entities, conformational changes may occur at the active site of the enzyme induced by the presence of the substrate. These changes may include the approximation of essential catalytic groups of the enzyme to the substrate functional groups capable of reacting, thereby increasing the binding of the enzyme with the substrate and improving the catalytic capacities of the enzyme.⁶

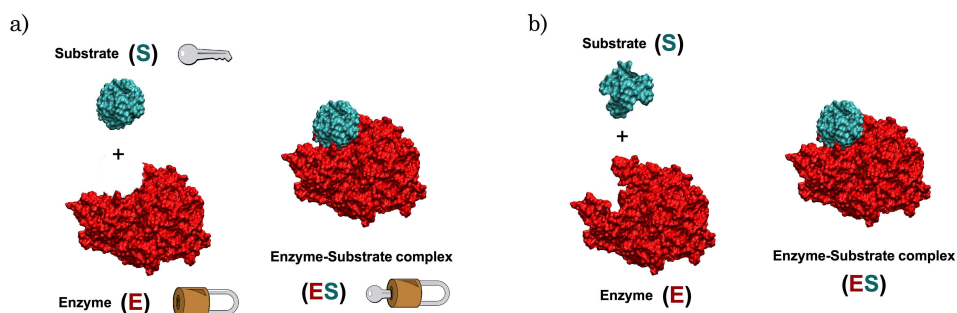


Figure 1.5.: a) Fischer's "Lock and key" model, b) Koshland's "Induced fit" model.

But, in fact, the modern notion of enzymatic catalysis was first proposed by Michael Polanyi²⁹ (1921) and J.B.S. Haldane²¹ (1930) and then developed by Linus Pauling³⁰ in 1946. Pauling proposed that in order to catalyze biological reactions, the active site of an enzyme must be complementary to the reaction's *transition state* (TS) and not to the substrate in its fundamental state. This means that the optimal interactions between enzyme and substrate take place only in the TS.¹ In

Fig. 1.6, both the energetic profile of a catalyzed and non-catalyzed reaction can be observed; the former corresponds to the kinetic scheme presented in equation 1.2.

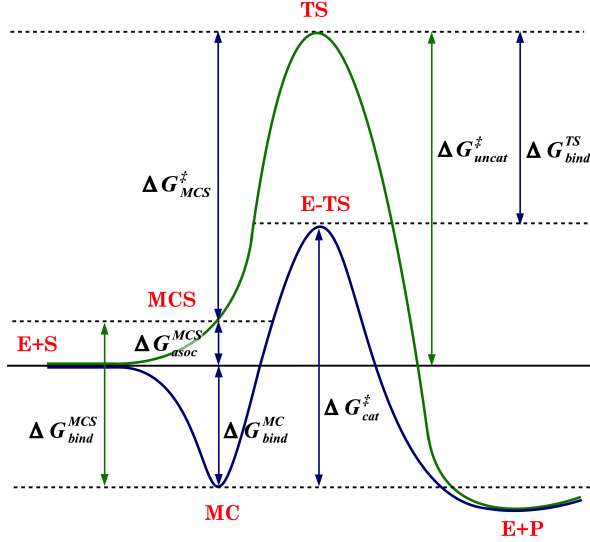


Figure 1.6.: A comparison of free energy profiles for enzyme-catalyzed reaction (blue line) and uncatalyzed reaction (green line). Figure adapted from reference [31].

As can be seen in the profile corresponding to the non-catalyzed reaction, this reaction proceeds in only one step with an activation energy of $\Delta G_{uncat}^{\ddagger}$, leading to the products. Conversely, the enzyme-catalyzed reaction of the example takes place in two different steps: the first one corresponds to the formation of the enzyme-substrate complex, known as the Michaelis-Menten complex (MC), in a process that involves binding energy (ΔG_{bind}^{MC}) loss. Subsequently, the chemical reaction itself takes place, with an activation energy of $\Delta G_{cat}^{\ddagger}$, which leads to the formation of the products and the release of the enzyme. The energetic cost of the catalyzed reaction is smaller than the uncatalyzed one and it can be related to the reaction rate, k_{cat} , according to the Transition State Theory^{32,33} (see Chapter 2):

$$k_{cat} = \frac{k_B T}{h} e^{(-\Delta G_{cat}^{\ddagger}/RT)} \quad (1.13)$$

The catalytic power of an enzyme, which measures the increase in velocity (acceleration) of the chemical process, is usually expressed by the ratio k_{cat}/k_{uncat} , that is, it is related with the difference between the activation free energy of the uncatalyzed reaction ($\Delta G_{uncat}^{\ddagger}$) and of the catalyzed one ($\Delta G_{cat}^{\ddagger}$). According to Fig. 1.6, the decrease in activation free energy can be expressed in terms of the binding free energy of the reactants and of the TS as:

$$-\Delta G_{bind}^{TS} + \Delta G_{cat}^{\ddagger} = -\Delta G_{bind}^{MC} + \Delta G_{uncat}^{\ddagger} \quad (1.14)$$

$$\Delta G_{uncat}^{\ddagger} - \Delta G_{cat}^{\ddagger} = \Delta G_{bind}^{MC} - \Delta G_{bind}^{TS} \quad (1.15)$$

If the enzyme speeds up the reaction, the left side of equation 1.15 should be positive and therefore the absolute value of the magnitude ΔG_{bind}^{TS} should be greater than ΔG_{bind}^{MC} . That is, the enzyme must show a higher affinity for the *TS* than for the reactants, thereby demonstrating that the catalytic role of the enzyme lies in the stabilization of the *TS* rather than the *MC* complex.^{31,34-36} However, from the viewpoint of the formation of *MC*, ΔG_{bind}^{MC} may contain an important energy contribution due to the change in the substrate geometry when moving from the structure of a reactant, where the fragments are found separated and completely solvated, to the structure of another reactant, where the fragments are maintaining contact and are in a proper orientation to react. If we now consider an imaginary intermediate complex in solution (*MCS*), which corresponds to a similar structure to the *MC* complex present in the enzyme, the binding free energy of the *MC* complex can be expressed as:

$$\Delta G_{bind}^{MC} = \Delta G_{asoc}^{MCS} + \Delta G_{bind}^{MCS} \quad (1.16)$$

where ΔG_{asoc}^{MCS} corresponds to the free energy required to obtain the *MCS* complex in solution from the reactants separated by solvent molecules, and ΔG_{bind}^{MCS} corresponds to the binding energy of the *MCS* structure. If we replace equation 1.16 in equation 1.15:

$$\Delta G_{uncat}^{\ddagger} - \Delta G_{cat}^{\ddagger} = \Delta G_{asoc}^{MCS} + \Delta G_{bind}^{MCS} - \Delta G_{bind}^{TS} \quad (1.17)$$

Since ΔG_{asoc}^{MCS} is always positive, a high catalytic power ($\Delta G_{uncat}^{\ddagger} - \Delta G_{cat}^{\ddagger} > 0$) can be obtained even when the enzyme shows greater affinity to the *MC* structure than to the *TS*. Therefore, according to this assumption, the key to catalysis would lie in both the pre-organization of the substrate (which will be favorable in the enzyme but energetically expensive in solution) and the greater affinity of the enzyme for the *TS*.

Hence, there is no single theory regarding enzymatic catalysis, and indeed several have been reported in the literature. These theories typically fall into two different groups: those that are related with the formation of the *MC* complex and those that are related with the stabilization of the *TS*. Several factors that may contribute to the catalytic power of enzymes have been proposed, some of which are described below:

1. Acid-base catalysis⁶

This type of catalytic mechanism employed by enzymes may occur through general acid catalysis, general base catalysis or both processes simultaneously: a concerted general acid–base catalyzed reaction.

Some enzymatic reactions are stimulated by both the partial proton transfer from a Brønsted acid (general acid catalysis) or partial proton abstraction by a Brønsted base (general base catalysis). Both processes decrease the free energy of a reaction's TS and therefore increase its rate.

2. Covalent catalysis⁶

According to this type of catalysis, the transient formation of a catalyst–substrate covalent bond may reduce the energy of the TS, which leads to reaction rate acceleration.

3. Metal ion catalysis⁶

Approximately 33% of known enzymes need the presence of metal ions to perform any catalytic activity. These metals can be tightly bound to the enzyme, as in the metalloenzymes (commonly transition metal ions such as Fe^{2+} , Fe^{3+} , Cu^{2+} , Zn^{2+} , Mn^{2+} , Co^{3+}), or loosely bound to the enzyme from a solution in order to activate it, (usually the alkali and alkaline metal ions Na^+ , K^+ , Mg^{2+} or Ca^{2+}). The role of the metals in enzymatic catalysis consists mainly of the following functions: to orientate the substrates in a proper way for the reaction once they have been bound to them, to mediate in oxidation–reduction reactions, and to electrostatically stabilize the negative charges.

4. Low-Barrier Hydrogen Bonds^{37–40}

As can be seen in Fig. 1.7, between two heteroatoms with similar proton affinities (X and Y), three types of hydrogen-bonded systems can be found: the weak, the strong and the very strong. In the weak hydrogen bonds (Fig. 1.7a), the energy wells are interacting but separate, so that the proton is covalently bonded to X and electrostatically attracted to Y . In the case of the very strong hydrogen bonds (Fig. 1.7c), the heteroatoms are identical and the proton is shared between them. Both $X-H$ bond lengths are equal and are much longer than the typical covalently bonded proton. Finally, in the case of strong hydrogen bonds (Fig. 1.7b), the distance between the two heteroatoms (X and Y) is less than the sum of their van der Waals radii and the pKa values of the hydrogen bond partners are similar to each other. For this reason, they exhibit a covalent nature that is absent in an ordinary hydrogen bond. In this way, the proton can move freely between both heteroatoms, but may be more strongly attracted to one of them. Bonds of this type are known as *low-barrier hydrogen bonds* (LBHBs) and it has been reported that they appear to stabilize TSs or metastable intermediates.^{38,41–43}

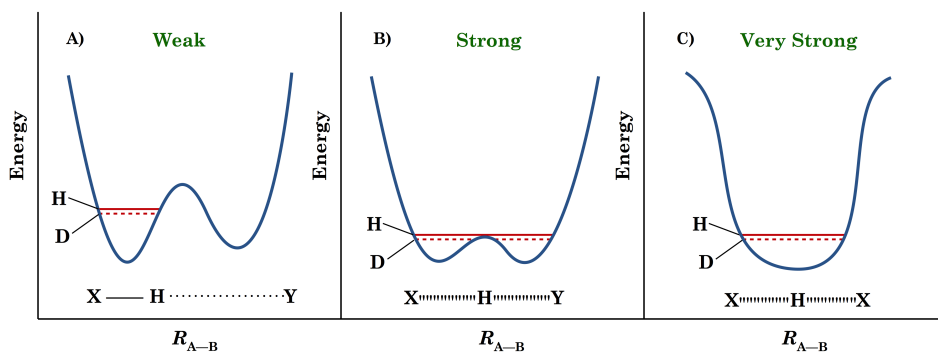


Figure 1.7.: Energy diagrams of the hydrogen-bonded heteroatoms (X and Y) as a function of R_{A-B} . Hydrogen bonds have been classified in three types: (a) weak or conventional, (b) strong or low barrier, and (c) very strong or single-well. Figure adapted from reference [39].

5. Electrostatic effects^{34–36,44}

Enzymatic catalysis is attributed to the electrostatic stabilization of the TS generated by the residues present in the active site of the enzyme. In the enzymatic environment, the dipoles created by the amino acids of the active site are relatively fixed and well oriented toward the substrate. This pre-organization involves a minor change in the environment on going from reactants to the transition state, which facilitates the electrostatic stabilization of the TS (see Fig. 1.8b). Conversely, in solution, the reorganization of the solvent dipoles toward the substrate is greater, which implies an energetic cost to orient the solvent according to the TS configuration (see Fig. 1.8a).

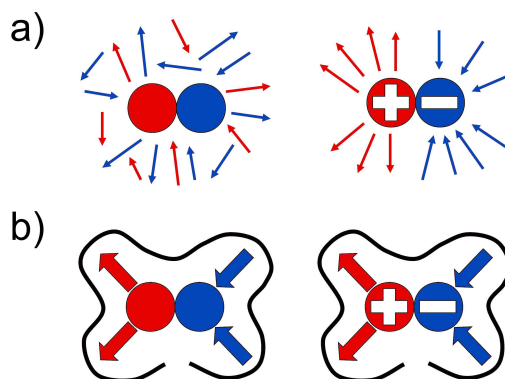


Figure 1.8.: Variation of charge distribution on going from reactants to the transition state in aqueous solution (a) and in the enzymatic environment (b). In solution, the reorganization of the dipoles brings an associated energy cost, although the pre-oriented dipoles in an enzymatic environment facilitate the transformation of apolar reactants into a strongly polar transition state. Figure adapted from reference [31].

6. Strain or distortion⁴⁵

The enzyme may induce a strain or a distortion in a bond of the substrate molecule, leading to the destabilization of this bond in the *MC* complex, thus facilitating its cleavage. This means that the enzyme favors the formation of the *MC* complex.

7. Desolvation⁴⁵⁻⁵¹

When the substrate binds to the enzyme, water molecules or molecules of other solvents are usually excluded from the active site (process of desolvation). This process reduces the local dielectric constant and hence improves the interactions between the enzyme active site and the substrate.

8. Orbital steering^{52,53}

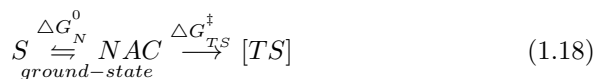
In order to reach the transition state, the substrate and the catalytic group of the enzyme must not only be close to each other, but must also have their molecular orbitals well oriented and aligned in a precise way.

9. Entropic factors^{45,54-59}

The entropy of a system is related with its degrees of freedom or its degree of disorder, in such a way that the more disordered the system is, the higher the entropy will be. In this way, given that the reactant is fixed to the active site of the enzyme due to specific interactions, its degrees of freedom are reduced so that the chemical reaction is favored. Conversely, the reaction in solution is slower because the rapprochement of the reactants involves a loss in rotational and translational entropy when going from reactants to the transition state. It is worth noting that the rotational and translational entropies of the substrate are lost in the formation of the enzyme-substrate complex, and not during the chemical steps.

10. Near-attack conformations effect⁶⁰⁻⁶²

In a chemical reaction involving the formation of a covalent bond there is a moment in which the two reacting atoms are very close each other (at a van der Waals contact distance), and at an angle similar to that in the *TS*. This situation is called a near-attack conformation (NAC) and the reaction of interest must proceed to the *TS* via NAC formation according to the following scheme:



$$\Delta G^\ddagger = \Delta G_N^0 + \Delta G_{TS}^\ddagger \quad (1.19)$$

The formation of near-attack conformers, which are ground-state conformers

that can convert directly to the transition state, involves standard free energies (ΔG_N^0). The role of the enzyme would be to reduce the free energy of activation by increasing the probability of finding such NAC structures.

11. Tunneling effect^{63,64}

The tunneling effect is a quantum-mechanical phenomenon in which a light particle tunnels through an energetic barrier that classically it could not overcome. This effect plays an important role in some enzymatic processes, which involves the transfer of light particles such as hydrogen atoms (hydrogen transfer reactions). The probability of the tunneling effect occurring in a chemical reaction is much greater in the enzyme environment than in aqueous solution.

12. Dynamic effects

The relationship between protein motions and enzyme catalysis is currently hotly debated and a matter of scientifically fruitful contention.⁶⁵ Both theoretical and experimental studies about the role of the dynamic effects in catalysis have been carried out during recent years without any consensus being reached.^{65,66} On the one hand, there are some computational⁶⁷⁻⁶⁹ and experimental⁷⁰ studies that support the fact that the dynamic effects are very small or negligible in enzymes. On the other hand, there are scientists that support the idea that the quantification of the protein's motion can be estimated as a deviation of the transition state theory, which is an equilibrium theory. According to the latter, the deviation of the rate constant, obtained by means of a transmission coefficient due to the recrossing effect, can be calculated through the study of rare event trajectories⁷¹⁻⁷⁴ or by the application of the Grote-Hynes theory.^{71,73,75-77} To some extent, the role of the enzyme would consist in reducing the recrossing effect when compared with the uncatalyzed reaction in solution.

In view of the large number of factors that attempt to explain the origin of the catalytic power of enzymes, it is easy to assume that there cannot be only one factor responsible for the overall activity for each and every one of the enzymes. For each type of enzyme there will probably be some factors that are more relevant than others in terms of their catalytic activity, or simply the origin of the catalytic power of the enzymes would lie in a specific combination of several of these factors.

What does seem clear is that the understanding of how the enzymes work is not an easy task. For this reason, the exploration of the enzymatic mechanisms at atomic level may shed some light on the behavior of these macromolecular biological entities, with significant implications for many fields of science and technology.

1.2. Cytochrome P450 Aromatase

1.2.1. Cytochromes P450

The cytochromes P450 (CYP or P450) form a superfamily of heme-metalloenzymes that are present in the three domains of life, Eukarya, Bacteria, and Archaea. This superfamily is composed of a great variety of isoforms, which share structural features, distributed among a wide selection of families and subfamilies. Currently there are more than 21,039 listed members, including 57 that belong to the species homo sapiens.⁷⁸ P450 enzymes are characterized by the presence of an iron protoporphyrin IX prosthetic group at the active site, which is coordinated by two axial ligands: on the proximal side, by a thiolate from a cysteine residue of the protein, whereas on the distal side the ligand is variable, since it is constantly changing along the catalytic cycle.^{79–81}

The discovery of cytochrome P450 dates back to the late 50s, when in 1958 Klingenberg⁸² reported for first time the existence of an unknown carbon monoxide-binding pigment, whose spectrum showed a prominent peak at 450 nm, which disappeared completely upon the addition of detergents to microsomes.⁸³ A few years later, in 1964, Omura and Sato^{84–86} identified the heme-protein nature of this pigment, which was present in liver microsomes from different mammal species. The term P450, where P stands for pigment, derives from the wavelength peak that the enzyme shows when its ferrous (reduced) form binds CO, thus generating a ferrous iron-CO complex.⁸⁴ This complex displays an unusual Soret ($\pi - \pi^*$) absorption maximum at a wavelength of approximately 450 nm (see Fig. 1.9). The function of cytochrome P450 was discovered by Estabrook et al. in 1963,⁸⁷ when they were studying the C_{21} -hydroxylation of 17-hydroxy-progesterone catalyzed by adrenal cortex microsomes, and were able to observe the involvement of cytochrome P450 in this reaction.⁸⁸



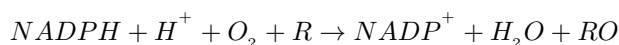
Figure 1.9.: Tsuneo Omura (left) and Ryo Sato (center) (both pictures are taken from www.issx.org). On the right, the title of the paper published by Omura and Sato in 1964^[85] and a figure obtained from reference^[84], showing the carbon monoxide and ethyl isocyanide difference spectra of liver microsomes.

In 1987, the system of nomenclature and classification of P450 enzymes was established,⁸⁹ obeying phylogenetic criteria. This nomenclature system is still utilized today and includes all cytochromes P450, both prokaryotic and eukaryotic. According to this criterion, the isoforms of the superfamily are identified by the acronym CYP, followed by a number designating the family, a letter that identifies the subfamily, and another number that corresponds to the gene (e.g., CYP1A1, CYP2C9).

CYP	3	A	4
Cytochrome	Family	Subfamily	Individual
Similarity	> 40 %	> 55 %	> 55 % < 97%

All the isoforms whose amino acid sequence has a similarity greater 40% belong to the same family. Within a family, all the enzymes sharing more than 55% of the sequence are grouped in the same subfamily. Finally, all the enzymes within the same subfamily that differ by more than 3% in their respective sequences are considered individual enzymes.

The principal function of cytochrome P450 is the mono-oxygenation of various substrates, which requires molecular oxygen and a supply of reducing equivalents from NADPH or NADH for the reactions.⁸⁸ The general reaction catalyzed by P450s, known as the monooxygenase reaction, is as follows:^{90,91}



In this reaction, P450 activates dioxygen (O_2) by means of its prosthetic group, an iron-heme porphyrin, and one of its oxygen atoms is inserted into a substrate (R). Conversely, the second oxygen is reduced to a water molecule (H_2O), utilizing two electrons that are provided by $NAD(P)H$ via a reductase protein.^{92–95} Because only one of the two oxygen atoms, initially present in O_2 , prevails in the oxidized substrate (RO), P450s are called monooxygenases.⁸⁰ The most common reactions catalyzed by P450 enzymes that involve oxygen insertions are: hydroxylation of unactivated $C-H$ bonds, alkene epoxidations, N - and O -dealkylations, heteroatom oxidations such as sulfoxidations, and so forth.^{81,90,96}

These enzymes play an important role in different biological processes in all eukaryotic organisms, plants, animals and fungi, and in some prokaryotes.⁸⁸ P450s can metabolize a wide variety of both endogenous and exogenous substrates,^{79,97} and actively participate in oxidation and reduction processes occurring in the metabolism of xenobiotics. In humans, these are responsible for the oxidative

metabolism for a large number of pharmaceuticals ($\sim 75\%$), and only five isoforms (CYP3A4, CYP2C9, CYP2D6, CYP2C19, CYP1A2) account for $\sim 90\%$ of the metabolism.⁹⁸

Catalytic cycle of P450 enzymes^{79–81,99,100}

Cytochrome P450 works as an assembly line at nano scale that proceeds according to the catalytic cycle depicted in Fig. 1.10.

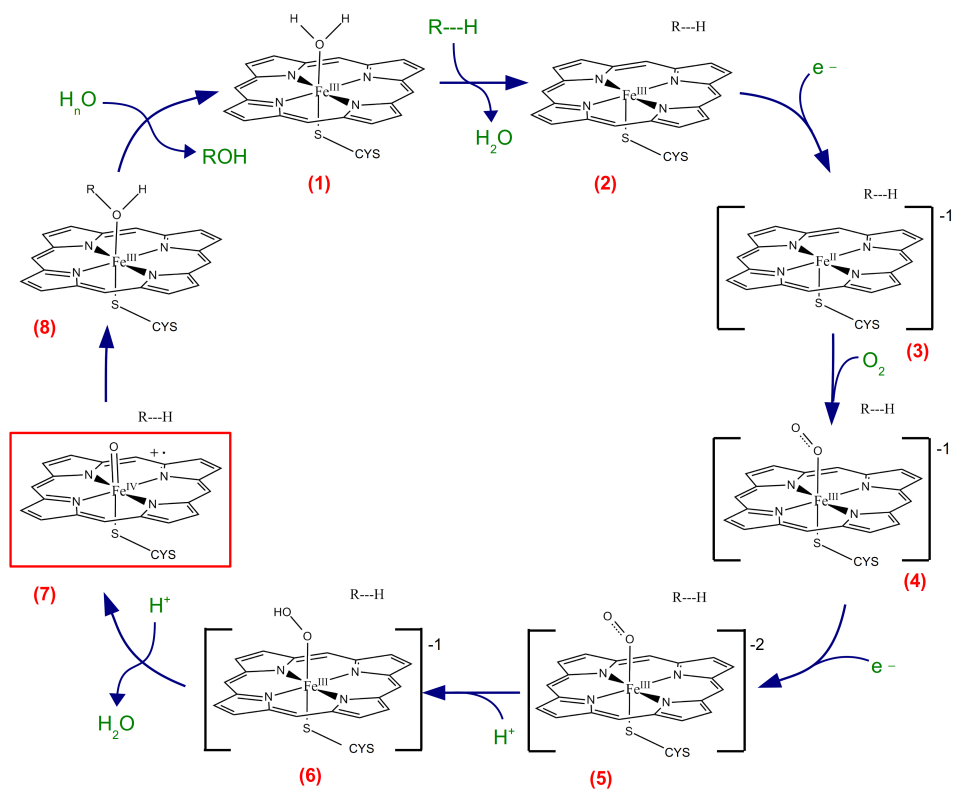


Figure 1.10.: Catalytic cycle of Cytochrome P450.

The cycle begins with the resting state of the enzyme (1), where a hexacoordinated ferric (Fe^{III}) complex is found in a low-spin doublet state. In this state, the central iron atom is coordinated by the sulfur atom of a cysteine at the proximal position, and by a water molecule at the distal position. When a substrate achieves the active site, the water molecule is displaced leading to a pentacoordinated ferric-complex found in a sextet spin state (2). This complex is a good electron acceptor and thus obtains an electron from a reductase protein, so that it becomes a high-spin ferrous complex (3). The latter binds molecular oxygen (O_2)

yielding the oxy-ferrous complex (**4**), which has a singlet spin state. This complex is again reduced by obtaining another electron, thus generating the ferric-peroxo anion species (**5**). Thereafter, two consecutive protonation stages of the peroxo species occur. The first one forms the ferric-hydroperoxide species (**6**), also known as *Compound 0* (Cpd 0); the second one facilitates the $O-O$ bond cleavage, leading to the formation of a water molecule and the high-valent iron-oxo species, the so-called *Compound I* (Cpd I) (**7**). Stage (**8**) corresponds to the oxidation of the substrate by means of Cpd I as well as to the restoring of the resting state (**1**).

1.2.2. The Active Species of CYP Enzymes: Compound I

The nature of the reactive species located at the active site of CYP enzymes has been extensively discussed in the literature by both theoretical⁸¹ and experimental⁹² studies. Currently, the accepted oxidant is Cpd I (see Fig. 1.11), a high-valent iron-(IV) oxo-porphyrin π -cation radical ($(S-Cys)Por^{+\bullet}-Fe^{IV}=O$), whose ability to activate inert $C-H$ bonds has been amply demonstrated.^{93,101-104}

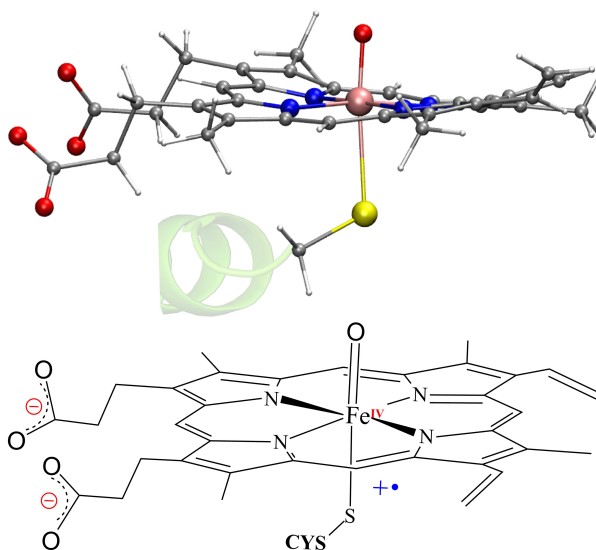


Figure 1.11.: Ball-and-stick model (top) and Lewis structure (bottom) of Compound I.

This species participates in the key step in the oxidation of substrates (step (**8**) of the cycle), although it is an elusive species because it does not accumulate in the catalytic cycle of the P450 enzymes.¹⁰⁵ In fact, due to the elusiveness of Cpd I and its high reactivity, making it very difficult to detect and capture, other oxidants such as the ferric hydroperoxo complexes ($Por-Fe^{III}-OOH^-$, Cpd 0),¹⁰⁶⁻¹¹⁰ Compound II ($Por-Fe^{IV}=O$, Cpd II),¹¹¹ and perferryl-based species ($Por-Fe^V=O$)¹¹²

have been proposed. However, Rittle and Green¹⁰⁴ succeeded in preparing Cpd I on CYP119 P450 cytochrome in a high yield, and were able to characterize it spectroscopically and kinetically, thereby settling the matter on the hydroxylating agent in P450 enzymes. This compound is based on a protoporphyrin-IX prosthetic group with a central iron atom linked to two axial ligands, the first of which is a cysteine amino acid called proximal axial ligand, and the second is an oxo anion called the distal ligand, which forms a ferryl $Fe^{IV} = O$ moiety. Cpd I is also formed in the catalytic cycle of other proteins, such as peroxidase or catalase,¹¹³ whose proximal axial ligands are amino acids histidine^{114,115} and tyrosine,¹¹⁶ respectively. In fact, the ability of Cpd I to adapt to the specific environments present in a given enzyme (proximal ligand and active site) has led to it being known as a “chameleon” species.¹¹⁷

Since the pioneering work of Loew and Harris at the end of the 1990s and early 2000s,^{118–124} several theoretical studies have been carried out to calculate the electronic structure and spectra of Cpd I, both in the gas phase and in the enzyme environment. A few years ago, Shaik et al.⁸¹ published a review on the theoretical works devoted to study the structure, reactivity, and selectivity of different P450 enzymes. In this review, the electronic structure of Cpd I is analyzed by means of different levels of theory, such as the orbital approach using density functional theory (DFT) calculations, multireference *ab initio* methods, as well as the valence bond theory. According to the accepted electronic structure of this species, Cpd I has a radical nature and possesses three singly occupied orbitals (see Fig. 1.12).

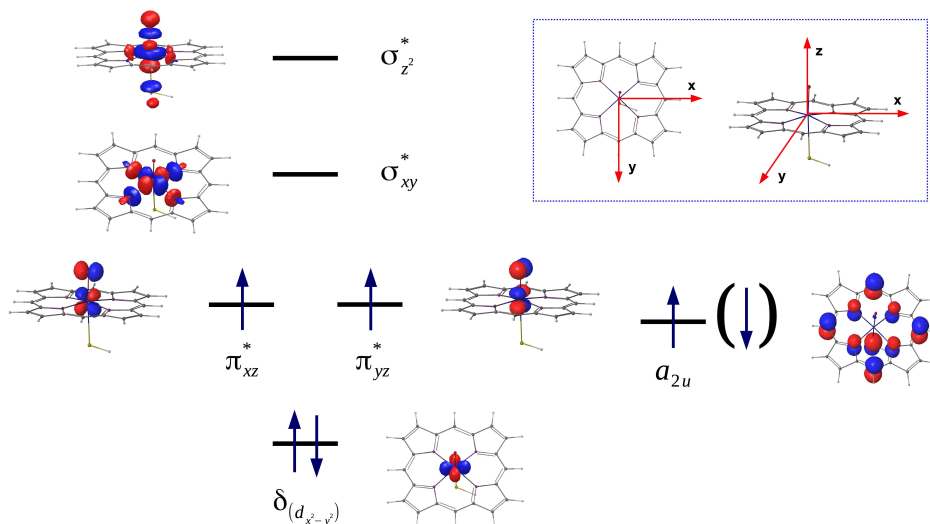


Figure 1.12.: Molecular orbitals diagram for Compound I.

Two of them are the π_{xz}^* and π_{yz}^* orbitals of the $Fe = O$ moiety that form a local π -anti-bonding triplet. The third one is an odd electron of a porphyrinic nature, which is delocalized between the a_{2u} orbital of the porphyrin and the π_s of the sulfur atom. A ferromagnetic coupling of the three electrons provides a high-spin quartet state ($S=3/2$), whereas an antiferromagnetic one leads to a low-spin doublet state ($S=1/2$). The resulting two lowest electronic states, ${}^2A_{2u}$ and ${}^4A_{2u}$, are nearly degenerate;¹²⁰ therefore, the two low-lying spin states (doublet and quartet) are virtually degenerate.^{80,81,100} As a result of this dual spin nature of Cpd I, Shaik and coworkers¹²⁵ proposed a two-state reactivity mechanism for the hydroxylation and the epoxidation of alkanes, in which product distribution of the reaction is determined by the combination of both spin states.¹²⁶

1.2.3. Human Aromatase Enzyme (CYP19A1)

There are 57 sequenced isoforms of cytochrome P450 in humans, which are distributed among 18 families of genes and 43 subfamilies.^{78,127,128} Human Aromatase (P450arom) is one of them, and it is the product of the CYP19A1 gene (Cytochrome P450, family 19, subfamily A, member 1), located on chromosome 15q21.1.¹²⁹ This enzyme consists of a polypeptide chain made up of 503 amino acid residues and a heme group (cofactor) (see Fig. 1.13).¹²⁸

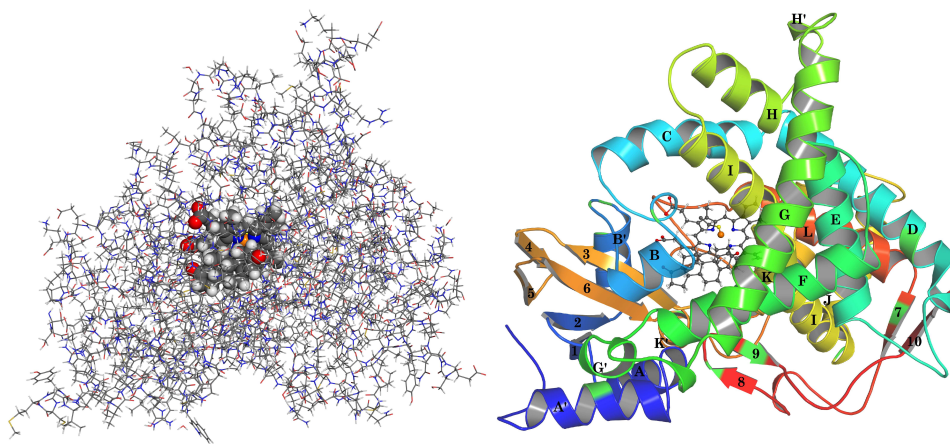
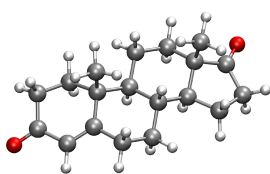
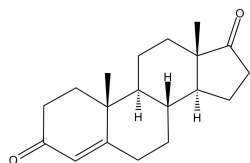


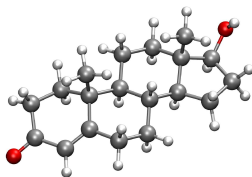
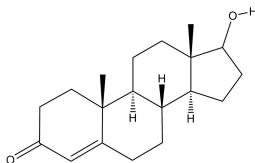
Figure 1.13.: Two different views of the enzyme CYP19A1 aromatase, generated from the pdb file with code 3EQM. On the left, the atomic positions of the enzyme are represented by sticks, and the substrate ASD and heme group are represented by spheres. On the right, the secondary structure of the aromatase is represented by a ribbon diagram, and the substrate ASD and heme group are represented by balls and sticks. The α -helices are labeled from A to L and the β -strands are numbered from 1 to 10. The figure on the right is adapted from reference [130].

Aromatase is located in a wide variety of tissues that include gonads, adrenal glands, ovaries, placenta, testes, adipose tissue, and numerous sites in the brain.^{131,132}

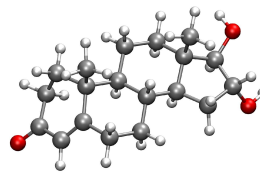
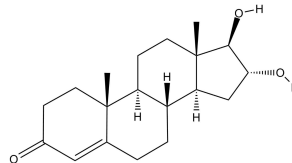
ANDROGENS



Androstenedione (ASD)
androst-4-ene-3,17-dione

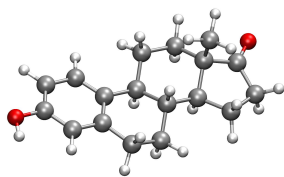
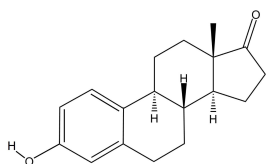


Testosterone (TST)
17 β -hydroxy-androst-4-ene-3-one

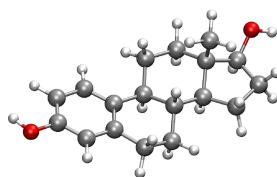
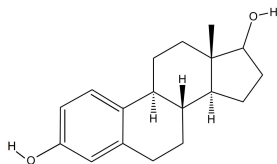


16 α -hydroxytestosterone (HTST)
16 α ,17 β -dihydroxy-androst-4-ene-3-one

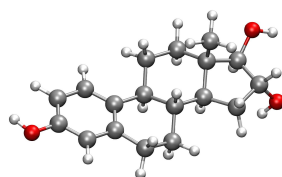
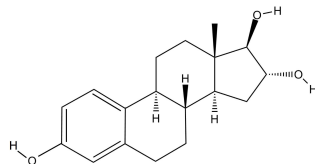
ESTROGENS



Estrone (E₁)
3-hydroxy-estra-1,3,5(10)-triene-17-one



17 β -estradiol (E₂)
17 β -estra-1,3,5(10)-triene-3,17-diol



17 β ,16 α -estriol (E₃)
16 α ,17 β -estra-1,3,5(10)-triene-3,16,17-triol

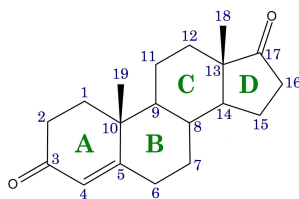


Figure 1.14.: Representation of androgens and estrogens in both the Lewis structure and the balls-and-sticks model. At the bottom of the figure, the atom numbering and the rings composing the steroid compounds are shown.

This enzyme, also known as estrogen biosynthetase, is responsible for the last (and key) step of the biosynthesis of steroid hormones from cholesterol. Specifically, aromatase is involved in the formation of C_{18} -estrogens, estrone (E_1), 17β -estradiol (E_2), and $17\beta, 16\alpha$ -estriol (E_3), from the C_{19} -androgens, androstenedione (ASD), testosterone (TST), and 16α -hydroxytestosterone (HTST), respectively (see Fig. 1.14).^{133,134} In fact, this is the only known enzyme responsible for the biosynthesis of estrogens from androgens in vertebrates.^{130,135,136} The history of aromatase, including its discovery and its characterization, has been exhaustively reviewed by Santen et al.¹³⁷ The initial discovery of estrogens and their bioactivity in urinary extracts in the 1920s encouraged members of the scientific community to better understand the structure, biosynthesis and function of these hormones.¹³⁷ In the 1930s, androgens and estrogens were isolated and characterized biochemically,^{138–142} showing similarities between the two classes of compounds. This fact suggested that C_{19} -steroids might be directly converted to C_{18} -estrogens.^{138–141,143} From that moment on, an exhaustive research process on estrogen biosynthesis began, aromatase being the subject of intense investigation, both biochemical and biophysical, especially in the last 50 years.^{128,137} Currently, we know that the biosynthesis of estrogens consists in the aromatization of the sterol *A*–ring of androgens, thus forming the aromatic estrogens. This conversion occurs through a process of three consecutive oxidations of the angular C_{19} -methyl group of the androgens, which is removed as formic acid during the course of the reaction.^{144–148} In this catalytic process, each oxidation step consumes 1 mol of NADPH, 1 mol of molecular oxygen, and requires the presence of the cytochrome P450 reductase (*CPR*) as a source of electrons.^{149–155}

Catalytic cycle of Aromatase

The overall process of aromatization of androgens via the enzyme aromatase has been depicted in Fig. 1.15. As can be observed in this figure, the first oxidation step (in blue) occurs through the hydroxylation of the C_{19} -methyl group of the androgen substrates that produces the C_{19} -*hydroxymethyl* derivative. This catalytic subcycle, which occurs with retention of configuration,^{156,157} is believed to proceed through the conventional hydrogen abstraction/oxygen rebound mechanism originally proposed by Groves and coworkers^{158,159} (see Fig. 1.16). This mechanism comprises two distinct stages: In the first one, Cpd I activates a *C*–*H* bond of the substrate which results in abstraction of the hydrogen atom by means of this oxidant agent. This leads to the formation of an alkyl radical on the carbon atom and the iron–hydroxo porphyrin complex. Later, a reorientation of both the alkyl and the *OH* groups is produced to facilitate the second stage (oxygen rebound step), where a new *C*–*O* bond is formed. In this step, the alkyl radical is recombined with the iron-bound hydroxyl radical to form the corresponding hydroxylated substrate and the iron(III) porphyrin complex.

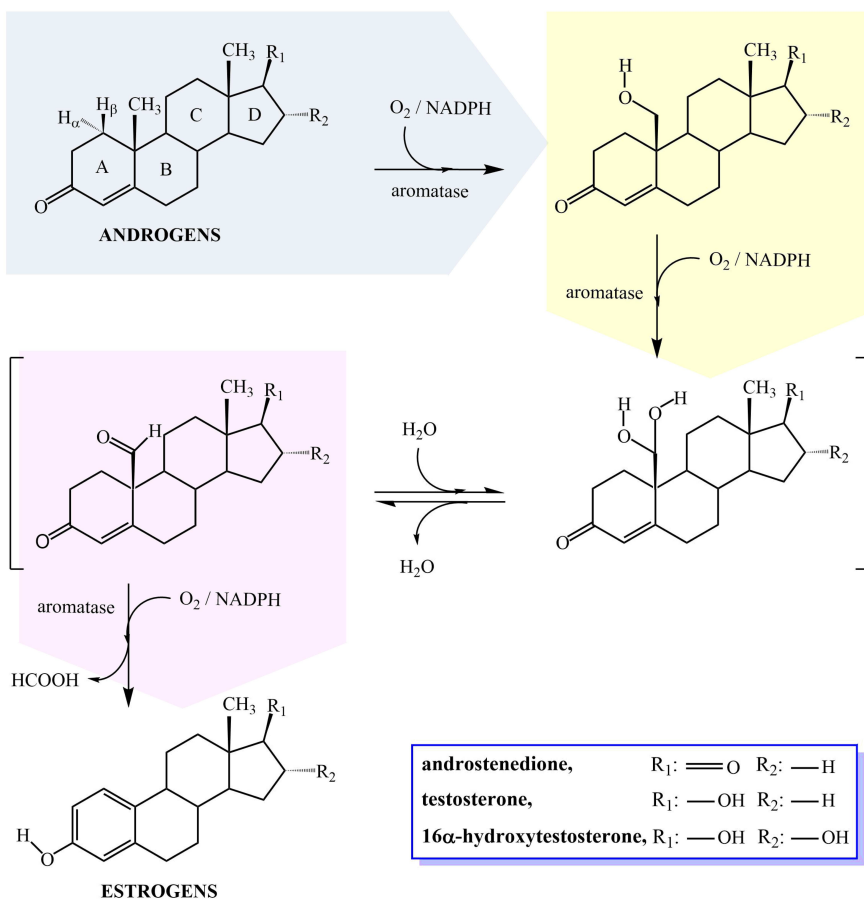


Figure 1.15.: Proposed catalytic cycle for the aromatization of androgens into estrogens via the enzyme aromatase (CYP19A1).

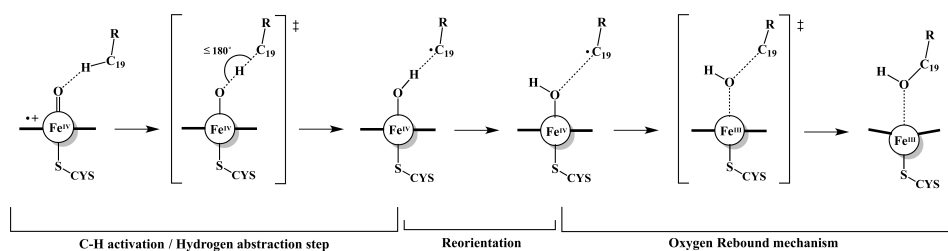


Figure 1.16.: Groves mechanism for the hydroxylation of substrates by cytochrome P450.

The second oxidation step (in yellow) is also believed to proceed through a second stereospecific hydroxylation of the C_{19} -methyl group of the androgen substrates. In this catalytic subcycle, the 19-pro-R hydrogen ($H_{19,R}$) is removed without an apparent kinetic isotopic effect to yield the C_{19} -gem-diol species.^{147,160} It

has been proposed that the C_{19} -*gem*-diol is then dehydrated to the aldehyde intermediate (C_{19} -*oxo*-androgen); however this dehydration step remains under dispute even today.

Finally, the third step consists of a lyase reaction, in which the $C_{10}-C_{19}$ bond of the androgens is cleaved, resulting in the aromatization of the phenolic A-ring of the androgens. The mechanism corresponding to this step has been studied extensively over the years, and several mechanisms have been proposed, including 1β -¹⁶¹ and 2β -hydroxylation,¹⁶²⁻¹⁶⁴ 4,5-epoxidation,^{165,166} the Baeyer-Villiger oxidation of C_{19} and 10β -hydroxyestr-4-ene-3,17-dione formation,¹⁴⁵ or with the intermediation of a C_{19} -peroxide.^{145,167} However, these mechanisms were rejected by later experimental works. It is experimentally proven that the 1β - and 2β -hydrogens of the A-ring are eliminated selectively in this catalytic subcycle.^{168,169} Likewise, it has been shown that the C_{19} carbon is expelled as formic acid in which both oxygens derive from molecular oxygen.^{145,170} This sequence of chemical events is supported by kinetic analysis of the reaction and its intermediates.¹⁷¹ The currently most cited mechanism^{146,172} has been depicted in Fig. 1.17.

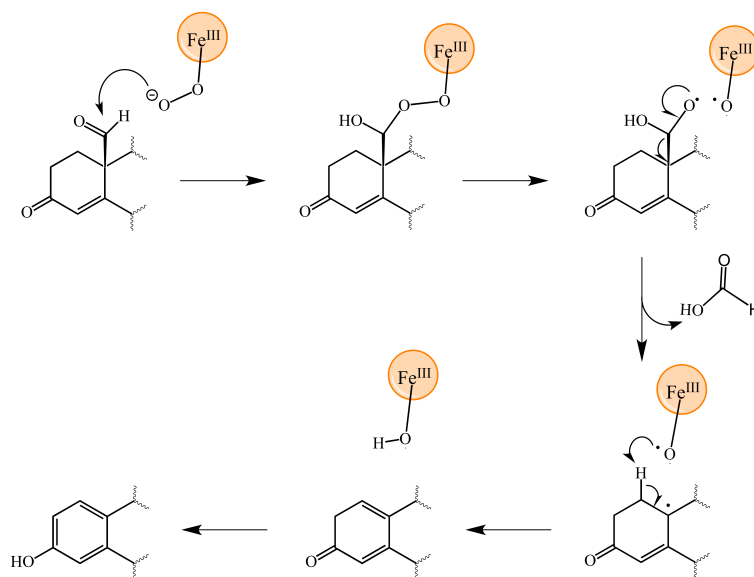


Figure 1.17.: The most commonly cited mechanism for the third oxidation step ($C_{10}-C_{19}$ bond cleavage reaction) that occurs in the conversion of androgens to estrogens by CYP19A1.

According to this mechanism, a peroxohemiacetal is formed after the nucleophilic attack of the ferric-peroxo anion ((5) in Fig. 1.10) on the C_{19} carbon of the C_{19} -*oxo*-androgen. Then, a homolytic cleavage of the peroxybond produces $Fe^{III}-O\cdot$ and an alkoxy radical, which decomposes, thereby expelling formic acid and forming a radical on the C_{10} carbon. Later, the $Fe^{III}-O\cdot$ species abstracts

the 1β -hydrogen, which leads to the formation of a C_1 - C_{10} double bond. Finally, an enolization of the 3-keto oxygen converts the A-ring into the aromatic phenol present in estrogens. A study on the oxidation of dihydrotestosterone by aromatase¹⁷³ indicated that the enolization of the 3-keto function is not a prerequisite for the C - C bond-cleaving function of aromatase, and thus enolization may occur prior to C - C bond cleavage.

Some theoretical studies have been performed regarding the third oxidation step catalyzed by aromatase. In 2005, a gas-phase DFT analysis of the C - C bond cleavage suggested that the Akhtar mechanism discussed above is not favored.¹⁷⁴ According to these authors, the 1β -hydrogen abstraction in this sequence shows a high-energy barrier, and thus they proposed an alternative mechanism. In this mechanism, the Cpd I abstracts the 1β -hydrogen to the 3-keto enolized form of the C_{19} -*gem*-diol. This leads initially to the formation of a radical on the C_1 carbon that evolves then to a cationic species, resulting in extrusion of formic acid and aromatization of the phenolic A-ring. However, a subsequent study performed by one of the authors,¹⁷⁵ this time including the enzymatic environment, revised the former mechanism proposed. In this study, the mechanism begins with a nucleophilic attack of the ferric-peroxo anion to the C_{19} -*oxo*-androgen, as in Akhtar's mechanism, but they introduce some subtleties into the subsequent steps.

Nonetheless, the role of Cpd I as the oxidant species in the third catalytic step is still thriving, as illustrated by the examples set out below. Recent experimental studies performed by Mak et al.¹⁷⁶ using resonance Raman spectroscopy and by Khatri et al.¹⁷⁷ using the kinetic solvent isotope effect implicate the involvement of Cpd I in this catalytic subcycle. Moreover, Yoshimoto and Guengerich¹⁷⁸ reproduced the ¹⁸O-labeling experiments previously conducted by Akhtar et al.^{145,179} and Caspi et al.,¹⁷⁰ but making use of more sensitive techniques and instruments. Their results proved that the aromatization mechanism could be rationalized only in terms of Cpd I, because unlike in the earlier experiments, the formic acid did not contain any labeled oxygen atom (¹⁸O) from molecular oxygen (¹⁸O₂). In this way, the ferric-peroxide mechanism, in which ¹⁸O is inevitably incorporated into the formic acid, might not explain this experimental observation.

As can be observed, the mechanism that takes place in the third and last oxidation of androgens via aromatase is still open and it remains under intense debate, which has still not reached consensus.

Crystal Structure and Active site of Aromatase

Even though the biosynthesis of estrogens has been extensively studied in the last few decades, many issues regarding this mechanism remain unresolved, partly because of the lack of a good crystallographic model. The crystallization of aromatase was challenging because of its strong hydrophobic character and suscepti-

bility to rapid denaturation in the absence of the protective lipid bilayer.¹⁸⁰ This fact prompted a number of homology models based on other P450 isoforms and site-directed mutagenesis data were proposed.^{181–185} Finally in 2009, the crystal structure of aromatase purified from human placenta in complex with its natural substrate ASD was determined by Ghosh et al. at a 2.9 Å resolution (protein data bank [PDB] code 3EQM).¹³⁰ The tertiary structure of aromatase displayed in the crystal structure is similar to other P450 enzymes, and is composed of 12 α -helices (A-L) and 10 β -strands (1-10) distributed into 1 major and 3 minor sheets (see Fig. 1.13). This X-ray structure also shows how the natural substrate is accommodated at the active site of this enzyme, and thus reveals significant details on the nature of the binding site of aromatase. The understanding of the most important interactions taking place at the active site is not only useful to better understand the aromatization reaction, but also for the development of new drugs based on aromatase inhibitors. The aromatase active site is enclosed near the geometric center of the enzyme and is formed by a heme cofactor and a series of amino acids at its distal position. These residues are as follows: Ile-133 and Phe-134 from B–C loop, Phe-221 and Trp-224 from F-helix, Ile-305, Ala-306, Asp-309, and Thr-310 from I-helix, Val-369, Val-370, Leu-372, and Val-373 from the K- β 3 loop, Met-374 from β 3, and Leu-477 and Ser-478 from the β 8- β 9 loop (see Fig. 1.18).¹³⁰ The cavity formed by the active site has a volume of 400 Å³, which is smaller than that found in other human isoforms, such as CYP3A4, CYP2D6, CYP11A1, CYP11B2, CYP17A1, CYP21A1, CYP24A1, and CYP51A1.¹²⁸ The small pocket formed by the active site reveals an androgen-specific cleft, in which the natural substrates (steroids) fit perfectly.¹⁸⁶ Specifically, ASD binds with its β -face toward the heme group, and its C₁₉-methyl group, which will be oxidized, is positioned very close to the iron atom (4 Å). The binding pocket where ASD is enclosed is mainly of a hydrophobic nature, with the exception of the polar residues Ser-478 and Thr-310. ASD in turn forms two hydrogen bonds between its polar moieties (17-keto and 3-keto oxygens) and the residues Met-374 and Asp-309, respectively. It has been suggested that the latter residue participates in the enolization reaction of ASD as a proton donor in the third catalytic subcycle of the enzyme aromatase.¹³⁰ Furthermore, there is an access channel to the active site that hosts the proton delivery network, whereby water molecules or steroid compounds may be moved, either inwardly or outwardly thereof. This channel is defined by residues Asp-309, Thr-310, Val-313, Phe-221, Ser-478, His-480, Arg-192, and Glu-483.¹³³ However, a study conducted by Sgrignani and Magistrato¹⁸⁷ using several computational techniques, and considering the influence of the membrane lipophilic environment, suggests an access/egress channel of human aromatase delimited by residues Phe-221, Trp-224, Gln-225, Val-313, Thr-320, Val-369, Val-370, Ser-478, and His-480.

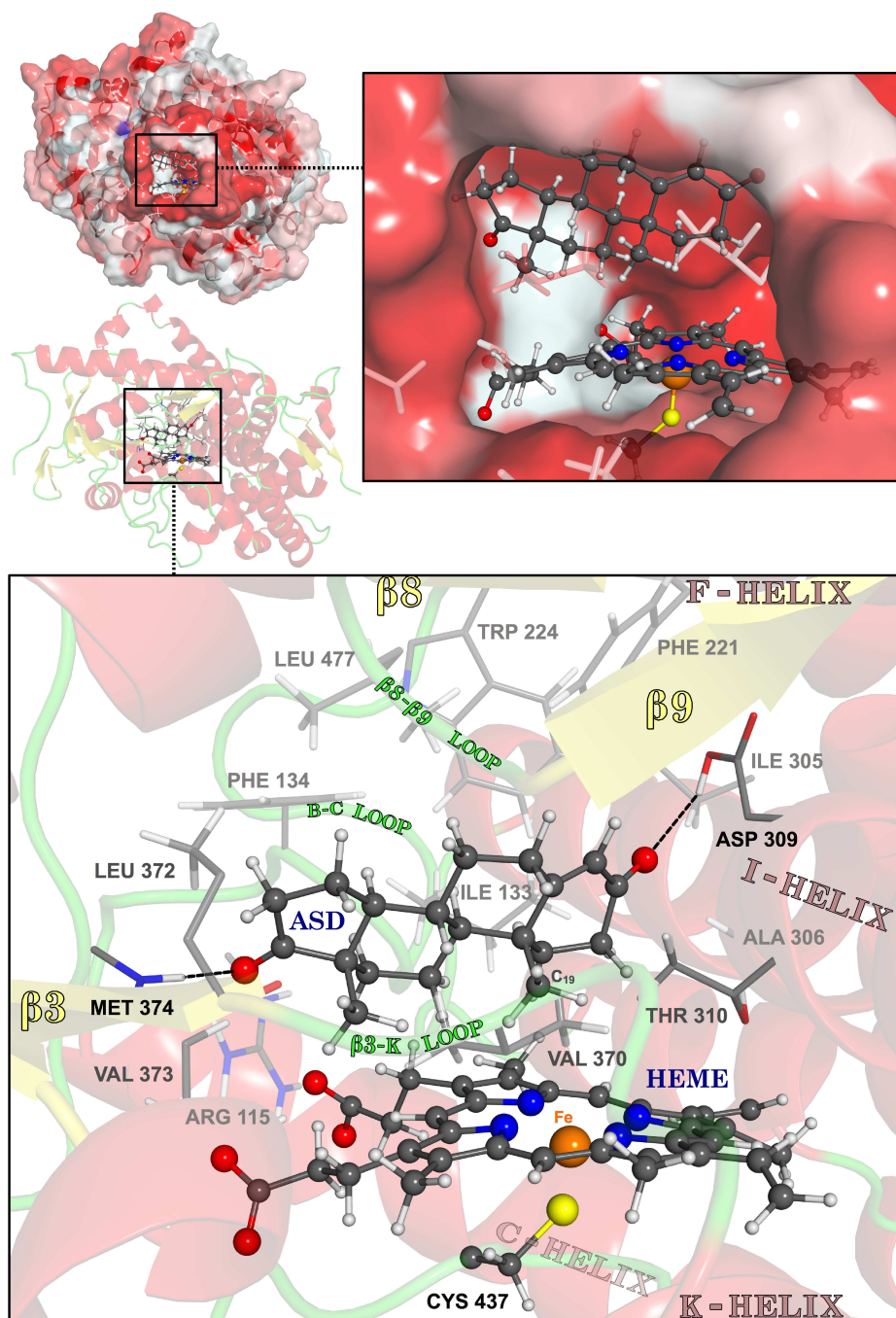


Figure 1.18.: At the top of the figure, the binding pocket of the active site of aromatase is shown. The van der Waals interaction surface shows a cavity complementary in shape to the bound ASD substrate, which has been colored white for the polar interactions and red for hydrophobic interactions. At the bottom of the figure, the residues comprising the catalytic cleft of aromatase, as well as the enzymatic environment at the active site, have been depicted.

1.2.4. Aromatase and Breast Cancer

Estrogens are important in humans because of the role they play in sexual and reproductive development. Particularly, estrogens are essential for women, because these are the primary female sex hormones. In premenopausal women, the main source of estrogens is the ovary,¹⁸⁸ which produces mostly estrone (E_1). Conversely, in postmenopausal women, the source of estrogen production is attributed to the aromatization of adrenal and ovarian androgens (ASD and TST) to estrogens (E_1 and E_2) via the enzyme aromatase in adipose tissue (fat).^{189,190} These steroid hormones, apart from being essential in female sexual development, also play a significant role in the growth and proliferation of hormone-dependent breast cancer. Breast cancer (BC) is the most common type of cancer among women (25% of the total).¹⁹¹ It is estimated that almost 1.7 million new cases of BC occurred among women worldwide in 2012 (most recent data available); in addition, that year, 522,000 women died as a result of this disease.¹⁹² Around two-thirds of all cases of BC are hormone-dependent and estrogen receptor-positive (ER+);¹⁹³⁻¹⁹⁷ yet this ratio increases up to 75% when we refer to postmenopausal BC.¹⁹⁸ The process involving estrogen in the proliferation of cancer cells in ER+ breast cancer is shown in Fig. 1.19.

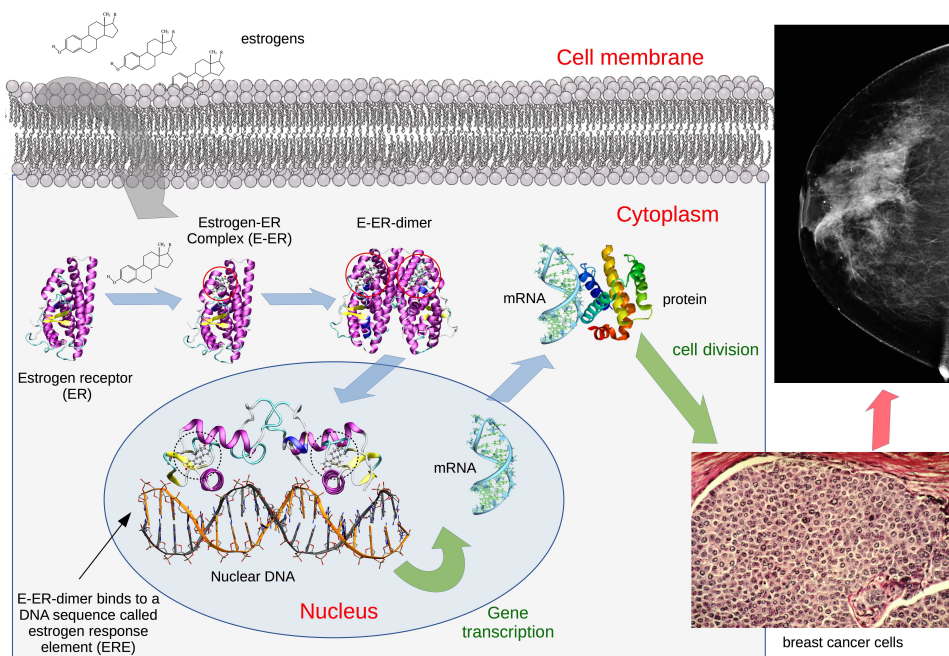


Figure 1.19.: Role of estrogens in the proliferation of ER+ breast cancer.

The estrogen receptor (ER) is a ligand-activated enhancer protein and it is a

member of the steroid/nuclear receptor superfamily.^{199,200} Mammalian *ER* is encoded in two genes: alpha (*ER* α) and beta (*ER* β), and they both work as signal transducers and transcription factors to modulate the expression of target genes.²⁰¹ Estrogens like E_2 act as ligands that bind to the *ER* in the cytoplasm of the cell, thereby forming the *E* – *ER* complex. This binding induces conformational changes in *ER* (activation), giving rise to the formation of a ligand-occupied *ER* dimer (*E* – *ER* dimer). The latter has a high affinity to bind to a specific sequence of DNA called estrogen response element (*ERE*). Once bound to an *ERE*, *E* – *ER* interacts with coactivator proteins and RNA polymerase II, resulting in increased gene transcription (synthesis of RNAm).^{200,201} This enhanced transcription stimulates proliferation of human mammary cells,^{202,203} and the high division rate allows less time for DNA repair. As a result, there is an increased likelihood of errors during DNA replication, which may lead to mutations and thus to carcinogenesis.¹⁹⁰

Based on the connection between *ER* and proliferation of cancer cells,¹⁹² two different strategies have been developed for the prevention and treatment of ER-positive breast cancer.²⁰⁴ The first one is to block the estrogen action by means of antiestrogens, such as the drug tamoxifen. Antiestrogens compete to bind to the estrogen receptors, so that when they bind to these, the number of receptors available for binding to endogenous estrogen is reduced. In this way, these “estrogen inhibitors” interfere with the transcription of estrogen-responsive genes, thus preventing estrogen-induced cell proliferation.^{194,205} The second strategy consists in blocking estrogen biosynthesis so that ultimately there are not sufficient levels of circulating estrogen to bind to the receptors.²⁰⁶ This can be accomplished by the inhibition of the enzyme aromatase by means of drugs known as *aromatase inhibitors* (AIs).^{133,137,194} For this reason, aromatase provides one of the first molecular targets for rational drug development in the treatment of this disease. AIs have proved to be effective not only in the first-line therapy for postmenopausal breast cancer, but also for the treatment of endometriosis,²⁰⁷ ovarian,²⁰⁸ and lung cancers.²⁰⁹

Aromatase inhibitors

The AIs can be classified according to their structures into Type I (steroidal) and Type II (nonsteroidal) inhibitors.^{210,211} The former, such as exemestane (EXE) or formestane, have a structure similar to natural substrate ASD, and hence they act as pseudo substrates of aromatase.^{212,213} In this way, these inhibitors irreversibly bind to the enzyme active site and thus act as mechanism-based inactivators or suicidal inhibitors. The latter are mostly azole-based compounds, such as anastrozole or letrozole, which are triazole derivatives. These inhibitors compete reversibly with the natural substrate to bind the enzyme, thus interfering with the heme–iron moiety of the cofactor.^{189,203,214,215} Currently, the first-line therapy for advanced hormone-dependent breast cancer is based on the third generation of AIs.

This family of inhibitors consists of three compounds, EXE, letrozole (LTZ), and anastrozole (ANZ), which are marketed under the brand names AROMASIN®, FEMARA®, and ARIMIDEX®, respectively (see Fig. 1.20).^{197,189} However, these AIs exhibit some drawbacks, such as the estrogenic weakness of EXE through its agonistic interaction with ER,²¹⁶ or cross-reactivity with other P450 isoforms. In fact, it is known that ANZ and LTZ, apart from inhibiting aromatase, also react with CYP1A2, CYP2C9, CYP2A6,^{217,218} and CYP3A4,²¹⁹ while EXE is metabolized by the latter.^{194,220–223} Work is currently underway on the discovery of new AIs that incorporate aromatase specificity, while improving some of the drawbacks of the AIs currently available. This can be accomplished because the crystal structure of human placental aromatase has revealed the architecture of the active site, thus making it possible to know the characteristic interactions that are established between substrates and their androgen-specific cleft.^{128,133,180}

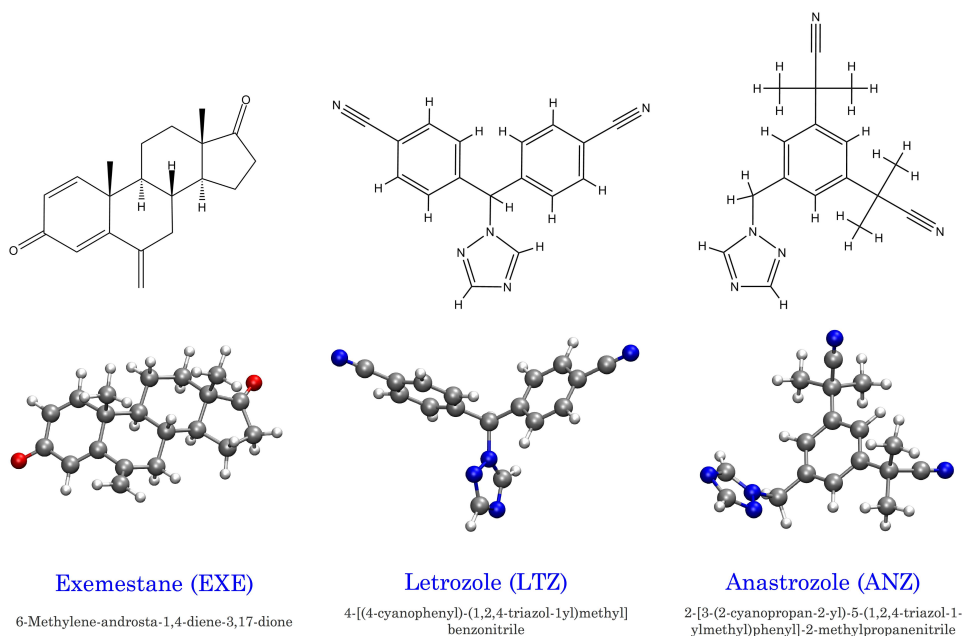


Figure 1.20.: Third generation aromatase inhibitors.

Bibliography

- [1] Nelson, D. L.; Lehninger, A. L.; Cox, M. M. *Lehninger principles of biochemistry*, 5th ed.; W. H. Freeman and Company: New York, NY, 2008.
- [2] Kohen, A. *Prog. React. Kinet. Mech.* **2003**, *28*, 119–156.
- [3] Berg, J. M.; Tymoczko, J. L.; Stryer, L. *Biochemistry*, 5th ed.; W. H. Freeman and Company: New York, NY, 2002.
- [4] Murray, R. k.; Bender, D.; Botham, K. M.; Kennelly, P. J.; Rodwell, V.; Weil, P. A. *Harper's Illustrated Biochemistry*, 29th ed.; McGraw Hill Professional, 2012.
- [5] Price, N. C.; Dwek, R. G. W. M. R., Raymond A.; Ratcliffe *Principles and problems in physical chemistry for biochemists*, 3rd ed.; Oxford University Press, 2001.
- [6] Voet, D.; Voet, J. G. *Biochemistry*, 4th ed.; John Wiley & Sons: New York, NY, 2011.
- [7] Elliott, W. H.; Elliott, D. C. *Biochemistry and Molecular Biology*, 4th ed.; Oxford University Press, 2009.
- [8] Royal Society of Chemistry *Chemistry for Biologists*; Royal Society of Chemistry: <http://www.rsc.org/Education/Teachers/Resources/cfb/enzymes.htm>, 2004.
- [9] Price, N.; Stevens, L. *Fundamentals of enzymology: the cell and molecular biology of catalytic proteins*, 3rd ed.; Oxford University Press: Oxford, 1999.
- [10] Buchner, E. *Ber. Dtsch. Chem. Ges.* **1897**, *30*, 117–124.
- [11] Kühne, W. *Verhandlungen des naturhistorisch-medicinischen Vereins zu Heidelberg* **1876**, *1*, 190–193.
- [12] Sumner, J. B. *J. Biol. Chem.* **1926**, *69*, 435–441.
- [13] Northrop, J. H. *J. Gen. Physiol.* **1930**, *13*, 739–766.
- [14] Northrop, J. H.; Kunitz, M. *J. Gen. Physiol.* **1932**, *16*, 267–294.
- [15] Hirs, C. H. W.; Moore, S.; Stein, W. H. *J. Biol. Chem.* **1960**, *235*, 633–647.
- [16] Blake, C. C. F.; Koenig, D. F.; Mair, G. A.; North, A. C. T.; Phillips, D. C.; Sarma, V. R. *Nature* **1965**, *206*, 757–761.
- [17] Gutte, B.; Merrifield, R. B. *J. Am. Chem. Soc.* **1969**, *91*, 501–502.
- [18] Michaelis, L.; Menten, M. L. *Biochem. Z* **1913**, *49*, 352.
- [19] Henri, V. *Lois Générales de l'action des diastases*; Librairie Scientifique A. Hermann: Paris, 1903.
- [20] Segel, I. H. *Enzyme Kinetics: Behavior and analysis of rapid equilibrium and steady-state enzyme systems*, wiley classics library ed.; John Wiley & Sons, Inc: New York, 1993.
- [21] Haldane, J. *Enzymes*; Longmans, Green and co: London, 1930.
- [22] Gutfreund, H. *Kinetics for the life sciences: receptors, transmitters and catalysts*; Cambridge University Press: Cambridge, UK, 1995.
- [23] Berberan-Santos, M. N. *Match-Commun. Math. Comput. Chem.* **2010**, *63*, 283–318.
- [24] Briggs, G. E.; Haldane, J. B. S. *Biochem. J.* **1925**, *19*, 338–339.

- [25] Humphrey, W.; Dalke, A.; Schulten, K. *J. Mol. Graphics* **1996**, *14*, 33–38.
- [26] Fischer, E. *Ber. Dtsch. Chem. Ges.* **1890**, *23*, 2611–2624.
- [27] Fischer, E. *Ber. Dtsch. Chem. Ges.* **1894**, *27*, 2985–2993.
- [28] Koshland, D. E. *Proc. Natl. Acad. Sci. U. S. A.* **1958**, *44*, 98–104.
- [29] Polanyi, M. Z. *Elektrochem. Angew. Phys. Chem.* **1921**, *27*, 142–150.
- [30] Pauling, L. *Chem. Eng. News* **1946**, *24*, 1375–1377.
- [31] Marti, S.; Roca, M.; Andres, J.; Moliner, V.; Silla, E.; Tunon, I.; Bertran, J. *Chem. Soc. Rev.* **2004**, *33*, 98–107.
- [32] Eyring, H. *J. Chem. Phys.* **1935**, *3*, 107–115.
- [33] Evans, M. G.; Polanyi, M. *Trans. Faraday Soc.* **1935**, *31*, 875–894.
- [34] Warshel, A. *Proc. Natl. Acad. Sci.* **1978**, *75*, 5250–5254.
- [35] Villà, J.; ; Warshel, A. *J. Phys. Chem. B* **2001**, *105*, 7887–7907.
- [36] Warshel, A. *J. Biol. Chem.* **1998**, *273*, 27035–27038.
- [37] Cleland, W.; Kreevoy, M. *Science* **1994**, *264*, 1887–1890.
- [38] Frey, P. A.; Whitt, S. A.; Tobin, J. B. *Science* **1994**, *264*, 1927–1930.
- [39] Frey, P. A.; Hegeman, A. D. *Enzymatic Reaction Mechanisms*; Oxford University Press: New York, 2007.
- [40] Jang, D. S.; Choi, G.; Cha, H. J.; Shin, S.; Hong, B. H.; Lee, H. J.; Lee, H. C.; Choi, K. Y. *Mol. Cells* **2015**, *38*, 409–415.
- [41] Golubev, N.; Denisov, G.; Gindin, V.; Ligay, S.; Limbach, H.-H.; Smirnov, S. *J. Mol. Struct.* **1994**, *322*, 83 – 91.
- [42] Choi, G.; Ha, N.-C.; Kim, S. W.; Kim, D.-H.; Park, S.; Oh, B.-H.; Choi, K. Y. *Biochemistry* **2000**, *39*, 903–909.
- [43] Zhao, Q.; Abeygunawardana, C.; Talalay, P.; Mildvan, A. S. *Proc. Natl. Acad. Sci.* **1996**, *93*, 8220–8224.
- [44] Warshel, A. *Acc. Chem. Res.* **1981**, *14*, 284–290.
- [45] Jencks, W. P. *Catalysis in chemistry and enzymology*; Dover Publications: New York, 1987.
- [46] Cohen, S. G.; Vaidya, V. M.; Schultz, R. M. *Proc. Natl. Acad. Sci.* **1970**, *66*, 249–256.
- [47] Wolfenden, R. *Science* **1983**, *222*, 1087–1093.
- [48] Dewar, M. J. S.; Storch, D. M. *Proc. Natl. Acad. Sci.* **1985**, *82*, 2225–2229.
- [49] Dewar, M. J. S.; Dieter, K. M. *Biochemistry* **1988**, *27*, 3302–3308.
- [50] Dewar, M. J. S. *Enzyme* **1986**, *36*, 8–20.
- [51] Weiner, S. J.; Singh, U. C.; Kollman, P. A. *J. Am. Chem. Soc.* **1985**, *107*, 2219–2229.
- [52] Storm, D. R.; Jr., D. E. K. *J. Am. Chem. Soc.* **1972**, *94*, 5815–5825.
- [53] Dafforn, A.; Koshland, D. *Biochem. Biophys. Res. Commun.* **1973**, *52*, 779 – 785.

- [54] Page, M. I.; Jencks, W. P. *Proc. Natl. Acad. Sci.* **1971**, *68*, 1678–1683.
- [55] Kollman, P. A.; Kuhn, B.; Donini, O.; Peräkylä, M.; Stanton, R.; ; Bakowies, D. *Acc. Chem. Res.* **2001**, *34*, 72–79.
- [56] Westheimer, F. H. *Mechanisms Related to Enzyme Catalysis*; John Wiley & Sons, Inc., 2006; pp 441–482.
- [57] Hermans, J.; ; Wang, L. *J. Am. Chem. Soc.* **1997**, *119*, 2707–2714.
- [58] Stanton, R. V.; Peräkylä, M.; Bakowies, D.; ; Kollman, P. A. *J. Am. Chem. Soc.* **1998**, *120*, 3448–3457.
- [59] Kuhn, B.; ; Kollman, P. A. *J. Am. Chem. Soc.* **2000**, *122*, 2586–2596.
- [60] Hur, S.; Bruice, T. C. *Proc. Natl. Acad. Sci. U. S. A.* **2003**, *100*, 12015–12020.
- [61] Hur, S.; ; Bruice, T. C. *J. Am. Chem. Soc.* **2003**, *125*, 10540–10542.
- [62] Bruice, T. C.; Benkovic, S. J. *Biochemistry* **2000**, *39*, 6267–6274.
- [63] Cha, Y.; Murray, C. J.; Klinman, J. P. *Science* **1989**, *243*, 1325–1330.
- [64] Kohen, A.; Cannio, R.; Bartolucci, S.; Klinman, J. P.; Klinman, J. P. *Nature* **1999**, *399*, 496–499.
- [65] Luk, L. Y. P.; Loveridge, E. J.; Allemann, R. K. *Phys. Chem. Chem. Phys.* **2015**, *17*, 30817–30827.
- [66] Kohen, A. *Acc. Chem. Res.* **2015**, *48*, 466–473.
- [67] Warshel, A.; Sharma, P. K.; Kato, M.; Xiang, Y.; Liu, H.; ; Olsson, M. H. M. *Chem. Rev.* **2006**, *106*, 3210–3235.
- [68] Pisljakov, A. V.; Cao, J.; Kamerlin, S. C. L.; Warshel, A. *Proc. Natl. Acad. Sci.* **2009**, *106*, 17359–17364.
- [69] Kamerlin, S. C. L.; Warshel, A. *Proteins: Struct., Funct., Bioinf.* **2010**, *78*, 1339–1375.
- [70] Doshi, U.; McGowan, L. C.; Ladani, S. T.; Hamelberg, D. *Proc. Natl. Acad. Sci.* **2012**, *109*, 5699–5704.
- [71] Soriano, A.; Silla, E.; Tuñón, I.; Ruiz-López, M. F. *J. Am. Chem. Soc.* **2005**, *127*, 1946–1957.
- [72] Bergsma, J. P.; Gertner, B. J.; Wilson, K. R.; Hynes, J. T. *J. Chem. Phys.* **1987**, *86*, 1356–1376.
- [73] Ruiz-Pernía, J. J.; Tuñón, I.; Moliner, V.; Hynes, J. T.; Roca, M. *J. Am. Chem. Soc.* **2008**, *130*, 7477–7488.
- [74] Castillo, R.; Roca, M.; Soriano, A.; Moliner, V.; Tuñón, I. *J. Phys. Chem. B* **2008**, *112*, 529–534.
- [75] Luk, L. Y. P.; Ruiz-Pernía, J. J.; Dawson, W. M.; Roca, M.; Loveridge, E. J.; Glowacki, D. R.; Harvey, J. N.; Mulholland, A. J.; Tuñón, I.; Moliner, V.; Allemann, R. K. *Proc. Natl. Acad. Sci.* **2013**, *110*, 16344–16349.
- [76] Gertner, B. J.; Wilson, K. R.; Hynes, J. T. *J. Chem. Phys.* **1989**, *90*, 3537–3558.
- [77] Grote, R. F.; Hynes, J. T. *J. Chem. Phys.* **1980**, *73*, 2715–2732.

- [78] Nelson, D. R. *Hum. Genomics* **2009**, *4*, 59–65.
- [79] Ortiz de Montellano, P. R. *Cytochrome P450 Structure, Mechanism, and Biochemistry*, 3rd ed.; Kluwer Academic/Plenum Publishers: New York, 2005.
- [80] Meunier, B.; de Visser, S. P.; Shaik, S. *Chem. Rev.* **2004**, *104*, 3947–3980.
- [81] Shaik, S.; Cohen, S.; Wang, Y.; Chen, H.; Kumar, D.; Thiel, W. *Chem. Rev.* **2010**, *110*, 949–1017.
- [82] Klingenberg, M. *Arch. Biochem. Biophys.* **1958**, *75*, 376 – 386.
- [83] Yamazaki, H. *Fifty Years of Cytochrome P450 Research*; Springer Japan: Tokio, 2014.
- [84] Omura, T.; Sato, R. *J. Biol. Chem.* **1962**, *237*, PC1375–PC1376.
- [85] Omura, T.; Sato, R. *J. Biol. Chem.* **1964**, *239*, 2370–2378.
- [86] Omura, T.; Sato, R. *J. Biol. Chem.* **1964**, *239*, 2379–2385.
- [87] Estabrook, C. D. Y., R. W.; Rosenthal, O. *Bio-chem. Z.* **1963**, *338*, 741–755.
- [88] Omura, T. *Biochem. Biophys. Res. Commun.* **1999**, *266*, 690 – 698.
- [89] Nebert, D. W.; Adesnik, M.; Coon, M. J.; Estabrook, R. W.; Gonzalez, F. J.; Guengerich, F. P.; Gunsalus, I. C.; Johnson, E. F.; Kemper, B.; Levin, W.; R., P. I.; Sato, R.; Waterman, M. R. *DNA* **1987**, *6*, 1–11.
- [90] Guengerich, F. P. *Chem. Res. Toxicol.* **2001**, *14*, 611–650.
- [91] Guengerich, F. P.; Macdonald, T. L. *Acc. Chem. Res.* **1984**, *17*, 9–16.
- [92] de Montellano, P. R. O. *Chem. Rev.* **2010**, *110*, 932–948.
- [93] Sligar, S. G.; Makris, T. M.; Denisov, I. G. *Biochem. Biophys. Res. Commun.* **2005**, *338*, 346–354.
- [94] Hrycay, E. G.; Bandiera, S. M. *Arch. Biochem. Biophys.* **2012**, *522*, 71–89.
- [95] Hrycay, E. G.; Bandiera, S. M. *Monoxygenase, Peroxidase and Peroxygenase Properties and Mechanisms of Cytochrome P450*, 1st ed.; Advances in Experimental Medicine and Biology, Vol. 851; Springer International Publishing, 2015.
- [96] Guengerich, F. P. *J. Biochem. Mol. Toxicol.* **2007**, *21*, 163–168.
- [97] Guengerich, F. P. *Chem. Res. Toxicol.* **2007**, *21*, 70–83.
- [98] Guengerich, F. P. *Aaps J.* **2006**, *8*, E101–E111.
- [99] Davydov, R.; Makris, T. M.; Kofman, V.; Werst, D. E.; Sligar, S. G.; Hoffman, B. M. *J. Am. Chem. Soc.* **2001**, *123*, 1403–1415.
- [100] Shaik, S.; Kumar, D.; de Visser, S. P.; Altun, A.; Thiel, W. *Chem. Rev.* **2005**, *105*, 2279–2328.
- [101] Makris, T. M.; von Koenig, K.; Schlichting, I.; Sligar, S. G. *J. Inorg. Biochem.* **2006**, *100*, 507–518.
- [102] Sligar, S. G. *Science* **2010**, *330*, 924–925.
- [103] Schlichting, I.; Berendzen, J.; Chu, K.; Stock, A. M.; Maves, S. A.; Benson, D. E.; Sweet, B. M.; Ringe, D.; Petsko, G. A.; Sligar, S. G. *Science* **2000**, *287*, 1615–1622.
- [104] Rittle, J.; Green, M. T. *Science* **2010**, *330*, 933–937.

- [105] Schoneboom, J. C.; Lin, H.; Reuter, N.; Thiel, W.; Cohen, S.; Ogliaro, F.; Shaik, S. *J. Am. Chem. Soc.* **2002**, *124*, 8142–8151.
- [106] Denisov, I. G.; Makris, T. M.; Sligar, S. G.; Schlichting, I. *Chem. Rev.* **2005**, *105*, 2253–2277.
- [107] Denisov, I. G.; Mak, P. J.; Makris, T. M.; Sligar, S. G.; Kincaid, J. R. *J. Phys. Chem. A* **2008**, *112*, 13172–13179.
- [108] Newcomb, M.; Shen, R.; Choi, S.-Y.; Toy, P. H.; Hollenberg, P. F.; Vaz, A. D. N.; Coon, M. J. *J. Am. Chem. Soc.* **2000**, *122*, 2677–2686.
- [109] Newcomb, M.; Aebischer, D.; Shen, R.; Chandrasena, R. E. P.; Hollenberg, P. F.; Coon, M. J. *J. Am. Chem. Soc.* **2003**, *125*, 6064–6065.
- [110] Newcomb, M.; Zhang, R.; Chandrasena, R. E. P.; Halgrimson, J. A.; Horner, J. H.; Makris, T. M.; Sligar, S. G. *J. Am. Chem. Soc.* **2006**, *128*, 4580–4581.
- [111] Shaik, S.; Kumar, D.; de Visser, S. P. *J. Am. Chem. Soc.* **2008**, *130*, 10128–10140.
- [112] Isobe, H.; Yamanaka, S.; Okumura, M.; Yamaguchi, K.; Shimada, J. *J. Phys. Chem. B* **2011**, *115*, 10730–10738.
- [113] Rydberg, P.; Sigfridsson, E.; Ryde, U. *JBIC, J. Biol. Inorg. Chem.* **2004**, *9*, 203–223.
- [114] Goodin, D. B.; McRee, D. E. *Biochemistry* **1993**, *32*, 3313–3324.
- [115] Gajhede, M. *Handbook of metalloproteins*; Messerschmidt, A.; Huber, R.; Poulos, T.; Wieghart, K., Eds.; Wiley: Chichester, 2001; Chapter Horseradish Peroxidase, pp 195–209.
- [116] Maté, M.; Bravo, J.; Fita, I.; Murshudov, G.; Melik-Adamyan, W.; Loewen, P. *Handbook of metalloproteins*; Messerschmidt, A.; Huber, R.; Poulos, T.; Wieghart, K., Eds.; Wiley: Chichester, 2001; Chapter Heme-Catalases, pp 486–502.
- [117] de Visser, S. P.; Ogliaro, F.; Sharma, P. K.; Shaik, S. *Angew. Chem. Intl. Ed.* **2002**, *41*, 1947–1951.
- [118] Harris, D. L.; Loew, G. H. *J. Am. Chem. Soc.* **1998**, *120*, 8941–8948.
- [119] Harris, D.; Loew, G.; Waskell, L. *J. Am. Chem. Soc.* **1998**, *120*, 4308–4318.
- [120] Harris, D. L. *Curr. Opin. Chem. Biol.* **2001**, *5*, 724–735.
- [121] Harris, D. L.; Loew, G. H. *J. Porphyr. Phthalocyanines* **2001**, *5*, 334–344.
- [122] Harris, D.; Loew, G.; Waskell, L. *J. Inorg. Biochem.* **2001**, *83*, 309–318.
- [123] Loew, G. H.; Harris, D. L. *Chem. Rev.* **2000**, *100*, 407–419.
- [124] Loew, G. *Int. J. Quantum Chem.* **2000**, *77*, 54–70.
- [125] Shaik, S.; de Visser, S. P.; Ogliaro, F.; Schwarz, H.; Schroder, D. *Curr. Opin. Chem. Biol.* **2002**, *6*, 556–567.
- [126] Harris, N.; Cohen, S.; Filatov, M.; Ogliaro, F.; Shaik, S. *Angew. Chem. Intl. Ed.* **2000**, *39*, 2003–2007.
- [127] Guengerich, P. F. *Cytochrome P450: Structure, Mechanism, and Biochemistry*; Ortiz de Montellano, R. P., Ed.; Springer International Publishing, 2015; Chapter Human Cytochrome P450 Enzymes, pp 523–785.

- [128] Ghosh, D.; Lo, J.; Egbuta, C. *Resistance to Aromatase Inhibitors in Breast Cancer*; Larionov, A., Ed.; Springer International Publishing, 2015; Chapter Structure, Function and Inhibition of Aromatase, pp 33–61.
- [129] HGNC-2594; www.genenames.org.
- [130] Ghosh, D.; Griswold, J.; Erman, M.; Pangborn, W. *Nature* **2009**, *457*, 219–223.
- [131] Simpson, E. R.; Davis, S. R. *Endocrinology* **2001**, *142*, 4589–4594.
- [132] Simpson, E. R.; Clyne, C.; Rubin, G.; Boon, W. C.; Robertson, K.; Britt, K.; Speed, C.; Jones, M. *Annu. Rev. Physiol.* **2002**, *64*, 93–127.
- [133] Ghosh, D.; Lo, J.; Morton, D.; Valette, D.; Xi, J.; Griswold, J.; Hubbell, S.; Egbuta, C.; Jiang, W.; An, J.; Davies, H. M. L. *J. Med. Chem.* **2012**, *55*, 8464–8476.
- [134] Hong, Y.; Cho, M.; Yuan, Y. C.; Chen, S. *Biochem. Pharmacol.* **2008**, *75*, 1161–1169.
- [135] Thompson, E. A.; Siiteri, P. K. *J. Biol. Chem.* **1974**, *249*, 5364–5372.
- [136] Simpson, E. R.; Mahendroo, M. S.; Means, G. D.; Kilgore, M. W.; Hinshelwood, M. M.; Grahamlence, S.; Amarneh, B.; Ito, Y. J.; Fisher, C. R.; Michael, M. D.; Mendelson, C. R.; Bulun, S. E. *Endocr. Rev.* **1994**, *15*, 342–355.
- [137] Santen, R. J.; Brodie, H.; Simpson, E. R.; Siiteri, P. K.; Brodie, A. *Endocr. Rev.* **2009**, *30*, 343–375.
- [138] Thayer, S. A.; Levin, L.; Doisy, E. A. *J. Biol. Chem.* **1931**, *91*, 655–665.
- [139] Veler, C. D.; Thayer, S.; Doisy, E. A. *J. Biol. Chem.* **1930**, *87*, 357–371.
- [140] Butenandt, A. *Angew. Chem.* **1931**, *44*, 905–908.
- [141] Butenandt, A. *Die Naturwissenschaften* **1929**, *17*, 879–879.
- [142] Fieser, L.; Fieser, M. *Steroids*; Reinhold Publishing Corporation: Baltimore, MD, 1959.
- [143] Zondek, B. *Nature* **1934**, *133*, 494–494.
- [144] Akhtar, M.; Skinner, S. J. *Biochem. J.* **1968**, *109*, 318–321.
- [145] Akhtar, M.; Calder, M. R.; Corina, D. L.; Wright, J. N. *Biochem. J.* **1982**, *201*, 569–580.
- [146] Akhtar, M.; Wright, J. N.; Lee-Robichaud, P. *J. Steroid Biochem. Mol. Biol.* **2011**, *125*, 2–12.
- [147] Arigoni, D.; Battaglia, R.; Akhtar, M.; Smith, T. *J. Chem. Soc. Chem. Commun.* **1975**, *0*, 185–186.
- [148] Skinner, S. J. M.; Akhtar, M. *Biochem. J.* **1969**, *114*, 75–81.
- [149] Hong, Y. Y.; Li, H. Z.; Yuan, Y. C.; Chen, S. A. *J. Steroid Biochem. Mol. Biol.* **2010**, *118*, 203–206.
- [150] Iyanagi, T. *Biochem. Biophys. Res. Commun.* **2005**, *338*, 520–528.
- [151] Vatsis, K. P.; Coon, M. J. *Biochem. Biophys. Res. Commun.* **2005**, *337*, 1107–1111.
- [152] Surawatanawong, P.; Tye, J. W.; Hall, M. B. *Inorg. Chem.* **2010**, *49*, 188–198.
- [153] Pandey, A. V.; Kempna, P.; Hofer, G.; Mullis, P. E.; Fluck, C. E. *Mol. Endocrinol.* **2007**, *21*, 2579–2595.

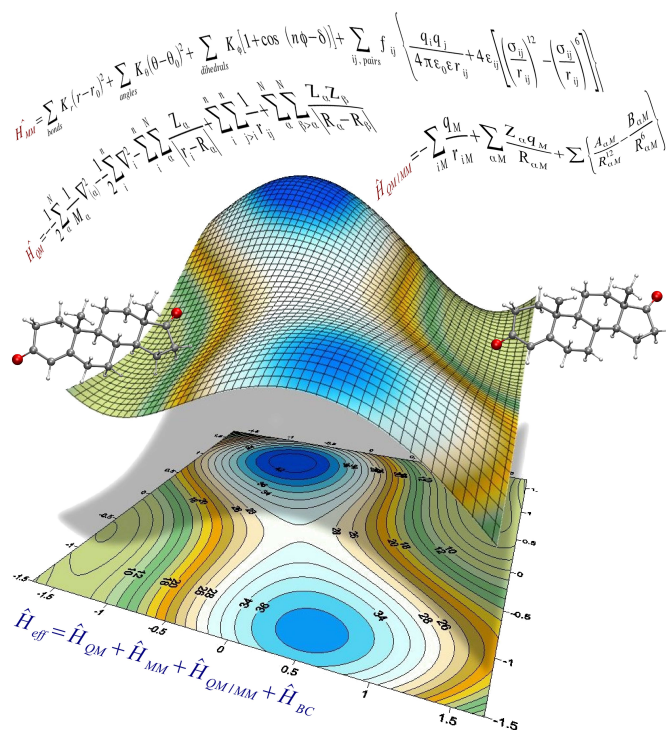
- [154] Lu, A. Y. H.; Junk, K. W.; Coon, M. J. *J. Biol. Chem.* **1969**, *244*, 3714–3721.
- [155] Ghosh, D.; Jiang, W. H.; Lo, J.; Egbuta, C. *Steroids* **2011**, *76*, 753–758.
- [156] Caspi, E.; Arunachalam, T.; Nelson, P. A. *J. Am. Chem. Soc.* **1983**, *105*, 6987–6989.
- [157] Caspi, E.; Arunachalam, T.; Nelson, P. A. *J. Am. Chem. Soc.* **1986**, *108*, 1847–1852.
- [158] Groves, J. T.; McClusky, G. A. *J. Am. Chem. Soc.* **1976**, *98*, 859–861.
- [159] Groves, J. T.; Watanabe, Y. *J. Am. Chem. Soc.* **1988**, *110*, 8443–8452.
- [160] Osawa, Y.; Shibata, K.; Rohrer, D.; Weeks, C.; Duax, W. L. *J. Am. Chem. Soc.* **1975**, *97*, 4400–4402.
- [161] Townsley, J. D.; Brodie, H. J. *Biochemistry* **1968**, *7*, 33–40.
- [162] Hosoda, H.; Fishman, J. *J. Am. Chem. Soc.* **1974**, *96*, 7325–7329.
- [163] Hahn, E. F.; Fishman, J. *J. Biol. Chem.* **1984**, *259*, 1689–1694.
- [164] Goto, J.; Fishman, J. *Science* **1977**, *195*, 80–81.
- [165] Morand, P.; Williamson, D. G.; Layne, D. S.; Lompa-Krzymien, L.; Salvador, J. *Biochemistry* **1975**, *14*, 635–638.
- [166] Mastalerz, H.; Morand, P. *J. Chem. Soc. Perkin Trans. 1* **1982**, 2611–2615.
- [167] Covey, D. F.; Hood, W. F. *Cancer Res.* **1982**, *42*, 3327–3333.
- [168] Brodie, H. J.; Kripalani, K. J.; Possanza, G. *J. Am. Chem. Soc.* **1969**, *91*, 1241–1242.
- [169] Fishman, J.; Guzik, H.; Dixon, D. *Biochemistry* **1969**, *8*, 4304–4309.
- [170] Caspi, E.; Wicha, J.; Arunachalam, T.; Nelson, P.; Spiteller, G. *J. Am. Chem. Soc.* **1984**, *106*, 7282–7283.
- [171] Sohl, C. D.; Guengerich, F. P. *J. Biol. Chem.* **2010**, *285*, 17734–17743.
- [172] Di Nardo, G.; Gilardi, G. *Biotechnol. Appl. Biochem.* **2013**, *60*, 92–101.
- [173] Cheng, Q.; Sohl, C. D.; Yoshimoto, F. K.; Guengerich, F. P. *J. Biol. Chem.* **2012**, *287*, 29554–29567.
- [174] Hackett, J. C.; Brueggemeier, R. W.; Hadad, C. M. *J. Am. Chem. Soc.* **2005**, *127*, 5224–5237.
- [175] Sen, K.; Hackett, J. C. *Biochemistry* **2012**, *51*, 3039–3049.
- [176] Mak, P. J.; Luthra, A.; Sligar, S. G.; Kincaid, J. R. *J. Am. Chem. Soc.* **2014**, *136*, 4825–4828.
- [177] Khatri, Y.; Luthra, A.; Duggal, R.; Sligar, S. G. *FEBS Lett.* **2014**, *588*, 3117–3122.
- [178] Yoshimoto, F. K.; Guengerich, F. P. *J. Am. Chem. Soc.* **2014**, *136*, 15016–15025.
- [179] Akhtar, M.; Corina, D.; Pratt, J.; Smith, T. *J. Chem. Soc. Chem. Commun.* **1976**, 854–856.
- [180] Ghosh, D.; Lo, J.; Egbuta, C. *J. Med. Chem.* **2015**, *ASAP*.
- [181] Szklarz, G. D.; Halpert, J. R. *Life Sci.* **1997**, *61*, 2507 – 2520.
- [182] Laughton, C.; Zvelbil, M.; Neidle, S. *J. Steroid Biochem. Mol. Biol.* **1993**, *44*, 399 – 407.

- [183] Favia, A. D.; Cavalli, A.; Masetti, M.; Carotti, A.; Recanatini, M. *Proteins: Struct., Funct., Bioinf.* **2006**, *62*, 1074–1087.
- [184] Karkola, S.; Holtje, H. D.; Wahala, K. *J. Steroid Biochem. Mol. Biol.* **2007**, *105*, 63–70.
- [185] Graham-Lorence, S.; Amarneh, B.; White, R. E.; Peterson, J. A.; Simpson, E. R. *Protein Sci.* **1995**, *4*, 1065–1080.
- [186] Ghosh, D.; Griswold, J.; Erman, M.; Pangborn, W. *J. Steroid Biochem. Mol. Biol.* **2010**, *118*, 197–202.
- [187] Sgrignani, J.; Magistrato, A. *J. Chem. Inf. Model.* **2012**, *52*, 1595–1606.
- [188] Miller, W. *Cancer Treat. Rev.* **1997**, *23*, 171–187.
- [189] Lombardi, P. *Biochim. Biophys. Acta, Mol. Basis Dis.* **2002**, *1587*, 326–337.
- [190] Pecorino, L. *Molecular Biology of Cancer*, 3rd ed.; Oxford University Press: Oxford, UK, 2012.
- [191] Stewart, B. W.; Wild, C. P. *World cancer report 2014*; International Agency for Research on Cancer (IARC), 2014.
- [192] *International Agency for Research on Cancer (IARC) and World Health Organization (WHO). GLOBOCAN 2012: Estimated cancer incidence, mortality and prevalence worldwide in 2012*; 2012. http://globocan.iarc.fr/Pages/fact_sheets_cancer.aspx.
- [193] Brueggemeier, R. W.; Richards, J. A.; Joomprabutra, S.; Bhat, A. S.; Whetstone, J. L. *J. Steroid Biochem. Mol. Biol.* **2001**, *79*, 75 – 84.
- [194] Brueggemeier, R. W.; Hackett, J. C.; Diaz-Cruz, E. S. *Endocr. Rev.* **2005**, *26*, 331–345.
- [195] Robinson, A. *Ther. Clin. Risk Manage.* **2009**, *5*, 91–98.
- [196] Masood, S. *Diagn. Cytopathol.* **1992**, *8*, 475–491.
- [197] Mohibi, S.; Mirza, S.; Band, H.; Band, V. *J. of Carcinog.* **2011**, *10*, 35.
- [198] Chen, S. *Front. Biosci.* **1998**, *3*, d922 – d933.
- [199] Mangelsdorf, D. J.; Thummel, C.; Beato, M.; Herrlich, P.; Schütz, G.; Umesono, K.; Blumberg, B.; Kastner, P.; Mark, M.; Chambon, P.; Evans, R. M. *Cell* **1995**, *83*, 835 – 839.
- [200] Klinge, C. M. *Nucleic Acids Res.* **2001**, *29*, 2905–2919.
- [201] Ramsey, T.; Klinge, C. *J. Mol. Endocrinol.* **2001**, *27*, 275–292.
- [202] Ciocca, D. R.; Fanelli, M. A. *Trends Endocrinol. Metab.* **1997**, *8*, 313 – 321.
- [203] Galeazzi, R.; Massaccesi, L. *J. Mol. Model.* **2011**, *18*, 1–14.
- [204] Thiantanawat, A.; Long, B. J.; Brodie, A. M. *Cancer Res.* **2003**, *63*, 8037–8050.
- [205] Bruno, R. D.; Njar, V. C. O. *Bioorg. Med. Chem.* **2007**, *15*, 5047–5060.
- [206] Brodie, A. M. H.; Njar, V. C. O. *Steroids* **2000**, *65*, 171–179.
- [207] Bulun, S. E.; Attar, E.; Gurates, B.; Chen, Y.-H.; Tokunaga, H.; Monsivais, D.; Pavone, M. E. *Endometriosis: Science and Practice*; Wiley-Blackwell, 2012; Chapter Medical Therapies: Aromatase Inhibitors, pp 357–365.

- [208] AlHilli, M. M.; Long, H. J.; Podratz, K. C.; Bakkum-Gamez, J. N. *J. Obstet. Gynaecol. Res.* **2012**, *38*, 340–344.
- [209] Miki, Y.; Abe, K.; Suzuki, S.; Suzuki, T.; Sasano, H. *Mol. Cell. Endocrinol.* **2011**, *340*, 168 – 174.
- [210] Dowsett, M. *Endocr.-Relat. Cancer* **1999**, *6*, 181–185.
- [211] Ahmad, I.; Shagufta *Eur. J. Med. Chem.* **2015**, *102*, 375 – 386.
- [212] Hong, Y. Y.; Yu, B.; Sherman, M.; Yuan, Y. C.; Zhou, D. J.; Chen, S. A. *Mol. Endocrinol.* **2007**, *21*, 401–414.
- [213] Tahara, M.; Nomura, S.; Hashimoto, M. *Folia Pharmacol. Jpn.* **2003**, *122*, 345–354.
- [214] Hong, Y. Y.; Rashid, R.; Chen, S. A. *Steroids* **2011**, *76*, 802–806.
- [215] Cepa, M.; da Silva, E. J. T.; Correia-da Silva, G.; Roleira, F. M. F.; Teixeira, N. A. A. *Steroids* **2008**, *73*, 1409–1415.
- [216] Ariazi, E. A.; Leitao, A.; Oprea, T. I.; Chen, B.; Louis, T.; Bertucci, A. M.; Sharma, C. G. N.; Gill, S. D.; Kim, H. R.; Shupp, H. A.; Pyle, J. R.; Madrack, A.; Donato, A. L.; Cheng, D.; Paige, J. R.; Jordan, V. C. *Mol. Cancer Ther.* **2007**, *6*, 2817–2827.
- [217] Jeong, S.; Woo, M. M.; Flockhart, D. A.; Desta, Z. *Cancer Chemother. Pharmacol.* **2009**, *64*, 867–875.
- [218] Wirz, B.; Valles, B.; Parkinson, A.; Madan, A.; Probst, A.; Zimmerman, A. *Proceedings of 7th North American ISSX Meeting* **1996**, *10*, 359.
- [219] Bahar, I.; Atilgan, A. R.; Erman, B. *Folding Des.* **1997**, *2*, 173 – 181.
- [220] Brueggemeier, R. W. *Expert Opin. Pharmacother.* **2006**, *7*, 1919–1930.
- [221] Brueggemeier, R. W. *Am. J. Ther.* **2001**, *8*, 333–344.
- [222] Buzdar, A. U.; Robertson, J. F. R.; Eiermann, W.; Nabholz, J.-M. *Cancer* **2002**, *95*, 2006–2016.
- [223] Miller, W. L. *Endocrinology* **2005**, *146*, 2544–2550.

Chapter 2

Theoretical Background and Methodology



Contents

2.1. Theoretical Chemistry and Computational Chemistry	45
2.2. Quantum Mechanics Methods	46
2.2.1. Born-Oppenheimer approximation	46
2.2.2. Hartree-Fock method	49
2.2.3. Basis Sets	60
2.2.4. Post Hartree-Fock methods	62
2.2.5. Density Functional Theory	64
2.2.5.1. The Hohenberg-Kohn Theorems	64
2.2.5.2. The Kohn-Sham (KS) Method	67
2.2.5.3. Exchange-Correlation functionals	70
2.3. Molecular Mechanics Methods	72
2.3.1. Force fields	72
2.3.1.1. Bonding Terms	73
2.3.1.2. Non-Bonded Terms	77
2.3.1.3. The OPLS Force Field	79
2.3.2. Treatment of the Non-Bonding Interactions:	80
2.3.2.1. Cut-off Methods	80
2.3.3. Periodic Boundary Conditions	83
2.3.4. Ewald Summation Method	84
2.4. Hybrid Quantum Mechanics / Molecular Mechanics	85
2.4.1. Frontier Bonds	89
2.5. Potential Energy Surface	90
2.5.1. Chemical structure, properties and reactivity	91
2.5.1.1. Characteristic Points of the PES	91
2.5.1.2. Methods to locate stationary points on the PES	95
2.5.2. Kinetic Isotope Effects	97
2.5.2.1. An Introduction to Transition State Theory	97
2.5.2.2. Kinetic Isotope Effects	99
2.5.2.3. Transition State Theory Limitations	103
2.5.2.4. Semi-Classical Tunneling Effect Corrections	105
2.5.3. Molecular Dynamics	106
2.5.4. Free Energy	114
2.5.4.1. Free Energy in statistical mechanics	114
2.5.4.2. Free energy perturbation	115
2.6. Topological Analysis of the Electron Localization Function	118
2.6.1. Electron Localization Function	119
2.6.2. Gradient Dynamical System	122
2.6.3. Topological Analysis of the ELF	124
2.6.4. Properties derived from the ELF	126
2.6.5. Thom's Catastrophe Theory	128
2.6.6. Bonding Evolution Theory	129
Bibliography	132

2.1. Theoretical Chemistry and Computational Chemistry

According to the scientific journal *Nature*, *theoretical chemistry* “seeks to provide theories and explanations for chemical observations whilst also posing questions to be answered by future experiments. Playing a key role in *physical chemistry*, it uses the laws of physics to predict molecular structure, dynamics, bonding, reactivity, physical properties and spectroscopic response”.¹ According to the same journal, *computational chemistry* “describes the use of computer modeling and simulation – including *ab initio* approaches based on quantum chemistry, and empirical approaches – to study the structures and properties of molecules and materials. Computational chemistry is also used to describe the computational techniques aimed at understanding the structure and properties of molecules and materials”.² Summarizing, theoretical chemistry may be defined as the mathematical description of chemistry, while computational chemistry is usually employed when a method or a theory is automated so that it can be implemented as software running on computers.³ Computational chemistry could be considered as a “*dry lab*”, in which the chemical processes are simulated “*in silico*” by means of software packages, where both theoretical methods and mathematical algorithms have been programmed. The binary codes resulting from these programs run in an environment formed by a network of personal computers and supercomputers linked together, in order to provide chemical information. Furthermore, computational chemistry provides an alternative to chemical laboratories, whereby we can get physicochemical data that would be very difficult to obtain in real laboratories. However, computational chemistry should not be viewed as an alternative to experimental chemistry, but rather as a complement to it. Recently, Martin Karplus, Michael Levitt, and Arieh Warshel were awarded the 2013 Nobel Prize in Chemistry for the investigations that they conducted in the 1970s on chemical computational models (“Development of Multiscale Models for Complex Chemical Systems”). These researchers combined the principles of quantum theory (quantum mechanics) and classical physics (Newtonian mechanics) to reproduce the chemical processes that occur in nature. However, this is not the first time that computational chemistry has received the highest award in chemistry. Previously, Walter Kohn and John Pople won the 1998 Nobel Prize in Chemistry for their work on computational methods in quantum chemistry and the density functional theory (“Development of the density functional theory (DFT)”). These events confirm the current importance of this branch of chemistry, the use of which will, as time goes by, become increasingly necessary in fields such as materials, biochemistry, pharmaceutical industry, and so on.

2.2. Quantum Mechanics Methods

Quantum mechanics (*QM*) is essential for the study of the electronic rearrangements involved in the breaking and forming of chemical bonds, which take place in the course of chemical reactions. This is because electrons are very light particles and, hence, classical mechanics is not able to correctly describe the electronic structure of molecules. For this reason, the quantum mechanics-based methods provide the most suitable and realistic framework to describe microscopic systems at the atomic scale. However, these theoretical methods entail complex mathematical models and thus require a large number of computational resources. Indeed, depending on the method chosen, the improved accuracy comes at the price of expense of increased calculation. The cost of performing an *ab initio* calculation scales formally between the fourth power of the number of basis functions ($O(N^4)$), in the simplest methods, to the eighth power ($O(N^8)$) in the more accurate methods. Therefore, the use of QM methods provides a high degree of accuracy in molecular simulations but their usability is limited to systems containing a few hundred atoms.

2.2.1. Born-Oppenheimer approximation

In quantum mechanics, the time-dependent Schrödinger equation gives a description of a system evolving in time and determines the complete behavior of a non-relativistic molecular system.⁴ It has the form:

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

where \hat{H} is the *Hamiltonian operator*, which operates on the *wave function* Ψ , and $\hbar = h/2\pi$, where h is the Planck constant. The Hamiltonian operator is the operator of the total energy of the system and is given by the sum of the total kinetic energy plus the total potential energy:

$$\hat{H} = \hat{T} + \hat{V} \quad (2.2)$$

Thus, eq. 2.1 can be rewritten as:

$$\left[\frac{-\hbar^2}{2m} \frac{\partial^2 \Psi(x,t)}{dx^2} + V(x,t) \right] \Psi(x,t) = i\hbar \frac{\partial}{\partial t} \Psi(x,t) \quad (2.3)$$

where the first term in the square brackets accounts for the kinetic energy, m being the mass of the electron, while the second accounts for the potential energy. The separation of the wave function $\Psi(x,t)$ in terms of their variables x and t :

$$\Psi(x, t) = \Psi(x)f(t) \quad (2.4)$$

allows us to use the simplest non-relativistic time-independent Schrödinger equation:

$$\left[\frac{-\hbar^2}{2m} \frac{\partial^2 \Psi(x)}{dx^2} + V(x) \right] \Psi(x) = E\Psi(x) \quad (2.5)$$

or

$$\hat{H}\Psi(x) = E\Psi(x) \quad (2.6)$$

to determine the properties of the stationary states of a system. In this equation, \hat{H} , $\Psi(x)$, and E are the Hamiltonian, the wave function and the energies of the stationary states of the system, respectively.

Assuming that both nuclei and electrons are point masses, and neglecting the spin-orbit coupling as well as other relativistic interactions, the “exact” non-relativistic, time-independent molecular Hamiltonian can be written as:

$$\hat{H} = \hat{T}_n + \hat{T}_e + \hat{V}_{nn} + \hat{V}_{ee} + \hat{V}_{ne} \quad (2.7)$$

where the two subscripts e and n refer to the electrons and to the nuclei, respectively.

The first and second terms of eq. 2.7 correspond to the kinetic energy operator of the nuclei and of the electrons, respectively. The third and fourth terms of this equation correspond to the potential energy of repulsion between nuclei and between electrons, respectively. Finally, the last term of the equation involves the potential energy of attraction between the nuclei and the electrons. In addition, the molecular Hamiltonian of eq. 2.7 can be expressed in atomic units of N nuclei and n electrons as follows:

$$\hat{H} = -\frac{1}{2} \sum_{\alpha}^N \frac{1}{M_{\alpha}} \nabla_{\alpha}^2 - \frac{1}{2} \sum_i^n \nabla_i^2 + \sum_{\alpha}^N \sum_{\beta > \alpha}^N \frac{Z_{\alpha} Z_{\beta}}{r_{\alpha\beta}} + \sum_i^n \sum_{j > i}^n \frac{1}{r_{ij}} - \sum_i^n \sum_{\alpha}^N \frac{Z_{\alpha}}{r_{i\alpha}} \quad (2.8)$$

where the subscripts α and β refer to the nuclei, and the subscripts i and j refer to the electrons. $r_{\alpha\beta} = |R_{\alpha} - R_{\beta}|$ is the distance between nuclei α and β , $r_{ij} = |r_i - r_j|$ is the distance between electrons i and j , and $r_{i\alpha} = |r_i - R_{\alpha}|$ is the distance between electron i and nuclei α . $Z_{\alpha/\beta}$ is the atomic number of either nucleus α or nucleus β , and M_{α} is the ratio of the mass of nucleus α .

Therefore, the time-independent Schrödinger equation that we have to solve is the following:

$$\left(\hat{T}_n + \hat{T}_e + \hat{V}_{nn} + \hat{V}_{ee} + \hat{V}_{ne}\right) \Psi(R; r) = E\Psi(R; r) \quad (2.9)$$

Since the Hamiltonian of eq. 2.8 is very complex, eq. 2.9 is very difficult to solve, partly due to the difficulty involved in treating the correlation between the nuclei and the electrons of the system. Fortunately, there are some assumptions that may simplify this Hamiltonian but at the cost of less accuracy. The Born-Oppenheimer (BO) approximation,^{5,6} in basic terms, considers that the nuclei are much heavier than electrons: $M_\alpha \gg m_e$ (approximately 1836 times more in the hydrogen atom). Therefore, the electrons are moving much faster than the nuclei, so that the latter can be considered static, while the electrons move at a specific position around them. As a result of the mass difference, when the nuclei move through molecular vibrations, the electrons immediately readjust their positions according to the new nuclear configuration. That is, this approximation decouples the nuclear movements of the electronic movements, thus eliminating the correlation in the attractive electron-nuclear potential energy term.

In this way, the total wave function can be separated into two different parts, the electronic wave function and the nuclear wave function, and is represented as the product of the two parts:

$$\Psi(R; r) = \Psi_{el}(R; r) \Psi_N(R) \quad (2.10)$$

where r and R represent the coordinates of all the electrons and nuclei, respectively. As can be noted, the electronic wave function depends on both electronic and nuclear coordinates.

In addition, according to this approximation, the kinetic energy of the nuclei (the term \hat{T}_n in eq. 2.7) can be neglected in eq. 2.9 due to the low speed of the nuclei when compared with electrons. This leads to an equation that describes the electronic movement:

$$\left(\hat{H}_{el} + \hat{V}_{nn}\right) \Psi_{el}(R; r) = E_{el} \Psi_{el}(R; r) \quad (2.11)$$

where \hat{H}_{el} is the purely electronic Hamiltonian:

$$\hat{H}_{el} = \hat{T}_e + \hat{V}_{ee} + \hat{V}_{ne} \quad (2.12)$$

Moreover, the repulsive nuclear-nuclear potential energy term $\left(\hat{V}_{nn}\right)$ is reduced to a constant parameter for a given geometry. Thus, the equation that describes the electronic motion for a given nuclear configuration (q_k) is as follows:

$$\hat{H}_{el} \Psi_{el}(q_i; q_k) = E_{el}(q_k) \Psi_{el}(q_i; q_k) \quad (2.13)$$

where E_{el} is the purely electronic energy that depends parametrically on the nuclear coordinates (q_k). When \hat{V}_{nn} is added to the electronic energy term, we obtain the so-called *Potential Energy Surface* (PES) :

$$U(q_k) = E_{el}(q_k) + V_{nn}(q_k) \quad (2.14)$$

The concept of PES emerges from the BO approximation and is central to computational chemistry. In fact, the exploration of the PES (which will be discussed further in sec.2.5) is one of the fundamental pillars for obtaining the molecular properties and chemical information of a system. However, it is important to emphasize that although we are neglecting the electron-nuclear correlation through the BO approximation, the correlation between electrons is still a problem when we are interested in dealing with poly-electronic systems. In order to solve this electronic problem, several computational methods based on the Hartree-Fock method have emerged.

2.2.2. Hartree-Fock method

The *Hartree-Fock* (HF) method⁷⁻¹⁰ was developed to solve the electronic Schrödinger equation that results from applying the BO approximation to the time-independent Schrödinger equation. This method is the basis of the molecular orbital (MO) theory and is considered an *ab initio* electronic structure method. The term “*ab initio*” (from the beginning) refers to quantum chemical calculations obtained solely by means of theoretical principles, without including experimental data, as in the semi-empirical methods. In fact, the HF method deserves special attention, since it is not only a very useful approximation to describe the electronic structure, but also forms the basis of more accurate and sophisticated models of molecular electronic structure.

The starting point of the HF theory is based on the idea that, in a many-electron system, the electrons do not interact with each other (i.e., that $\hat{V}_{ee} = 0$). This model is known as the *independent particle model* and, although it is a strong approximation, it allows us to consider the wave function as a product of mono-electronic wave functions, which is known as the *Hartree product* (HP) .⁷

$$\Psi_{HP}(\vec{r}_1, \vec{r}_2, \dots, \vec{r}_N) = \phi_1(\vec{r}_1) \phi_2(\vec{r}_2) \dots \phi_N(\vec{r}_N) \quad (2.15)$$

Electrons are fermions and, thus, they are identical and indistinguishable particles. For this reason, they must satisfy the principle of antisymmetry, also known as

*Pauli exclusion principle.*¹¹ This principle states that fermions must be described by wave functions that are antisymmetric with respect to the interchange of the coordinates, both spatial (\vec{r}) and spin (ω), of any pair of electrons. Since electrons are described by a set of spatial and spin coordinates ($\mathbf{x} = \{\vec{r}, \omega\}$), it is advisable to change our notation for orbitals from spatial orbitals ($\phi(\vec{r})$) to spin orbitals ($\chi(\mathbf{x})$). Spin orbitals are just the product of a spatial orbital and one of the two spin functions, $\alpha(\omega)$ or $\beta(\omega)$, so that $\chi(\mathbf{x}) = \phi(\vec{r})\sigma(\omega)$, where $\sigma = \alpha, \beta$. Thus, the HP with the spin orbital notation is as follows:

$$\Psi_{HP}(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N) = \chi_1(\mathbf{x}_1)\chi_2(\mathbf{x}_2) \dots \chi_N(\mathbf{x}_N) \quad (2.16)$$

However, the HP fails to satisfy the antisymmetry principle and it is necessary to find another functional form that ensures its compliance. The solution was proposed by John Slater,^{12,13} who found that a determinant of spin orbitals ensures the indistinguishability of all electrons and compliance with the antisymmetry principle. This determinant is called a *Slater determinant* (SD), and the example for an N electron system is as follows:

$$\Psi_{SD} = \frac{1}{\sqrt{N!}} \begin{vmatrix} \chi_1(\mathbf{x}_1) & \chi_2(\mathbf{x}_1) & \cdots & \chi_N(\mathbf{x}_1) \\ \chi_1(\mathbf{x}_2) & \chi_2(\mathbf{x}_2) & \cdots & \chi_N(\mathbf{x}_2) \\ \vdots & \vdots & \ddots & \vdots \\ \chi_1(\mathbf{x}_N) & \chi_2(\mathbf{x}_N) & \cdots & \chi_N(\mathbf{x}_N) \end{vmatrix} \quad (2.17)$$

where $1/\sqrt{N!}$ is a normalization constant. The SD can be rewritten in a shorthand representation, using the Bracket notation, as a ket:

$$\Psi_{SD}(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N) = |\chi_1(\mathbf{x}_1)\chi_2(\mathbf{x}_2) \dots \chi_N(\mathbf{x}_N)\rangle \quad (2.18)$$

Now that we have selected the functional form of the wave function in HF theory, it is time to focus on the Hamiltonian operator. To this end, we can invoke the electronic Hamiltonian arising from the BO approximation, and then rewrite it in a simple manner as:

$$\hat{H}_{el} = \sum_i^N \hat{h}(i) + \sum_i^N \sum_{j>i}^N \hat{v}(i, j) + V_{nn} \quad (2.19)$$

where the first term, $h(i)$, is the one-electron operator, also known as one-electron Hamiltonian or *core Hamiltonian*:

$$\hat{h}(i) = -\frac{1}{2}\nabla_i^2 - \sum_A^N \frac{Z_A}{r_{iA}} \quad (2.20)$$

the second term, $v(i, j)$, is the two-electron operator:

$$\hat{v}(i, j) = \frac{1}{r_{ij}} \quad (2.21)$$

and the third term, V_{nn} , is the nuclear-nuclear potential, which is a constant for the fixed nuclear coordinates (q_k). Therefore, it can be ignored because it does not change the eigenfunctions, but only the eigenvalues.

Assuming that the wave function (SD) is normalized, the expectation value of energy is given by the following expression:

$$E_{el} = \langle \Psi_{SD} | \hat{H}_{el} | \Psi_{SD} \rangle \quad (2.22)$$

According to the variational principle, this energy is always an upper bound to the true energy, i.e., an approximate wave function has an energy above or equal to the exact energy. Hence, by constructing a trial wave function, Ψ_{SD} , containing a number of parameters, we can generate the “best” trial function of the given form by minimizing the energy as a function of these parameters:

$$E_{HF} = \min E[\Psi_{SD}] \quad (2.23)$$

The Hartree-Fock energy electronic energy (E_{HF}) can be rewritten in terms of integrals of one- and two-electron operators as:

$$E_{HF} = \langle \Psi_{SD} | \hat{H}_{el} | \Psi_{SD} \rangle = \sum_i^N \langle i | \hat{h} | j \rangle + \frac{1}{2} \sum_i^N \sum_{j>i}^N \langle ii | jj \rangle - \langle ij | ji \rangle \quad (2.24)$$

where the first term defines the contribution due to the kinetic energy and the electron-nucleus attraction and corresponds to the one-electron integrals:

$$\langle i | \hat{h} | j \rangle = \int \chi_i^*(\mathbf{x}_1) \hat{h}(r_1) \chi_j(\mathbf{x}_1) d\mathbf{x}_1 \quad (2.25)$$

and the second and third terms correspond to two-electron integrals, which have the following general form:

$$\langle ij | kl \rangle = \int \int \chi_i^*(\mathbf{x}_1) \chi_j(\mathbf{x}_1) \frac{1}{r_{12}} \chi_k^*(\mathbf{x}_2) \chi_l(\mathbf{x}_2) d\mathbf{x}_1 d\mathbf{x}_2 \quad (2.26)$$

In particular, $\langle ii | jj \rangle$ is known as the *Coulomb* integral:

$$\langle ii | jj \rangle = \int \int |\chi_i(\mathbf{x}_1)|^2 \frac{1}{r_{12}} |\chi_j(\mathbf{x}_2)|^2 d\mathbf{x}_1 d\mathbf{x}_2 \quad (2.27)$$

and $\langle ij|ji\rangle$ is known as the *exchange* integral:

$$\langle ij|ji\rangle = \int \int \chi_i^*(\mathbf{x}_1)\chi_j(\mathbf{x}_1)\frac{1}{r_{12}}\chi_j^*(\mathbf{x}_2)\chi_i(\mathbf{x}_2)d\mathbf{x}_1d\mathbf{x}_2 \quad (2.28)$$

E_{HF} from eq. 2.24 is obviously a functional of the spin orbitals ($E_{HF} = E[\{\chi_i\}]$). Therefore, we need to minimize the E_{HF} with respect to the changes in the orbitals, but with the constraint that they remain orthogonal and normalized throughout the minimization. This can be accomplished by means of the method of *Lagrangian multipliers* (ε_{ij}).

$$L = E - \sum_{i=1}^N \sum_{j=1}^N \varepsilon_{ij} (\langle \chi_i | \chi_j \rangle \delta_{ij}) \quad (2.29)$$

The minimization of the Lagrangian function leads to a set of equations known as the HF equations, which determine the “best” spin orbitals for which E_{HF} obtains its lowest value:

$$\hat{f}_i(\mathbf{x}_1)\chi_i(\mathbf{x}_1) = \sum_{j=1}^N \varepsilon_{ij}\chi_j(\mathbf{x}_1) \quad i = 1, 2, \dots, N \quad (2.30)$$

The set of N HF equations have the appearance of eigenvalue equations. In these equations, the Lagrangian multipliers ε_{ij} are the eigenvalues of a new operator \hat{f} , and have the physical interpretation of the *orbital energies*. The *Fock operator* \hat{f} is an effective one-electron operator defined as:

$$\hat{f}_i(\mathbf{x}_1) = \hat{h}_i(\mathbf{x}_1) + V_{HF} \quad \text{or} \quad \hat{f}_i(\mathbf{x}_1) = -\frac{1}{2}\nabla_i^2 - \sum_A \frac{Z_A}{r_{iA}} + V_{HF} \quad (2.31)$$

where $V_{HF}(i)$ is the average repulsive potential experienced by the i_{th} electron due to the remaining $N - 1$ electrons, and is known as the *Hartree-Fock potential*:

$$V_{HF}(\mathbf{x}_1) = \sum_{j=1}^N \left(\hat{J}_j(\mathbf{x}_1) - \hat{K}_j(\mathbf{x}_1) \right) \quad (2.32)$$

In this way, the complicated two-electron repulsion operator $\hat{v}(i, j) = 1/r_{ij}$ is replaced by a simpler one-electron operator $V_{HF}(i)$, which considers the electron-electron repulsion only in an average way. As can be noted, two new operators \hat{J}_j and \hat{K}_j have emerged.

The *Coulomb operator* \hat{J}_j is defined as:

$$\hat{J}_j(\mathbf{x}_1)\chi_i(\mathbf{x}_1) = \left[\int |\chi_j(\mathbf{x}_2)|^2 \frac{1}{r_{12}} d\mathbf{x}_2 \right] \chi_i(\mathbf{x}_1) \quad (2.33)$$

and represents the average “local” potential that an electron at point \mathbf{x}_1 experiences due to the charge distribution from another electron in the spin-orbital χ_j .

The *exchange operator* \hat{K}_j is defined as:

$$\hat{K}_j(\mathbf{x}_1)\chi_i(\mathbf{x}_1) = \left[\int \chi_j^*(\mathbf{x}_2) \frac{1}{r_{12}} \chi_i(\mathbf{x}_2) d\mathbf{x}_2 \right] \chi_j(\mathbf{x}_1) \quad (2.34)$$

This term emerges from the antisymmetry requirement of the wave function and has no alternative analog in classical mechanics. The main function of this operator is to permute the labels on any two spin orbitals χ_i and χ_j . Moreover, this operator is considered a “non-local” operator, since when it operates on the spin orbital χ_i the resulting potential depends not only on its value at a point, but in all the space.

The set of spin orbitals that minimize the energy is not unique, since any unitary transformation on them leaves the energy value unchanged. One of these unitary transformations is that which gives place to a diagonal Lagrange multipliers matrix. The spin orbitals that fulfill this condition are known as *canonical spin orbitals*. Based on these, the HF equation is as follows:

$$\hat{f}_i(\mathbf{x}_1)\chi'_i(\mathbf{x}_1) = \varepsilon'_i \chi'_i(\mathbf{x}_1) \quad i = 1, 2, \dots, N \quad (2.35)$$

From now on, only the canonical orbitals will be considered; however, the primes that appear in eq. 2.35 will be omitted in order to simplify the notation.

The equations 2.35 also present the appearance of a system of eigenvalue equations. However, note that the spin orbitals which are the solution of these equations are also included in the definition of the Fock operator through the Coulomb (eq. 2.33) and exchange (eq. 2.34) operators. Therefore, we have what looks rather like a pseudo-eigenvalue problem that has to be worked out iteratively. The technique used to solve this kind of equations is called *Self-consistent field* (SCF) and thus the HF method is considered a self-consistent method.⁹

In this way, each spin-orbital, χ_i , has its own constant, ε_i , coming from the Lagrange multipliers. ε_i has energy units, so that each spin orbital may have an associated energy, which is known as *orbital energy*. In this way, the energy of the i_{th} spin orbital is given by the expression:

$$\varepsilon_i = h_{ii} + \sum_{j=1}^N (J_{ij} - K_{ij}) \quad (2.36)$$

and the total electronic energy will be:

$$E_{el} = \langle \Psi_{SD} | \hat{H}_{el} | \Psi_{SD} \rangle = \sum_{i=1}^N \varepsilon_i - \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N (J_{ij} - K_{ij}) \quad (2.37)$$

Note that the total electronic energy is not the sum of the orbital energies for all the electrons, since the repulsion between each pair of electrons would be counted twice.

When the valence shell is completely filled with electrons, we talk about closed-shell systems. In this case, a single molecular orbital is often used to describe two electrons with different spin functions. This gives rise to what is known as the *Restricted Hartree-Fock* (RHF) method.

For a system of $2N$ electrons in N_{occ} occupied molecular orbitals, the corresponding Slater orbital is:

$$\Psi(1, 2, \dots, 2N) = |\Psi_1(1)\bar{\Psi}_1(2)\Psi_2(3)\bar{\Psi}_2(4)\dots\Psi_{N_{occ}}(2N-1)\bar{\Psi}_{N_{occ}}(2N)| \quad (2.38)$$

and the electronic energy is given by:

$$E_{el} = 2 \sum_{i=1}^{N_{occ}} h_{ii} + \sum_{i=1}^{N_{occ}} \sum_{j=1}^{N_{occ}} (2J_{ij} - K_{ij}) \quad (2.39)$$

The Fock operator becomes:

$$\hat{f}(\mathbf{x}_1) = \hat{h}(\mathbf{x}_1) + \sum_{j=1}^{N_{occ}} (2\hat{J}_j(\mathbf{x}_1) - \hat{K}_j(\mathbf{x}_1)) \quad (2.40)$$

Finally, the HF equations are as follows:

$$\hat{f}_i(\mathbf{x}_1)\Psi_i(\mathbf{x}_1) = \varepsilon_i\Psi_i(\mathbf{x}_1) \quad i = 1, 2, \dots, N_{occ} \quad (2.41)$$

Since the solution of eq. 2.41 implies the resolution of a very complex system, another approximation is necessary. For this reason, Roothaan and Hall proposed to express the molecular orbitals, Ψ_i , as a *linear combination of atomic orbitals* (LCAO).^{14,15} These atomic orbitals are in turn described by a set of K functions of basis $\{\phi_\nu\}$:

$$\Psi_i = \sum_{\nu=1}^K c_{\nu i} \phi_\nu \quad (2.42)$$

Thus, substituting eq. 2.42 in eq. 2.41:

$$\sum_{\nu=1}^K c_{\nu i} \hat{f}(\mathbf{x}_1) \phi_{\nu}(\mathbf{x}_1) = \sum_{\nu=1}^K c_{\nu i} \varepsilon_i \phi_{\nu}(\mathbf{x}_1) \quad (2.43)$$

Multiplying both sides by $\phi_{\mu}^*(\mathbf{x}_1)$ and integrating all the electronic space we obtain:

$$\sum_{\nu=1}^K c_{\nu i} \langle \phi_{\mu}(\mathbf{x}_1) | \hat{f}(\mathbf{x}_1) | \phi_{\nu}(\mathbf{x}_1) \rangle = \sum_{\nu=1}^K c_{\nu i} \varepsilon_i \langle \phi_{\mu}(\mathbf{x}_1) | \phi_{\nu}(\mathbf{x}_1) \rangle \quad (2.44)$$

On the one hand, the integrals appearing on the left of eq. 2.44 correspond to the matrix elements that the \hat{f} operator represents in the basis $\{\phi_{\nu}\}$:

$$F_{\mu\nu} = \langle \phi_{\mu}(\mathbf{x}_1) | \hat{f}(\mathbf{x}_1) | \phi_{\nu}(\mathbf{x}_1) \rangle = \int \phi_{\mu}(\mathbf{x}_1) \hat{f}(\mathbf{x}_1) \phi_{\nu}(\mathbf{x}_1) d\mathbf{x}_1 \quad (2.45)$$

On the other hand, the integrals appearing on the right of eq. 2.44 are the overlap integrals between the functions of basis $\{\phi_{\nu}\}$:

$$S_{\mu\nu} = \langle \phi_{\mu}(\mathbf{x}_1) | \phi_{\nu}(\mathbf{x}_1) \rangle = \int \phi_{\mu}(\mathbf{x}_1) \phi_{\nu}(\mathbf{x}_1) d\mathbf{x}_1 \quad (2.46)$$

Thus, the Roothaan-Hall equations are as follows:

$$\sum_{\nu=1}^K F_{\mu\nu} c_{\nu i} = \varepsilon_i \sum_{\nu=1}^K S_{\mu\nu} c_{\nu i} \quad \mu = 1, 2, 3, \dots, K \quad (2.47)$$

which can be rewritten in a simpler form as:

$$FC = SC\varepsilon \quad (2.48)$$

The Fock operator is often written as:

$$F_{\mu\nu} = H_{\mu\nu}^{core} + \sum_{\lambda=1}^K \sum_{\sigma=1}^K P_{\lambda\sigma} \left[\left\langle \phi_{\mu}(\mathbf{x}_1) \phi_{\lambda}(\mathbf{x}_2) \left| \frac{1}{r_{12}} \right| \phi_{\nu}(\mathbf{x}_1) \phi_{\sigma}(\mathbf{x}_2) \right\rangle - \frac{1}{2} \left\langle \phi_{\mu}(\mathbf{x}_1) \phi_{\lambda}(\mathbf{x}_2) \left| \frac{1}{r_{12}} \right| \phi_{\sigma}(\mathbf{x}_1) \phi_{\nu}(\mathbf{x}_2) \right\rangle \right] \quad (2.49)$$

where $H_{\mu\nu}^{core}$ is the *core* Hamiltonian:

$$H_{\mu\nu}^{core} = T_{\mu\nu} + V_{\mu\nu}^{nucl} = \langle \phi_{\mu}(\mathbf{x}_1) | \hat{h}(\mathbf{x}_1) | \phi_{\nu}(\mathbf{x}_1) \rangle \quad (2.50)$$

and $P_{\lambda\sigma}$ is the density matrix, which represents the electron density in the overlap region between ϕ_λ and ϕ_σ :

$$P_{\lambda\sigma} = 2 \sum_{i=1}^N c_{\lambda i} c_{\sigma i} \quad (2.51)$$

From now on, the two-electron integrals, which are within square brackets, will be represented in the following simplified form:

$$\left[\langle \mu\nu | \sigma\lambda \rangle - \frac{1}{2} \langle \mu\lambda | \sigma\nu \rangle \right]$$

In fact, a new matrix, $G_{\mu\nu}$, can be defined as the two-electron part of the Fock matrix. $G_{\mu\nu}$ depends on the density matrix $P_{\lambda\sigma}$ and a set of two-electron integrals:

$$G_{\mu\nu} = \sum_{\lambda=1}^K \sum_{\sigma=1}^K P_{\lambda\sigma} \left[\langle \mu\nu | \sigma\lambda \rangle - \frac{1}{2} \langle \mu\lambda | \sigma\nu \rangle \right] \quad (2.52)$$

In this way, the Fock matrix can be rewritten in a more compact notation as:

$$F_{\mu\nu} = H_{\mu\nu}^{core} + G_{\mu\nu} \quad (2.53)$$

For its part, the electronic energy of eq. 2.39 becomes the following expression:

$$E = \sum_{\mu=1}^K \sum_{\nu=1}^K P_{\mu\nu} H_{\mu\nu}^c + \sum_{\mu=1}^K \sum_{\nu=1}^K \sum_{\lambda=1}^K \sum_{\sigma=1}^K P_{\mu\nu} P_{\lambda\sigma} \left[\langle \mu\nu | \sigma\lambda \rangle - \frac{1}{2} \langle \mu\lambda | \sigma\nu \rangle \right] \quad (2.54)$$

To solve eq. 2.48 it is appropriate to perform a transformation of the atomic basis functions, so that these are orthonormal and the S matrix is the unitary matrix, U . Once the basis functions have been orthogonalized, the Roothan-Hall equations become:

$$F' C' = C' E \quad (2.55)$$

These equations are solved by means of the SCF procedure according to the algorithm depicted in Fig. 2.1.

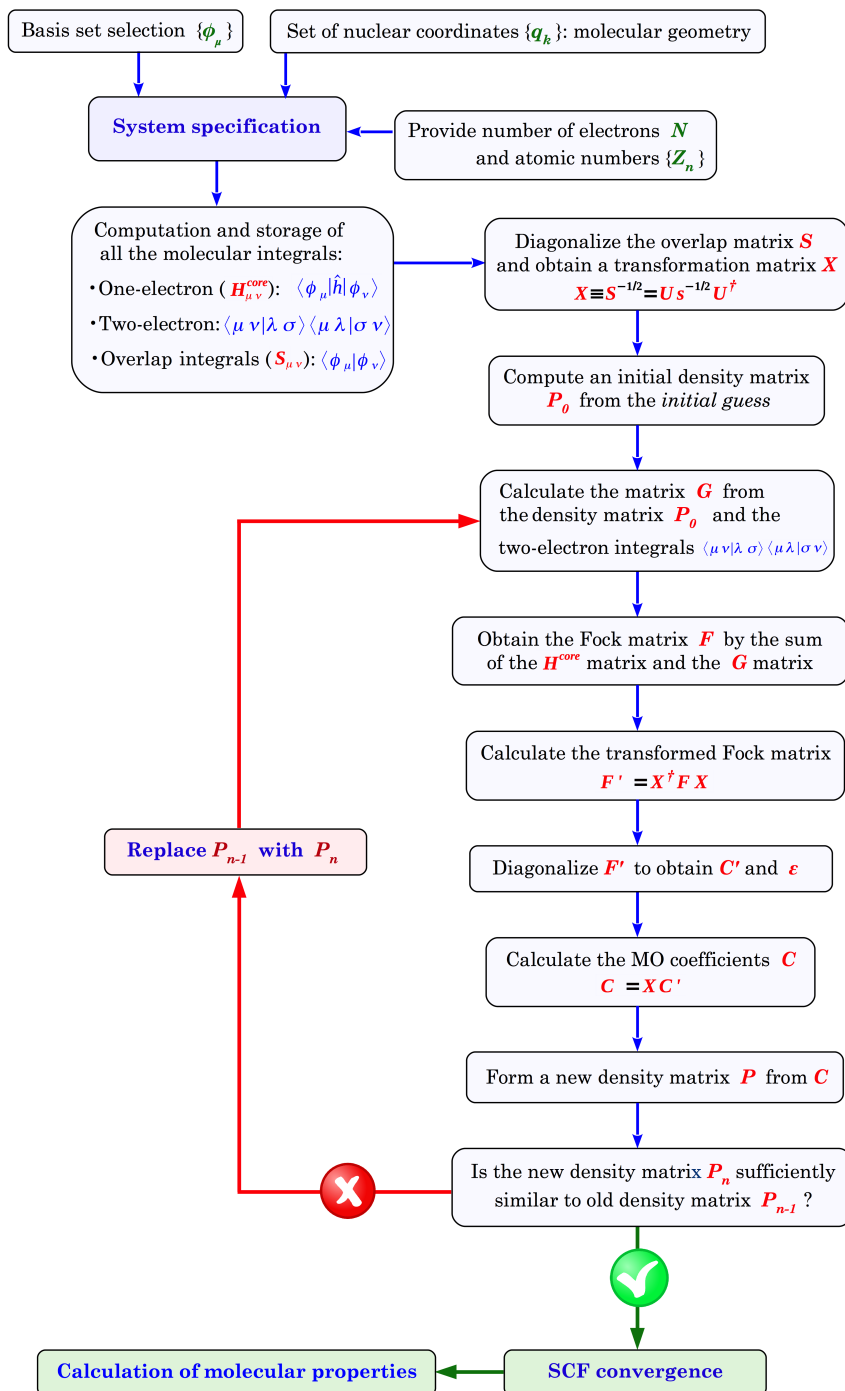


Figure 2.1.: General scheme of the SCF procedure.

Open-Shell Systems

In the case of a system with six electrons, the singlet state is described by the following Slater determinant (see Fig. 2.2a for the representation of the MO diagram):

$$\Psi_{ROHF} = |\Psi_1(1)\bar{\Psi}_1(2)\Psi_2(3)\bar{\Psi}_2(4)\Psi_3(5)\bar{\Psi}_3(6)| \quad (2.56)$$

However, when the valence shell is not completely filled with electrons, we have an open-shell system. In order to study this kind of systems, there will be two possible treatments: The *Restricted Open-shell Hartree-Fock* (ROHF) method and the *Unrestricted Hartree-Fock* (UHF) method.

Restricted Open-shell Hartree-Fock method

In this method we consider, on the one hand, the doubly occupied molecular orbitals and, on the other hand, the open shells formed by singly occupied molecular orbitals. In the example of a system of six electrons that has been previously considered, the Slater determinant corresponding to one of the components of the first triplet would be as follows:

$$\Psi_{ROHF} = |\Psi_1(1)\bar{\Psi}_1(2)\Psi_2(3)\bar{\Psi}_2(4)\Psi_3(5)\Psi_4(6)| \quad (2.57)$$

Fig. 2.2b shows the representation of the MO diagram corresponding to this example. The ROHF wave functions are eigenfunctions of $\langle \hat{S}^2 \rangle$ and, thus, they describe pure spin states. As a consequence, the ROHF method does not introduce spin contamination. However, this method is not perfect and it has other disadvantages. On the one hand, the calculations require more computing time to treat the singly occupied and doubly occupied molecular orbitals correctly, as well as the interaction between them. On the other hand, the ROHF wave function is less flexible than the unrestricted approach, since it fails to account for spin polarization in the doubly occupied orbitals.

Unrestricted Hartree-Fock method

This method makes use of different orbitals to describe electrons with different spin functions. The two sets of orbitals are defined with the same basis set, but with different coefficients. In this way, we have to solve two sets of Roothaan-Hall equations, one for the electrons with the spin function α and another one for the electrons with the spin function β :

$$\begin{aligned} F^\alpha C^\alpha &= SC^\alpha \varepsilon^\alpha \\ F^\beta C^\beta &= SC^\beta \varepsilon^\beta \end{aligned} \quad (2.58)$$

For example, in the above case of a system with six electrons, one of the com-

ponents of the first triplet would be described by the following Slater determinant:

$$\Psi_{UHF} = |\Psi_1^\alpha(1)\bar{\Psi}_1^\beta(2)\Psi_2^\alpha(3)\bar{\Psi}_2^\beta(4)\Psi_3^\alpha(5)\Psi_4^\alpha(6)| \quad (2.59)$$

where the superscripts α and β denote that the MO are different depending on the electron spin function. Fig. 2.2c shows the representation of the MO diagram corresponding to this example. The main advantage of the UHF method is the flexibility afforded by treating orbitals independently, which confer more flexibility on this method than on the ROHF method. The main disadvantage of the UHF method is the fact that the functions obtained are usually not eigenfunctions of $\langle \hat{S}^2 \rangle$. This happens inasmuch as the α and β orbitals are slightly different, which may lead to an effect known as *spin polarization*. As a result, the UHF wave function does not describe pure spin states and, thus, what is known as *spin contamination* appears. Spin contamination refers to the inclusion in the wave function of contributions from states of higher spin that mix in with the state which is being studied. With the aim of reducing or eliminating these contamination problems in unrestricted wave functions, we can use *spin projection methods*. These methods have been developed to annihilate the contributions of certain spin states higher than the desired one.

On the other hand, the UHF can also be applied in closed-shell systems. In most cases, the result matches that provided by the RHF methods; however, in certain cases, the UHF method can provide results lower in energy than the RHF method.

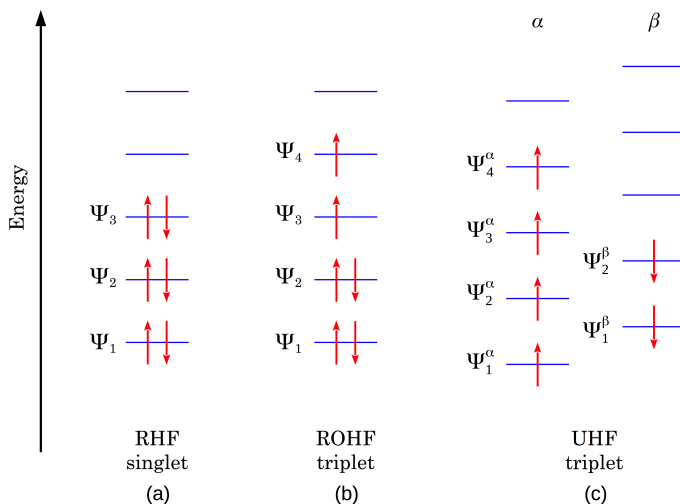


Figure 2.2.: Molecular orbital diagram for a six-electron system: RHF closed-shell singlet (a), ROHF open-shell triplet (b), and UHF open-shell triplet (c).

2.2.3. Basis Sets

The basis set is the collection of mathematical functions from which the wave function is built. The linear combination of these basis functions (also called atomic orbitals (AO)) gives rise to MOs, which configure the electronic structure of the system. In the context of the HF methodology, the best possible solution is at the HF limit. This limit is achieved by using a complete basis set, which necessarily permits an optimal description of the electron probability density. However, a complete basis set requires an infinite number of basis functions, which is impossible to obtain in practice. Therefore, we must select a finite set of mathematical functions that allow us to construct wave functions that approach the HF limit as much as possible, but also in a manner that is as efficient as possible. There are two types of basis functions commonly used in electronic structure calculations:

- **Slater-type orbitals**

The *Slater-type orbitals* (STO) were introduced by Slater¹⁶ in 1930 and have the following functional form:

$$\phi_{\zeta,n,l,m}^{STO}(r, \theta, \varphi) = N_{\zeta} r^{n-1} e^{-\zeta r} Y_{l,m}(\theta, \varphi) \quad (2.60)$$

where N_{ζ} is a normalization constant, ζ is a parameter that can be determined variationally, and $Y_{l,m}$ are spherical harmonic functions. These orbitals have the attractive feature that they closely resemble *hydrogenic atomic orbitals* (HAO). STOs include the angular function as in the HAO, but differ from the latter in the radial part, where the polynomial term in r has been replaced by an r power. The main problem of these orbitals lies in the fact that the calculation of multi-center integrals, i.e., three- and four- center, cannot be performed analytically, its numerical calculation being very costly.

- **Gaussian-type orbital**

Gaussian-type orbitals (GTO) were introduced by Boys¹⁷ in 1950 as an alternative to STOs. These functions can be expressed in terms of polar coordinates, i.e., spherical GTOs:

$$\phi_{\alpha,n,l,m}^{GTO}(r, \theta, \varphi) = N_{\alpha} r^{2n-2-l} e^{-\alpha r^2} Y_{l,m}(\theta, \varphi) \quad (2.61)$$

where N_{α} is a normalization constant, α is a variational parameter, and $Y_{l,m}$ are spherical harmonic functions.

But they can also be expressed in terms of Cartesian coordinates, i.e., Cartesian GTOs:

$$\phi_{\alpha,l_x,l_y,l_z}^{GTO}(x, y, z) = N_{\alpha} x^{l_x} y^{l_y} z^{l_z} e^{-\alpha r^2} \quad (2.62)$$

where x , y , z are the Cartesian coordinates and l_x , l_y , l_z are non-negative integers that dictate the nature of the orbital in a Cartesian sense. For example, $l_x + l_y + l_z = 0$ is an s -type orbital, $l_x + l_y + l_z = 1$ is a p -type orbital, $l_x + l_y + l_z = 2$ is a d -type orbital, and so forth. It is worth pointing out that both spherical and Cartesian functions are the same for up to $l = 1$ (p functions), but they differ slightly for $l = 2$ or higher. In quantum chemistry, a single Gaussian function with a certain exponent is called a *primitive Gaussian function*, or primitive GTO (g_p^{GTO}). This function presents a poorer description of the atomic orbitals than the STOs, especially at short and large nuclear distances, because of the r^2 dependence in the exponential function. To solve this problem, a linear combination of primitive GTOs is used to obtain a function as similar as possible to an STO.

$$\phi_{\alpha, l_x, l_y, l_z}^{CGTO}(x, y, z) = \sum_{a=1}^M c_a \phi_{\alpha, l_x, l_y, l_z}^{GTO}(x, y, z) \quad (2.63)$$

The orbitals resulting from the linear combination of GTOs are known as *contracted Gaussian-type orbitals* (CGTO). For example, a combination of three GTOs is needed to model a single $1s$ -STO, which is known as an STO-3G basis set. The smallest number of functions possible is a *minimal* basis set, which uses only one basis function per atom in the description of its occupied orbitals. The STO-3G basis set is what is known as a ‘single- ζ ’ (single zeta), and represents the *minimal* basis set. However, the basis sets may be improved by including certain features, such as:

- **Split-Valence basis set**

In these basis sets, the core orbitals are represented by a single (contracted) basis function, while the valence orbitals are split arbitrarily into many functions. The most widely used split-valence basis sets are those of Pople et al.,^{18,19} whose nomenclature is based on the contraction scheme. In this way, the first number before the hyphen indicates the number of primitives used in the contracted core functions. The numbers after a hyphen indicate the numbers of primitives used in the valence functions. If there are two of such numbers, it is a valence-double- ζ basis, if there are three, valence-triple- ζ , and so forth. Examples of double- ζ are: 3-21G, 6-21G, 4-31G, 6-31G, and an example of triple- ζ is the 6-311G.

- **Polarization functions**

Polarized basis sets add orbitals to atoms with angular momenta higher than that required for the proper description of their ground state. For example, polarized basis sets add extra d -functions to carbon atoms or p -functions to hydrogen atoms. In this way, polarized functions add flexibility within the basis set, since they allow molecular orbitals to be more delocalized around the nucleus. This makes it easier for the atoms to form chemical bonds in any direction. Polarization

functions in Pople's basis sets are denoted by an asterisk (*), e.g., 6-31G*; two asterisks indicate that polarization functions are also added to light atoms (hydrogen and helium), e.g., 6-31G**.

- **Diffuse functions**

Basis sets with diffuse functions are important for systems where electrons may be far from the nucleus. Diffuse orbitals occupy a larger region of space and thus improve the description of species with extended electronic densities, such as anions or molecules forming hydrogen bonds. In the Pople family of basis sets, the presence of diffuse functions is indicated by one or two '+' in the basis set name. In this way, 6-31+G* indicates that heavy atoms have been augmented with an additional one *s* and one set of *p*-functions having small exponents, and 6-31++G* indicates the presence of diffuse *s*-functions on H.

- **Effective core potentials**

There are heavy atoms that have large numbers of electrons, such as the transition metals, which require the use of a large number of basis functions to describe them. The vast majority of these electrons are internal or core electrons, and thus a minimal representation will probably be adequate to describe the atom. However, these elements represent a great challenge in electronic structure calculations due to the large number of electrons involved. In 1935, Hellmann proposed a solution to this problem:²⁰ replace the core electrons by a potential parameterized by expansion into a suitable set of analytical functions of the nuclear–electron distance. Such functions are referred to as effective core potentials (ECPs), although they are also known as pseudopotentials (PP). In addition, in the ECPs, the valence orbitals with radial nodes are replaced by a set of pseudo-orbitals without the nodal structure in the core region in order to simplify the calculations. Popular currently used pseudopotentials are, for example, the Hay and Wadt (sometimes also called the Los Alamos National Laboratory or LANL) ECPs.²¹

2.2.4. Post Hartree-Fock methods

Limitations of the Hartree-Fock method

As was seen in sec. 2.2.2, in the HF method, the electronic wave function is expressed as an antisymmetrized product of spin orbitals (Slater determinant), and the energy is calculated by applying the variational principle. According to this method, electrons move in the field of the average potential created by the rest of the electrons. This is, obviously, a rough approximation, since the movement of the electrons depends on the instantaneous position of the other electrons, i.e., their movement is correlated.

To analyze the limitations of the HF method it is interesting to analyze the simultaneous probability of locating electron 1 in $d\vec{r}_1$ and electron 2 in $d\vec{r}_2$. Given a

Slater determinant, $|\Psi_p(1)\bar{\Psi}_p(2)|$, it is easy to demonstrate that if the two electrons have opposite spin, the probability density is the product of the probability densities of electron 1 and of electron 2.

$$P(\vec{r}_1, \vec{r}_2) = |\Psi_p(\vec{r}_1)|^2 |\Psi_p(\vec{r}_2)|^2 \quad (2.64)$$

That is, the probability of finding an electron located at a certain point of the space is independent of the position of the other electron and, thus, the movement of the two electrons is not correlated. It can be observed that $P(\vec{r}_1, \vec{r}_2) \neq 0$, i.e., there is a finite probability that the two electrons of opposite spin are located at the same point of the space.

By contrast, if two electrons with the same spin are located in the orbitals Ψ_p and Ψ_q , it is easy to demonstrate that the probability density is:

$$P(\vec{r}_1, \vec{r}_2) = \frac{1}{2} \left\{ \left[|\Psi_p(\vec{r}_1)|^2 |\Psi_q(\vec{r}_2)|^2 + |\Psi_p(\vec{r}_2)|^2 |\Psi_q(\vec{r}_1)|^2 \right] - \left[\Psi_p^*(\vec{r}_1)\Psi_q(\vec{r}_1)\Psi_q^*(\vec{r}_2)\Psi_p(\vec{r}_2) + \Psi_p(\vec{r}_1)\Psi_q^*(\vec{r}_1)\Psi_q(\vec{r}_2)\Psi_p^*(\vec{r}_2) \right] \right\} \quad (2.65)$$

In this case the probability that the two electrons are located at the same point of the space is zero. In this way, when using a Slater determinant, the movement of electrons with parallel spin is correlated, although the movement of electrons with opposite spin is not.

The correlation of electrons with opposite spin is often called *Coulomb correlation*, while the correlation of electrons with the same spin is often called *Fermi correlation*. Likewise, the region of the space around an electron, where the probability of locating another electron with the opposite or the same spin is very small, is called a *Coulomb hole* or *Fermi hole*, respectively.

As discussed in sec. 2.2.3, if the set of basis functions used in the HF method was infinite, the energy and the function obtained would be the best within this monoconfigurational approximation and we would be at the HF limit. The correlation energy, E_{HF}^{corr} , is defined as the difference between the exact, non-relativistic ground state energy, E_0 , and the HF energy at the limit of an infinite basis set, E_{HF}^{limit} .

$$E_{HF}^{corr} = E_0 - E_{HF}^{limit} \quad (2.66)$$

Electron Correlation Methods

There are different methods that introduce the electron correlation. The conventional methods are based on the HF method and, thus, are known as *post-HF* methods. Among all of them, the most important are: the variational method of

Configuration Interaction (CI),^{22,23} the *Møller–Plesset* perturbation theory²⁴ (MP- n , where $n = 2, 3, 4, \dots$ stands for the order of the perturbation) and the *Coupled Cluster* (CC) method.^{25–29}

Other non-conventional methods that include the electron correlation, alternatively, are those based on the *Density Functional Theory* (DFT). These methods have experienced a significant boom due to their low computational cost compared with the conventional methods. Since the DFT has been used in this thesis, this methodology will be dealt with in further detail in the next section.

2.2.5. Density Functional Theory

The conventional correlation methods are *ab initio* methods based on the wave function. An alternative to these methods is the DFT methodology, which uses the *electron density*, $\rho(\vec{r})$, as the variable to introduce the correlation. The advantage of using the electron density is obvious, since it only depends on the three spatial coordinates (x, y, z) and the spin (ω), while the wave function depends on $3N$ variables ($4N$ if we include the spin), N being the number of electrons. Unlike traditional quantum chemical methods, in the DFT formalism we are not dealing with a system of N interacting electrons, but with a dynamically equivalent system composed of N non-interacting *fictitious* electrons, which have the same density as the real system. In this way, DFT is a non-interacting theory and does not yield a correlated N -body wave function. The algebraic implementation of DFT through the Kohn-Sham equations is a one-electron theory, thus sharing many similarities with the HF formulation.

DFT is a powerful, formally exact theory. However, the use of approximations is necessary in practice, although this does not prevent a high accuracy from being achieved in the DFT calculations. On the other hand, the methods based on the DFT are substantially simpler than those based on the wave function. In addition, DFT is *potentially* capable of providing similar or even better results than wavefunction-based methods, and with a lower computational cost.

2.2.5.1. The Hohenberg-Kohn Theorems

The DFT theory was developed by Pierre Hohenberg and Walter Kohn in 1964,³⁰ following the Thomas-Fermi (TF) model,^{31,32} which was the genuine precursor of the DFT theory. Hohenberg and Kohn demonstrated that for molecules with a non-degenerate ground state, the molecular energy, the wave function, and all other properties are determined by electron probability density. Consequently, it is established that the energy of the system is a density functional $E_0 = E_0[\rho_0]$. Hence, DFT allows the energy and other ground-state molecular properties to be calculated exactly from the electron density of the ground state.

First Hohenberg-Kohn theorem: Proof of existence

The first Hohenberg-Kohn (HK) theorem demonstrates that the electron density uniquely determines the Hamiltonian operator and thus all the properties of the system.

This first theorem states that *the external potential $V_{ext}(\vec{r})$ is (to within a constant) a unique functional of $\rho(\vec{r})$; since, in turn $V_{ext}(\vec{r})$ fixes \hat{H} , we see that the full many-particle ground state is a unique functional of $\rho(\vec{r})$.*

The proof that the ground-state density determines the external potential proceeds via *reductio ad absurdum*, that is, it is shown that an assumption to the contrary generates an impossible result.

Let us assume that there exist two potentials $V_{ext}(\vec{r})$ and $V'_{ext}(\vec{r})$ differing by more than a constant and giving rise to the same ground state density, $\rho(\vec{r})$. These two external potentials are part of two distinct Hamiltonians \hat{H} and \hat{H}' , which only differ in the external potential:

$$\hat{H} = \hat{T} + \hat{V}_{ee} + \hat{V}_{ext} \quad (2.67)$$

$$\hat{H}' = \hat{T} + \hat{V}_{ee} + \hat{V}'_{ext} \quad (2.68)$$

Obviously, the Hamiltonians above belong to two different ground state wave functions Ψ and Ψ' , and corresponding ground state energies, E_0 and E'_0 , respectively, with $E_0 \neq E'_0$. However, we assume that both wave functions, although different, give rise to the same electron density. Summarizing:

$$\hat{V}_{ext} \Rightarrow \hat{H} \Rightarrow \Psi \Rightarrow \rho(\vec{r}) \Leftarrow \Psi' \Leftarrow \hat{H}' \Leftarrow \hat{V}'_{ext}$$

Because of the variational principle, the expectation value of the Hamiltonian \hat{H} over the trial wave function Ψ' must be higher than the ground state energy E_0 , i.e.,

$$E_0 < \langle \Psi' | \hat{H} | \Psi' \rangle = \langle \Psi' | \hat{H} - \hat{H}' + \hat{H}' | \Psi' \rangle = \langle \Psi' | \hat{H}' | \Psi' \rangle + \langle \Psi' | \hat{H} - \hat{H}' | \Psi' \rangle \quad (2.69)$$

Since the two Hamiltonian operators only differ in the external potential:

$$E_0 < E'_0 + \langle \Psi' | \hat{T} + \hat{V}_{ee} + \hat{V}_{ext} - \hat{T} - \hat{V}_{ee} - \hat{V}'_{ext} | \Psi' \rangle \quad (2.70)$$

which yields:

$$E_0 < E'_0 + \int \rho(\vec{r}) [\hat{V}_{ext} - \hat{V}'_{ext}] d\vec{r} \quad (2.71)$$

Similarly, taking Ψ as a trial function for the problem \hat{H}' :

$$E'_0 < E_0 + \int \rho(\vec{r}) [\hat{V}_{ext} - \hat{V}'_{ext}] d\vec{r} \quad (2.72)$$

Adding eq. 2.71 and eq. 2.72, we would obtain:

$$E_0 + E'_0 < E'_0 + E_0 \quad (2.73)$$

which is a contradiction. This concludes the proof that there cannot be two different \hat{V}_{ext} that yield the same ground state electron density, ρ_0 , or, in other words, that the ground state density uniquely specifies the external potential \hat{V}_{ext} , and vice versa. Thus, ρ_0 contains the information about N ($\int \rho_0(\vec{r}) d\vec{r} = N$) and V_{ext} (uniquely determined by Z_α and R_α), from which \hat{H} can be constructed. From the latter we can obtain the ground state wave function Ψ_0 , which in turn enables the determination of the ground state energy, E_0 , and of all the other properties of the system.

Since the complete ground state energy is a functional of the ground state electron density, its individual components must also be, and thus we can write:

$$E_0[\rho_0] = T[\rho_0] + E_{ee}[\rho_0] + E_{Ne}[\rho_0] = \int \rho_0(\vec{r}) V_{Ne} d\vec{r} + T[\rho_0] + E_{ee}[\rho_0] \quad (2.74)$$

which results in:

$$E_0[\rho_0] = \int \rho_0(\vec{r}) V_{Ne} d\vec{r} + F_{HK}[\rho_0] \quad (2.75)$$

where:

$$F_{HK}[\rho_0] = T[\rho_0] + E_{ee}[\rho_0] = \langle \Psi | \hat{T} + \hat{V}_{ee} | \Psi \rangle \quad (2.76)$$

is the Hohenberg functional, and it is the holy grail of density functional theory. If it were known, we would have solved the Schrödinger equation exactly! And, since it is a universal functional completely independent of the system at hand, it applies equally well to small atoms as to gigantic molecules. $F_{HK}[\rho]$ contains the functional for the kinetic energy $T[\rho]$ and the one for the electron-electron interaction, $E_{ee}[\rho]$. The explicit form of both these functionals unfortunately remains completely unknown. However, from the latter we can extract at least the classical Coulomb part $J[\rho]$:

$$E_{ee}[\rho] = \frac{1}{2} \int \int \frac{\rho(\vec{r}_1)\rho(\vec{r}_2)}{r_{12}} d\vec{r}_1 d\vec{r}_2 + E_{nc}[\rho] = J[\rho] + E_{nc}[\rho] \quad (2.77)$$

$E_{nc}[\rho]$ is the non-classical contribution to the electron-electron interaction, which contains all the effects of the self-interaction correction, exchange and Coulomb correlation. It is worth pointing out that the explicit form of the functionals $T[\rho]$ and $E_{nc}[\rho]$ is the major challenge of DFT.

Second Hohenberg-Kohn theorem: Variational Theorem

How can we be sure that a certain density is the ground-state density that we are looking for?

The second HK theorem states that $F_{HK}[\rho]$, the functional that delivers the ground state energy of the system, delivers the lowest energy if and only if the input density is the true ground state density, ρ_0 . This is nothing but the variational principle:

$$E_0 \leq E[\tilde{\rho}] = T[\tilde{\rho}] + E_{Ne}[\tilde{\rho}] + E_{ee}[\tilde{\rho}] \quad (2.78)$$

In other words this means that for any trial density $\tilde{\rho}(\vec{r})$, which satisfies the necessary boundary conditions such as $\tilde{\rho}(\vec{r}) \geq 0$, $\int \tilde{\rho}(\vec{r}) d\vec{r} = N$, and which is associated with some external potential \hat{V}_{ext} , the energy obtained from the functional of eq. 2.74 represents an upper bound to the true ground state energy E_0 . The energy E_0 results if and only if the exact ground state density is inserted in eq. 2.75.

The proof of eq. 2.78 makes use of the variational principle established for wave functions. We should recall that any trial density $\tilde{\rho}$ defines its own Hamiltonian \hat{H} and hence its own wave function $\tilde{\Psi}$. This wave function can now be taken as the trial wave function for the Hamiltonian generated from the true external potential \hat{V}_{ext} . Thus:

$$\langle \tilde{\Psi} | \hat{H} | \tilde{\Psi} \rangle = T[\tilde{\rho}] + E_{ee}[\tilde{\rho}] + \int \tilde{\rho}(\vec{r}) V_{ext} d\vec{r} = E[\tilde{\rho}] \geq E_0[\rho] = \langle \tilde{\Psi}_0 | \hat{H} | \tilde{\Psi}_0 \rangle \quad (2.79)$$

2.2.5.2. The Kohn-Sham (KS) Method

It was not until 1965 when Kohn and Sham found a practical way to obtain the system properties directly from the density.³³ This method solves the problem concerning the poor approximation of the kinetic energy in the TF model.^{31,32} With the KS method it is possible to solve the equations in an exact way, but since, unfortunately, the $F_{HK}[\rho]$ functional is not known, it must be approximated. Thus, the KS formulation gives rise to approximate results.

The crucial idea behind this method is to take as a starting point a fictitious system of non-interacting electrons (built from a set of orbitals, i.e., one-electron functions) that have as their overall ground-state density the same density as some real system of interest where the electrons do interact. Note that since $\rho(\vec{r})$ determines Z_α and R_α , these quantities are necessarily identical in both the non-interacting and the real systems. Next, the energy functional is divided into specific components to facilitate further analysis. In this way, the energy functional is as follows:

$$E[\rho(\vec{r})] = T_{ni}[\rho(\vec{r})] + V_{ne}[\rho(\vec{r})] + V_{ee}[\rho(\vec{r})] + \Delta T[\rho(\vec{r})] + \Delta V_{ee}[\rho(\vec{r})] \quad (2.80)$$

where the first term, $T_{ni}[\rho(\vec{r})]$, corresponds to the kinetic energy of the non-interacting electrons; the second term, $V_{ne}[\rho(\vec{r})]$, corresponds to the nuclear–electron interaction; the third term, $V_{ee}[\rho(\vec{r})]$, corresponds to the classical electron–electron repulsion; the fourth term, $\Delta T[\rho(\vec{r})]$, is the *correction* to the kinetic energy deriving from the interacting nature of the electrons, and the fifth term, $\Delta V_{ee}[\rho(\vec{r})]$, accounts for all non-classical corrections to the electron–electron repulsion energy.

Within a spin orbital expression for the density, eq. 2.80 can be rewritten as:

$$\begin{aligned} E[\rho(\vec{r})] = & \sum_{i=1}^N \left(\left\langle \chi_i \left| -\frac{1}{2} \nabla_i^2 \right| \chi_i \right\rangle - \left\langle \chi_i \left| \sum_{\alpha=1}^{nuclei} \frac{Z_\alpha}{|r_i - R_\alpha|} \right| \chi_i \right\rangle \right) + \\ & + \sum_{i=1}^N \left\langle \chi_i \left| \frac{1}{2} \int \frac{\rho(\vec{r}')}{|r_i - r'|} d\vec{r}' \right| \chi_i \right\rangle + E_{xc}[\rho(r)] \end{aligned} \quad (2.81)$$

where N is the number of electrons, and the density for a Slater-determinantal wave function is given by:

$$\rho = \sum_{i=1}^N \langle \chi_i | \chi_i \rangle \quad (2.82)$$

Note that the last two terms of eq. 2.80 have been included in a term E_{xc} , typically referred to as the *exchange-correlation energy*, which is the functional that contains everything that is unknown.

In order to find the spin orbitals that minimize E in eq. 2.81, we have to solve the following pseudo-eigenvalue equations:

$$h_i^{KS} \chi_i = \varepsilon_i \chi_i \quad (2.83)$$

where the KS one-electron operator is defined as:

$$h_i^{KS} = -\frac{1}{2}\nabla_i^2 - \sum_{\alpha=1}^{nuclei} \frac{Z_\alpha}{|r_i - R_\alpha|} + \int \frac{\rho(\vec{r}')}{|r_i - r'|} d\vec{r}' + V_{XC} \quad (2.84)$$

V_{XC} being the so-called functional derivative, which is the one-electron operator for which the expectation value of the KS Slater determinant is E_{XC} .

$$V_{XC} = \frac{\delta E_{XC}}{\delta \rho} \quad (2.85)$$

Since the energy of eq. 2.80 that we want to minimize is exact, the spin orbitals χ_i must provide the exact density. The set of equations that determine the “best” spin orbitals for which E obtains its lowest value are:

$$\sum_{i=1}^N h_i^{KS} |\chi_1 \chi_2 \cdots \chi_N\rangle = \sum_{i=1}^N \varepsilon_i |\chi_1 \chi_2 \cdots \chi_N\rangle \quad (2.86)$$

As in the context of MO theory, we can express the KS orbitals within a basis set of functions $\{\phi_\nu\}$, and then determine the individual orbital coefficients by solving a secular equation entirely analogous to that employed for the HF theory. In this case, the matrix elements of the Fock operator $F_{\mu\nu}$ are replaced by elements $K_{\mu\nu}$ defined by:

$$K_{\mu\nu} = \left\langle \phi_\mu \left| -\frac{1}{2}\nabla_i^2 - \sum_{\alpha=1}^{nuclei} \frac{Z_\alpha}{|r_i - R_\alpha|} + \int \frac{\rho(\vec{r}')}{|r_i - r'|} d\vec{r}' + V_{XC} \right| \phi_\nu \right\rangle \quad (2.87)$$

As can be seen, the KS equations are closely related to the HF equations as far as their resolution is concerned, since from a mathematical standpoint, these equations are solved iteratively. However, the main difference between the HF and the KS equations is that the latter may be solved in an exact way. All that is required is the value of E_{XC} , i.e., the value of the exchange-correlation energy. Therefore, the greatest efforts to develop this theory are based on the search for an exchange-correlation potential whose expectation value, E_{XC} , approaches the exact form as accurately as possible. In other words, the DFT method optimizes the electron density and not the wave function. This implies that the validity of the results depends on how the property which is being calculated is described by the electron density, i.e., how this property is related with the electron density. The accuracy of any exchange-correlation functional can depend as much on the chemical system under study as on the intended property.

Within the range of exchange correlation potentials, it is possible to classify them into three major types according to the approximations used in their estimation.

2.2.5.3. Exchange-Correlation functionals

- **Local density approximation (LDA)**

At the center of this model is the idea of a hypothetical *uniform electron gas*, also known as homogeneous electron gas. This is a system in which electrons move in a positive background charge distribution such that the total ensemble is electrically neutral. The reason why the uniform electron gas has such a prominent place in DFT is that it is the only system for which we know the form of the exchange and correlation energy functionals exactly or at least to a very high degree of accuracy. The central idea of LDA is the assumption that we can write E_{xc} in the following form:

$$E_{xc}^{LDA}[\rho] = \int \rho(\vec{r}) E_{xc}(\rho(\vec{r})) d\vec{r} \quad (2.88)$$

where $E_{xc}(\rho(\vec{r}))$ is the exchange-correlation energy per particle of a uniform electron gas of density $\rho(\vec{r})$. This energy per particle is weighted with the probability $\rho(\vec{r})$ that there is an electron at this position in the space. The quantity $E_{xc}(\rho(\vec{r}))$ can be further split into exchange and correlation contributions:

$$E_{xc}(\rho(\vec{r})) = E_x(\rho(\vec{r})) + E_c(\rho(\vec{r})) \quad (2.89)$$

The exchange part, E_x , which represents the exchange energy of an electron in a uniform electron gas of a particular density:

$$E_x(\rho(\vec{r})) = -\frac{3}{4} \left(\frac{3}{\pi} \right)^{1/3} \int \rho^{1/3}(\vec{r}) \quad (2.90)$$

which is, apart from the pre-factor, equal to the form found by Slater in his approximation of the Hartree-Fock exchange:

$$E_x(\rho(\vec{r})) = -\frac{9\alpha}{8} \left(\frac{3}{\pi} \right)^{1/3} \int \rho^{1/3}(\vec{r}) \quad (2.91)$$

Inserting eq. 2.90 into eq. 2.88, we can see the $\rho^{4/3}$ dependence of the exchange energy:

$$E_x^{LDA}[\rho] = -\frac{3}{4} \left(\frac{3}{\pi} \right)^{1/3} \int \rho^{4/3}(\vec{r}) d\vec{r} \quad (2.92)$$

This exchange functional is frequently called the *Slater exchange* and is abbreviated to S . No such explicit expression is known for the correlation part, E_c , although various authors have presented analytical expressions of E_c based

on sophisticated interpolation schemes. The most widely used representations of E_C are the ones developed by Vosko, Wilk, and Nusair³⁴ (VWN, and its variant VWN5), while the most recent and probably also the most accurate has been given by Perdew–Wang 92 (PW92).³⁵

- **Generalized gradient approximation (GGA)**

The first logical step to go beyond LDA is the use of not only the information about the density $\rho(\vec{r})$ at a particular point \vec{r} , but to supplement the density with information about the gradient of the charge density, $\nabla\rho(\vec{r})$ in order to account for the non-homogeneity of the true electron density. Thus, we can write the exchange-correlation energy in the following form, termed the *generalized gradient approximation* (GGA):

$$E_{XC}^{GGA}[\rho_\alpha, \rho_\beta] = \int f(\rho_\alpha, \rho_\beta, \nabla\rho_\alpha, \nabla\rho_\beta) d\vec{r} \quad (2.93)$$

Note that approximate functionals are usually also expressed in an unrestricted version, where not the electron density $\rho(\vec{r})$, but the two spin densities, $\rho_\alpha(\vec{r})$ and $\rho_\beta(\vec{r})$, with $\rho(\vec{r}) = \rho_\alpha(\vec{r}) + \rho_\beta(\vec{r})$ are employed as the central input.

In practice, E_{XC}^{GGA} is usually split into its exchange and correlation contributions:

$$E_{XC}^{GGA} = E_X^{GGA} + E_C^{GGA} \quad (2.94)$$

where the exchange part of E_{XC}^{GGA} is given by:

$$E_X^{GGA} = E_X^{LDA} - \sum_\sigma \int F(s_\sigma) \rho_\sigma^{4/3}(\vec{r}) d\vec{r} \quad (2.95)$$

The argument of the function F is the *reduced density gradient* for spin σ :

$$s_\sigma(\vec{r}) = \frac{|\nabla\rho_\sigma(\vec{r})|}{\rho_\sigma^{4/3}(\vec{r})} \quad (2.96)$$

Some popular GGA exchange functionals are: Becke 88 (B),³⁶ Perdew–Wang (PW),³⁷ Perdew–Burke–Ernzerhof (PBE³⁸ and mPBE^{38,39}), and some popular GGA correlation methods are: Becke 88,⁴⁰ Perdew–Wang 91 (PW91),⁴¹ and Lee–Yang–Parr (LYP).⁴²

A new class of very promising DFT functionals based on the GGA, the so-called meta-GGA (M-GGA), was developed by including additional semi-local information beyond the first-order density gradient contained in the GGAs. These methods depend explicitly on higher-order density gradients, or typically on the kinetic energy density, which involves derivatives of the occupied Kohn-Sham orbitals. Examples of M-GGA functionals are TPSS⁴³ and VSXC.⁴⁴

- **Hybrid functionals**

Hybrid density functional methods combine the exchange-correlation of a conventional GGA and M-GGA methods with a percentage of Hartree-Fock (or exact) exchange. A certain degree of empiricism is used in optimizing the weight factor for each component and the functionals that are mixed. Inclusion of exact HF exchange is often found to improve the calculated results, although the optimum fraction to be included depends on the specific property of interest. In fact, the exact amount of HF exchange cannot be assigned from first-principles and is therefore fitted empirically.

There are a large number of exchange-correlation hybrid-GGA (H-GGA) and hybrid-meta-GGA (HM-GGA) functionals. The most widely used hybrid functional and the one used in this thesis is the *Becke three-parameter Lee-Yang-Parr* (B3LYP) :^{45,46}

$$E_{XC}^{B3LYP} = E_X^{LDA} + a_0(E_X^{HF} - E_X^{LDA}) + a_x(E_x^{GGA} - E_x^{LDA}) + E_C^{LDA} + a_c(E_C^{GGA} - E_C^{LDA}) \quad (2.97)$$

where $a_0=0.20$, $a_x=0.72$, and $a_c=0.81$. E_x^{GGA} is the exchange GGA functional Becke 88,³⁶ E_C^{GGA} is the correlation GGA functional LYP,⁴² and E_C^{LDA} is the correlation LDA functional VWN.³⁴

B3LYP is by far the most popular density functional in chemistry, representing 80% of the total of occurrences of density functionals in the literature, in the period 1990-2006.⁴⁷ However, in recent times other hybrid functionals such as the Minnesota M06 suite of density functionals (M06,⁴⁸ M06-2X,⁴⁸ and M06-HF⁴⁹) are being extensively used.

2.3. Molecular Mechanics Methods

2.3.1. Force fields

The calculation of the electronic energy for a given nuclear configuration to provide a PES is prohibitive by QM methods when a large number of electrons are involved. If we are interested in studying complex systems consisting of many atoms, such as biological systems, we cannot resort to these methods due to the large computational cost involved in solving the Schrödinger equation. For this reason, the obtaining and further exploration of the PES should be performed by other computationally less expensive methods such as molecular mechanics (MM). In the MM methods, the calculation of the electronic energy by means of the Schrödinger equation is neglected and it is replaced by parameters derived from experimental

data (generally obtained by vibrational spectroscopy techniques) or from *ab initio* calculations performed on model systems. In this way, the potential energy is obtained from a parameterized function where only the nuclear coordinates and parameters are included, thereby allowing these methods to be used in large systems ($>10^4$ atoms). The potential obtained from molecular mechanics, also known as “force field”, treats the dynamics of atoms with classical mechanics (Newton’s second law) instead of by the quantum-mechanical equations, the atoms being reduced to spheres and the bonds to springs. However, these force fields are not able to describe the electronic rearrangements, and thus the bond breaking/forming processes and electron pair rearrangements cannot be correctly described by MM.

The full Hamiltonian, or potential function, used in the force fields is described as the sum of several terms representing the energy required to distort a molecule in a specific way:

$$V_{FF}(r) = V_{stretch} + V_{bend} + V_{tors} + V_{impr} + V_{cross} + V_{VDW} + V_{elec} \quad (2.98)$$

In turn, the energetic terms of the force field can be classified into two distinct groups: the covalent or bonding energy terms ($V_{bonding}$) and the non-bonding terms ($V_{non-bonding}$):

$$V_{FF}(r) = V_{bonding} + V_{non-bonding} \quad (2.99)$$

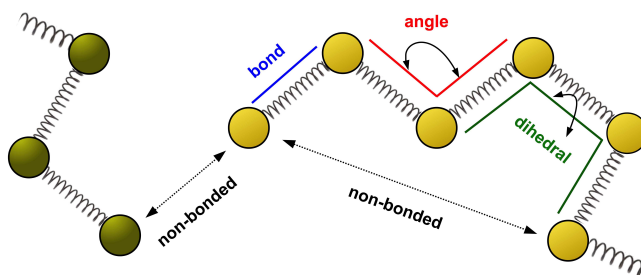


Figure 2.3.: Representation of the main interactions involved in a force field.

2.3.1.1. Bonding Terms

These terms describe the interactions within the structure of a molecule and define the interaction between atoms belonging to the same molecule. The bonding potential includes the contributions corresponding to the bonds (Fig. 2.4a), the angles (Fig. 2.4a), the dihedral angles (Fig. 2.4b) and the improper angles (Fig. 2.4c).

$$V_{bonding} = V_{stretching} + V_{bending} + V_{torsional} + V_{impropers} + V_{cross-terms} \quad (2.100)$$

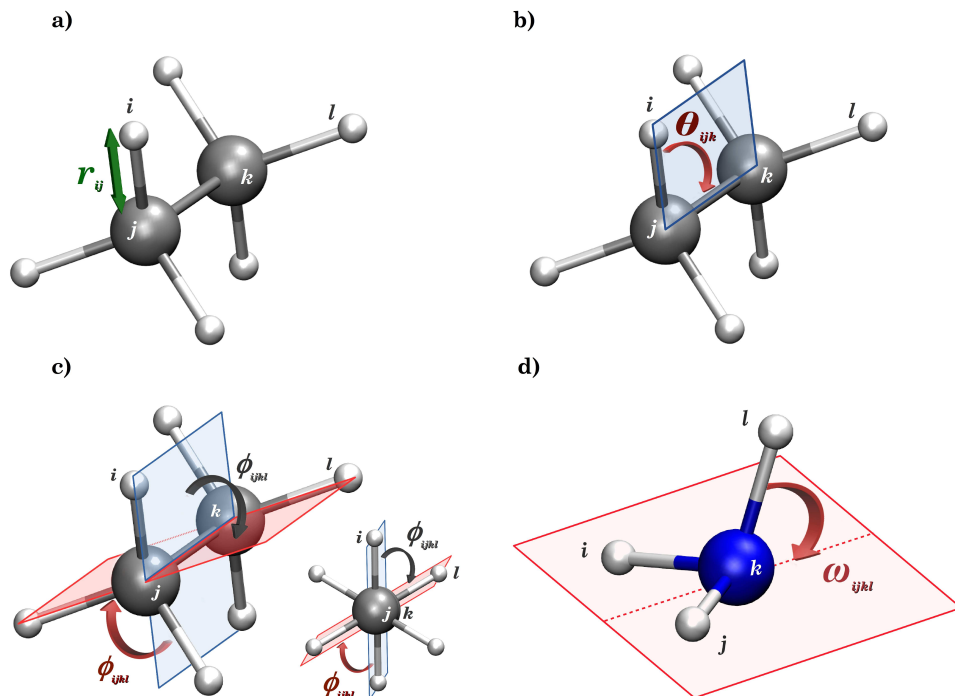


Figure 2.4.: Representation of the internal coordinates: (a) bond distance between two atoms (r_{ij}), (b) angle between three atoms (θ_{ijk}), (c) two different views of the dihedral angle between four atoms (ϕ_{ijkl}), (d) improper or out-of-plane angle between four atoms (ω_{ijkl}).

• Bond energy

This energy term, also known as *bond stretching energy*, represents the energy required to stretch a bond between two atoms (r_{ij} , see Fig. 2.4a). It comes from a second-order Taylor expansion of the energy along the equilibrium distance $r_{ij,0}$, which in its simplest form is reduced to a quadratic term of the distance: the harmonic oscillator form (see Fig. 2.5).

$$V_{stretching} = \sum^{bonds} \frac{1}{2} k_b (r_{ij} - r_{ij,0})^2 \quad (2.101)$$

where r_{ij} is the internuclear separation between two atoms, $r_{ij,0}$ is the equilibrium distance for the bond, and $k_b = \left(\frac{d^2 V}{dr^2} \right)_{r_{ij}=r_{ij,0}}$ is the bond force constant.

However, the quadratic function is only valid when small movements are made

with respect to the equilibrium position. Although the harmonic approximation is sufficient for many studies, if we want to calculate a bond energy that describes the dissociation of a bond correctly, we have to resort to other potentials, such as the Morse potential (see Fig. 2.5).

$$V_{MORSE} = \sum^{bonds} D_e \left(1 - e^{-a(r_{ij} - r_{ij,0})}\right)^2 \quad (2.102)$$

where two new parameters appear, D_e and a , which are related according to the expression $D_e = k_b/2a^2$. The first one is the the bond dissociation energy and the second is the width of the potential well.

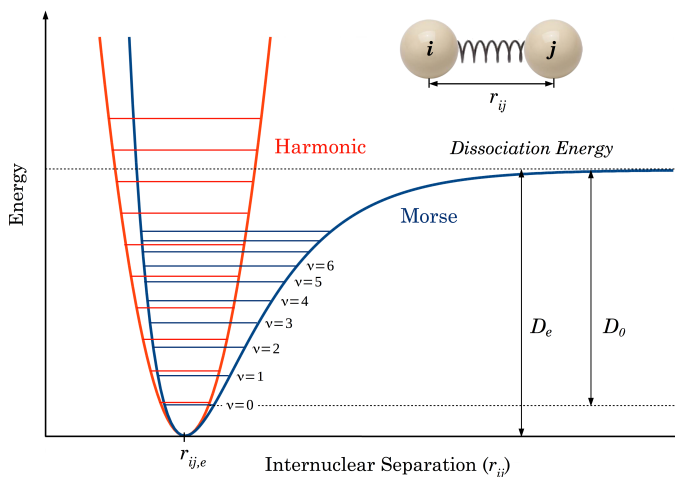


Figure 2.5.: The Harmonic oscillator potential (red) and Morse potential (blue) that describe the bond stretching energy.

- **Angle energy**

This energy term, also known as *angle bending energy*, represents the energy associated with the small deformations of the angle formed between three atoms (θ_{ijk} , see Fig. 2.4a). Like the stretching energy, this term comes from a Taylor series expanded to the second order and has a “harmonic” form. As in the previous case, this formula is designed to provide the energy change of the angle between three consecutive atoms when it is distorted away from its equilibrium distance.

$$V_{bending} = \sum^{angles} \frac{1}{2} k_\theta (\theta_{ijk} - \theta_{ijk,0})^2 \quad (2.103)$$

where θ_{ijk} is the actual angle between three atoms, $\theta_{ijk,0}$ is the equilibrium angle and k_θ is the force constant of the angle.

- **Torsional energy**

This energy term describes the energy changes of a molecule that undergoes a rotation about one of its bonds (see Fig. 2.6). In this way, this energy is related with the torsions originated in the dihedral angle formed between four atoms (ϕ_{ijkl} , see Fig. 2.4b). Unlike the previous energy terms, torsional energy does not have a “harmonic” form because this form is not able to describe the torsions adequately. The reason for this is that, for many dihedral angles, the full range of angles from 0° to 360° can be accessible with differences in energy that are not too large. These effects are corrected by introducing a periodic function, which is continuous throughout the range of possible angles. The resulting expression is:

$$V_{torsional} = \sum^{dihedral} \frac{1}{2} V_n [1 + \cos(n\phi_{ijkl} - \delta)] \quad (2.104)$$

where V_n is the force constant associated to the rotation of the dihedral, n is the periodicity of the angle taken (ϕ_{ijkl}), which determines the number of peaks and wells that appear in the potential, and δ is the phase angle that allows the position of the energy minimum to move (often it is restricted to taking the values 0° or 180°).

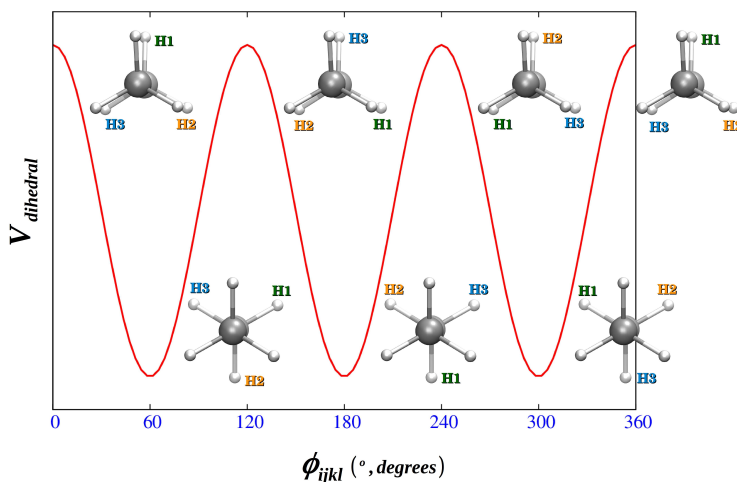


Figure 2.6.: Example of the potential generated by the rotation of the dihedral angle formed by the atoms $ijkl$ (ϕ_{ijkl} , see Fig. 2.4b), with a periodicity of 3. Both staggered and eclipsed conformations of the ethane molecule can be observed every 120° .

- **Improper energy**

This is the most complicated energy term of all bonding terms. It describes the energy of out-of-plane motions or improper torsion (distortion) on the flat structure

of minimum energy. This improper dihedral angle is generally used in cases where there are planar groups with sp^2 -hybridized carbon atoms, such as in carbonyl groups or aromatic systems. This is because the proper dihedral angle between four consecutive atoms ($ijkl$) cannot correct the deviations from planarity, and thus we define an improper dihedral angle between four atoms ($ijkl$) that are not directly bonded to each other (ω_{ijkl}), (see Fig. 2.4c). Some force fields such as the OPLS⁵⁰⁻⁵⁵ use a form similar to that of the torsional energy:

$$V_{improper} = \sum^{improvers} \frac{1}{2} V_n [1 + \cos(n\omega_{ijkl} - \delta)] \quad (2.105)$$

while others, such as the CHARMM⁵⁶⁻⁶² force field, take on a “harmonic” form:

$$V_{improper} = \sum^{improvers} \frac{1}{2} k_\omega (\omega_{ijkl} - \omega_{ijkl,0})^2 \quad (2.106)$$

where k_ω is the force constant for the energy term and $\omega_{ijkl,0}$ is the equilibrium value of the improper dihedral angle.

- **Cross-term energy**

Some force fields include the cross-term contribution to the bonding energy. Cross-terms refer to the coupling between different internal coordinates, such as angles and bond lengths. As in the other bonding terms, it comes from a Taylor series expanded to the second order. These terms are particularly important when describing the spectroscopic properties, such as vibrational frequencies (stretching, bending, etc.). For example, in the C=O and C–N bonds of amides, stretching-stretching coupling is very important. This term can be modeled as:

$$V_{cross-term}(r_{CN}, r_{CO}) = \frac{1}{2} k_{CN,CO} (r_{CN} - r_{CN,0}) (r_{CO} - r_{CO,0})$$

2.3.1.2. Non-Bonded Terms

The non-bonded (nb) terms describe the energy between atoms belonging to different molecules, or between atoms which are not directly bonded together within the same molecule, i.e., those separated by more than three chemical bonds (1-5 interactions). Some force fields, however, allow a total or partial (with a weighting factor) interaction between 1-4 atoms (see Fig. 2.7).

The non-bonding potential includes the contributions corresponding to the electrostatic (elec) and Van der Waals (VdW) interactions.

$$V_{non-bonding} = V_{elec} + V_{VdW} \quad (2.107)$$

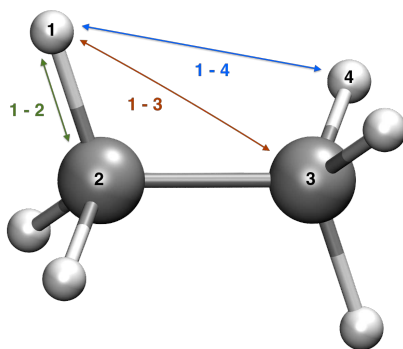


Figure 2.7.: Illustration of the 1-2, 1-3 and 1-4 non-bonding exclusions for a hydrogen atom in an ethane molecule. Figure adapted from reference [63].

- **Electrostatic energy**

The electrostatic energy, V_{elec} , represents the energy arising from the electrostatic interactions between two charge distributions. The most common way to describe electrostatic interactions is through interactions between partial charges assigned to each of the atoms forming molecules. The net partial charge on an atom is obtained by adding the nuclear charge of the atom and the charge coming from the electron cloud that surrounds it. The electrostatic energy is calculated as:

$$V_{elec} = \frac{1}{4\pi\epsilon_0\epsilon} \sum_{i,j}^{pairs} \frac{q_i q_j}{r_{ij}} \quad (2.108)$$

where q_i and q_j are the fractional charges on atoms i and j , and r_{ij} is the distance between the two particles. The prefactor $1/4\pi\epsilon_0\epsilon$ is Coulomb's constant, where ϵ is the dielectric constant, which will have a value of 1 when the system is in vacuum. The sum in eq.2.108 runs over all pairs of atoms for which the electrostatic interaction is to be calculated. It should be noted that the fractional charges on the atoms are constants and do not change during a calculation.

- **Van der Waals energy**

The VdW energy, V_{vdw} , mimics the long-range dispersion interactions and the short-range repulsive interactions. Since V_{vdw} takes the form of a Lennard-Jones potential (12-6 potential), it is also known as Lennard-Jones energy, V_{LJ} :

$$V_{LJ} = \sum_{i,j}^{pairs} \frac{B_{ij}}{r_{ij}^{12}} - \frac{C_{ij}}{r_{ij}^6} \quad (2.109)$$

where B_{ij} and C_{ij} are positive constants which depend on the types of the

atoms, i and j . As in the electrostatic energy, the sum is over all pairs of atoms for which the interaction is to be calculated. The term on the left is the repulsive part and the term on the right is the attractive part (see Fig. 2.8). The parameters B_{ij} and C_{ij} are usually defined in terms of the depth of the Lennard-Jones well, ε_{ij} , and either the position of the minimum energy (bottom of the well), $r_{ij,0}$, or the distance at which the energy of the interaction is zero, σ_{ij} .

$$V_{LJ} = \varepsilon_{ij} \left[\left(\frac{r_{ij,0}}{r_{ij}} \right)^{12} - 2 \left(\frac{r_{ij,0}}{r_{ij}} \right)^6 \right] = 4\varepsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^6 \right] \quad (2.110)$$

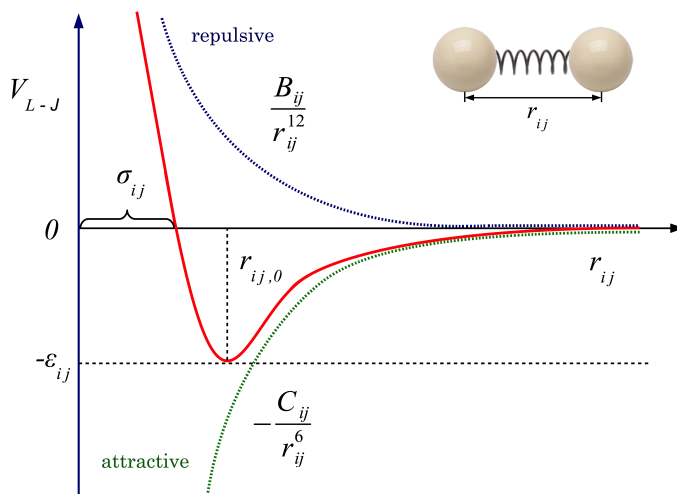


Figure 2.8.: Variation of the Lennard-Jones potential with intermolecular distance (red) and the attractive (green) and repulsive (blue) terms.

2.3.1.3. The OPLS Force Field

There are several parameterized force fields used in the MM, the most commonly used are AMBER,⁶⁴⁻⁶⁶ GROMOS,⁶⁷ CHARMM, and OPLS. In the present thesis, the force field employed throughout all the simulations has been the OPLS (Optimized Potentials for Liquid Simulations) and, more specifically, the OPLS-AA (all atoms).⁵⁰⁻⁵² This force field was developed by William L. Jorgensen at Yale University in 1996, and its functional form is as follows:

$$V(r^N) = V_{bond} + V_{angles} + V_{dihedral} + V_{non-bonding} \quad (2.111)$$

where:

$$V_{bond} = \sum_i^{bonds} k_r (r - r_e)^2 \quad (2.112)$$

$$V_{angle} = \sum_i^{angles} k_\theta (\theta - \theta_e)^2 \quad (2.113)$$

$$V_{dihedral} = \frac{V_1}{2} [1 + \cos(\phi - \delta_1)] + \frac{V_2}{2} [1 + \cos(2\phi - \delta_2)] + \frac{V_3}{2} [1 + \cos(3\phi - \delta_3)] + \frac{V_4}{2} [1 + \cos(4\phi - \delta_4)] \quad (2.114)$$

$$V_{nb} = \sum_{i,j}^{pairs} f_{ij} \left\{ \frac{q_i q_j}{4\pi\epsilon_0\epsilon r_{ij}} + 4\epsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^6 \right] \right\} \quad (2.115)$$

The OPLS-AA force field uses geometrical combination rules for the Lennard-Jones parameters ϵ_{ij} and σ_{ij} in the non-bonding interactions. Moreover, in eq. 2.115 the variable f_{ij} is a weighting factor for the interactions so that:

- For 1-2 and 1-3 interactions f_{ij} takes the value 0 and so these interactions are excluded from the sum.
- For the 1-4 interactions the value $f_{ij} = 0.5$.
- For all the other interactions $f_{ij} = 1$.

All the interaction sites are centered on the atoms, so the “lone pairs” are not taken into account.

It should be pointed out that OPLS simulations in aqueous solution typically use the TIP3P⁶⁸ or the TIP4P⁶⁹ water model potentials.

2.3.2. Treatment of the Non-Bonding Interactions:

2.3.2.1. Cut-off Methods

The principal problem with the non-bonding energy is the long-range electrostatic interaction. As discussed above, the non-bonding energy should be evaluated over all the atom pairs of the system. However, to do this, a very expensive computational cost is required, and a significant part of the effort is used to evaluate minor interactions. For this reason, to perform calculations with a reasonable computational cost, but without lose of accuracy, we must establish a cut-off distance beyond which the non-bonding interactions are no longer taken into account. With this strategy, only the interaction among pairs within a distance lower than

the cut-off distance will be counted. There are different functions describing cut-off strategies to calculate the non-bonding interactions:⁶³

1.- Truncated function

This function neglects all the interactions beyond a certain distance, the cut-off distance (r_c). Each term in the non-bonding energy expression (eq.2.107) is multiplied by the function:

$$S(r) = \begin{cases} 1 & r \leq r_c \\ 0 & r > r_c \end{cases} \quad (2.116)$$

This option is not the most convenient because small displacements of the nuclei may change the total number of interactions, and may introduce discontinuities in the energy function and its derivatives.

2.- Smoothing function

This truncation method tapers the interaction continuously to zero at a given distance. In this case, the first derivative of the energy is continuous, although the second derivative is still discontinuous. Many smoothing functions have been proposed, the most common types being:

- **Switch function**

This computes all the interactions inside a radius (r_{on}), from which it is smoothed to zero at the outer cut-off (r_{off}).

$$S(r) = \begin{cases} 1 & r \leq r_{on} \\ \frac{(r_{off}^2 - r^2)^2 (r_{off}^2 + 2r^2 - 3r_{on}^2)}{(r_{off}^2 - r_{on}^2)^3} & r_{on} < r \leq r_{off} \\ 0 & r > r_{off} \end{cases} \quad (2.117)$$

- **Shift function**

This relies on just one cut-off distance (r_c) and modifies the form of the interaction throughout its entire range.

$$S(r) = \begin{cases} \left[1 - \left(\frac{r}{r_c} \right)^2 \right]^2 & r \leq r_c \\ 0 & r > r_c \end{cases} \quad (2.118)$$

In Fig.2.9, we can compare the behavior of these three different methods.

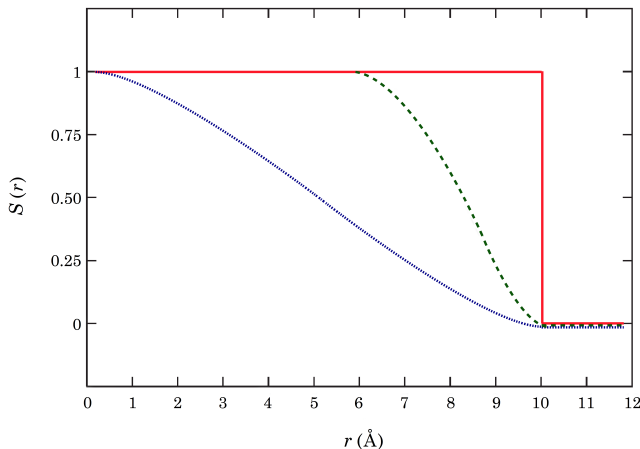


Figure 2.9.: Plots of different truncation functions, $S(r)$. The direct truncation function (plotted in red), the switch function (plotted in green) and the shift function (plotted in blue). In this chart, the cut-offs are as follows: $r_c = r_{off} = 10\text{\AA}$ and $r_{on} = 6\text{\AA}$. Figure adapted from reference [63].

The truncation method used in our calculations, the *atom-based force-switching truncation scheme*, was described by P. Steinbach and B. R. Brooks⁷⁰ and is a mixture of the switch and shift functions. In this method, it is not the interaction energy that is truncated directly, but its first derivative (i.e., its force). The truncation function used for the electrostatic interactions is as follows:

$$v(r) = \begin{cases} v_{true} + \frac{8c}{\gamma} \left[r_{on}^2 r_{off}^2 (r_{off} - r_{on}) - \frac{1}{5}(r_{off}^5 - r_{on}^5) \right] & r \leq r_{on} \\ c \left[A \left(\frac{1}{r} - \frac{1}{r_{off}} \right) + B(r_{off} - r) + \right. \\ \quad \left. + C(r_{off}^3 - r^3) + D(r_{off}^5 - r^5) \right] & r_{on} < r \leq r_{off} \\ 0 & r > r_{off} \end{cases} \quad (2.119)$$

where the constants are:

$$c = \frac{q_i q_j}{4\pi\epsilon_0\epsilon} \quad \gamma = (r_{off}^2 - r_{on}^2)^3 \quad A = \frac{r_{off}^4 (r_{off}^2 - 3r_{on}^2)}{\gamma}$$

$$B = \frac{6r_{on}^2 r_{off}^2}{\gamma} \quad C = -\frac{r_{off}^2 + r_{on}^2}{\gamma} \quad D = \frac{2}{5\gamma} \quad (2.120)$$

and the function used for the Lennard-Jones interactions is:

$$v(r) = \begin{cases} v_{true} - \frac{c}{(r_{on} r_{off})^{n/2}} & r \leq r_{on} \\ \frac{cr_{off}^{n/2}}{r_{off}^{n/2} - r_{on}^{n/2}} \left[\left(\frac{1}{2}\right)^{n/2} - \left(\frac{1}{r_{off}}\right)^{n/2} \right]^2 & r_{on} < r \leq r_{off} \\ 0 & r > r_{off} \end{cases} \quad (2.121)$$

$n = 6 \text{ or } 12$

This scheme is effective as long as we use a reasonably long inner cut-off distance and a broad switching region. In our calculations, we use an r_{off} of 16 Å, and an r_{on} of 14.5 Å. Furthermore, we introduce an additional term, r_{list} , in order to create a temporary list of non-bonding interactions. Since only the atoms on this list are taken into account, the calculation of the pairwise non-bonded interactions is accelerated. In our calculations, we use an r_{list} of 18 Å.

2.3.3. Periodic Boundary Conditions

The vast majority of biochemical processes occur in the condensed phase, e.g., a crystal or a solvated protein. To mimic the presence of a solvent in the enzymatic system some boundary approximations are needed, because the condensed phase systems are effectively infinite in extent. A large number of boundary approximations exist for use in molecular simulations, but the most commonly used representation is periodic boundary conditions (PBC).⁷¹

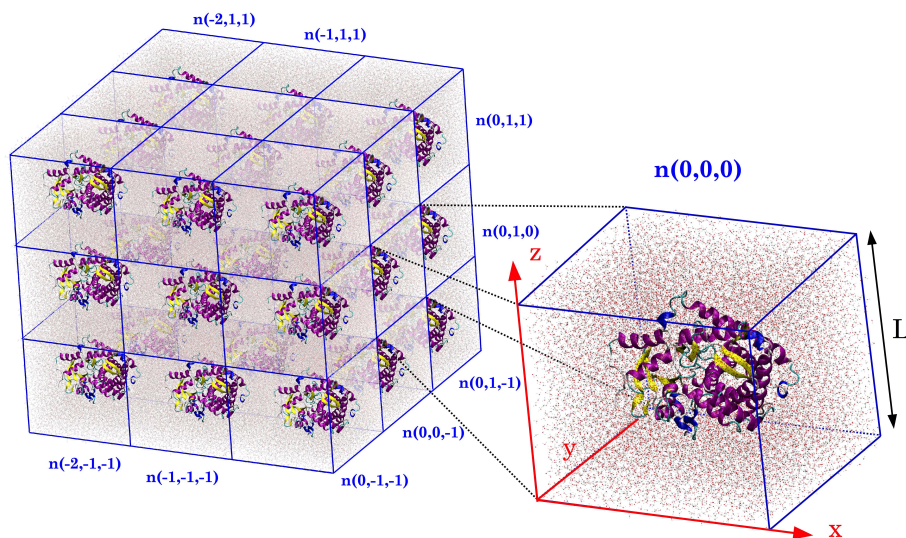


Figure 2.10.: Periodic Boundary Conditions in a 3D representation. The unit cell ($n(0,0,0)$) which is replicated in all directions in space is shown on the right of the image.

In this approximation, the system is introduced into a unit cell, usually a cube or a parallelepiped, which can be infinitely replicated in the three dimensions of space (see Fig. 2.10). Within this approximation, equivalent atoms in each of the copies behave identically and therefore they do not need to be treated in a different way throughout the simulation. Actually, this model does not modify the energy, it only changes the coordinates. The PBC approximation implies the use of the *minimum image convention* in order to prevent, during the calculation of the non-bonding energy, each particle from being able to interact with its nearest copy in the system. This convention makes its implementation computationally more economical, but in turn the truncation of non-bonding interactions is constrained to a cut-off distance less than half the length of the side of the unit cell ($L/2$).

2.3.4. Ewald Summation Method

Ewald summation is a method to compute the long-range electrostatic interactions under PBC in a more precise manner than by using cut-off schemes. This scheme was first developed by Ewald⁷² for calculating electrostatic energies in ionic crystals, although now, because of advances in the algorithms, it is commonly used in simulations of macromolecular systems.

Let us consider a system composed of N atoms or ions in a cubic unit cell of length L . The atoms within the unit cell have point charges $q_1, q_2, q_3, \dots, q_N$, at positions $r_1, r_2, r_3, \dots, r_N$, respectively. The cell-coordinate vector is $n = (n_x, n_y, n_z) = n_x L_x + n_y L_y + n_z L_z$, where x, y, z are the Cartesian coordinate unit vectors (see Fig. 2.10). The origin cell is located at $n = (0, 0, 0)$ with image cells located at nL intervals in all three dimensions as n goes to infinity, thereby forming a supercell. The distance between a particle in the origin cell and another at an image cell is $r_{ij} + nL = r_i - r_j + nL$, so that whenever there is a point charge q_i at position r_i , there also exists a point charge at $r_i + n_x L_x + n_y L_y + n_z L_z$.

The total Coulomb interaction energy of this system and its infinite replicas in PBC is given by:

$$V_{elec,PBC} = \frac{1}{4\pi\epsilon_0} \frac{1}{2} \sum_n \sum_{i=1}^N \sum_{\substack{j=1 \\ j \neq i, \text{ if } n=0}}^N \frac{q_i q_j}{|r_{ij} + nL|} \quad (2.122)$$

Notice that the sum over all pairs has been replaced by sums over all atoms, with a factor $1/2$, to remove the double counting. The problem comes from the fact that the infinite sum in eq. 2.122 not only converges very slowly, but is also *conditionally convergent* in 3D space. The Ewald summation is based on splitting the slowly convergent eq. 2.122 into two series of summations that converge not

only more rapidly but also absolutely. The trick basically consists in splitting the interaction $1/r$ as:

$$\frac{1}{r} = \frac{f(r)}{r} - \frac{1-f(r)}{r} \quad (2.123)$$

An usual choice is $f(r) = \text{erfc}(\alpha r)$, where α is called the *Ewald splitting parameter*, which results in the Ewald formula for the energy of the main cell:

$$V_{\text{Ewald}} = V^{(r)} + V^{(k)} + V^{(\text{self})} \quad (2.124)$$

The term $V^{(r)}$ represents the short-range interactions and its summation quickly converges in real space. This term takes only the interactions in the primary simulation box. The term $V^{(k)}$ represents the long-range term and cannot be directly computed by a sum in real space, since this has to run over all lattice vectors (n). However, the summation of this term quickly converges if its Fourier transform is taken and it is performed in reciprocal space. Finally, the term $V^{(\text{self})}$ is a correction term that cancels out the interaction of each of the introduced artificial counter-charges with itself.

As can be seen, the Ewald sum is a huge improvement with respect to direct sum, although it is not as fast as might be expected. Ewald summation scales with $N^{3/2}$, and thus the efficiency decreases as the system size increases. Nonetheless, there are implementations of this method that improve the reciprocal sum, thus accelerating its evaluation. One of the most widely used in molecular dynamics simulations is the *Particle Mesh Ewald* method (PME).⁷³ This method scales as $N \log N$ and is substantially faster, since it employs the fast Fourier transform (FFT) to evaluate the reciprocal sum, which is not computed for atom positions, but for fixed positions on a grid.

2.4. Hybrid Quantum Mechanics / Molecular Mechanics

As discussed above in sec.2.2, the QM methods provide a very accurate way to describe the chemical reactions (i.e., bond-forming and bond-breaking) and other processes that involve changes in the electronic structure, such as charge transfer or electronic excitation.⁷⁴ However, the application of these methods is confined to systems containing a discrete number of atoms. On the other hand, the MM force fields, which are discussed in sec.2.3, have amply demonstrated their ability to treat systems including a large amount of atoms. However, since the MM methods are based on classical empirical potentials, they cannot describe the bond-forming

and bond-breaking processes. As a result, these methods are computationally very cheap, but they are not suitable for studying chemical reactivity.

When we study enzymatic reactions, it can be clearly observed that the protein has an influence on the overall reaction mechanism. For this reason, the protein environment effects should be included in the computational simulations, because the specific interactions they generate can modify the biochemical process. Therefore, what if we are interested in modeling very large biomolecular systems, such as proteins, nucleic acids, membranes, or other biological macromolecules, which not only include hundreds or thousands of atoms, but also involve a chemical process?

The most useful and intelligent solution lies in describing the region where the reaction takes place by means of QM methods (e.g., substrates and cofactors in an enzymatic reaction), while the MM methods are chosen to treat the surrounding amino acids of the protein environment, and/or the solvent molecules.⁷⁴ The resulting scheme of the combination of both techniques is referred to as the hybrid quantum mechanics/molecular mechanics method (QM/MM). This methodology allows large-scale biomolecular systems involving a chemical reaction to be simulated, with a reasonable computational cost and with an acceptable degree of accuracy.

The seminal contribution on this computational technique is due to Warshel and Levitt,⁷⁵ who introduced the QM/MM approach exactly forty years ago, in 1976, and applied it to the enzymatic reaction corresponding to the cleavage of a glycosidic bond by lysozyme, to form the carbonium ion intermediate. Other major contributions to the development of this methodology were presented in 1986 by Singh and Kollman,⁷⁶ and in 1990 by Field, Bash and Karplus.⁷⁷ The first authors combined the *ab initio* QM method (Hartree-Fock) with a force field to optimize a QM/MM structure, while the second work gave a detailed description of the coupling of a semi-empirical *ab initio* method with the CHARMM force field, and compared their results with *ab initio* and experimental data.

The general scheme of the partitioning of the entire system into its different subsystems is shown in Fig. 2.11. As can be observed, the system is divided into an inner subsystem (QM region), an outer subsystem (MM region), and a boundary region. The first includes all the atoms involved in the chemical reaction and is treated by the quantum mechanical methods, the second includes the remaining atoms and is described by a force field, and, finally, the boundary conditions are a set of restrictions that must be applied due to the fact that it is not possible to simulate an infinite system. In addition, it is worth noting that there is a strong interaction between the QM and MM regions, and that is the reason why it is not possible to calculate the total energy of the entire system simply as the sum of both subsystems.

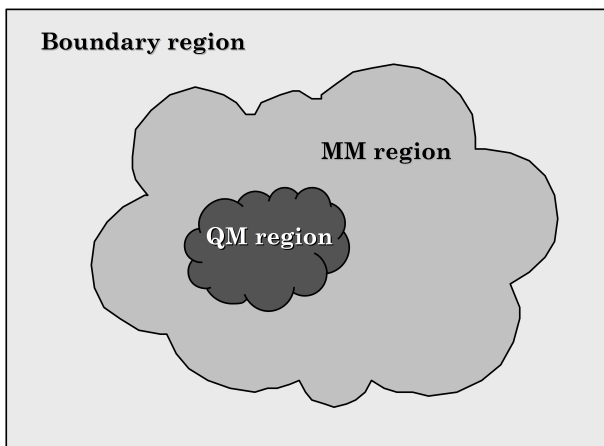


Figure 2.11.: Partitioning of the entire system into its different subsystems: QM region, MM region and the Boundary region.

The total energy of the system comes from the effective Hamiltonian (\hat{H}_{eff}), which is composed of the following terms:

$$\hat{H}_{eff} = \hat{H}_{QM} + \hat{H}_{MM} + \hat{H}_{QM/MM} \quad (2.125)$$

- The term \hat{H}_{QM} calculates the energy of the atoms included in the QM region and depends on the quantum method selected (see sec. 2.2).
- The term \hat{H}_{MM} calculates the energy of the atoms included in the MM region and depends on the force field selected (see sec. 2.3).
- The term $\hat{H}_{QM/MM}$ is the coupling or interaction Hamiltonian and accounts for the interaction between the QM and the MM subsystems. This Hamiltonian can be expressed as the sum of three contribution terms: electrostatic, Van der Waals, and polarization:

$$\hat{H}_{QM/MM} = \hat{H}_{QM/MM}^{elec} + \hat{H}_{QM/MM}^{VdW} + \hat{H}_{QM/MM}^{pol} \quad (2.126)$$

Generally, the term corresponding to the polarization between subsystems, $\hat{H}_{QM/MM}^{pol}$, is neglected in most implementations. This is because the polarization of the MM subsystem would result in having to solve the total Hamiltonian in a self-consistent way, as a result of the variation of the quantum wave function in the presence of the varying classical charges. Therefore, if we express eq. 2.126 in atomic units and dismiss the term corresponding to the polarization, the QM/MM interaction Hamiltonian takes the following form:

$$\hat{H}_{QM/MM} = - \sum_i^n \sum_m^{MM} \frac{q_m}{r_{im}} + \sum_\alpha^N \sum_m^{MM} \frac{Z_\alpha q_m}{r_{\alpha m}} + \sum_\alpha^N \sum_m^{MM} \left\{ \frac{B_{\alpha m}}{r_{\alpha m}^{12}} - \frac{C_{\alpha m}}{r_{\alpha m}^6} \right\} \quad (2.127)$$

where the subscripts i , α and m refer to electrons, nuclei and MM atoms, respectively. The first term of eq. 2.127 is the electrostatic interaction between the partial charges of the MM atoms (q_m) and the electrons of the QM region. The second term is the electrostatic interaction between the partial charges of the MM atoms (q_m) and the nuclei of the QM region. The last term of the equation corresponds to the VdW interaction between the MM and QM atoms, B and C being the parameters of a Lennard-Jones potential, which models the VdW interactions. These parameters depend only on the type of atoms involved, so that they are fully transferable from one system to another.⁷⁸ It is worth noting that the only operator present in eq. 2.127 is the first term, as it contains the coordinates of the electrons. This operator, as well as the QM electronic Hamiltonian (\hat{H}_{QM}), operates on the electronic wave function, and thus, both are treated with the SCF procedure. The other terms in the equation remain constant, for a given set of atomic coordinates, just like the MM Hamiltonian (\hat{H}_{MM}).

Once the Hamiltonian is defined it is straightforward to calculate the energy of the system as the expectation value of the effective Hamiltonian over the wave function:

$$E_{QM/MM} = \frac{\langle \Psi | \hat{H}_{eff} | \Psi \rangle}{\langle \Psi | \Psi \rangle} \quad (2.128)$$

For an MO-based method the total energy of the system is as follows:

$$E_{QM/MM} = \frac{\langle \Psi | \hat{H}_{QM} + \hat{H}_{QM/MM}^{elec} | \Psi \rangle}{\langle \Psi | \Psi \rangle} + E_{QM/MM}^{VdW} + E_{MM} \quad (2.129)$$

where it can be noticed that the parts of the energy that are independent of QM particles have been separated out from the expectation value. Finally, the forces and higher derivatives (e.g., Hessian matrix) on the QM and MM atoms can be obtained by differentiating the total energy expression (eq. 2.129) with respect to the coordinates of the QM nuclei and the MM atoms respectively.

$$F_\alpha = -\nabla_{r_\alpha} E = -\frac{\partial E}{\partial r_\alpha} \quad ; \quad F_m = -\nabla_{r_m} E = -\frac{\partial E}{\partial r_m} \quad (2.130)$$

It should be emphasized that the QM/MM scheme presented here is the one used in the present thesis but it is far from unique. The energy expression pre-

sented above follows the *additive QM/MM scheme*, which may vary depending on the exact form of the QM/MM coupling term $\hat{H}_{QM/MM}$. Furthermore, there are energy expressions that follow the *subtractive QM/MM scheme*, as is the case of the *integrated molecular-orbital/molecular-mechanics* (IMOMM) scheme developed by Maseras and Morokuma,⁷⁹ or the more sophisticated *our N-layered integrated molecular orbital and molecular mechanics* (ONIOM) method.^{80,81}

2.4.1. Frontier Bonds

As explained above, in the study of very large systems, it is often necessary to split the system between different regions, usually a QM part and an MM part. When studying reactions in biological systems, it is quite common to find some amino acids present in the active site that actively participate in the reaction (cofactors, ligands, etc.). Since the reaction process involves the formation or breaking of covalent bonds, these residues should be introduced in the QM part. Sometimes, it is also desirable to include certain amino acids in this part in order to take into account possible charge transfer phenomena from the reaction system. In these cases, there will be covalent bonds between the atoms of the QM part and the MM part, and thus we face the problem of where to cut or split between these two subsystems. To overcome this delicate step, it is necessary to resort to an approach that allows us to satisfy the valences of the atoms that define the bond. There are several methods to perform this partition, which could be categorized into the link atom formalism,^{76,77} the local self-consistent field (LSCF) formalism,⁸²⁻⁸⁶ and the generalized hybrid orbital (GHO) formalism.^{82,83,87} In this thesis, only the first one has been used in the calculations and, therefore, only the link atom will be explained.

Link Atoms

The link atom method is conceptually simple and it is the most widely used boundary scheme (see Fig. 2.12). It is based on introducing an additional atomic center that is covalently bonded to the QM atom, so that it saturates the QM atom free valence. In this way, the function of the link atom is to cap the QM subsystem, thus substituting the MM region of the frontier bond. Normally, this link atom is taken as a hydrogen atom (HQ) with one electron in a single 1s orbital, although halogens and other group types have also been used.⁸⁸ The covalent bond being cut should be non-polar and not involved in conjugative interactions, for this reason, the preferred place to cut is an aliphatic $C_{sp^3}-C_{sp^3}$ bond.⁷⁴ Since the hydrogen atom has similar electronegativity to the carbon atom, it is not expected to alter the original environment of the QM atom significantly. This link atom is treated by means of the QM methodology and is invisible to the MM atoms, because the interaction between the link atom and the atoms of the classical partition is not taken into account.

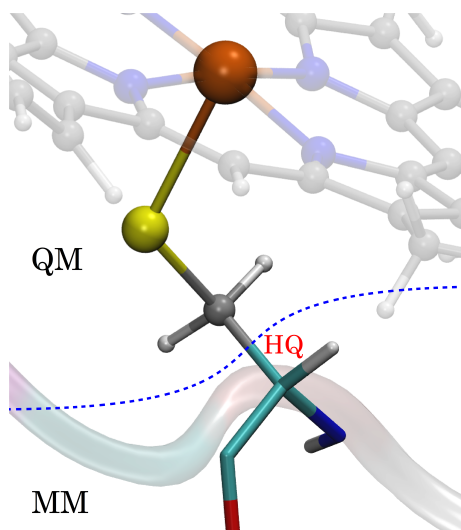


Figure 2.12.: Representation of the link atom scheme for treating frontier bonds between the QM and MM regions.

However, the presence of the link atom introduces some distortions in the electric field, i.e., over-polarization of the QM density by the close rigid point charges. To overcome this problem, several strategies have been proposed, some of which have been reviewed by Senn and Thiel.⁷⁴ The most important approaches used to reduce over-polarization are: 1) deletion of one-electron integrals associated with link atoms, 2) removal of MM point charges close to the QM region, 3) redistribution of the point MM charges in the link region, and 4) replacement of the charges close to the QM region by charge distributions (e.g., Gaussian functions).

2.5. Potential Energy Surface

As we discussed above, the concept of Potential Energy Surface emerges from the Born-Oppenheimer approximation. The energy for a given nuclear configuration ($V_i(\vec{R})$) is obtained by solving the electronic Schrödinger equation (eq. 2.13). The value of this Potential Energy for all possible nuclear configurations (q_k) is called the Potential Energy Hyper-surface (PEH). Obtaining PEH is often not straightforward due to the large number of existing dimensions that are available. In the case of a molecule composed of N atoms, either $3N$ Cartesian coordinates or $3N - 6(5)$ internal coordinates are needed to define a geometry. For example, a non-linear molecule composed of eight atoms has $(3 \times 8) - 6 = 18$ degrees of freedom. If we consider that at least 10 points per degree of freedom are needed to define the surface reasonably, we would need 10^{18} calculations to scan the PEH completely. The strategy employed to analyze this hyper-surface, in a chemical system containing

a considerable number of atoms, lies in simplifying the analysis to a few specific internal coordinates; this means reducing the PEH to a PES. The set of internal coordinates is usually chosen so that they belong to a PEH region, in which the system contains useful chemical information. For example, intramolecular hydrogen transfer requires only the internal coordinates corresponding to the distance of this hydrogen with two nuclei, the bond which is being broken and the new bond which is being formed.

The motion of the nuclei on this PES can be solved by means of classical methods (Newton mechanics), quantum methods (Schrödinger equation) or hybrid QM/MM methods. As will be seen below, from a computational point of view, many aspects of chemistry can be reduced to questions about PES.

2.5.1. Chemical structure, properties and reactivity

The PES can be viewed as a “landscape” where there are valleys, mountain passes, depressions and peaks. The exploration of the PES topology (the “landscape”) by means of computational tools provides information about molecular structure, properties, reactivity and spectra.

2.5.1.1. Characteristic Points of the PES

There are two mathematical concepts that are essential to explore the PES: the gradient, and the Hessian.

- The gradient (\vec{g}) is a vector, or column matrix, composed of the first derivatives of the potential energy with respect to the nuclear coordinates.

$$\vec{g} = \begin{pmatrix} \frac{\partial V}{\partial x_1} \\ \frac{\partial V}{\partial y_1} \\ \vdots \\ \frac{\partial V}{\partial z_N} \end{pmatrix}_{3Nx1} \quad (2.131)$$

As can be noted, there are $3N$ Cartesian coordinates or degrees of freedom, where N is the number of atoms.

- The Hessian (or force constants) matrix (H) is a symmetric matrix containing as its elements the second derivatives of the potential energy with respect to the nuclear coordinates.

$$H = \begin{pmatrix} \frac{\partial^2 \vec{q}}{\partial x_1^2} \\ \frac{\partial^2 \vec{q}}{\partial y_1^2} \\ \vdots \\ \frac{\partial^2 \vec{q}}{\partial z_N^2} \end{pmatrix} \equiv \begin{pmatrix} \frac{\partial^2 V}{\partial x_1^2} & \frac{\partial^2 V}{\partial x_1 \partial y_1} & \cdots & \frac{\partial^2 V}{\partial x_1 \partial z_N} \\ \frac{\partial^2 V}{\partial y_1 \partial x_1} & \frac{\partial^2 V}{\partial y_1^2} & \cdots & \frac{\partial^2 V}{\partial y_1 \partial z_N} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 V}{\partial z_N \partial x_1} & \frac{\partial^2 V}{\partial z_N \partial y_1} & \cdots & \frac{\partial^2 V}{\partial z_N^2} \end{pmatrix}_{3N \times 3N} \quad (2.132)$$

The H matrix contains topological information about the system, although that is often hard to interpret. For this reason, we diagonalize this matrix in order to facilitate the management of the information it contains. The diagonalization of H is carried out by solving the following eigenvalue equation:

$$(H - \lambda_i I) \vec{u}_i = 0 \quad (2.133)$$

The eigenvalues (λ_i) obtained in eq. 2.133 allow us to characterize the nature of the geometry associated with given nuclear coordinates, i.e., to characterize the stationary points of the PES. Likewise, the associated eigenvectors (\vec{u}_i) define the principal curvature directions and are related with the vibrational normal modes.

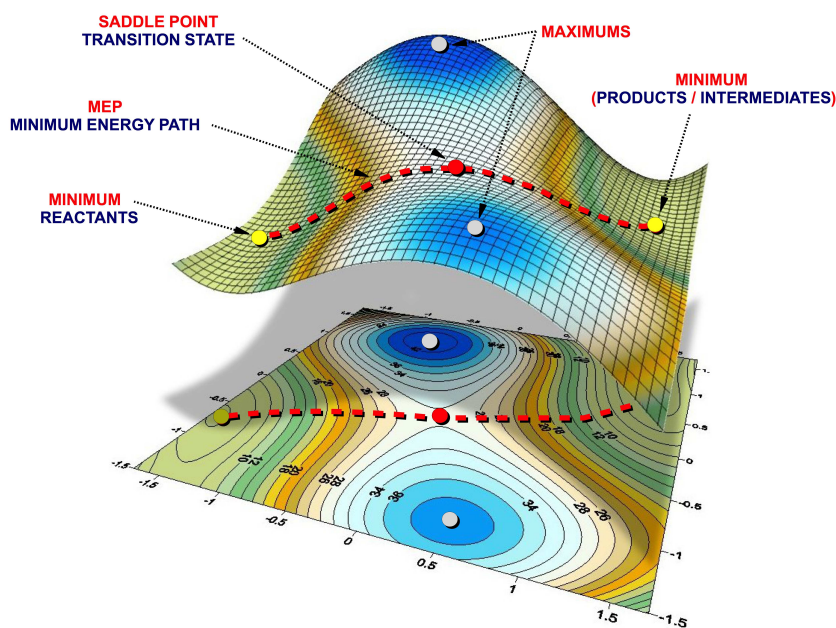


Figure 2.13.: Potential Energy Surface (PES) where the main stationary structures have been depicted.

Stationary points of the PES

The stationary (or critical) points of the PES (see Fig. 2.13) are characterized by zero first derivatives with respect to the nuclear coordinates, in other words, when the gradient is equal to zero ($\vec{g} = 0$). However, depending on the curvature of the stationary points (provided by the Hessian matrix), the corresponding structures can be classified as:

- **Minimum**

This stationary point is located when all the eigenvalues obtained from eq. 2.133 have positive values. This means that an infinitesimal displacement along the eigenvector of any of its eigenvalues will increase the energy. Such structures are the reactants, products or reaction intermediates; however, each one may present many different conformations. Each of these conformations is minimal because they are at the bottom of a valley on the *PES*, and correspond to a local minimum. The global minimum is the lowest energy point on the *PES*.

- **Saddle point**

When H is diagonalized and provides eigenvalues with both positive and negative values, an n -order saddle point has been located on the PES, n being the number of negative eigenvalues. If $n = 1$, it is a first-order saddle point, and thus characterizes a transition state or transition structure (TS) of the reaction. This point is a maximum in one direction and a minimum in all other directions. Topologically speaking, the saddle point is the highest point between two valleys and the lowest point along a ridge. This point takes its name from its shape, which is similar to that of a saddle. The structure associated with the TS will exhibit one imaginary frequency and its corresponding normal mode of vibration should imitate the motion of the atoms along the reaction coordinate. The eigenvector associated to this single negative eigenvalue is called the transition vector (TV), and also determines the direction and symmetry of the reaction path (RP), at the point of maximum energy of passage through the region of the TS, which connects the two local minima.

- **Maximum**

This stationary point is located when all the eigenvalues obtained from eq. 2.133 have negative values. This means that an infinitesimal displacement along the eigenvector of any of its eigenvalues will decrease the energy. These stationary points on the PES do not provide any relevant chemical information.

In addition to the previously mentioned stationary points, there are other interesting concepts concerning the Potential Energy Surface, which are related to the chemical reactivity:

- **Reaction Coordinate**

In chemistry, a reaction coordinate is an arbitrarily chosen parameter that describes the evolution of the chemical reaction from reactants to products. A *distinguished reaction coordinate* (ξ_i) is usually used, where only those geometric parameters that change substantially during the chemical conversion are considered, such as a bond length, a bond angle, etc., or a combination of some of them. However, it is sometimes approximated by a non-geometric parameter, such as the bond order of some specified bond. Accordingly, the concept of reaction coordinate is intuitive and not too accurate, but is widely used to represent the course followed by a chemical reaction.

- **Reaction Path**

The Reaction Path can be defined as the path that connects the different chemical species formed during the evolution of the chemical process. The simplest definition is a line in coordinate space that connects two minima, passing through the saddle point, along the minimum energy path; for this reason it is also referred to as *minimum energy path* (MEP).⁸⁹ However, the RP is calculated from the surface, whose topology is dependent on the coordinates chosen as variables, thus, it is noteworthy that the reaction path is not unique. For this reason, a more accurate definition of the RP of an adiabatic PES is the set of trajectories that an ensemble of molecular entities take through the PES, at a given temperature, until the reaction is completed.⁹⁰

- **Intrinsic Reaction Coordinate**

Intrinsic reaction coordinate (IRC) was proposed by Fukui^{91,92} in 1970. IRC represents the path covered by a particle when it descends slowly from the saddle point, following the direction of the gradient vector to both sides in the direction of the two energy minima, thus connecting reactants with products. Because the initial point, the TS, has zero gradient, the initial direction of the IRC must be specified in advance and is given by the direction of the TV. Mathematically, the IRC is obtained in a first approach (Euler) as a solution to the following system of differential equations:

$$\frac{d\vec{x}}{ds} = \frac{-\vec{g}}{|\vec{g}|} \quad (2.134)$$

where s is the reaction path length measured from the transition vector or simply the reaction coordinate. s is calculated using the expression:

$$(ds)^2 = \sum_i (dx_i)^2 + (dy_i)^2 + (dz_i)^2 \quad (2.135)$$

where the subscript i refers to each atom and s takes positive or negative values depending on whether the displacement is towards products or reactants,

respectively. The problem is that the reaction path obtained from eq. 2.134 depends on the coordinate system chosen. Fukui demonstrated^{93,94} that in the special case of using the mass-weighted Cartesian coordinates:

$$q_i = x_i \sqrt{m_i} \quad (2.136)$$

the requirement of following the gradient coincides with the path that the system would follow if, starting from the TS, it was directed to reactants and products always following a zero kinetic energy trajectory, i.e., an infinitely slow trajectory. For this reason, the IRC is understood as the reaction path resulting from solving the system of differential equations of eq. 2.134, expressed in terms of mass-weighted Cartesian coordinates.⁹⁴

The IRC is calculated using the classical equations of motion. For a given PES, the movement of a classical particle must be the same regardless of the coordinate system chosen, i.e., either Cartesian coordinates or a set of internal coordinates. In this way, the trajectories can be obtained independently of the coordinate system used. We could make an analogy of the IRC with the fact of going down a slide.

2.5.1.2. Methods to locate stationary points on the PES

Locating Minima

The location of a minimum on a PES is a complex mathematical problem due to the multidimensional nature of the PES. There are a large number of standard methods available for performing such a task, but the choice of the optimization method depends on several factors, such as the nature of the function to be minimized, whether derivatives are available (and their order), and the number of variables. A classification of the optimization algorithms by the type of derivatives that they use is as follows:

- No derivatives

These algorithms, also known as zero-order methods, make use of the function values only (in our case, the energy). The most common algorithm in this group is the *simplex*⁹⁵ method. Simplex and other non-gradient methods are only used in desperate cases when gradient information is not available and when the function is very noisy or discontinuous.

- First derivatives

These algorithms, also known as first-order methods, make use of the energy and its first derivatives (the gradient). The most common algorithms in this group are the *steepest (or gradient) descent*⁹⁶ (SD) and the *conjugate gradient*⁹⁷ (CG). The SD is the simplest iterative algorithm based on the gradient

vector. It is highly appropriate when the system is far from the minimum (i.e., at the beginning of the system setup). In the vicinity of the minimum it is not very useful because it has a tendency to “make detours”, so that they fall into loops without reaching optimization. The CG is similar to SD, but with the difference that it makes use of the gradient history (information about the steps taken previously) to decide on a better direction for the next step. This method does not exhibit a zig-zagging behavior during convergence, but still tends to be quite slow in very non-linear cases. Furthermore, the CG algorithm usually performs a “line search” which allows an adjustable step length, thereby accelerating convergence. Some algorithms associated to the CG method are *Fletcher-Reeves*⁹⁸ (FR) or *Polak-Ribiere*⁹⁹ (PR).

- First and exact second derivatives

These algorithms, also known as second-order or quadratic methods, use energy values and both their first (gradient) and second (Hessian) derivatives. The most common algorithms in this group are *exact Newton* and *Newton-Raphson*^{100,101} (NR) methods. NR method approximates the objective function by a quadratic surface at each step and moves to the minimum of that surface. The advantage of Newton methods is that the convergence is second-order near a stationary point; its problem lies in the computation of the Hessian matrix, which is very expensive in electronic structure calculations.

- First and approximate second derivatives

These algorithms, also known as pseudo second-order or quasi-Newton methods, use energy values and the exact first derivatives of the energy. They do not compute the second-derivative matrix (Hessian matrix), but they use approximations thereof. In these methods, the Hessian matrix is usually extracted from the gradient history, i.e., it is updated from the gradient vector. Some popular methods belonging to this group are: *DFP update*¹⁰² (Davidon-Fletcher-Powell), *BFGS update*¹⁰³⁻¹⁰⁶ (Broyden-Fletcher-Goldfarb-Shanno), and *Limited-memory BFGS*¹⁰⁷ (L-BFGS).

Locating Saddle points

Locating transition structures is usually quite a lot more difficult than locating minima, since we are trying to find a point that is a maximum in one direction but a minimum in all the others $3N - 7(6)$. As a result, the algorithm has to perform a delicate balancing act between the two conflicting types of search. The algorithms for the location of saddle points have been studied much less widely than the algorithms for locating minima. In fact, many of these algorithms have been developed by chemists interested in looking at potential energy surfaces and not by mathematicians. One of the most efficient methods available and the most widely

used is the mode-following algorithm developed by J. Baker.^{108,109} This method makes use of both first- and second-derivative information in the diagonal space and so its application is restricted to relatively small systems.

2.5.2. Kinetic Isotope Effects

2.5.2.1. An Introduction to Transition State Theory

After the development of the PES concept, several quantitative theories of reaction rates were developed based on the idea of the critical role of the TS in controlling the rate. The most remarkable theory thereon was developed simultaneously in 1935 by Henry Eyring,¹¹⁰ and by Meredith Gwynne Evans and Michael Polanyi.¹¹¹ In this theory, the chemical species present in the TS region of the reaction coordinate was named the activated complex, which is why this theory is sometimes known as the activated complex theory. In an alternative terminology, TS is interpreted to mean both the region at the maximum energy on the reaction coordinate and the species inhabiting this region; thus we have the transition state theory (TST). TST has also been called the absolute reaction rate theory.¹¹²

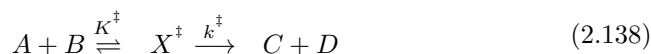
The TST is based on the following three postulates:¹¹²

1. The reacting system, in passing from the initial state (reactants) to the final state (products) along the PES, must go through a region of the reaction path, known as “transition state”, whose potential energy is the highest energy of the path.
2. The chemical species in the transition state are in equilibrium with the reactant state.
3. The rate of the reaction is equal to the product of the concentration of transition state species formed from the reactant state and the frequency with which this species passes on to the product state.

Consider the following bimolecular chemical reaction:



One of the most important features introduced by Eyring, Polanyi and Evans was the notion that activated complexes are in quasi-equilibrium with the reactants. This equilibrium is different from the classical chemical equilibrium, but can be described using the same thermodynamic treatment.



where X^\ddagger is the activated complex of relatively high energy.

This complex should not be thought of as an unstable isolatable intermediate because it is assumed to always be in the process of decomposing. The equilibrium constant may be written as:

$$K^\ddagger = \frac{[X^\ddagger]}{[A][B]} \quad (2.139)$$

and, thus, the concentration of the activated complex is:

$$[X^\ddagger] = K^\ddagger [A][B] \quad (2.140)$$

The rate of the reaction may be expressed as follows:

$$v_{reaction} = k^\ddagger [X^\ddagger] \quad (2.141)$$

where $[X^\ddagger]$ is the concentration of the activated complex at the top of the barrier. The rate constant, k^\ddagger , for the activated complex converting to products is related to a vibration in the activated complex:

$$v_{reaction} = k^\ddagger [X^\ddagger] = \nu [X^\ddagger] \quad (2.142)$$

where ν is the frequency of crossing the activation energy barrier, i.e., the frequency of the vibrational mode (ν) responsible for converting the activated complex to the product (third postulate). Hence, the reaction rate may also be written as:

$$v_{reaction} = \nu K^\ddagger [A][B] \quad (2.143)$$

but it can also be written as:

$$v_{reaction} = k [A][B] \quad (2.144)$$

From equations 2.143 and 2.144 we obtain that:

$$k = \nu K^\ddagger \quad (2.145)$$

Considering that ν is the frequency of the vibrational mode associated to the movements of the activated complex along the reaction coordinate, it can be derived that:

$$\nu = \frac{k_B T}{h} \quad (2.146)$$

where k_B is the Boltzmann constant, h is Planck's constant, and T is the temperature. Thus, substituting eq. 2.146 in eq. 2.145:

$$k = \frac{k_B T}{h} K^\ddagger \quad \left(M^{1-m} s^{-1} \right) \quad (2.147)$$

where M is the molarity, m is the molecularity of the reaction and s are seconds.

Thermodynamic formulation of the TST

The standard molar Gibbs Energy of activation is given by:

$$\Delta G^\ddagger = G(X^\ddagger) - G(\text{reactants}) = -RT \ln K^\ddagger \quad (2.148)$$

Thus,

$$K^\ddagger = e^{\left(\frac{-\Delta G^\ddagger}{RT} \right)} \quad (2.149)$$

Substituting eq. 2.149 in eq. 2.147, the rate constant may be expressed as follows:

$$k = \frac{k_B T}{h} e^{\left(\frac{-\Delta G^\ddagger}{RT} \right)} \quad (2.150)$$

This is the so-called Eyring equation, which allows us to estimate the reaction rate if we are able to determine the free energy between reactants and the transition state, ΔG^\ddagger . In addition, knowing the relationship of the Gibbs energy with the enthalpy and the entropy :

$$\Delta G^\ddagger = \Delta H^\ddagger - T\Delta S^\ddagger \quad (2.151)$$

The reaction rate can be rewritten as:

$$k = \frac{k_B T}{h} e^{\left(\frac{-\Delta H^\ddagger}{RT} \right)} e^{\left(\frac{-\Delta S^\ddagger}{R} \right)} \quad (2.152)$$

2.5.2.2. Kinetic Isotope Effects

One of the most widely used methods for the determination of molecular mechanisms in chemical reactions is the study of kinetic isotope effects (KIE). This method is based on the study of the differences in the reaction rate constant when one or several atoms involved in the chemical reaction are replaced by one of their corresponding isotopes. When an isotopic substitution is carried out, e.g., substitution of hydrogen atoms by deuterium atoms, this effect does not alter the chemical reaction but rather modifies the mass of the molecules and their vibrational frequencies. In other words, the PES of the chemical system is independent of the isotope substitution, although a change in the reaction rate occurs. Therefore,

the KIEs may be used to compare theoretical results with the experimental data available on reaction rate or equilibrium constants, thereby making it possible to validate the various mechanisms proposed and the methodology used. In addition, KIEs provide useful information about the TS of the rate limiting step of the chemical process. The absence of KIE would demonstrate that an isotopically substituted atom is not involved in the rate-limiting step, or that the changes are offset, giving rise to a KIE with a value equal to unity.

According to the TST discussed in the subsection above, the reaction rate can be defined with eq. 2.150, where K^\ddagger is the equilibrium constant between reactants and transition state. Assuming a reaction where reactants (R) react to give rise to products (P) through a transition state (TS), we can define the equilibrium constant depending on the partition functions as:

$$K^\ddagger = \frac{[TS]}{[R]} = \frac{Q_{TS}}{Q_R} \exp\left(\frac{-\Delta E^\ddagger}{k_B T}\right) \quad (2.153)$$

where ΔE^\ddagger is the potential energy barrier between R and TS, and Q_{TS} and Q_R are the partition functions of TS and products, respectively. The partition function has the following functional form:

$$Q = \sum_i^{\text{microstates}} \exp\left(-\frac{\varepsilon_i}{k_B T}\right) \quad (2.154)$$

In the case of an ideal gas, the partition function can be expressed as the product of the independent contributions of the translational (Q_t), rotational (Q_r), vibrational (Q_v) and electronic (Q_e) energies:

$$Q = Q_t Q_r Q_v Q_e \quad (2.155)$$

where:

$$Q_t = \frac{V}{h^3} (2\pi M k_B T)^{\frac{3}{2}} \quad (2.156)$$

$$Q_r = \frac{8\pi^2}{\sigma h^3} (2\pi k_B T)^{\frac{3}{2}} (I_a I_b I_c)^{\frac{1}{2}} \quad (2.157)$$

$$Q_v = \prod_i^{N_v} \frac{e^{-\frac{h\nu_i}{2k_B T}}}{\left(1 - e^{-\frac{h\nu_i}{k_B T}}\right)} \quad (2.158)$$

$$Q_e = \sum_j^{\text{levels}} g_j \exp\left(-\frac{E_j}{k_B T}\right) \quad (2.159)$$

In equations 2.154 to 2.159, ε_i represents the molecular energetic level, V is the volume of the system, M is the molecular mass of the system, and σ corresponds to a symmetry factor which characterizes the type of inertia momentum with respect to the principal axes ($I_{a,b,c}$), as a function of the molecular geometry. Notice that the rotational partition function has been considered for the case of a non-linear molecule. N_v is the number of vibrational normal modes (which will be $3N - 6$ for reactants and $3N - 7$ for the TS, due to the fact that the frequency associated to the reaction coordinate is imaginary and, thus, it is omitted), ν_i are the values of the frequencies associated to the vibrational normal modes of the molecule, E_j are the different electronic energy levels, and g_j is the degeneracy of the energy level j .

According to the last equations, we can write the KIE as a ratio between the partition functions of the TS and of the reactants. Assuming that the isotopic substitution involves the replacement of a hydrogen atom by a deuterium atom, the mathematical expression is given by the following expression:

$$\frac{k_H}{k_D} = \frac{K_H^\ddagger}{K_D^\ddagger} = \left(\frac{Q_{H,t} Q_{H,r} Q_{H,v}}{Q_{D,t} Q_{D,r} Q_{D,v}}\right)_{TS} \left(\frac{Q_{D,t} Q_{D,r} Q_{D,v}}{Q_{H,t} Q_{H,r} Q_{H,v}}\right)_R \quad (2.160)$$

As can be noted, the term $(-\Delta E^\ddagger/k_B T)$ of eq. 2.153 is canceled out, because when the Born-Oppenheimer approximation is applied, ΔE^\ddagger does not depend on the mass of the molecule. In the same way, the electronic term (Q_e) is also canceled out under this approximation, since $Q_{H,e} = Q_{D,e}$. In this way, the quotient of the contributions of the partition function, either for R or TS, is as follows:

$$\begin{aligned} & \left(\frac{Q_{H,t} Q_{H,r} Q_{H,v}}{Q_{D,t} Q_{D,r} Q_{D,v}}\right)_{R/TS} = \\ & = \underbrace{\left(\frac{M_H}{M_D}\right)^{3/2} \left(\frac{I_{H,a} I_{H,b} I_{H,c}}{I_{D,a} I_{D,b} I_{D,c}}\right)^{1/2}}_{MMI} \underbrace{\prod_i^{N_v} \frac{e^{-\frac{h\nu_{i,H}}{2k_B T}}}{e^{-\frac{h\nu_{i,D}}{2k_B T}}}}_{ZPE} \underbrace{\left(\frac{1 - e^{-\frac{h\nu_{i,D}}{k_B T}}}{1 - e^{-\frac{h\nu_{i,H}}{k_B T}}}\right)}_{EXC} \end{aligned} \quad (2.161)$$

where the factor MMI is a function of the *masses and moments of inertia*, ZPE is associated to the differences in the *zero-point vibrational energy* levels, and

the term EXC takes into account the fact that an isotopic substitution affects the population of *excited vibrational energy* levels. In this way, the KIE can be expressed by the product of these three terms:

$$KIE = \frac{k_H}{k_D} = MMI \cdot ZPE \cdot EXC \quad (2.162)$$

which is the Bigeleisen^{113,114} equation and where:

$$MMI = \frac{MMI_{TS}}{MMI_R} \quad ZPE = \frac{ZPE_{TS}}{ZPE_R} \quad EXC = \frac{EXC_{TS}}{EXC_R} \quad (2.163)$$

In practice, the product of the MMI and EXC terms is usually close to unity, hence, the KIE value is determined by the ZPE term (see Fig. 2.14). This is due to the fact that, for medium-sized molecules, the changes in the total mass and in the components of inertia momentum are usually small. The ZPE factor depends on the frequencies associated to the vibrational modes of the molecule, which are defined in the harmonic-oscillator approximation as:

$$v = \frac{1}{2\pi} \sqrt{\frac{k'}{\mu}} \quad (2.164)$$

where k' is the force constant of the harmonic oscillator and μ is the reduced mass associated to the normal mode. Therefore, we can generalize that, the bigger the mass is, the lower the frequency will be, which means that the vibrational partition constant is also lower.

The KIEs can be classified as *normal* or *inverse*. When the TS is more relaxed than the R, then $\Delta ZPE_{light}^\ddagger < \Delta ZPE_{heavy}^\ddagger$ and eq. 2.160 shows a value higher than unity, $k_{light}/k_{heavy} > 1$, which is known as *normal* KIE. In this case, the reaction with the lightest isotope will be faster. Conversely, if the R is more relaxed than the TS, then $\Delta ZPE_{light}^\ddagger > \Delta ZPE_{heavy}^\ddagger$ and eq. 2.160 shows a value lower than unity, $k_{light}/k_{heavy} < 1$, which is known as *inverse* KIE. In this case, the reaction with the heaviest isotope is favored. In the example represented in Fig. 2.14, $\Delta ZPE_H^\ddagger < \Delta ZPE_D^\ddagger$, thus, there is a normal KIE ($k_H/k_D > 1$).

The KIEs can also be classified as primary KIEs, when a bond to an isotopically labeled atom is broken or formed during the reaction, or secondary KIEs, when no bond fission to the labeled atom takes place (the isotopic substitution is remote from the bond which is being broken or formed). Generally the latter are smaller than the primary KIEs.

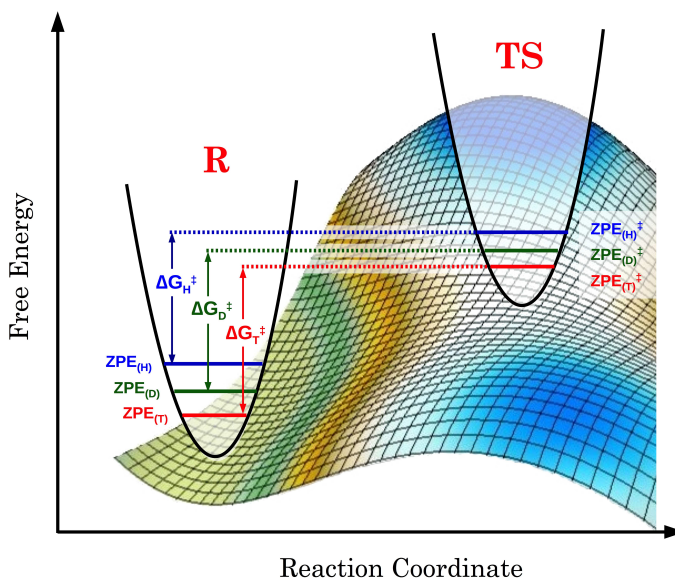


Figure 2.14.: Different free energies of activation (ΔG^\ddagger) for protium, deuterium and tritium, resulting from their different zero-point energy (ZPE) at the ground state (R) and at the transition state (TS).

2.5.2.3. Transition State Theory Limitations

As mentioned in section sec.2.5.2.1, TST is based on three postulates. Unfortunately, the three of them have been proven to be non-realistic, since they are not able to interpret some phenomena. In order to overcome these limitations a new theory arises, which is known as the generalized transition state theory (GTST).^{115–119} GTST provides a framework for understanding chemical reactions and allows rate constants to be obtained which can be compared with the experimental ones. With this purpose, the *generalized transmission factor*, $\gamma(T)$, is included, which relates the actual rate of the reaction to that obtained from simple TST (eq. 2.150).

$$k = \gamma(T) \frac{k_B T}{h} e^{\left(\frac{-\Delta G^\ddagger}{RT}\right)} \quad (2.165)$$

This *transmission factor* is defined as:¹²⁰

$$\gamma(T) = \Gamma(T) \cdot \kappa(T) \cdot g(T) \quad (2.166)$$

where $\Gamma(T)$ is the *quasi-classical transmission factor* and arises from classical dynamical recrossing; its value is less than or equal to 1. $\kappa(T)$ is the *semi-classical transmission factor* and arises from the contribution of quantum mechanical tun-

neling; its value is almost always greater than or equal to 1. Finally, $g(T)$ arises from deviations of the equilibrium distribution in the phase space; its value can be either less than or greater than 1. Notice that in conventional TST $\gamma(T)$ is set as equal to one.

The two most important contributions affecting conventional TST, the tunneling effect and recrossings, are commented on below. However, it should be noted that only the contribution of tunneling has been studied in this thesis.

- **Tunneling Effect**

In the TST, the movement along the reaction coordinate has been treated classically. However, for a more adequate treatment, it is necessary to consider the quantum description of this movement, which leads us to consider the tunneling effect. Indeed, ignoring this effect in reactions involving the movement of very light atoms (such as the hydrogen atom) leads to an underestimation of the rate constant when the TST is used. This is because the classical treatment of this theory does not take into account situations that, even without having enough energy to pass over the energy barrier, give rise to a reaction. That is, trajectories that, according to the TST, have zero probability of reaching products from reactants, because they do not have sufficiently high energy, have a non-zero probability of reaching the products, from the quantum point of view (see Fig. 2.15a). Experimentally, the tunneling effect becomes evident when observing larger KIEs than those obtained with the TST. This is because deuterium is heavier than hydrogen, and thus it suffers less tunneling effect. Hence, the rate constant also decreases because of this effect and not only due to the increase in the difference of fundamental energies.

- **Recrossings**

Another important limitation of TST arises from the fact of assuming that all trajectories that reach the dividing surface are reactive. The dividing surface is a surface (or hypersurface) defined in the space of the nuclear configurations, which includes the saddle point and divides the reaction PES into two zones: one corresponding to reactants and the other corresponding to products. Studies on reaction dynamics have shown that recrossings sometimes occur, i.e., there are classical trajectories that, once crossing the dividing surface from the reactants' valley towards that of the products, recross the dividing surface by turning towards reactants (see Fig. 2.15b). Usually this type of trajectories can amount to around 10% of the total, although in some cases they can reach more than 50%, especially for reactions in solution. Ignoring recrossings leads to overestimation of the rate constant when the TST is used. This defect may be corrected by considering the *transmission coefficient* discussed above.

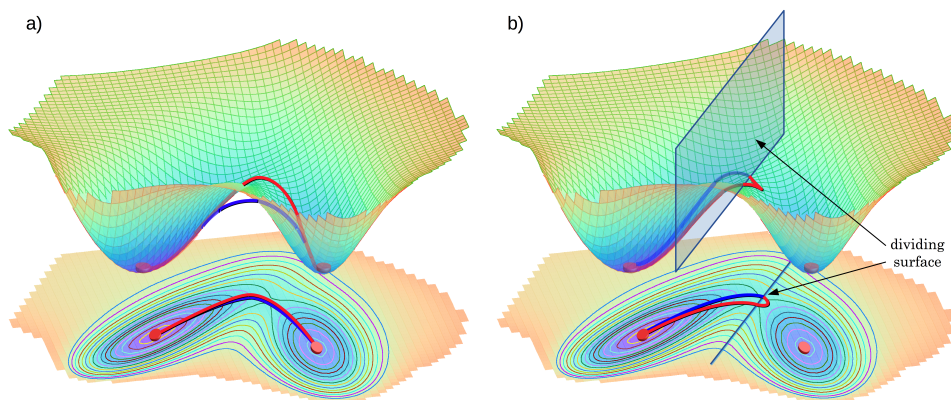


Figure 2.15.: (a) Minimum Energy Path from Reactants to Products on a three-dimensional PES (red) and tunneling through the same PES (blue). (b) A classical trajectory recrossing the dividing surface and turning towards reactants.

2.5.2.4. Semi-Classical Tunneling Effect Corrections

The most straightforward way to compute the KIEs is by using a semi-classical approximation. The first QM “correction” to the classical understanding of the hydrogen transfer arises from the differences in vibrational frequencies for isotopically labeled bonds. According to the resulting semi-classical theory for KIEs, ground-state vibrational differences (ZPE) among H, D and T establish the maximum kinetic difference in rate among the isotopic bonds, which in the case of k_H/k_D for C–H abstraction is ca. 7.¹²¹ In this way, values of $k_H/k_D > 7$ exceed the semi-classical Eyring limit and suggest an amount of tunneling effect.

Accurate prediction of transmission coefficients including many degrees of freedom is a very difficult QM problem. A straightforward approximation is to consider tunneling only in the degree of freedom corresponding to the reaction coordinate, i.e., to reduce the problem to a one-dimensional formalism.¹²² Different approaches have been developed to calculate the transmission tunneling coefficient, including classical and semi-classical approximations. Two of the most widely used semi-classical tunneling effect corrections are:

- **Wigner correction**

In 1932 Wigner developed an early formulation for incorporating the effect of tunneling.¹²³ The Wigner correction for tunneling assumes a parabolic potential for the nuclear motion near the transition state,¹²⁴ and corrects the rate constant with a multiplicative factor:

$$\kappa(T) = 1 + \frac{1}{24} \left(\frac{h\nu^\ddagger}{k_B T} \right)^2 \quad (2.167)$$

where v^\ddagger is the imaginary frequency of the asymmetric stretch along the reaction coordinate, k_B is the Boltzmann constant, h is Planck's constant, and T is the absolute temperature. This approximation works well as long as $hv^\ddagger \ll k_B T$.

- **Bell correction**

It is well known that quantum-mechanical tunneling participates in chemical reactions at very low temperatures, i.e., at temperatures close to absolute zero. Bell, in his classic text, "The Tunnel Effect in Chemistry",¹²⁵ emphasized the fact that tunneling may also be expected to participate in a reaction near room temperature. The transmission coefficient calculated according to the Bell methodology is:

$$\kappa(T) = \frac{hv^\ddagger/2k_B T}{\sin(hv^\ddagger/2k_B T)} \quad (2.168)$$

where, as in the Wigner correction, v^\ddagger is the imaginary frequency of the asymmetric stretch along the reaction coordinate, k_B is the Boltzmann constant, h is Planck's constant, and T is the absolute temperature. It is worth mentioning that the value of $\kappa(T)$ inflates with the value of v^\ddagger , such that when $v^\ddagger = 1000i$, then $\kappa(T) > 2$ at room temperature.

2.5.3. Molecular Dynamics

When we seek to study chemical reactions in condensed media, the PES provides very useful information on the system; however, this is not enough to describe the entire system in a representative manner. This is because, unlike in the gas-phase, in condensed media there is a large number of degrees of freedom, owing to the effect of the environment. In this way, each of the stationary points (minima or saddle point) will be composed of many conformations, which differ in small geometric and energetic changes (see Fig. 2.16a and b). In Fig. 2.16c, an example of the energy profile of a reaction in condensed media has been depicted. As can be seen, if we focus on just one reactant structure and one transition structure, we will obtain only one energetic barrier ΔE_1^\ddagger . Conversely, if we select other structures for reactants and transition structure, the energy barrier might be different (ΔE_2^\ddagger), even though the latter structures provide virtually the same chemical information as the former. Moreover, as observed in Fig. 2.16c, the same argument can be used for reaction energies (ΔE_1 and ΔE_2).

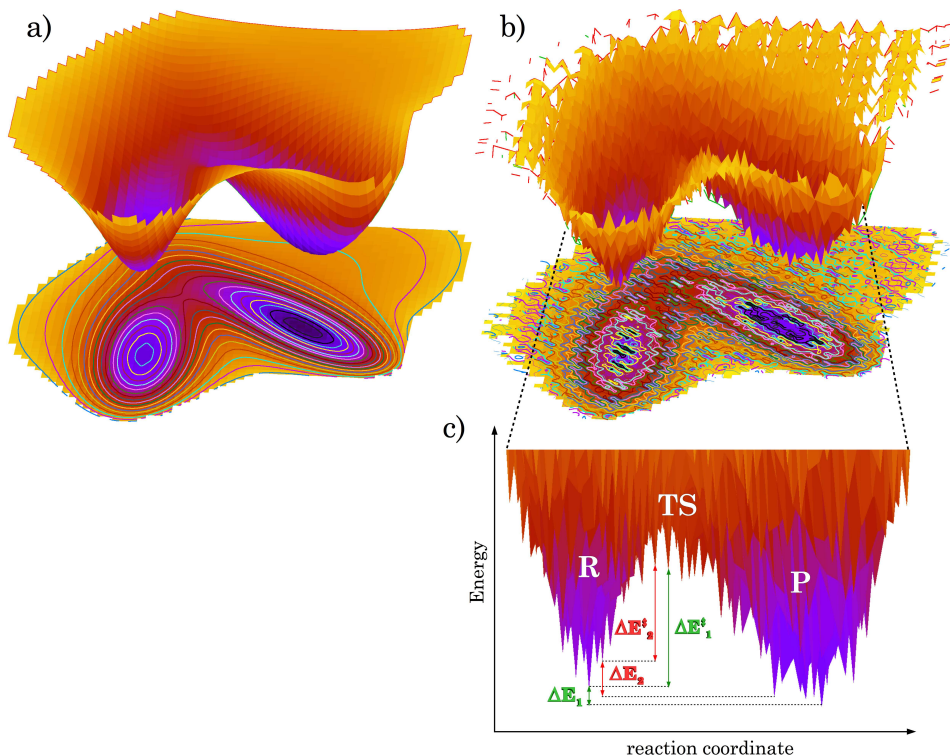


Figure 2.16.: Potential Energy Surface in the gas-phase (a) and condensed media (b), and schematic energy profile of a reaction in a condensed media (c).

For this reason, to calculate the properties of the system, it is necessary to explore the ensembles composed of the multiple conformations associated to the different stationary points (reactants, transition structures, intermediates and products). In this way, both the kinetic and the thermodynamic properties of the system are obtained as an average value over all the possible conformations, so that they can be compared with the experimental values. There are different ways to evaluate the contribution of the different configurations of the system, although the method used in the present thesis is based on the study of dynamics of the atomic nuclei that form the system, and it is known as *Molecular Dynamics* (MD).

MD is a computer simulation method which tries to solve the classical equations of motion (Newton equations) for the N particles forming a system. These equations generate sets of coordinates and velocities at different times (microstates), from which any macroscopic mechanical property can be calculated, mainly potential and kinetic energies, by performing a time average over the microscopic properties:¹²⁶

$$\bar{X} = \langle X \rangle_t \quad (2.169)$$

This method was developed following the success previously reached by another simulation method, the Monte Carlo method,¹²⁷ which makes use of stochastic methods. However, unlike the latter, the MD allows the calculation of time-dependent properties (non-equilibrium properties), since the sequence of coordinates is not obtained randomly, but from the equations of motion. The first MD simulation was performed by Alder and Wainwright in 1957,¹²⁸ who carried out a simulation of perfectly elastic collisions on a system formed by hard spheres. The first MD on a realistic system was performed in 1964 by A. Rahman,¹²⁹ who studied the interaction of liquid argon molecules with a Lennard-Jones potential.

To derive the equations of motion, we take as a starting point the $6N$ -dimensional classical Hamiltonian of the system, which is a function of the $3N$ linear momenta and the $3N$ particle positions. This Hamiltonian can be expressed as the sum of the kinetic (\mathcal{K}) and potential (V) energies:

$$\hat{H}(\vec{r}_i, \vec{p}_i) = \sum_{i=1}^N \mathcal{K}(\vec{p}_i) + V(\vec{r}_i) = \sum_{i=1}^N \frac{1}{2m_i} \vec{p}_i^2 + V(\vec{r}_i) \quad (2.170)$$

where \vec{p}_i is the linear momentum of the particle i_{th} , \vec{r}_i is its position vector, m_i is its mass and V is the effective potential (which is computed by means of QM methods, MM methods or QM/MM methods). On the other hand, the time evolution of the system is uniquely defined by Hamilton's equations:

$$\frac{\partial \hat{H}}{\partial \vec{r}_i} = -\frac{d\vec{p}_i}{dt} = \frac{\partial V}{\partial \vec{r}_i} = -\vec{F}_i \quad (2.171)$$

$$\frac{\partial \hat{H}}{\partial \vec{p}_i} = \frac{d\vec{r}_i}{dt} = \frac{\vec{p}_i}{m_i} = \vec{v}_i \quad (2.172)$$

The substitution of eq. 2.172 in eq. 2.171 leads to Newton's equation, which can describe the nuclear motion of our system:

$$\vec{F}_i(t) = -\nabla_i V(\vec{r}_i) = m_i \frac{d^2 \vec{r}_i(t)}{dt^2} = m_i \vec{a}_i(t) \quad (2.173)$$

Therefore, to perform an MD simulation or to study the dynamics of the system, the equations of motion (either equations 2.171 and 2.172 or eq. 2.173) must be solved for each particle of the system. Once the force acting on a nucleus ($\vec{F}_i(t)$) is known, as well as its position ($\vec{r}_i(t)$) and its velocity ($\vec{v}_i(t)$) in the previous instant, it is possible to calculate the new position it will have after a finite time increment (δt). There are several algorithms that can be used to integrate these equations, but in this thesis we have focused on the *Verlet* methods. The standard Verlet method¹³⁰ considers that if the positions of the atoms in the system at time t are $\vec{r}_i(t)$, then the position of the atoms at time $t + \delta t$ can be calculated from a Taylor

expansion in terms of the time-step, δt , and the positions and their derivatives at time t :

$$\vec{r}_i(t + \delta t) = \vec{r}_i(t) + \delta t \frac{d\vec{r}_i}{dt} + \frac{\delta t^2}{2} \frac{d^2 \vec{r}_i}{dt^2} + \dots \simeq \vec{r}_i(t) + \delta t \vec{v}_i(t) + \frac{\delta t^2}{2} \vec{a}_i(t) \quad (2.174)$$

Likewise, the positions at time $t - \delta t$ are obtained from the expansion:

$$\vec{r}_i(t - \delta t) = \vec{r}_i(t) - \delta t \frac{d\vec{r}_i}{dt} + \frac{\delta t^2}{2} \frac{d^2 \vec{r}_i}{dt^2} - \dots \simeq \vec{r}_i(t) - \delta t \vec{v}_i(t) + \frac{\delta t^2}{2} \vec{a}_i(t) \quad (2.175)$$

Adding eq. 2.174 and eq. 2.175, rearranging, and making use of eq. 2.173:

$$\vec{r}_i(t + \delta t) + \vec{r}_i(t - \delta t) = 2\vec{r}_i(t) + \delta t^2 \vec{a}_i(t) = 2\vec{r}_i(t) + \delta t^2 \frac{\vec{F}_i(t)}{m_i} \quad (2.176)$$

from where the equation describing the positions of the particles at time $t + \delta t$ can be obtained:

$$\vec{r}_i(t + \delta t) = 2\vec{r}_i(t) - \vec{r}_i(t - \delta t) + \delta t^2 \frac{\vec{F}_i(t)}{m_i} \quad (2.177)$$

The equation describing the velocities of the particles, \vec{v}_i , at the current time t , is obtained by subtracting eq. 2.175 from eq. 2.174:

$$\vec{v}_i(t) = \frac{1}{2\delta t} (\vec{r}_i(t + \delta t) - \vec{r}_i(t - \delta t)) \quad (2.178)$$

The equations 2.177 and 2.178 would be enough to integrate the equations of motion; however, the velocities at time t can only be computed once the positions at time $t + \delta t$ have been obtained. Consequently, when the simulation starts ($t = 0$), a different equation is needed. In order to address this difficulty, a slight modification of the Verlet-type algorithm is often used, which is known as the *Velocity Verlet* method,¹³¹ whose equations are as follows:

$$\vec{r}_i(t + \delta t) = \vec{r}_i(t) + \delta t \vec{v}_i(t) + \frac{\delta t^2}{2} \frac{\vec{F}_i(t)}{m_i} \quad (2.179)$$

$$\vec{v}_i(t + \delta t) = \vec{v}_i(t) + \frac{\delta t}{2} \frac{(\vec{F}_i(t) + \vec{F}_i(t + \delta t))}{m_i} \quad (2.180)$$

It should be emphasized that, whichever method or algorithm is employed to

carry out the MD, it is very important that certain properties of the system are conserved during dynamics, notably the total momentum (M), angular momentum (\mathcal{L}), and the energy.

$$M = \sum_{i=1}^N \vec{p}_i = cst \quad \mathcal{L} = \sum_{i=1}^N \vec{r}_i \wedge \vec{p}_i = cst \quad (2.181)$$

In addition, there are a number of issues we have to highlight when conducting MD simulations. The first key point to consider is what value should be chosen for the time-step. The limiting factor in the choice of this value lies in the nature of the highest frequency motions in the system. As a general criterion, it is considered that δt must be small enough so that the fastest vibration that is described in the system can be followed with sufficient detail during the integration of the equations of motion. In the case of organic molecules, the highest vibration mode is around 3000 cm^{-1} (stretching vibration of carbon-hydrogen bond), which requires timescales of the order of a few femtoseconds ($1 \text{ fs} = 10^{-15} \text{ s}$). In practice, the time increment selected is 1 fs.

Another point of discussion is how to determine the initial velocities for the atoms of the system. One of the most convenient criteria consists in selecting the velocities in such a way that the system will have a particular temperature at the beginning of the simulation. From the standpoint of statistical thermodynamics it is known that the velocities of the atoms in a classical system are distributed according to the *Maxwell-Boltzman distribution*. Hence, if the temperature of the system is T , the probability of finding the i_{th} atom with a velocity in the infinitesimal element $d\vec{v}_i [dv_{i,x}, dv_{i,y}, dv_{i,z}]$ about velocity $\vec{v}_i = [v_{i,x}, v_{i,y}, v_{i,z}]$ is given by:

$$g(\vec{v}_i) d\vec{v}_i = \left(\frac{m_i}{2\pi k_B T} \right)^{3/2} \exp \left(-\frac{m_i \vec{v}_i^2}{2k_B T} \right) d\vec{v}_i \quad (2.182)$$

where the velocity distribution for a single direction is:

$$g(v_x) = \sqrt{\frac{m_i}{2\pi k_B T}} \exp \left(-\frac{m_i v_x^2}{2k_B T} \right) \quad (2.183)$$

Each component of the velocity vector has a normal distribution with a mean value of zero ($\mu_{v_x} = \mu_{v_y} = \mu_{v_z} = 0$) and a standard deviation of $\sigma = \sqrt{k_B T/m_i}$. The distribution function in each velocity component is a Gaussian function centered at the origin ($v_i = 0$) and, thus, the values of the velocities of the atoms can be assigned by treating them as independent *Gaussian random variables*. The temperatures obtained from this expression are not exact because they have been computed randomly, and hence the temperature is calculated as the average kinetic

energy ($\langle \mathcal{K} \rangle$), through the equipartition principle. This principle establishes a relationship between the translational kinetic energies and the temperature according to:

$$\langle \mathcal{K} \rangle = \left\langle \frac{1}{2} m_i v_i^2 \right\rangle = \frac{1}{2} k_B T N_{df} \quad (2.184)$$

where N_{df} is the number of degrees of freedom of the system ($3N - 3$ for a system of N particles with fixed total momentum). Therefore, the temperature can be computed as follows:

$$T = \frac{2}{N_{df} k_B} \langle \mathcal{K} \rangle = \sum_{i=1}^N \frac{m_i v_i^2(t)}{N_{df} k_B} \quad (2.185)$$

The instantaneous temperature, \mathcal{T} , can be defined by means of the same equation but by removing the average. Thus:

$$\mathcal{T} = \frac{2}{N_{df} k_B} \mathcal{K} \quad (2.186)$$

This expression allows the instantaneous temperature to be defined exactly once the initial velocity values have been chosen.

Furthermore, before performing any simulation, the conditions in which there will be MD must be chosen. Among the different working conditions available, there is one which is known in thermodynamic terms as microcanonical ensemble (or N, V, E ensemble), where the number of particles (N), the volume (V), and the energy (E) are conserved. These conditions allow the dynamics to be performed in a easy way, although they are not the most appropriate conditions for an experimental procedure. The most commonly used working conditions are those where N , V , and the temperature (T) are constant, i.e., the canonical ensemble (or N, V, T ensemble). There are other conditions available, although with less applicability, such as the isothermal-isobaric ensemble (N, P, T) or the isobaric-isoenthalpic ensemble (N, P, H).

If we perform an MD simulation under the working conditions of the canonical ensemble (N, V, T), we need to maintain the temperature constant throughout the simulation. A method to achieve this consists in making the system that we want to simulate interact (or couple) with an external bath, such as that originally proposed by Nosé and Hoover.¹³²⁻¹³⁴ Another algorithm to keep the temperature constant was developed by H.C. Andersen,¹³⁵ who suggested that, at intervals during a normal simulation, the velocities of a randomly chosen particle (or group of nuclei) could be reassigned from a Maxwell-Boltzmann distribution. This would be equivalent to the elastic collision of a particle with one of the particles in a heat bath. Thus,

as a result of the reassignment of velocities, this algorithm produces discontinuous trajectories. In order to avoid the discontinuities in the velocities, another method was proposed by Berendsen et al.,¹³⁶ which includes an extra term added to the equations of motion, and which acts as a *frictional force*.

$$\vec{a}_i(t) = \frac{\vec{F}_i(t)}{m_i} + \vec{v}_i(t) \frac{1}{2\tau_i} \left(\frac{T_{bath}}{\mathcal{T}} - 1 \right) \quad (2.187)$$

where T_{bath} is the reference temperature, which is the temperature of the external thermal bath, \mathcal{T} is the instantaneous temperature defined in eq. 2.186 and τ_i is a coupling constant (with units of time). When the actual temperature of the system is higher than the desired temperature (temperature of the bath), the force is negative, and this leads to a decrease in the kinetic energy and, hence, in the temperature. Conversely, if the actual temperature of the system is lower than the temperature of the bath, then the frictional force is positive and energy is supplied to the system, thus increasing the temperature. In this way, control is exerted automatically over all the nuclei of the system, thereby obtaining continuous trajectories over time.

Once the equations of motion are known, and all the factors influencing molecular dynamics have been considered, the steps to be followed throughout an MD simulation can be schematized in the algorithm of Fig. 2.17.

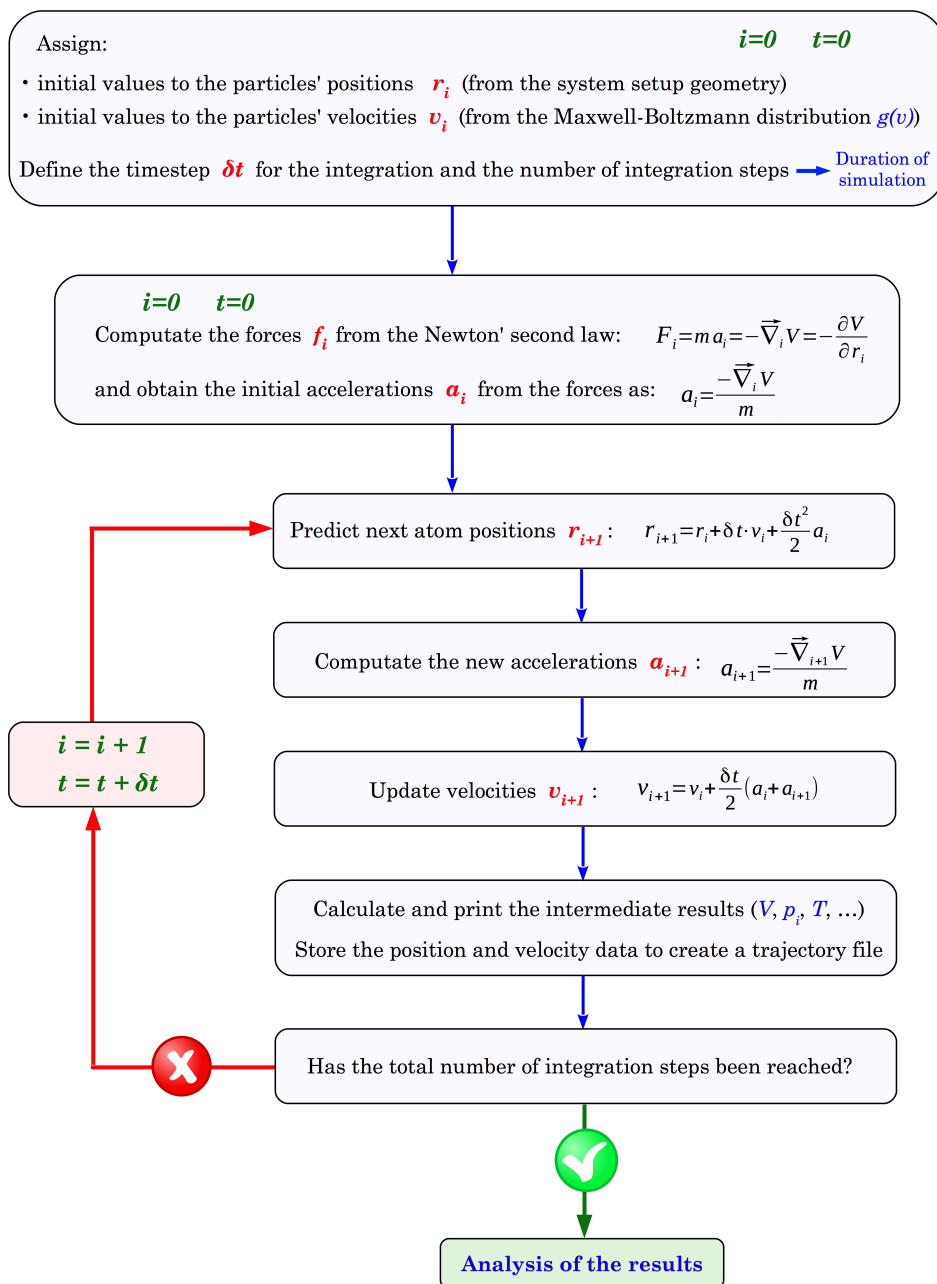


Figure 2.17.: Scheme of the algorithm used to perform a molecular dynamics simulation.

2.5.4. Free Energy

2.5.4.1. Free Energy in statistical mechanics

In the canonical ensemble (N, V, T) , the Helmholtz free energy A is the fundamental thermodynamic potential of the system, and is defined by:

$$A = -k_B T \ln Q_{(N,V,T)} \quad (2.188)$$

where k_B is the Boltzmann constant, T is the temperature of the N -particle system, and $Q_{(N,V,T)}$ is the partition function of the system. In the classical formulation, the $6N$ -dimensional partition function is specified by the set of $3N$ atomic coordinates, $r^N = \{r_1, r_2, \dots, r_N\} = \{x_1, y_1, z_1, \dots, x_N, y_N, z_N\}$, and the $3N$ momenta, $p^N = \{p_1, p_2, \dots, p_N\} = \{p_{x1}, p_{y1}, p_{z1}, \dots, p_{xN}, p_{yN}, p_{zN}\}$, of the particles composing the system. If the energy is defined by the classical Hamiltonian describing the system, $H(r^N, p^N) = H(r_1, r_2, \dots, r_N, p_1, p_2, \dots, p_N)$, then the corresponding expression of the partition function is given by:

$$Q_{(N,V,T)} = \frac{1}{h^{3N} N!} \int \int \exp \left[-\frac{H(r^N, p^N)}{k_B T} \right] dr^N dp^N \quad (2.189)$$

In eq. 2.189, the normalization factor incorporates the factorial term $N!$, which takes into account the indistinguishable nature of the N particles, and the term h^{3N} , which represents the “minimum” size of the microstate and is related with the uncertainty principle. The integration is carried out over all atomic coordinates and momenta.

Substituting eq. 2.189 in eq. 2.188, the Helmholtz free energy is as follows:

$$A = -k_B T \ln \left(\frac{1}{h^{3N} N!} \int \int \exp \left[-\frac{H(r^N, p^N)}{k_B T} \right] dr^N dp^N \right) \quad (2.190)$$

if we make use of the identity:

$$\int \int \exp \left[\frac{H(r^N, p^N)}{k_B T} \right] \exp \left[-\frac{H(r^N, p^N)}{k_B T} \right] dr^N dp^N = h^{3N} N! \quad (2.191)$$

it follows that:

$$A = -k_B T \ln \frac{\int \int \exp \left[-\frac{H(r^N, p^N)}{k_B T} \right] dr^N dp^N}{\int \int \exp \left[\frac{H(r^N, p^N)}{k_B T} \right] \exp \left[-\frac{H(r^N, p^N)}{k_B T} \right] dr^N dp^N} =$$

$$\begin{aligned}
&= k_B T \ln \frac{\int \int \exp \left[\frac{H(r^N, p^N)}{k_B T} \right] \exp \left[-\frac{H(r^N, p^N)}{k_B T} \right] dr^N dp^N}{\int \int \exp \left[-\frac{H(r^N, p^N)}{k_B T} \right] dr^N dp^N} = \\
&= k_B T \ln \int \int \exp \left[\frac{H(r^N, p^N)}{k_B T} \right] \rho(r^N, p^N) dr^N dp^N \quad (2.192)
\end{aligned}$$

where $\rho(r^N, p^N) dr^N dp^N = P_{(N,V,T)}$, which is the probability distribution of finding the system in the unique microstate characterized by positions r^N and momenta p^N and with volume $dr^N dp^N$, and $\rho(r^N, p^N)$ is the probability density of the microstate:

$$\rho(r^N, p^N) = \frac{1}{h^{3N} N!} \frac{\exp \left[-\frac{H(r^N, p^N)}{k_B T} \right]}{Q_{(N,V,T)}} = \frac{\exp \left[-\frac{H(r^N, p^N)}{k_B T} \right]}{\int \int \exp \left[-\frac{H(r^N, p^N)}{k_B T} \right] dr^N dp^N} \quad (2.193)$$

Eq. 2.192 can be rewritten as the ensemble average, which has the following form:

$$A = k_B T \ln \left\langle \exp \left[\frac{H(r^N, p^N)}{k_B T} \right] \right\rangle \quad (2.194)$$

A consequence of eq. 2.193 is that the low-energy regions of the phase space (lower-energy configurations) will be sampled preferentially during the trajectories, according to their respective Boltzmann weight.¹³⁷ On the other hand, eq. 2.194 shows that the weight of the high-energy regions of the space will contribute more to the average. Therefore, an infinite simulation would be required to evaluate the free energy of the system. This fact means that, in most cases, evaluation of accurate absolute free energies from statistical simulations is not possible. The latter may, however, give access to free energy differences between two well-delineated thermodynamic states.¹³⁸

2.5.4.2. Free energy perturbation

When we study a chemical process, we are interested in the free energy difference between the initial and final state of the process. There are different methodologies for calculating free energies, among which we highlight the free energy perturbation^{139–143} (FEP) and thermodynamic integration¹⁴⁴ (TI). These methods allow the alchemical transformation from one state into another so that, although the in-

intermediate states are devoid of any physical significance, we can computationally calculate the free energy change associated with each of these chimeric states and, thus, free energy differences between the initial and final states. Here we present only the FEP methodology, because it is what has been used in this thesis.

Let us consider two thermodynamic states, a and b . To calculate the free energy differences between both states, the Hamiltonian, $H(r^N, p^N)$, describing the transformation is made a function of the reaction coordinate, or “coupling parameter”, λ .

$$H(r^N, p^N; \lambda) = (1 - \lambda) H_a(r^N, p^N) + \lambda H_b(r^N, p^N) \quad (2.195)$$

In a conventional way, λ takes values between 0 and 1, so that the Hamiltonian in state a is given by $H(r^N, p^N; \lambda_a) = H(r^N, p^N; \lambda = 0) = H_a$ and the Hamiltonian of state b is given by $H(r^N, p^N; \lambda_b) = H(r^N, p^N; \lambda = 1) = H_b$.

Within this framework, the canonical partition function of eq. 2.189 depends explicitly on the parameter λ ; therefore, the free energy difference between a and b can be calculated as follows:

$$\Delta A_{a \rightarrow b} = A(\lambda_b) - A(\lambda_a) = -k_B T \ln \frac{Q_{(N,V,T)}(\lambda_b)}{Q_{(N,V,T)}(\lambda_a)} \quad (2.196)$$

Introducing the definition of the partition function (eq. 2.189) into eq. 2.196:

$$\Delta A_{a \rightarrow b} = -k_B T \ln \frac{\int \int \exp \left[-\frac{H(r^N, p^N; \lambda_b)}{k_B T} \right] dr^N dp^N}{\int \int \exp \left[-\frac{H(r^N, p^N; \lambda_a)}{k_B T} \right] dr^N dp^N}$$

If we now use the following identity:

$$\exp \left[-\frac{H(r^N, p^N; \lambda_a)}{k_B T} \right] \exp \left[\frac{H(r^N, p^N; \lambda_a)}{k_B T} \right] \equiv 1 \quad (2.197)$$

Then $\Delta A_{a \rightarrow b} =$

$$-k_B T \ln \frac{\int \int \exp \left[-\frac{H(r^N, p^N; \lambda_b)}{k_B T} \right] \exp \left[-\frac{H(r^N, p^N; \lambda_a)}{k_B T} \right] \exp \left[\frac{H(r^N, p^N; \lambda_a)}{k_B T} \right] dr^N dp^N}{\int \int \exp \left[-\frac{H(r^N, p^N; \lambda_a)}{k_B T} \right] dr^N dp^N} =$$

$$\begin{aligned}
&= -k_B T \ln \int \exp \left[\frac{-\left(H(r^N, p^N; \lambda_b) - H(r^N, p^N; \lambda_a) \right)}{k_B T} \right] \rho(r^N, p^N; \lambda_a) dr^N dp^N \\
&= -k_B T \ln \left\langle \exp \left[\frac{-\left(H(r^N, p^N; \lambda_b) - H(r^N, p^N; \lambda_a) \right)}{k_B T} \right] \right\rangle_{\lambda_a} \quad (2.198)
\end{aligned}$$

In the same way, the free energy difference between b and a can be calculated as:

$$\Delta A_{b \rightarrow a} = A(\lambda_a) - A(\lambda_b) = k_B T \ln \left\langle \exp \left[\frac{-\left(H(r^N, p^N; \lambda_a) - H(r^N, p^N; \lambda_b) \right)}{k_B T} \right] \right\rangle_{\lambda_b} \quad (2.199)$$

So that:

$$\Delta A_{a \rightarrow b} = -\Delta A_{b \rightarrow a} \quad (2.200)$$

In eq. 2.198 $\langle \dots \rangle_{\lambda_a}$ denotes an ensemble average over configurations representative of the initial state (a). However, the validity of this perturbation formula is only satisfied for small changes between the initial state (a) and the final state (b) of the transformation.¹⁴¹ To overcome this difficulty, the reaction pathway is split into a number of a physically meaningless intermediate states connecting a to b , so that between any two contiguous states, the condition of overlapping ensembles is satisfied.¹⁴⁵ The interval separating these intermediate states, which corresponds to selected fixed values of the coupling parameter, λ , is typically referred to as a “window”.

For a series of N intermediate states (windows), the total free energy change for the transformation from a to b is expressed as a sum of $N - 1$ free energy differences:¹⁴⁵

$$\Delta A_{a \rightarrow b} = -K_b T \sum_{k=1}^{N-1} \ln \left\langle \exp \left[\frac{-\left(H(r^N, p^N; \lambda_{k+1}) - H(r^N, p^N; \lambda_k) \right)}{k_B T} \right] \right\rangle_{\lambda_k} \quad (2.201)$$

To satisfy the condition of overlapping ensembles between any two contiguous states, it is advisable to choose a number of windows that ensure similar free energy

changes between contiguous intermediate states. This choice may lead to different widths of the consecutive windows connecting a to b .¹³⁸ In addition, to verify convergence of the FEP calculations it is convenient to perform sampling of the transformation of one state, a , to another, b , in both directions (forward ($\Delta A_{a \rightarrow b}$) and reverse ($\Delta A_{b \rightarrow a}$)). Performing bidirectional FEP calculations allows us to test the self-consistency of the simulations and shows the possible presence of the hysteresis effect. Hysteresis is the difference in the free-energy change resulting from forward and backward perturbation ($hysteresis = | \langle \Delta A_{\lambda_k \rightarrow \lambda_{k+1}} \rangle - \langle \Delta A_{\lambda_{k+1} \rightarrow \lambda_k} \rangle |$), i.e., when the equality of eq. 2.200 is not fulfilled. A strategy to reduce the computational burden incurred in performing independent calculations in both directions consists in using the double-wide sampling method,¹⁴⁶ wherein the free energy changes between forward sampling ($\lambda_k \rightarrow \lambda_{k+1}$) and backward sampling ($\lambda_{k+1} \rightarrow \lambda_k$) is determined concomitantly.

2.6. Topological Analysis of the Electron Localization Function

As seen in previous sections, most of the theoretical studies on chemical reactivity are based on the localization of minimum energy reaction pathways between stable configurations on a multidimensional PES. This energy-based representation provides a wealth of information on the chemical reactivity; however, energy as a global quantity misses mechanistic details such as bond breaking/forming processes, which are the essence of chemical reactivity. From a quantum mechanics standpoint, the interpretation of chemical data is directly related to the extraction of information contained in the wave function, Ψ , whose square, Ψ^2 , provides the electron density, $\rho(r)$. Chemical reactions are essentially movements of pairs of electrons, which proceed through electron density redistributions, and hence $\rho(r)$ appears to be an appropriate physical observable to understand the electron structure and reactivity. Bader¹⁴⁷ noted that $\rho(r)$ must also contain the essence of a molecule's structure, which can be described topologically in terms of its critical points (CP), the zeros of the gradient of this field. With the aim of building a rigorous theory of chemical bonds and identifying distinct bonding patterns, Bader investigated the gradient field of the electron density and pioneered the quantum theory of atoms in molecules (QTAIM).¹⁴⁷⁻¹⁵¹ QTAIM provides a partition of the geometrical space occupied by the chemical system into adjacent non-overlapping volumes called “*basins*”, which can be thought of as chemical entities such as atoms in molecules (AIM).

One of the most popular alternative partitions of the molecular space is provided by the topological analysis of the electron localization function (ELF), $\eta(\vec{r})$, which

was originally introduced by Becke and Edgecombe.¹⁵² ELF is a mathematical function that measures the likelihood of finding an electron in the neighboring space of a reference electron located at a given point and with the same spin. This function is formally a measure of the excess of kinetic energy density due to the Pauli exclusion principle and takes values between 0 and 1 ($0 \leq \eta(\vec{r}) \leq 1$).¹⁵³ In this way, low ELF values ($\eta(\vec{r}) \rightarrow 0$) indicate the presence of high-velocity electrons because of a low Pauli repulsion; thus, the probability of a second electron with parallel spin being close to the reference electron increases. Conversely, high ELF values ($\eta(\vec{r}) \rightarrow 1$) indicate a region that contains low-velocity electrons owing to a low Pauli repulsion; hence, the probability of a second electron with parallel spin being close to the reference electron decreases, or, in other words, the probability of finding paired electrons is high.¹⁵⁴

In addition, ELF values can be treated as a continuous and differentiable scalar field capable of inducing a topology in a three-dimensional space, similarly to the topological analysis of the electron density in Bader's theory. The topological analysis of the ELF gradient field was proposed by Silvi and Savin^{155,156} and provides the partitioning of the molecular space into basins of attractors, which provide information on the structure of atomic shells, location, size of bonding, and lone electron pairs.^{157,158} In this way, ELF allows recovery of the Lewis picture of bonding¹⁵⁹ and the electronic domains of the VSEPR approach.^{160,161}

2.6.1. Electron Localization Function

Original definition of the ELF

The electron localization function, denoted as $\eta(\vec{r})$, was introduced by Becke and Edgecombe with the aim of "measuring" the localization of electrons in atomic and molecular systems.¹⁵² The original derivation of this function was based on Hartree–Fock theory. In this theoretical framework, the like-spin pair probability density is the probability density of simultaneously finding an electron with σ -spin at \vec{r} and another one with the same spin at \vec{r}' , and is given by:

$$P_2^{\sigma\sigma}(\vec{r}, \vec{r}') = \rho^\sigma(\vec{r})\rho^\sigma(\vec{r}') - |\rho_1^\sigma(\vec{r}, \vec{r}')|^2 \quad (2.202)$$

$\rho^\sigma(\vec{r})$ and $\rho^\sigma(\vec{r}')$ being the electron densities at \vec{r} and \vec{r}' , respectively, and $\rho_1^\sigma(\vec{r}, \vec{r}')$ is the σ -spin one-particle density matrix of the HF determinant:

$$\rho_1^\sigma(\vec{r}, \vec{r}') = \sum_i^\sigma \psi_i^*(\vec{r}')\psi_i(\vec{r}) \quad (2.203)$$

where the summation runs over all occupied HF orbitals ψ_i with σ -spin (i.e., either alpha or beta spin).

The conditional like-spin pair probability density, $P_{cond}^{\sigma\sigma}(\vec{r}, \vec{r}')$, is the probability density of finding a σ -spin electron at some position \vec{r}' if a like-spin reference electron is located with certainty at position \vec{r} .

$$P_{cond}^{\sigma\sigma}(\vec{r}, \vec{r}') = \frac{P_2^{\sigma\sigma}(\vec{r}, \vec{r}')}{\rho^\sigma(\vec{r})} = \rho^\sigma(\vec{r}') - \frac{|\rho_1^\sigma(\vec{r}, \vec{r}')|^2}{\rho^\sigma(\vec{r})} \quad (2.204)$$

The probability density of finding a like-spin electron at a distance s from the reference point \vec{r} is given by the Taylor expansion of the spherically averaged conditional like-spin pair probability density, $P_{cond}^{\sigma\sigma}(\vec{r}, s)$:

$$P_{cond}^{\sigma\sigma}(\vec{r}, s) = \frac{1}{3} \left[\tau^\sigma(\vec{r}) - \frac{1}{4} \frac{|\nabla \rho^\sigma(\vec{r})|^2}{\rho^\sigma(\vec{r})} \right] s^2 + \dots \quad (2.205)$$

In this expression, (\vec{r}, s) denotes the spherical average on a shell of radius s around the reference point \vec{r} . It is worth pointing out that only the second-order term (quadratic) of the Taylor expansion has been taken into account, since the first term is neglected because of the Pauli principle. The term $\tau^\sigma(\vec{r})$ is the positive definite kinetic energy density and is defined by:

$$\tau^\sigma(\vec{r}) = \sum_i^\sigma |\nabla \psi_i(\vec{r})|^2 \quad (2.206)$$

Becke and Edgecombe considered that the Taylor expansion of eq. 2.205 contains all the information about the localization of an electron, and they associated this information with the probability density of finding a second like-spin electron near the reference point. The smaller this probability density (the quadratic term) is, the more highly localized the reference electron will be. The expression in brackets is directly related to the curvature of the electron pair density for an electron of identical spins (HF Fermi hole) at \vec{r} .¹⁶²

$$D^\sigma(\vec{r}) = \sum_i^\sigma |\nabla \psi_i(\vec{r})|^2 - \frac{1}{4} \frac{|\nabla \rho^\sigma(\vec{r})|^2}{\rho^\sigma(\vec{r})} \quad (2.207)$$

Hence, the smallness of $D^\sigma(\vec{r})$ describes the Pauli repulsion between two like-spin electrons, and it is related to the electron localization. In fact, $D^\sigma(\vec{r})$ is taken as a measure of the electronic localization. Using the corresponding factor found for homogeneous electron gas with a spin-density equal to the local value of $\rho_\sigma(\vec{r})$:

$$D_0^\sigma(\vec{r}) = \frac{3}{5} \left(6\pi^2 \right)^{\frac{2}{3}} \rho_\sigma(\vec{r})^{\frac{5}{3}} \quad (2.208)$$

Becke and Edgecombe defined the ELF function as:

$$\eta(\vec{r}) = \frac{1}{1 + \chi_{BE}^2(\vec{r})} \quad (2.209)$$

where $\chi(\vec{r})$ was attributed to a dimensionless localization index arbitrarily calibrated with respect to the homogeneous electron gas as a reference.

$$\chi_{BE}(\vec{r}) = \frac{D^\sigma(\vec{r})}{D_0^\sigma(\vec{r})} \quad (2.210)$$

The *Lorentzian* form of eq. 2.209 makes it possible to confine the ELF values in the [0,1] interval. In fact, the quotient $\chi(\vec{r})$ has the significant value of the ELF, while the remaining mathematical artifacts have the sole purpose of confining the function within this interval.

ELF for density functionals

The pair density and its curvature ($D^\sigma(\vec{r})$) are not explicitly defined in local density functionals. Therefore, the original formulation of the ELF derived from the pair density (HF) cannot be applied to DFT. In an attempt to use the ELF in density functional calculations, Savin et al. utilized the observation that the Kohn-Sham orbital representation of the Pauli kinetic energy density has the same formal structure as the expression $D^\sigma(\vec{r})$ of Becke and Edgecombe. In Savin's formulation, the ELF is defined as a property based on the diagonal elements of the one-particle density matrix, i.e., the electron density.

In the Kohn-Sham method, the kinetic energy of N non-interacting electrons is:

$$T_s = \frac{1}{2} \int \sum_{i=1}^N |\nabla \psi_i(\vec{r})|^2 d\tau \quad (2.211)$$

with the Kohn-Sham orbitals $\psi_i(\vec{r})$. However, the pertinent information is contained in the three-dimensional kinetic energy density:

$$t(\vec{r}) = \frac{1}{2} \sum_{i=1}^N |\nabla \psi_i(\vec{r})|^2 \quad (2.212)$$

The function $t(\vec{r})$ is positive definite and cannot exceed a certain threshold:

$$t(\vec{r}) \geq \frac{1}{8} \frac{|\nabla \rho(\vec{r})|^2}{\rho(\vec{r})} \quad (2.213)$$

which occurs when the orbitals are proportional to $\sqrt{\rho}$ (i.e., as in a bosonic system). The Pauli kinetic energy is the energy resulting from the redistribution of the electrons according to the Pauli principle, so that it is the excess of local

kinetic energy with respect to that of the bosonic system. The Pauli kinetic energy density is given by the expression:

$$t_p(\vec{r}) = t(\vec{r}) - \frac{1}{8} \frac{|\nabla\rho(\vec{r})|^2}{\rho(\vec{r})} \quad (2.214)$$

This magnitude itself does not resolve the bonding situation. The complete electronic localization information is provided by the relatively arbitrary quotient of $t_p(\vec{r})$ and kinetic energy density for a homogeneous electron gas with the same electronic density, ($t_{p,h}(\vec{r})$), the latter being expressed as:

$$t_{p,h}(\vec{r}) = c_F \rho_\sigma(\vec{r})^{\frac{5}{3}} \quad (2.215)$$

where $c_F = \frac{3}{10} (3\pi^2)^{2/3} = 2.871$ is the Fermi constant.

For a closed-shell system, the quotient:

$$\chi_s(\vec{r}) = \frac{t_p(\vec{r})}{t_{p,h}(\vec{r})} \quad (2.216)$$

is formally identical to the quotient $\chi_{BE}(\vec{r})$ in the HF approximation. Such an identity continues to be applicable for an open-shell system, when the kinetic energy densities are calculated for the corresponding spin part only. In this way, Savin's ELF function for density functionals, computed according to the following formula:

$$\eta(\vec{r}) = \frac{1}{1 + \chi_s^2(\vec{r})} \quad (2.217)$$

is also identical to that obtained in the HF approximation, and it also yields ELF values in the [0,1] interval.

2.6.2. Gradient Dynamical System

Of the large number of topologies that can be described in a space such as \mathbb{R}^3 , it is of great interest for our purposes to examine the topology induced by a gradient *dynamical system* (DS).

With the term DS we are referring to a vector field y defined on an n -dimensional manifold M . If the field is differentiable, then the system of differential equations $d\vec{r}/dt = y$ uniquely defines the trajectories $\vec{r}(t)$ of the DS.

Let us consider the *scalar function* V :

$$\begin{aligned} \mathbb{R}^3 &\rightarrow \mathbb{R} \\ \vec{r}(x, y, z) &\rightarrow V(\vec{r}) \end{aligned} \quad (2.218)$$

which we call the potential function.

The gradient DS associated to V is the vector field $\nabla V = \vec{f}$, which is defined by the action of the gradient operator on the scalar field V :

$$\nabla V(\vec{r}) = \frac{\partial V(\vec{r})}{\partial x} \vec{i} + \frac{\partial V(\vec{r})}{\partial y} \vec{j} + \frac{\partial V(\vec{r})}{\partial z} \vec{k} = V_x \vec{i} + V_y \vec{j} + V_z \vec{k} \quad (2.219)$$

The trajectories of this DS are defined by the system of differential equations:

$$\frac{d\vec{r}}{dt} = \vec{f} = \nabla V(\vec{r}) \quad (2.220)$$

whose solution can be written in terms of the the following curves, parameterized in \mathbb{R}^3 :

$$\vec{r}(t) = \vec{r}(t_0) + \int_{t_0}^t \nabla V(\vec{r}(t)) dt \quad (2.221)$$

which are also known as flow lines, force lines, field lines or gradient lines. Some characteristics of these flow lines are:¹⁴⁷

1. Since the gradient vector of a scalar points in the direction of the greatest increase in the scalar, the trajectories of ∇V are perpendicular to the isoscalar contour lines.
2. The vector $\nabla V(\vec{r})$ is tangent to its trajectory at each point \vec{r} .
3. Every trajectory must originate or terminate at a point where $\nabla V(\vec{r})$ vanishes, i.e., at a critical point in V .
4. Trajectories cannot cross, since $\nabla V(\vec{r})$ defines but one direction at each point \vec{r} . That is to say, the paths do not intersect at any point. The only exception to this rule occurs at so-called special points of the field.

The expression 2.221 allows the limit points to be defined when $t \rightarrow -\infty$ (α) and $t \rightarrow \infty$ (ω). The point ω is the maximum of the trajectory or *attractor*; the set of points that lead to this attractor define the *basin* of the attractor, which has an *open volume*. The separatrix is the set of frontier points between two or more basins. Fig. 2.18b illustrates the phase portrait of a gradient system in two dimensions. As can be seen, there are two basins delimited by the *separatrix* (green); each basin is composed of a single attractor (red dot) and the set of trajectories leading to it. The point located at the maximum of the separatrix is the α limit of two trajectories that lead each of them to one attractor, and it is known as the *saddle point*.

The α and ω limit points are *critical points*, which are characterized by having a zero potential gradient, i.e., they determine a stationary point of the dynamical system:

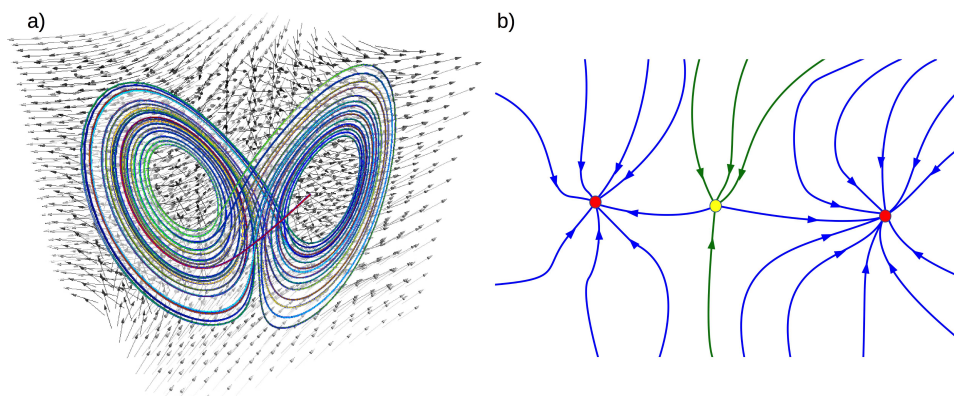


Figure 2.18.: (a) An example of a dynamical system: the Lorenz attractor; (b) Representation of a dynamical system where the attractors (red dots) and the trajectories belonging to their basins (blue lines) have been depicted. The yellow dot corresponds to the separatrix of the dynamical system and the green lines are its corresponding trajectories.

$$\nabla V(\vec{r})_{\alpha/\beta} = 0 \quad (2.222)$$

These critical points correspond to maxima, minima and saddle points of the potential function, similarly to that described for the PES in sec.2.5.1.1. In this way:

- in a *minimum* of $\nabla V(\vec{r})$, all flow lines converge at the critical point leading to a *sinkhole*.
- in a *maximum* of $\nabla V(\vec{r})$, all flow lines leave the critical point leading to a *source* of the field.
- in a *first-order saddle point* of $\nabla V(\vec{r})$, the flow lines are directed toward the critical point in two directions, and escape in one.
- in a *second-order saddle point* of $\nabla V(\vec{r})$, the flow lines are directed toward the critical point in one direction, and escape in two.

2.6.3. Topological Analysis of the ELF

Within the framework of the gradient dynamical system theory, the points corresponding to the local maxima of the ELF function are called attractors. The ELF attractors are located in regions where excess kinetic energy due to the Pauli exclusion principle is minimal, i.e., where the paired electrons are found. In the same way, the basin of an attractor is a region of the space containing the points by which a steepest ascent leads to the attractor, thus yielding a non-overlapping partition of the molecular space. The topological partition of the ELF gradient field^{155,156} yields basins of attractors that can be thought of as corresponding to

atomic cores, bonds, and lone pairs, thereby allowing recovery of the Lewis picture of bonding and the electronic domains of the VSEPR approach. Since the ELF is totally symmetrical, the attractors can be single points (in most systems), circles (off-axis attractors of linear molecules), or even spheres (off-center attractors of atoms), according to their location and to the molecular symmetry.

In the description of the chemical bond by Silvi et al.,^{155,163} two types of basins can be found in a molecule: *core basins* and *valence basins*. On the one hand, core basins surround nuclei with atomic number $Z > 2$ and are labeled as $C(A)$, A being the atomic symbol of the element. It should be noted that for atoms heavier than neon, there are several core shells and therefore several core basins that are, in practice, unified in a single superbasin also labeled as $C(A)$. On the other hand, valence basins are characterized by the number of atomic valence shells in which they participate or, in other words, by the number of core basins with which they share a boundary. This number is called the *synaptic order*.¹⁶³ Thus, there are *monosynaptic*, *disynaptic*, *trisynaptic* basins, and so on (see Tab. 2.1). In this way, it is possible to find *monosynaptic* basins, labeled as $V(X_i)$, which correspond to lone pairs of the Lewis model,¹⁵⁹ and *polysynaptic* basins that correspond to the shared pairs of the Lewis model. In particular, *disynaptic* basins, labeled as $V(X_i, Y_j)$,¹⁵⁹ are the ELF-localization basins with common surfaces between two $C(X_i)$ and $C(Y_j)$ core basins, and thus correspond to two-center bonds; *trisynaptic* basins, labeled as $V(X_i, Y_j, Z_k)$, correspond to the three-center bonds, and so on $V(X_i, Y_j, Z_k, \dots)$ for higher polysynaptic basins. Note that the full valence shell of a molecule is the union of its valence basins.

Hydrogen is a particular case because it is, along with helium, the only coreless atom ($Z \leq 2$). Therefore, as in Lewis theory, it has to be considered an exception. Hydrogenated attractors are intermediate between core and valence attractors.¹⁶⁴ In the description of the chemical bond by Silvi et al., an attractor whose basin contains a proton is considered a valence attractor and will be designated as a *protonated basin*.¹⁶⁴

synaptic order	nomenclature	label
0	asynaptic	V
1	monosynaptic	$V(X_i)$
2	disynaptic	$V(X_i, Y_j)$
3	trisynaptic	$V(X_i, Y_j, Z_k)$
≥ 3	polysynaptic	$V(X_i, Y_j, Z_k, \dots)$

Table 2.1.: Nomenclature of the valence basins according to the synaptic order (connectivity to core basins)

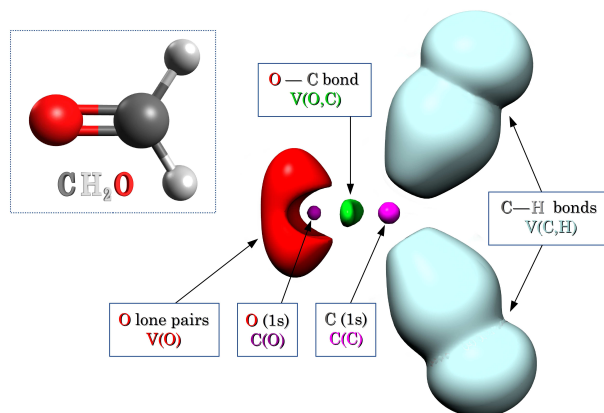


Figure 2.19.: Electron Localization Function (ELF) of the formaldehyde molecule (CH_2O). The internal or core electrons are represented in magenta, the lone pairs or monosynaptic basins are represented in red, the bonding electrons or disynaptic basins are represented in green, and finally the hydrogen basins are represented in light blue.

The topological analysis of the ELF function for the formaldehyde molecule, CH_2O , can be viewed in the example depicted in Fig. 2.19. On the one hand, the formaldehyde molecule shows two core basins, one of them corresponding to the carbon atom $C(C)$ and another corresponding to the oxygen atom $C(O)$. On the other hand, different types of valence basins can be observed, which are listed as follows: a disynaptic basin is found between the oxygen and carbon cores $V(O,C)$, which is related with the bond between both atoms; two protonated basins, $V(C,H)$, are found between the carbon core and the two hydrogen atoms; and, finally, a monosynaptic basin is surrounding the oxygen core $V(O)$, which corresponds to the lone pairs of the oxygen atom.

2.6.4. Properties derived from the ELF

As with the electron density, the ELF function enables us to calculate atomic properties quantitatively. Here, some of them are highlighted.

- **Basin population**

The average population of a basin of volume Ω_i , i.e., the number of electrons in each basin $\bar{N}(\Omega_i)$, is given by the integrated electron density over the basin volumes.

$$\bar{N}(\Omega_i) = \int_{\Omega_i} \rho(\vec{r}) d\vec{r} \quad (2.223)$$

In fact, the population is defined as the sum of alpha and beta spin contributions.

$$\bar{N}(\Omega_i) = \int_{\Omega_i} \rho^\alpha(\vec{r}) d\vec{r} + \int_{\Omega_i} \rho^\beta(\vec{r}) d\vec{r} = \bar{N}^\alpha(\Omega_i) + \bar{N}^\beta(\Omega_i) \quad (2.224)$$

and the sum of all the basin populations leads to the total number of electrons:

$$\sum_i \bar{N}(\Omega_i) = N \quad (2.225)$$

- **Spin density**

Calculation of the spin density in open-shell systems, from the contribution of the α and β electrons. It should be stressed that in closed-shell systems, $\bar{N}^\alpha(\Omega_i) = \bar{N}^\beta(\Omega_i)$.

$$\bar{S}(\Omega_i) = \frac{1}{2} \int_{\Omega_i} (\rho^\alpha(\vec{r}) - \rho^\beta(\vec{r})) = \frac{1}{2} (\bar{N}^\alpha(\Omega_i) - \bar{N}^\beta(\Omega_i)) \quad (2.226)$$

- **Pair population**

This population is obtained by integration of the pair functions.

$$\bar{N}(\Omega_i, \Omega_j) = \langle N(\Omega_i) \cdot N(\Omega_j) \rangle = \int_{\Omega_i} \int_{\Omega_j} P_2(\vec{r}, \vec{r}') d\vec{r} d\vec{r}' \quad (2.227)$$

where $P_2(\vec{r}, \vec{r}')$ is the spinless pair function. If we now consider the spin, this population can be calculated as:

$$\bar{N}^{\sigma\sigma}(\Omega_i, \Omega_j) = \int_{\Omega_i} \int_{\Omega_j} P_2^{\sigma\sigma}(\vec{r}, \vec{r}') d\vec{r} d\vec{r}' \quad (2.228)$$

$$\bar{N}^{\sigma\sigma'}(\Omega_i, \Omega_j) = \int_{\Omega_i} \int_{\Omega_j} P_2^{\sigma\sigma'}(\vec{r}, \vec{r}') d\vec{r} d\vec{r}' \quad (2.229)$$

- **Variance**

The variance or fluctuation of the population represents the quantum uncertainty over a basin population $\bar{N}(\Omega_i)$.

$$\begin{aligned} \sigma^2(\bar{N}; \Omega_i) &= \langle N^2(\Omega_i) \rangle - \langle N(\Omega_i) \rangle^2 \\ &= \int_{\Omega_i} \int_{\Omega_i} P_2(\vec{r}, \vec{r}') d\vec{r} d\vec{r}' + \bar{N}(\Omega_i) - \bar{N}(\Omega_i)^2 \end{aligned} \quad (2.230)$$

- **Covariance**

Covariance is very useful for a practical analysis of electron delocalization. The pair covariance indicates how much of the population fluctuation of two given basins are correlated. This term can be derived from the variance:

$$\sigma^2(\bar{N}; \Omega_i) = -\sum_{j \neq i} V(\Omega_i, \Omega_j) \quad (2.231)$$

where $V(\Omega_i, \Omega_j)$ is the covariance and is calculated as follows:

$$\sum_{j \neq i} V(\Omega_i, \Omega_j) = \bar{N}(\Omega_i, \Omega_j) - \bar{N}(\Omega_i) \bar{N}(\Omega_j) \quad (2.232)$$

where $\bar{N}(\Omega_i, \Omega_j)$ is the actual number of pairs obtained by integration of the pair function over the basins Ω_i and Ω_j , whereas $\bar{N}(\Omega_i) \bar{N}(\Omega_j)$ is the number of electron pairs classically expected from the basin population.

2.6.5. Thom's Catastrophe Theory

Catastrophe theory (CT) is a mathematical method originally developed in the 1960s by René Thom, who expounded the philosophy behind the theory in his book published in 1972, "Stabilité Structurale et Morphogénèse".¹⁶⁵ In mathematics, CT is a branch of topology in the study of dynamical systems, and, more specifically, it is a branch of the bifurcation theory. The latter addresses the study and classification of phenomena characterized by sudden and dramatic shifts in behavior, e.g., a landslide, arising from small changes in circumstances. CT attempts to study how the qualitative nature of the solutions of equations depends on the parameters that appear in the equations. Small changes in certain parameters of a nonlinear system can cause equilibria to appear or disappear, or to change from attracting to repelling and vice versa, leading to large and sudden changes in the behavior of the system. Thom categorized seven fundamental catastrophe types: *Fold catastrophe*, *Cusp catastrophe* (see Fig. 2.20b), *Swallowtail catastrophe*, *Butterfly catastrophe*, *Hyperbolic umbilic catastrophe*, *Elliptic umbilic catastrophe*, and *Parabolic umbilic catastrophe*.

In the case of a gradient dynamical system, the elementary CT studies how the equilibria of the system change as the control parameters change, by considering the behavior of the Hessian matrix (H_{ij}). According to this theory, the set of points at which the determinant of the Hessian matrix is non-zero, $\det|H_{ij}| \neq 0$, are hyperbolic critical points; in contrast, the set of points at which the $\det|H_{ij}| = 0$ are non-hyperbolic critical points or bifurcation points. The number of critical points satisfies the Poincaré-Hopf formula:

$$\sum_P (-1)^{I(P)} = \chi(M) \quad (2.233)$$

in which the sum is performed over the critical points, $I(P)$ is the index of the

critical point labeled by P , and $\chi(M)$ is the Euler characteristic of the manifold on which the gradient field is bound, i.e., 1 for a molecule, 0 for a periodic system.

The set of values for which the determinant of the Hessian matrix of a given critical point is non-zero defines the domain of stability of the critical point also known as the *structural stability domain* (SSD). When a small perturbation of the function (by means of changes in the control parameters) leads to a zero value of the determinant of the Hessian matrix (non-hyperbolic point), the system changes from one SSD to another through a bifurcation point (catastrophic point).

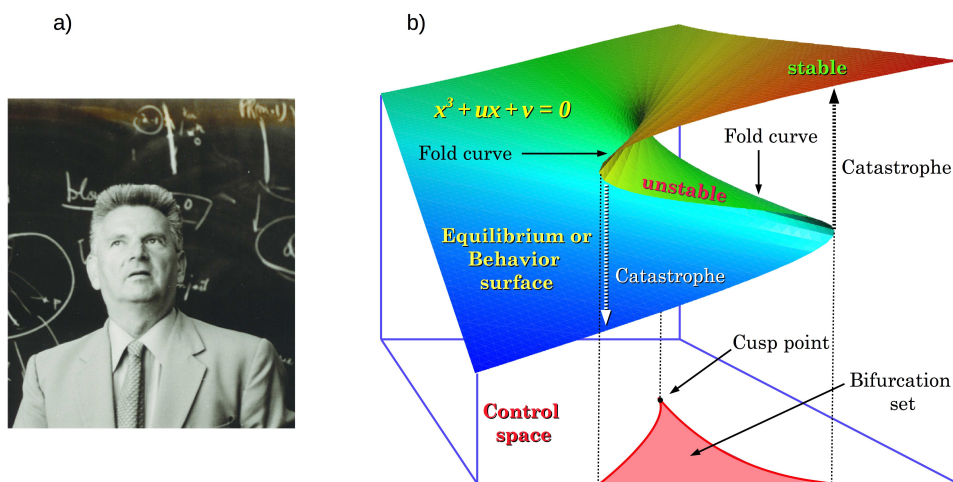


Figure 2.20.: (a) René Thom (picture taken from <http://images.math.cnrs.fr>); (b) Cusp catastrophe model obtained from the representation of the equation $\frac{d}{dx} V_{uv}(x) = 0$, where $V_{uv}(x) = \frac{1}{4}x^4 + \frac{1}{2}ux^2 + vx$.

2.6.6. Bonding Evolution Theory

The bonding evolution theory (BET) is a methodology developed by Krokidis et al.¹⁶⁶ which consists in using Thom's CT to study the evolution along a reaction path of the ELF topologies. BET classifies the elementary chemical processes according to the variation in either the synaptic order (σ) or the morpic number (μ) of at least one basin. Notice that μ is the number of basins of the molecular partition of the system. The mathematical foundations, as well as a number of cases where BET has been successfully applied, have been reviewed comprehensively in recent times.^{167–171} Since the topological behavior of the ELF gradient field can be studied within the framework provided by CT, BET allows a direct relationship to be found between the ELF topology and the evolution of the bond breaking/forming processes and electron pair rearrangements as the chemical reaction proceeds.

According to the mathematical model, a chemical reactive process corresponds

to changes in the number and types of the critical points of the dynamical systems occurring when the control space parameters evolve from the initial subset of values of the reactants $\{\alpha_r\}$ to the final subset of the products $\{\alpha_F\}$. Since the critical points always obey the Poincaré-Hopf formula, this latter appears to be a very strong constraint governing the chemical mechanisms. The control space can be considered as the union of subsets within which all the critical points remain hyperbolic, the structural stability domains, and which correspond to given chemical structures. At the turning points between these domains at least one critical point becomes non-hyperbolic, thereby enabling its index to change. Such a discontinuity is called a bifurcation catastrophe.

In short, the changes in the control parameters defining the reaction pathway (such as the nuclear coordinates and the electronic state) can lead to different topologies of the ELF. Along the reaction pathway, the chemical system goes from a given ELF-SSD to another by means of bifurcation catastrophes occurring at the turning points (TPs). In this way, a chemical reaction can be viewed as a dynamical system where a sequence of elementary chemical processes occur, being characterized by catastrophes. The bifurcation catastrophes occurring at these turning points may be identified according to Thom's classification. Only three elementary catastrophes have been recognized so far in chemical reactions:¹⁷²⁻¹⁷⁶

(i) the *fold catastrophe*, corresponding to the creation or the annihilation of two critical points of different parities. Specifically, it transforms a wandering point (i.e., a point which is not a critical one) into two critical points of different parity.¹⁷⁷ Its unfolding is:

$$V_u(x) = x^3 + ux \quad (2.234)$$

where x is the direction of the eigenvector corresponding to the eigenvalue of the Hessian matrix, which changes sign, and u is the control space parameter, which governs the discontinuity. For $u > 0$, the first derivative is positive for all x , the catastrophe takes place at $u = 0$, for which both the first and second derivatives are zero, and for $u < 0$ there are two critical points at:

$$x = \pm \sqrt{\frac{u}{3}} \quad (2.235)$$

(ii) the *cusp catastrophe* (see Fig. 2.20b), which transforms a critical point of a given parity into three critical points (and vice versa): two of them of the same parity and one of the opposite parity.¹⁷⁷ The standard unfolding is:

$$V_{uv}(x) = \frac{1}{4}x^4 + \frac{1}{2}ux^2 + vx \quad (2.236)$$

and the point where the derivative of the function is equal to zero are the critical

points:

$$\frac{d}{dx}V_{uv}(x) = x^3 + ux + v = 0 \quad (2.237)$$

This type of catastrophe can be observed in the formation or the breaking of a covalent bond.

(iii) the *elliptic umbilic*, which changes the index of one critical point to two. Its unfolding is:

$$V_{uvw}(x, y) = \frac{1}{3}x^3 - xy^2 + u(x^2 + y^2) + vx + wy \quad (2.238)$$

The characterization of the TPs connecting the ELF-SSDs along the reaction pathway allows us to catalog the different chemical events occurring during the course of the reaction as a sequence of electron pair rearrangements taking place during a chemical transformation.

Bibliography

- [1] <http://www.nature.com/subjects/theoretical-chemistry>.
- [2] <http://www.nature.com/subjects/computational-chemistry>.
- [3] Young, D. C. *Computational Chemistry: A Practical Guide for Applying Techniques to Real World Problems*; John Wiley & Sons, Inc., 2001.
- [4] Schrödinger, E. *Phys. Rev.* **1926**, *28*, 1049–1070.
- [5] Born, M.; Oppenheimer, R. *Ann. Phys. (Berlin, Ger.)* **1927**, *389*, 457–484.
- [6] Born, M.; Huang, K. *Dynamical Theory of Crystal Lattices, appendices VII and VIII*; Oxford University Press: London, 1954.
- [7] Hartree, D. R. *Math. Proc. Cambridge Philos. Soc.* **1928**, *24*, 89–110.
- [8] Fock, V. Z. *Phys.* **1930**, *61*, 126–148.
- [9] Fock, V. Z. *Phys.* **1930**, *62*, 795–805.
- [10] Slater, J. C. *Phys. Rev.* **1930**, *35*, 210–211.
- [11] Pauli, W. Z. *Phys.* **1925**, *31*, 765–783.
- [12] Slater, J. C. *Phys. Rev.* **1929**, *34*, 1293–1322.
- [13] Slater, J. C. *Phys. Rev.* **1951**, *81*, 385–390.
- [14] Roothaan, C. C. J. *Rev. Mod. Phys.* **1951**, *23*, 69–89.
- [15] Hall, G. G. *Proc. R. Soc. London, Ser. A* **1951**, *205*, 541–552.
- [16] Slater, J. C. *Phys. Rev.* **1930**, *36*, 57–64.
- [17] Boys, S. F. *Proc. R. Soc. London, Ser. A* **1950**, *200*, 542–554.
- [18] Ditchfield, R.; Hehre, W. J.; Pople, J. A. *J. Chem. Phys.* **1971**, *54*, 724–728.
- [19] Hehre, W. J.; Ditchfield, R.; Pople, J. A. *J. Chem. Phys.* **1972**, *56*, 2257–2261.
- [20] Hellmann, H. *J. Chem. Phys.* **1935**, *3*, 61–61.
- [21] Hay, P. J.; Wadt, W. R. *J. Chem. Phys.* **1985**, *82*, 299–310.
- [22] Shavitt, I. *Methods of Electronic Structure Theory*; Schaefer, H. F., Ed.; Springer US: Boston, MA, 1977; Chapter The Method of Configuration Interaction, pp 189–275.
- [23] Szabo, A.; Ostlund, N. S. *Modern Quantum Chemistry: Introduction to Advanced Electronic Structure Theory*; Dover Books on Chemistry; Dover Publications, 1989.
- [24] Møller, C.; Plesset, M. S. *Phys. Rev.* **1934**, *46*, 618–622.
- [25] Coester, F. *Nucl. Phys.* **1958**, *7*, 421–424.
- [26] Coester, F.; Kümmel, H. *Nucl. Phys.* **1960**, *17*, 477–485.
- [27] Čížek, J. *J. Chem. Phys.* **1966**, *45*, 4256–4266.
- [28] Čížek, J. *Adv. Chem. Phys.* **1969**, *14*, 35–91.
- [29] Paldus, J.; Čížek, J.; Shavitt, I. *Phys. Rev. A* **1972**, *5*, 50–67.
- [30] Hohenberg, P.; Kohn, W. *Phys. Rev.* **1964**, *136*, B864–B871.
- [31] Thomas, L. H. *Math. Proc. Cambridge Philos. Soc.* **1927**, *23*, 542–548.

- [32] Fermi, E. *Rend. Accad. Naz. Lincei* **6** **1927**, 602–D607.
- [33] Kohn, W.; Sham, L. J. *Phys. Rev.* **1965**, *140*, A1133–A1138.
- [34] Vosko, S. H.; Wilk, L.; Nusair, M. *Can. J. Phys.* **1980**, *58*, 1200–1211.
- [35] Perdew, J. P.; Wang, Y. *Phys. Rev. B* **1992**, *45*, 13244–13249.
- [36] Becke, A. D. *Phys. Rev. A* **1988**, *38*, 3098–3100.
- [37] Perdew, J. P.; Yue, W. *Phys. Rev. B* **1986**, *33*, 8800–8802.
- [38] Perdew, J. P.; Burke, K.; Ernzerhof, M. *Phys. Rev. Lett.* **1996**, *77*, 3865–3868.
- [39] Adamo, C.; Barone, V. *J. Chem. Phys.* **2002**, *116*, 5933–5940.
- [40] Becke, A. D. *J. Chem. Phys.* **1988**, *88*, 1053–1062.
- [41] Perdew, J. P. In *Electron. Struct. Solids '91*; Ziesche, P.; Eschrig, H., Eds.; Akademie Verlag: Berlin/DE, 1991; p 11.
- [42] Lee, C.; Yang, W.; Parr, R. G. *Phys. Rev. B* **1988**, *37*, 785–789.
- [43] Tao, J.; Perdew, J. P.; Staroverov, V. N.; Scuseria, G. E. *Phys. Rev. Lett.* **2003**, *91*, 146401.
- [44] Van Voorhis, T.; Scuseria, G. E. *J. Chem. Phys.* **1998**, *109*, 400–410.
- [45] Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 5648–5652.
- [46] Miehlich, B.; Savin, A.; Stoll, H.; Preuss, H. *Chem. Phys. Lett.* **1989**, *157*, 200–206.
- [47] Sousa, S. F.; Fernandes, P. A.; Ramos, M. J. *J. Phys. Chem. A* **2007**, *111*, 10439–10452.
- [48] Zhao, Y.; Truhlar, D. G. *Theor. Chem. Acc.* **2008**, *120*, 215–241.
- [49] Zhao, Y.; Truhlar, D. G. *J. Phys. Chem. A* **2006**, *110*, 13126–13130.
- [50] Jorgensen, W. L.; Maxwell, D. S.; Tirado-Rives, J. *J. Am. Chem. Soc.* **1996**, *118*, 11225–11236.
- [51] Jorgensen, W. L.; Tirado-Rives, J. *J. Am. Chem. Soc.* **1988**, *110*, 1657–1666.
- [52] Kaminski, G. A.; Friesner, R. A.; Tirado-Rives, J.; Jorgensen, W. L. *J. Phys. Chem. B* **2001**, *105*, 6474–6487.
- [53] Pranata, J.; Wierschke, S. G.; Jorgensen, W. L. *J. Am. Chem. Soc.* **1991**, *113*, 2810–2819.
- [54] Rizzo, R. C.; Jorgensen, W. L. *J. Am. Chem. Soc.* **1999**, *121*, 4827–4836.
- [55] Kony, D.; Damm, W.; Stoll, S.; Van Gunsteren, W. F. *J. Comput. Chem.* **2002**, *23*, 1416–1429.
- [56] Nilsson, L.; Karplus, M. *J. Comput. Chem.* **1986**, *7*, 591–616.
- [57] Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. *J. Comput. Chem.* **1983**, *4*, 187–217.
- [58] Momany, F. A.; Rone, R. *J. Comput. Chem.* **1992**, *13*, 888–900.
- [59] Foloppe, N.; MacKerell, A. D., Jr. *J. Comput. Chem.* **2000**, *21*, 86–104.
- [60] MacKerell, A. D.; Banavali, N. K. *J. Comput. Chem.* **2000**, *21*, 105–120.
- [61] Patel, S.; Brooks, C. L. *J. Comput. Chem.* **2004**, *25*, 1–16.

- [62] MacKerell, A. D.; Wiorkiewicz-Kuczera, J.; Karplus, M. *J. Am. Chem. Soc.* **1995**, *117*, 11946–11975.
- [63] Field, M. J. *A Practical Introduction to the Simulation of Molecular Systems*, 2nd ed.; Cambridge University Press, 2007; Cambridge Books Online.
- [64] Weiner, S. J.; Kollman, P. A.; Case, D. A.; Singh, U. C.; Ghio, C.; Alagona, G.; Profeta, S.; Weiner, P. *J. Am. Chem. Soc.* **1984**, *106*, 765–784.
- [65] Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. *J. Am. Chem. Soc.* **1995**, *117*, 5179–5197.
- [66] Weiner, S. J.; Kollman, P. A.; Nguyen, D. T.; Case, D. A. *J. Comput. Chem.* **1986**, *7*, 230–252.
- [67] van Gunsteren, W. F.; Berendsen, H. J. C. *Angew. Chem. Int. Ed. Engl.* **1990**, *29*, 992–1023.
- [68] Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926–935.
- [69] Jorgensen, W. L.; Madura, J. D. *Mol. Phys.* **1985**, *56*, 1381–1392.
- [70] Steinbach, P. J.; Brooks, B. R. *J. Comput. Chem.* **1994**, *15*, 667–683.
- [71] Born, M.; von Karman, T. *Physik Z.* **1912**, *13*, 297–309.
- [72] Ewald, P. P. *Ann. Phys.* **1921**, *64*, 253–287.
- [73] York, D. M.; Darden, T. A.; Pedersen, L. G. *J. Chem. Phys.* **1993**, *99*, 8345–8348.
- [74] Senn, H. M.; Thiel, W. *Atomistic Approaches in Modern Biology: From Quantum Chemistry to Molecular Simulations*; Reiher, M., Ed.; Springer Berlin Heidelberg: Berlin, Heidelberg, 2007; Chapter QM/MM Methods for Biological Systems, pp 173–290.
- [75] Warshel, A.; Levitt, M. *J. Mol. Biol.* **1976**, *103*, 227–249.
- [76] Singh, U. C.; Kollman, P. A. *J. Comput. Chem.* **1986**, *7*, 718–730.
- [77] Field, M. J.; Bash, P. A.; Karplus, M. *J. Comput. Chem.* **1990**, *11*, 700–733.
- [78] Tee, L. S.; Gotoh, S.; Stewart, W. E. *Ind. Eng. Chem. Fundam.* **1966**, *5*, 356–363.
- [79] Maseras, F.; Morokuma, K. *J. Comput. Chem.* **1995**, *16*, 1170–1179.
- [80] Svensson, M.; Humbel, S.; Froese, R. D. J.; Matsubara, T.; Sieber, S.; Morokuma, K. *J. Phys. Chem.* **1996**, *100*, 19357–19363.
- [81] Dapprich, S.; Komáromi, I.; Byun, K. S.; Morokuma, K.; Frisch, M. J. *J. Mol. Struct.: THEOCHEM* **1999**, *461-462*, 1–21.
- [82] Gao, J.; Truhlar, D. G. *Ann. Rev. Phys. Chem.* **2002**, *53*, 467–505.
- [83] Reuter, N.; Dejaegere, A.; Maignret, B.; Karplus, M. *J. Phys. Chem. A* **2000**, *104*, 1720–1735.
- [84] Théry, V.; Rinaldi, D.; Rivail, J.-L.; Maignret, B.; Ferenczy, G. G. *J. Comput. Chem.* **1994**, *15*, 269–282.
- [85] Ferenczy, G. G.; Rivail, J.-L.; Surján, P. R.; Náray-Szabó, G. *J. Comput. Chem.* **1992**, *13*, 830–837.

- [86] Monard, G.; Loos, M.; Théry, V.; Baka, K.; Rivail, J.-L. *Int. J. Quantum Chem.* **1996**, *58*, 153–159.
- [87] Gao, J.; Amara, P.; Alhambra, C.; Field, M. J. *J. Phys. Chem. A* **1998**, *102*, 4714–4721.
- [88] Field, M. J. *Computational Approaches to Biochemical Reactivity*; Náray-Szabó, G.; Warshel, A., Eds.; Springer Netherlands: Dordrecht, 2002; Chapter Hybrid Potentials for Molecular Systems in the Condensed Phase, pp 125–151.
- [89] Gonzalez, C.; Schlegel, H. B. *J. Chem. Phys.* **1989**, *90*, 2154–2161.
- [90] Bolhuis, P. G.; Chandler, D.; Dellago, C.; Geissler, P. L. *Annu. Rev. Phys. Chem.* **2002**, *53*, 291–318.
- [91] Fukui, K. *J. Phys. Chem.* **1970**, *74*, 4161–4163.
- [92] Fukui, K. *Acc. Chem. Res.* **1981**, *14*, 363–368.
- [93] Fukui, K. *Recl. Trav. Chim. PaysBas* **1979**, *98*, 75–78.
- [94] Melissas, V. S.; Truhlar, D. G.; Garrett, B. C. *J. Chem. Phys.* **1992**, *96*, 5758–5772.
- [95] Nelder, J. A.; Mead, R. *Computer J.* **1965**, *7*, 308–313.
- [96] Debye, P. *Mathematische Annalen* **1909**, *67*, 535–558.
- [97] Hestenes, M. R.; Stiefel, E. *J. Res. Natl. Bur. Stand. (U. S.)* **1952**, *49*, 409–436.
- [98] Fletcher, R.; Reeves, C. M. *Computer J.* **1964**, *7*, 149–154.
- [99] Polak, R. G., E. *ESAIM: Model. Math. Anal. Numer.* **1969**, *3*, 35–43.
- [100] Raphson, J. *Analysis Aequationum universalis*; typis T. B. prostant venales apud A. & I. Churchill: London, 1702.
- [101] Wallis, J. *A Treatise of Algebra, Both Historical and Practical: Shewing the Original, Progress, and Advancement Thereof ... with Some Additional Treatises*; Davis, R.: London, 1685.
- [102] Davidon, W. C. *SIAM J. Optimiz.* **1991**, *1*, 1–17.
- [103] Fletcher, R. *Computer J.* **1970**, *13*, 317–322.
- [104] Shanno, D. F. *Math. Comp.* **1970**, *24*, 647–656.
- [105] Broyden, C. G. *J. Inst. Math. Its Appl.* **1970**, *6*, 222–231.
- [106] Goldfarb, D. *Math. Comp.* **1970**, *24*, 23–26.
- [107] Byrd, R. H.; Lu, P.; Nocedal, J.; Zhu, C. *SIAM J. Sci. Comput.* **1995**, *16*, 1190–1208.
- [108] Baker, J.; Kessi, A.; Delley, B. *J. Chem. Phys.* **1996**, *105*, 192–212.
- [109] Baker, J. *J. Comput. Chem.* **1997**, *18*, 1079–1095.
- [110] Eyring, H. *J. Chem. Phys.* **1935**, *3*, 107–115.
- [111] Evans, M. G.; Polanyi, M. *Trans. Faraday Soc.* **1935**, *31*, 875–894.
- [112] Connors, K. A. *Chemical Kinetics: The Study of Reaction Rates in Solution*; VCH Publishers: New York, 1990.
- [113] Bigeleisen, J.; Mayer, M. G. *J. Chem. Phys.* **1947**, *15*, 261–267.

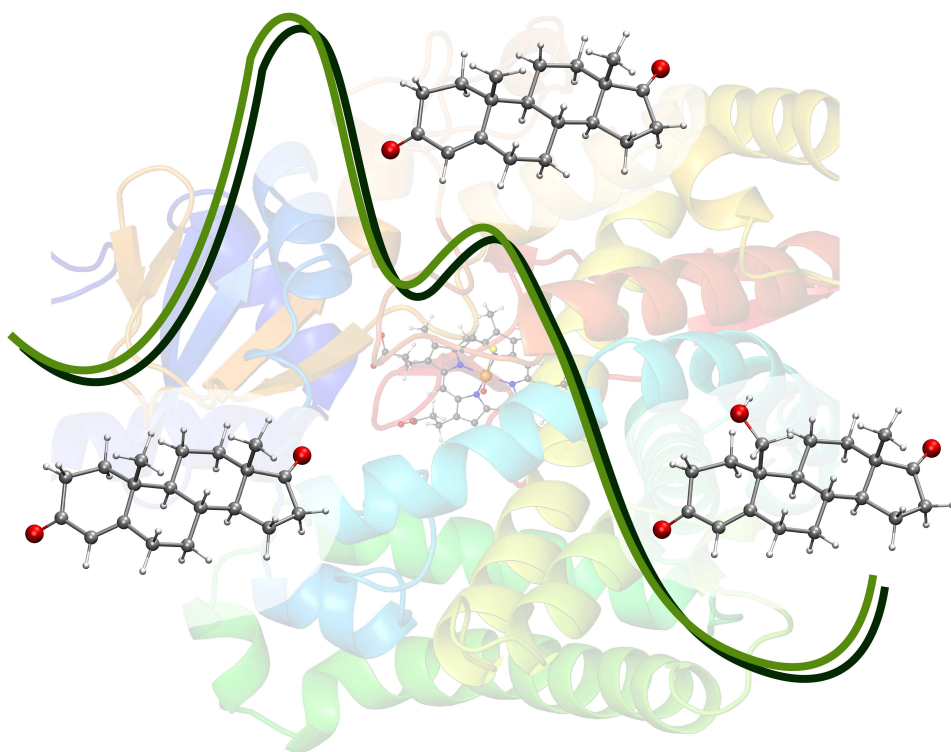
- [114] Bigeleisen, J.; Wolfsberg, M. *Theoretical and Experimental Aspects of Isotope Effects in Chemical Kinetics*; John Wiley & Sons, Inc., 1958; pp 15–76.
- [115] Eyring, H.; Stearn, A. E. *Chem. Rev.* **1939**, *24*, 253–270.
- [116] G., T. D.; Hase, W. L.; Hynes, J. T. *J. Phys. Chem.* **1983**, *87*, 2664–2682.
- [117] Anderson, J. B. *Advances in Chemical Physics*; Prigogine, I.; Rice, S. A., Eds.; John Wiley & Sons, Inc., 1995; Vol. 91, Chapter Predicting Rare Events in Molecular Dynamics, pp 381–431.
- [118] Truhlar, D. G.; Garrett, B. C.; Klippenstein, S. J. *J. Phys. Chem.* **1996**, *100*, 12771–12800.
- [119] Kreevoy, M. M.; Truhlar, D. G. *Investigation of Rates and Mechanisms of Reactions*, 4th ed.; John Wiley and Sons, inc.: New York, 1986; Vol. 6, Chapter Transition State Theory, pp 13–95.
- [120] Garcia-Viloca, M.; Gao, J.; Karplus, M.; Truhlar, D. G. *Science* **2004**, *303*, 186–195.
- [121] Klinman, J. P. In *Quantum Tunnelling in Enzyme-Catalysed Reaction*; Allemann, R. K.; Scrutton, N. S., Eds.; The Royal Society of Chemistry, 2009; Chapter 6, pp 132–160.
- [122] Cramer, C. J. *Essentials of Computational Chemistry: Theories and Models*, 2nd ed.; John Wiley & Sons Ltd, 2004.
- [123] Wigner, E. *Z. Phys. Chem. Abt. B.* **1932**, *19*, 203–216.
- [124] Wigner, E. *J. Chem. Phys.* **1937**, *5*, 720–725.
- [125] Bell, R. P. *The Tunnel Effect in Chemistry*; Chapman and Hall: London, 1980; Chapter The application of tunnel corrections in chemical kinetics, p 140.
- [126] Tuñón, I. n.; Silla, E. *Química molecular estadística*; Biblioteca De Químicas; Editorial Sintesis, 2008.
- [127] Metropolis, N.; Rosenbluth, A. W.; Rosenbluth, M. N.; Teller, A. H.; Teller, E. *J. Chem. Phys.* **1953**, *21*, 1087–1092.
- [128] Alder, B. J.; Wainwright, T. E. *J. Chem. Phys.* **1957**, *27*, 1208–1209.
- [129] Rahman, A. *Phys. Rev.* **1964**, *136*, A405–A411.
- [130] Verlet, L. *Phys. Rev.* **1967**, *159*, 98–103.
- [131] Swope, W. C.; Andersen, H. C.; Berens, P. H.; Wilson, K. R. *J. Chem. Phys.* **1982**, *76*, 637–649.
- [132] Nosé, S. *J. Chem. Phys.* **1984**, *81*, 511–519.
- [133] Hoover, W. G. *Phys. Rev. A* **1985**, *31*, 1695–1697.
- [134] Nosé, S.; Klein, M. *Molec. Phys.* **1983**, *50*, 1055–1076.
- [135] Feller, S. E.; Zhang, Y.; Pastor, R. W.; Brooks, B. R. *J. Chem. Phys.* **1995**, *103*, 4613–4621.
- [136] Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; DiNola, A.; Haak, J. R. *J. Chem. Phys.* **1984**, *81*, 3684–3690.

- [137] Allen, M. P.; Tildesley, D. J. *Computer simulation of liquids*, reprinted with corrections ed.; Oxford university press, 1989.
- [138] Chipot, C. *New Algorithms for Macromolecular Simulation*; Leimkuhler, B.; Chipot, C.; Elber, R.; Laaksonen, A.; Mark, A.; Schlick, T.; Schütte, C.; Skeel, R., Eds.; Springer Berlin Heidelberg: Berlin, Heidelberg, 2006; Chapter Free Energy Calculations in Biological Systems. How Useful Are They in Practice?, pp 185–211.
- [139] Born, M. *Z. Phys.* **1920**, *1*, 45–48.
- [140] Kirkwood, J. G. *J. Chem. Phys.* **1935**, *3*, 300–313.
- [141] Zwanzig, R. W. *J. Chem. Phys.* **1954**, *22*, 1420–1426.
- [142] Jorgensen, W. L. *Acc. Chem. Res.* **1989**, *22*, 184–189.
- [143] Kollman, P. *Chem. Rev.* **1993**, *93*, 2395–2417.
- [144] Kästner, J.; Senn, H. M.; Thiel, S.; Otte, N.; Thiel, W. *J. Chem. Theory Comput.* **2006**, *2*, 452–461.
- [145] Mark, A. E. *Encyclopedia of Computational Chemistry*; Schleyer, P. v. R.; Allinger, N. L.; Clark, T.; Gasteiger, J.; Kollman, P. A.; Schaefer III, H. F.; Schreiner, P. R., Eds.; John Wiley & Sons, Ltd, 1998; Vol. 2, Chapter Free Energy Perturbation Calculations, pp 1070–1083.
- [146] Jorgensen, W. L.; Ravimohan, C. *J. Chem. Phys.* **1985**, *83*, 3050–3054.
- [147] Bader, R. F. W. *Atoms in molecules : a quantum theory*; Clarendon Press ; Oxford University Press: New York, 1990.
- [148] Popelier, P. *Coord. Chem. Rev.* **2000**, *197*, 169 – 189.
- [149] Matta, C. F.; Boyd, R. J. *The Quantum Theory of Atoms in Molecules: From Solid State to DNA and Drug Design*; Wiley-VCH Verlag GmbH & Co. KGaA, 2007; Chapter An Introduction to the Quantum Theory of Atoms in Molecules, pp 1–34.
- [150] Popelier, P. *Atoms in Molecules: An Introduction*; Pearson Education; Prentice Hall: Harlow, 2000.
- [151] Bader, W. R. F. *Theor. Chem. Acc.* **2001**, *105*, 276–283.
- [152] Becke, A. D.; Edgecombe, K. E. *J. Chem. Phys.* **1990**, *92*, 5397–5403.
- [153] Berski, S.; Silvi, B.; Lundell, J.; Noury, S.; Latajka, Z. *New Trends in Quantum Systems in Chemistry and Physics*; Maruani, J.; Minot, C.; McWeeny, R.; Smeyers, Y.; Wilson, S., Eds.; Kluwer Academic Publishers, 2001; Vol. 1, Chapter The nature of binding in HRgY (Rg = Ar, Kr, Xe; Y = F, Cl) compounds based on the topological analysis of the electron localisation function (ELF), pp 259–279.
- [154] Flad, H.-J.; Schautz, F.; Wang, Y.; Dolg, M.; Savin, A. *Eur. Phys. J. D* **1999**, *6*, 243–254.
- [155] Silvi, B.; Savin, A. *Nature* **1994**, *371*, 683–686.
- [156] Häussermann, U.; Wengert, S.; Hofmann, P.; Savin, A.; Jepsen, O.; Nesper, R. *Angew. Chemie Int. Ed. English* **1994**, *33*, 2069–2073.
- [157] Nesper, R. *Angew. Chemie Int. Ed. English* **1991**, *30*, 789–817.
- [158] Savin, A.; Nesper, R.; Wengert, S.; Fässler, T. F. *Angew. Chemie Int. Ed. English* **1997**, *36*, 1808–1832.

- [159] Lewis, G. N. *J. Am. Chem. Soc.* **1916**, *38*, 762–785.
- [160] Gillespie, R. J.; Nyholm, R. S. *Q. Rev. Chem. Soc.* **1957**, *11*, 339–380.
- [161] Gillespie, R. J. *J. Chem. Educ.* **1970**, *47*, 18.
- [162] Dobson, J. F. *J. Chem. Phys.* **1991**, *94*, 4328–4333.
- [163] Silvi, B. *J. Mol. Struct.* **2002**, *614*, 3–10.
- [164] A., S.; Silvi, B.; Coionna, F. *Can. J. Chem.* **1996**, *74*, 1088–1096.
- [165] Thom, R. *Structural stability and morphogenesis; an outline of a general theory of models*; W. A. Benjamin: Reading, Mass, 1975; Translated from the French ed. by D. H. Fowler. With a foreword by C. H. Waddington.
- [166] Krokidis, X.; Noury, S.; Silvi, B. *J. Phys. Chem. A* **1997**, *101*, 7277–7282.
- [167] Andres, J.; Berski, S.; Domingo, L. R.; Polo, V.; Silvi, B. *Curr. Org. Chem.* **2011**, *15*, 3566–3575.
- [168] Andrés, J.; González-Navarrete, P.; Safont, V. S. *Int. J. Quantum Chem.* **2014**, *114*, 1239–1252.
- [169] Andrés, J.; Gracia, L.; González-Navarrete, P.; Safont, V. S. *Comput. Theor. Chem.* **2015**, *1053*, 15–17.
- [170] Gillet, N.; Chaudret, R.; Contreras-Garcia, J.; Yang, W.; Silvi, B.; Piquemal, J.-P. *J. Chem. Theory Comput.* **2012**, *8*, 3993–3997.
- [171] Polo, V.; Andres, J.; Berski, S.; Domingo, L. R.; Silvi, B. *J. Phys. Chem. A* **2008**, *112*, 7128–7136.
- [172] Krokidis, X.; Gonçalves, V.; Savin, A.; Silvi, B. *J. Phys. Chem. A* **1998**, *102*, 5065–5073.
- [173] Krokidis, X.; Vuilleumier, R.; Borgis, D.; Silvi, B. *Molec. Phys.* **1999**, *96*, 265–273.
- [174] Krokidis, X.; Silvi, B.; Dezarnaud-Dandine, C.; Sevin, A. *New J. Chem.* **1998**, *22*, 1341–1350.
- [175] Krokidis, X.; Silvi, B.; Alikhani, M. E. *Chem. Phys. Lett.* **1998**, *292*, 35–45.
- [176] Michelini, M. C.; Sicilia, E.; Russo, N.; Alikhani, M. E.; Silvi, B. *J. Phys. Chem. A* **2003**, *107*, 4862–4868.
- [177] Poston, T.; Stewart, I. *Catastrophe Theory and Its Applications*; Dover Publications, Inc., 1978.

Chapter 3

Aim and scope of the thesis



The enzyme Cytochrome P450 aromatase plays an essential role in the biosynthesis of estrogens, and its inhibition is an important target for the development of drugs for the treatment of breast cancer. The main purpose of the present thesis is to improve the understanding of the catalytic mechanism and the biochemistry of this enzyme from the standpoint of theoretical chemistry. With this aim, a theoretical study of the first catalytic subcycle of the enzyme aromatase, as well as an analysis of the electronic structure of the oxidant species, have been carried out by means of computational techniques. Specifically, the objectives that have been raised and addressed throughout this thesis are the following:

- **Study of the electronic structure of the reactive species of the enzyme aromatase (Cpd I).**

The aim of this study is to shed light on the electron structure and the bonding nature of Cpd I, which is very important to be able to understand the biochemistry of the enzyme. For this purpose, this study is based on two methods of quantum chemical topology analysis: electron localization function, ELF, and electron density.

- **Study of the first catalytic subcycle of the enzyme Aromatase.**

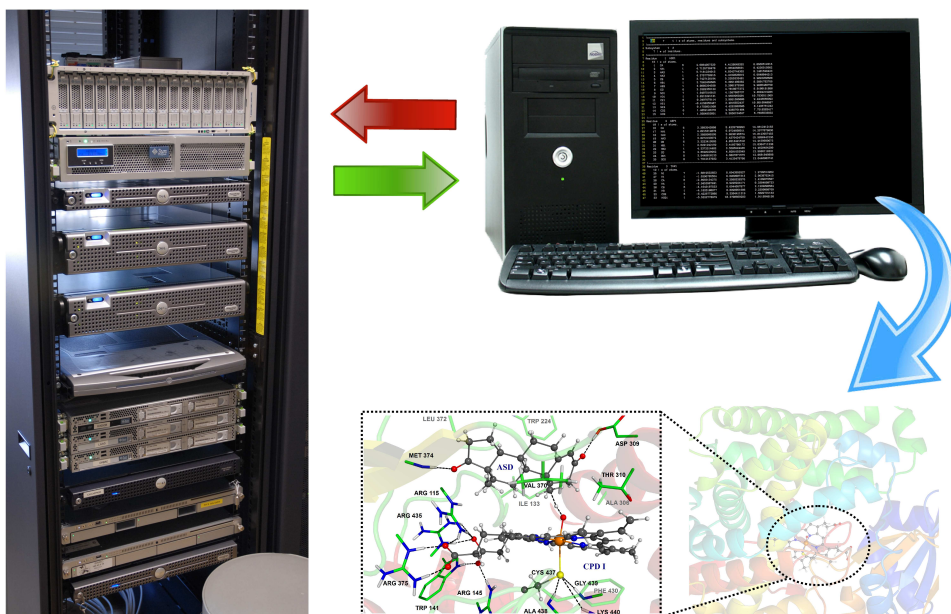
Study of the reaction mechanism of the first oxidation step of the natural substrate androstenedione, ASD, by means of the hybrid QM/MM methodology. This catalytic subcycle consists in the hydroxylation of the C_{19} methyl group of the substrate ASD in a two-step mechanism: a first hydrogen atom abstraction step through Cpd I, followed by a subsequent oxygen rebound step. In addition, further study on the mechanism of the hydrogen abstraction step will be performed by the joint use of the Bonding Evolution Theory and QM/MM techniques, thereby providing a new perspective on this step.

- **Study of the hydroxylation of the aromatase inhibitor exemestane during the first catalytic subcycle of Aromatase.**

The substrate ASD will be replaced by the substrate exemestane, EXE, a third-generation aromatase inhibitor used in breast cancer therapy, and the hydroxylation step of its C_{19} methyl group will be studied. Since both substrates, ASD and EXE, are steroidal derivatives and their molecular geometries are very similar, a comparison of the two substrates in the hydroxylation process will be performed.

Chapter 4

Results



Contents

4.1. Study of the Reactive Species of the Enzyme Aromatase: Compound I	145
4.1.1. New Insight into the Electronic Structure of Iron(IV)-oxo Porphyrin Compound I. A Quantum Chemical Topological Analysis	145
4.2. First Catalytic Subcycle of the Enzyme Aromatase .	165
4.2.1. QM/MM Modeling of the Hydroxylation of the Androstenedione Substrate Catalyzed by Cytochrome P450 Aromatase (CYP19A1)	165
4.2.2. Joint Use of Bonding Evolution Theory and QM/MM Hybrid Method for Understanding the Hydrogen Abstraction Mechanism via Cytochrome P450 Aromatase .	189
4.3. Study of the Hydroxylation of an Aromatase Inhibitor: Exemestane	211
4.3.1. Theoretical Study of the Mechanism of Exemestane Hydroxylation Catalyzed by Human Aromatase Enzyme .	211

4.1. Study of the Reactive Species of the Enzyme Aromatase: Compound I

4.1.1. New Insight into the Electronic Structure of Iron(IV)-oxo Porphyrin Compound I. A Quantum Chemical Topological Analysis

Ignacio Viciano, Slawomir Berski, Sergio Martí, and Juan Andres

Journal of Computational Chemistry **2013**, 34, 780–789

New Insight into the Electronic Structure of Iron(IV)-oxo Porphyrin Compound I. A Quantum Chemical Topological Analysis

Ignacio Viciano, Slawomir Berski, Sergio Martí,* and Juan Andrés

The electronic structure of iron-oxo porphyrin π -cation radical complex $\text{Por}^{\cdot+}\text{Fe}^{\text{IV}}=\text{O}$ (S—H) has been studied for doublet and quartet electronic states by means of two methods of the quantum chemical topology analysis: electron localization function (ELF) $\eta(r)$ and electron density $\rho(r)$. The formation of this complex leads to essential perturbation of the topological structure of the carbon-carbon bonds in porphyrin moiety. The double C=C bonds in the pyrrole anion subunits, represented by pair of bonding disynaptic basins $V_{i=1,2}(\text{C},\text{C})$ in isolated porphyrin, are replaced by single attractor $V(\text{C},\text{C})_{i=1-20}$ after complexation with the Fe cation. The iron-nitrogen bonds are covalent dative bonds, $\text{N}\rightarrow\text{Fe}$, described by the disynaptic bonding basins $V(\text{Fe},\text{N})_{i=1-4}$, where electron density is almost formed by the lone pairs of the N atoms. The nature of the iron-oxygen bond predicted by the ELF topological analysis, shows a

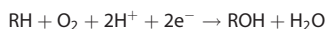
main contribution of the electrostatic interaction, $\text{Fe}^{\delta+}\dots\text{O}^{\delta-}$, as long as no attractors between the C(Fe) and C(O) core basins were found, although there are common surfaces between the iron and oxygen basins and coupling between iron and oxygen lone pairs, that could be interpreted as a charge-shift bond. The Fe—S bond, characterized by the disynaptic bonding basin $V(\text{Fe},\text{S})$, is partially a dative bond with the lone pair donated from sulfur atom. The change of electronic state from the doublet ($M = 2$) to quartet ($M = 4$) leads to reorganization of spin polarization, which is observed only for the porphyrin skeleton ($-0.43e$ to $0.50e$) and S—H bond ($-0.55e$ to $0.52e$). © 2012 Wiley Periodicals, Inc.

DOI: 10.1002/jcc.23201

Introduction

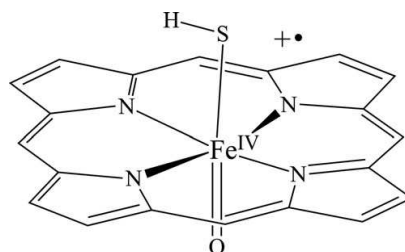
Cytochrome P450 enzymes is a superfamily of hemoproteins that plays an important role in some biological processes in several bioorganisms such as plants, mammalian, fungi, archaea, bacteria, or protists.^[1,2] This superfamily is composed by a large number of isoforms which can metabolize a wide variety of endogenous as well as exogenous substrates.^[3]

The most important catalyzed reaction by the cytochrome P450 enzymes is the monooxygenation of organic substrates:



in which an oxygen atom is inserted into the unactivated C—H bond of an organic substrate via oxygen activation process. In fact, this is the reason why Cytochrome P450 is also called monooxygenase and it is well known as one of the most versatile oxidant in the nature.^[3]

During the catalytic cycle of this enzyme, an iron-(IV) oxo-porphyrin π -cation radical specie ($\text{Por}^{\cdot+}\text{Fe}^{\text{IV}}=\text{O}(\text{S-Cys})$) also known as Compound I (Cpd I) is formed.^[4] This Cpd I (Scheme 1) is an elusive specie in the catalytic cycle of the cytochrome P450 and is the key step in the oxidation of substrates.^[5] It is based on a protoporphyrin-IX prosthetic group with a central iron atom linked to two axial ligands: the first one is a cysteine amino acid named as proximal axial ligand and the second one is an oxo anion named distal ligand, which forms a ferryl $\text{Fe}^{\text{IV}}=\text{O}$ moiety. Cpd I is also formed in the catalytic cycle of other proteins, such as peroxidase or catalase, whose proximal axial ligands are the histidine and tyrosine amino acids, respec-



Scheme 1. Lewis structure of Compound I.

tively. As a result, the elusiveness of Cpd I and due to the high reactivity, which makes it very difficult to detect, other alternative oxidant species such as Compound 0^[6] ($\text{Por}-\text{Fe}^{\text{III}}-\text{OOH}^-$, Cpd 0), Compound II^[7] ($\text{Por}-\text{Fe}^{\text{IV}}=\text{O}$, Cpd II) as well as perferryl species^[8,9] ($\text{Por}-\text{Fe}^{\text{V}}=\text{O}$) have been suggested that could be the reactive intermediates in the oxidation of the substrates by the

[a] I. Viciano, S. Martí, J. Andrés

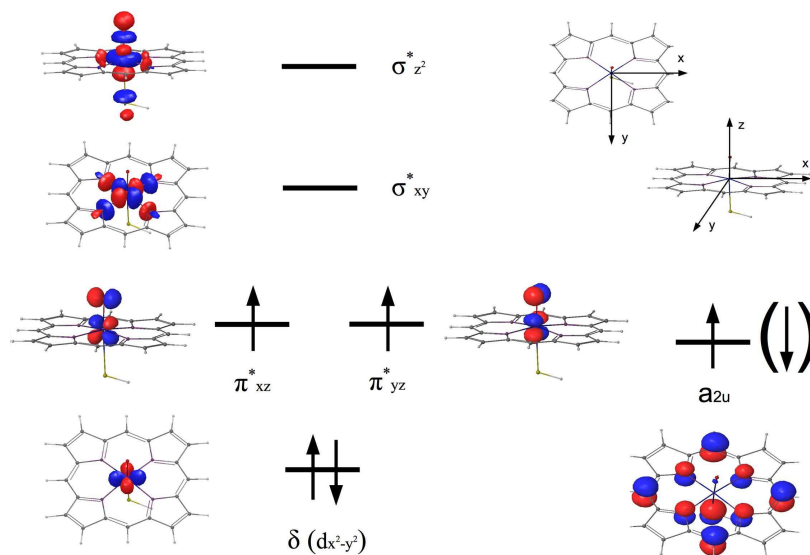
Departamento de Química-Física y Analítica, Universitat Jaume I, 12071, Castelló, Spain

[b] S. Berski

Faculty of Chemistry, University of Wrocław, F. Joliot-Curie 14, 50-383 Wrocław, Poland
E-mail: smarti@uji.es

Contract/grant sponsors: Spanish Ministerio de Ciencia e Innovación (CTQ2009-14541-C02), Generalitat Valenciana Prometeo/2009/053 project, Fundación Bancaixa-UJI for financial support during S. Berski research stay at University Jaume I (UJI).

© 2012 Wiley Periodicals, Inc.



Scheme 2. Molecular orbitals diagram of doublet and quartet spin states for Compound I. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

cytochrome P450. However, very recently, Rittle and Green prepared Cpd I in a high concentration on CYP119 P450 cytochrome, being able to confirm the ability of this specie to produce the oxidation of unactivated C–H bonds,^[10] confirming the important role of the Cpd I as active oxidant in P450 enzymes.

The Cpd I is a radical and it is well known that there are two low-lying spin states (doublet and quartet), due to the ferromagnetic/antiferromagnetic coupling between a triplet formed by the Fe=O moiety and a doublet formed by an odd electron, which has a porphyrinic nature and is delocalized toward the sulfur atom of the ligand. The accepted electronic structure of this compound (Scheme 2) consists in a π -antibonding triplet formed by the ferryl Fe^{IV}=O moiety (π^*_{xz} and π^*_{yz} orbitals) and a π character a_{2u} radical orbital which can couple ferromagnetically to provide a high-spin (HS) state with $S = 3/2$ or antiferromagnetically to provide a low-spin (LS) state with $S = 1/2$, being both states almost degenerated.^[11–13] The role of the axial ligand in the electronic structure of Cytochrome P450 Cpd I has been studied by several researchers using different kinds of thiolate models: SH, SMe, SCys, or Cysteinate, concluding that the ground state of this compound depends on the nature of the thiolate used due to the donor character of the sulfur of each model and the interaction of that atom with the environment.^[13] In fact, in the experiments where the SMe^[14] and S-cys^[15] were chosen as axial ligands, it was observed that the spin density of the unpaired radical resided mainly in a lone-pair of the sulfur atom of the ligand, being scarce the contribution of the porphyrin ring. Conversely, in the cases where the SH and Cysteinate models (which has internal hydrogen bonding) were chosen,^[16–18] it was observed that the a_{2u} porphyrin orbital is singly occupied being the spin density distributed between

the porphyrin and the sulfur atom of the axial ligand. As a result of the dual character in the spin state, Shaik et al.^[19] proposed that a two state reactivity (TSR) mechanism is operative along the most common reactions of Cpd I of Cytochrome P450, which are the hydroxylation of alkanes and the epoxidation of alkenes.

Since the end of the 1990s with the pioneer work of Loew,^[20–23] several theoretical studies have been carried out to calculate the electronic structure and spectra of the Cpd I, both in gas phase and in the enzyme environment. Very recently, Shaik et al.^[13] have published a review on the theoretical works devoted to study the structure, reactivity, and selectivity on different P450 enzymes. In this review,

the electronic structure of Cpd I is analyzed by means of different levels of theory, such as the orbital approach using density functional theory (DFT) calculations, multireference *ab initio* methods, as well as the valence bond theory.

The relationship between charge density topology and physical/chemical properties can be understood from the Hohenberg–Kohn theorem, which asserts that a system's ground-state properties are a consequence of its charge density, a scalar field denoted as $\rho(r)$.^[24] Indeed, Bader^[25] noted that the essence of a molecule's structure must be contained within the topology of $\rho(r)$. Thus, the quantum theory of atoms in molecules (QTAIM or in short AIM) developed by Bader et al.^[25] is a method based on the topology of the electron density $\rho(r)$, a physical observable that can be determined from electronic structure theory calculations or experimentally, and its Laplacian, $\nabla^2\rho(r)$, and it has enjoyed great success as a tool to analyze and rationalize chemical bonding in molecules. In addition, real-space partitioning of the molecular space may also be achieved by using functions of the electronic density and/or its derivatives. Conversely, the electron localization function (ELF), $\eta(r)$, proposed by Becke and Edgecombe^[26] in 1990, allows for the mapping of electron pair probability in multielectron systems based on the electron density. The topological analysis of $\eta(r)$, proposed in 1994 by Silvi and Savin,^[27–29] permits to perform a quantitative analysis of the three-dimensional function, providing useful information on the electronic structure. Both scalar fields can be considered within the quantum chemical topology framework^[30] as proposed by Popelier et al.^[30,31] to describe accurately chemical concepts of atoms, bonding, and molecular structure.

Topological analysis of the $\eta(r)$ function has been already used to study iron porphyrins. Novozhilova et al.^[32] and Xu

et al.^[33] illustrated “ferrous” {FeNO}⁷⁻-like electron pair localization on a nitrosyl N-atom in a “ferric” {FeNO}⁶⁻ porphyrin. In the Fe–NO region, the disynaptic bonding basin $V(\text{Fe},\text{N})$ has been observed which corresponds to the Fe–N bond and the $V(\text{N},\text{O})$ disynaptic bonding basin characterizing nitrogen–oxygen bond. Furthermore, this methodology has been applied to study the nature of metal–oxygen bond in other systems,^[34,35] which feature a high ionic character in this kind of bonds, due to the lack of disynaptic basins between both atom cores.

In this article we investigate, for the first time, the electronic structure of Cpd I by using the topology of both ELF, $\eta(r)$, and electron density, $\rho(r)$. This allows us to broaden the spectrum of information known about the nature of this compound with so much biological interest, providing new insight into the nature of the bonds and the electronic distributions of its two most characteristic spin states: doublet and quartet.

Computational Details

Full geometry optimizations of the $\text{Por}^+ \text{Fe}^{\text{IV}}=\text{O}(\text{S}-\text{H})$ complex in doublet and quartet electronic states have been performed using the unrestricted formalism with the UB3LYP density functional as well as the LACVP* basis set^[36] (consisting of the combination of the 6-31G(d) basis for all the atoms except for the iron one, which was represented by the LANL2DZ effective core potential). In addition, single-point energies of the resulting structures have been carried out at the UB3LYP/6-31G(d) level to obtain an accurate wave functions, avoiding the use of pseudopotentials on the iron atom. Conversely, geometrical structures of isolated pyrrole, pyrrole anion, and porphyrin ring have been optimized at the B3LYP/6-31G(d) level. With the aim of obtaining a more accurate description of porphyrin, the geometry optimization and topological analysis of $\eta(r)$ have been repeated at B3LYP/6-311++G(d,p) level. The minima on the potential energy surface have been confirmed by nonimaginary vibrational frequencies. All the calculations were performed using the Gaussian 03 program.^[37]

The ELF function [$\eta(r)$] provides regions of the space, where the probability of finding an electron pair in the molecular space is high. The topological analysis of this function yields basins of attractors corresponding to atomic cores, bonds, and lone pairs. The atomic cores coincide with the atomic nuclei and are labeled as $C(A)$, being A the atomic symbol of the element. The concept of *synapticity*, introduced by Silvi et al.^[27,38] allows the classification of η -localization basins in the $\eta(r)$ domain. In this way, it is possible to find monosynaptic basins, labeled as $V(A)$, which correspond to lone pairs of the Lewis model, as well as disynaptic basins labeled as $V(X,Y)$ which correspond with the η -localization basins with common surfaces between two $C(X)$ and $C(Y)$ core basins. Conversely, trisynaptic basins labeled as $V(X,Y,Z)$ correspond to the three-center bonds, as so on. The integration of the electron density over the different basins yields to the basin population, \bar{N} , allowing the quantification of possible charge transfer phenomena. In the present work, the topological analysis of the ELF $\eta(r)$ and electron density $\rho(r)$ has been carried out using Dgrid-4.5 program^[39] over a rectangular parallelepiped grid with a step of 0.05 bohr. Graphical representations have been generated by means of the

Jmol,^[40] visual molecular dynamics (VMD),^[41] and UCSF Chimera package^[42] programs.

Results and Discussion

The optimized geometry of the Cpd I model compound, the isolated porphyrin, the isolated pyrrole, and the pyrrole anion are depicted in Figure 1. As can be observed the C2–C3 and C4–C5 distances are shorter than the C1–C2 lengths in both the Cpd I and the isolated porphyrin, which means a larger bond order in the former bonds, thus exhibiting a large amount of double bond character. However, this behavior differs from the one observed in both the isolated pyrrole and the pyrrole anion, where the C2–C3 bond length is larger than the C1–C2 or C3–C4 ones (see Figs. 1c and 1d). Since the porphyrin ring can be seen as four pyrrole rings connected by methine bridges (=CH–), the splitting of this large macrocycle into its subunits leads to changes in the bond lengths and thus in the degree of delocalization. With respect to the C–N distances, the values obtained for Cpd I indicate a partial double bond character as well, similar to those found in the isolated porphyrin, the isolated pyrrole, and in the pyrrol anion ring.

Furthermore, all the values for the Fe–N distances of Cpd I (~2.02 Å) point out the fact that Fe–N interactions should be of the same nature. In Figure 1b, the Fe–O and Fe–S distances as well as the S–Fe–O angle are also depicted. The Fe–O distance in both spin states is around 1.63 Å, which is in good agreement with experimental x-ray absorption spectroscopy data found in the literature for Cpd I (1.65 Å).^[43] However, the Fe–S bond exhibits a larger distance (2.60 Å) than the found experimentally (2.48 Å using EXAFS spectroscopy^[43]). Nevertheless, this result is comparable to other theoretical studies, which also makes use of the SH axial ligand and DFT methodology, indicating that the strong interactions between the thiolate and oxo ligands provokes a large value for the Fe–S bond.^[5,14,18] Finally, the value of the S–Fe–O angle shows an almost colinear distribution of both the proximal SH and the distal Oxo ligands (~171° in average).

The electronic structure of Cpd I, in both doublet and quartet electronic states, is characterized by 76 local maxima (attractors) localized for the field of ELF $\eta(r)$. The spatial localization of all attractors is presented at Figure 2. The core electron density is characterized by a total of 27 core attractors $C(\text{Fe})$, $C(\text{S})$, $C(\text{O})$, $C(\text{N})_{i=1-4r}$, and $C(\text{C})_{i=1-20}$ which coincide with positions of atomic nuclei. Conversely, in the valence shell, the electrons are associated with three monosynaptic nonbonding basins (attractors): $V(\text{S})$ and $V_{i=1,2}(\text{O})$ corresponding to the lone pairs of the sulfur and oxygen atoms, respectively. Furthermore, a total of 33 valence disynaptic bonding attractors can be observed: $V(\text{C},\text{C})_{i=1-20r}$, $V(\text{C},\text{N})_{i=1-8r}$, $V(\text{Fe},\text{S})$, and $V(\text{Fe},\text{N})_{i=1-4r}$, corresponding to the C–C, C–N_{*i=1-8r*}, Fe–S, Fe–N_{*i=1-4r*} chemical bonds, respectively. The C–H and S–H bonds are reflected by 12 protonated disynaptic attractors $V(\text{H},\text{C})_{i=1-12}$ and one $V(\text{H},\text{S})$ attractor. According to interpretation proposed by Silvi and Savin,^[27] the localization of the valence bonding attractors $V(A,B)$ yields proof that considered A–B bonds are of the covalent type.

It is worth noting, that the same number and type of attractors are localized for the Cpd I system in both the doublet

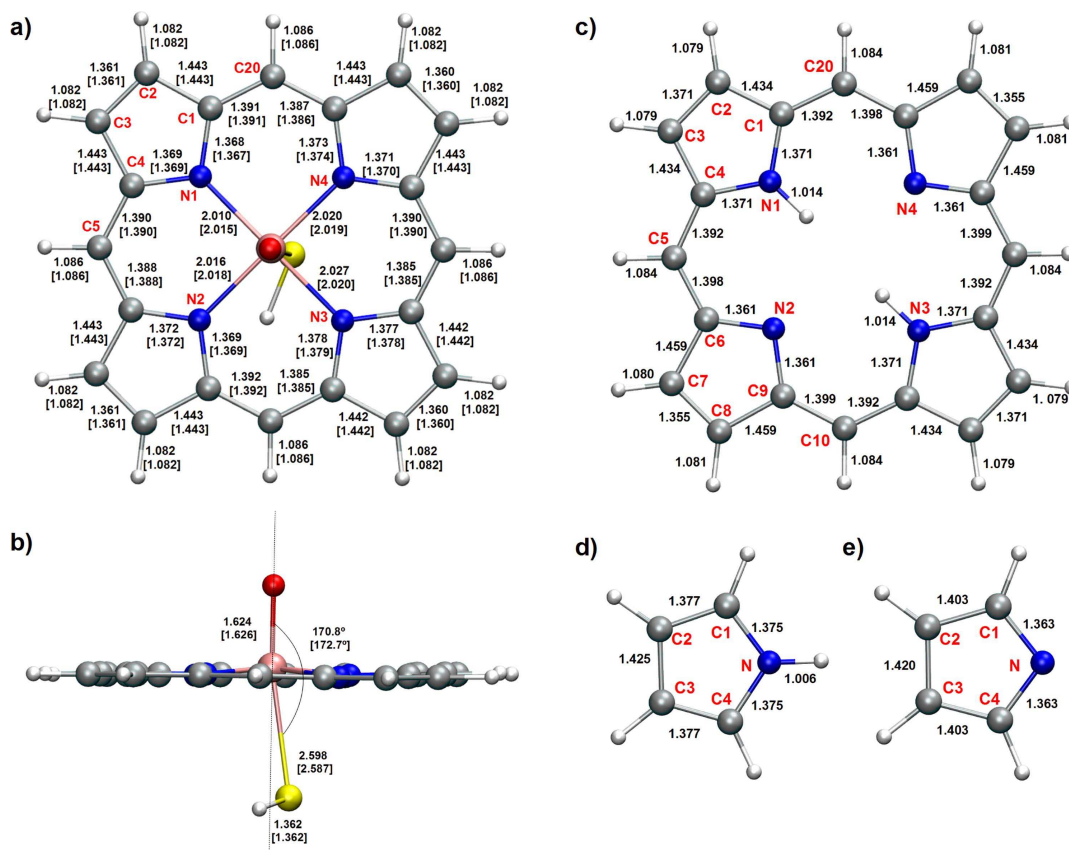


Figure 1. Geometrical parameters for: a) and b) the optimized Cpd I, in doublet and [quartet] spin states, c) the porphyrin ring, d) the pyrrole molecule, and e) the pyrrole anion. The distances are expressed in angstroms and angles in degrees. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

and quartet electronic states. The unique difference is the slightly different spatial orientation of the $V_{i=1,2}(O)$ attractors regarding the Fe—N bonds and the porphyrin plane which can be associated with occupation of different orbitals in both electronic states as shown in Figure 2.

The core attractors $C(Fe)$, $C(N)_{i=1-4}$, $C(C)_{i=1-20}$ and the valence attractors $V(H,C)_{i=1-12}$, $V(C,C)_{i=1-20}$, $V(C,N)_{i=1-8}$, $V(Fe,N)_{i=1-4}$ reflect the electronic structure of the iron-porphyrin complex: the planar geometry of porphyrin is associated with the aromatic character originated by the conjugation of the double C=C and C=N bonds. As a matter of fact, all attractors are observed lying approximately within the porphyrin plane. Finally, the lack of the $V_{i=1,2}(C,C)$ or $V_{i=1,2}(C,N)$ attractors above and below the macrocycle ring plane, resembles the topology of $\eta(r)$ observed in other molecules with large delocalization of the electron density, such as for instance benzene^[44] or pyridine.^[45] The porphyrin ring in Cpd I is built from four pyrrole subunits, where the hydrogen atoms from two N—H bonds (in parent porphyrin) have been removed. Thus, the differences between the electronic structure of Cpd I and its subunits may be obtained from the comparison with

the electronic structure of isolated pyrrole (C_4H_4NH), pyrrole anion ($[C_4H_4N]^-$), and porphyrin ring. The core and valence attractors localized in the three molecules are shown in Figure 3, whereas the basin populations are presented in Table 1.

The topology of the $\eta(r)$ function for the pyrrole anion—considering a number of attractors and their synpacticity^[38,46]—is the same as observed for the Cpd I. In both C_4H_4NH and $[C_4H_4N]^-$ molecules the valence attractors, corresponding to delocalized C=C and C—N bonds, are found in the molecular plane. The main difference is found for regions where nitrogen lone pairs are expected. The nonbonding electron density of the N atom in C_4H_4NH is reflected by two nonbonding $V_{i=1,2}(N)$ attractors localized below and above molecular plane, meanwhile in the $[C_4H_4N]^-$ anion only single nonbonding attractor $V(N)$ in the plane is localized.

The basin population of $V(N)$ increases from 1.08e for C_4H_4NH to 3.03e for $[C_4H_4N]^-$, value which is very similar to those calculated for the $V(Fe,N)_{i=1-4}$ basins in Cpd I (3.04e). This fact supports the observation that the $V(Fe,N)_{i=1-4}$ basin consists mainly of electrons from the lone pair of the nitrogen atom. However, the distance between the $V(Fe,N)_{i=1-4}$ and

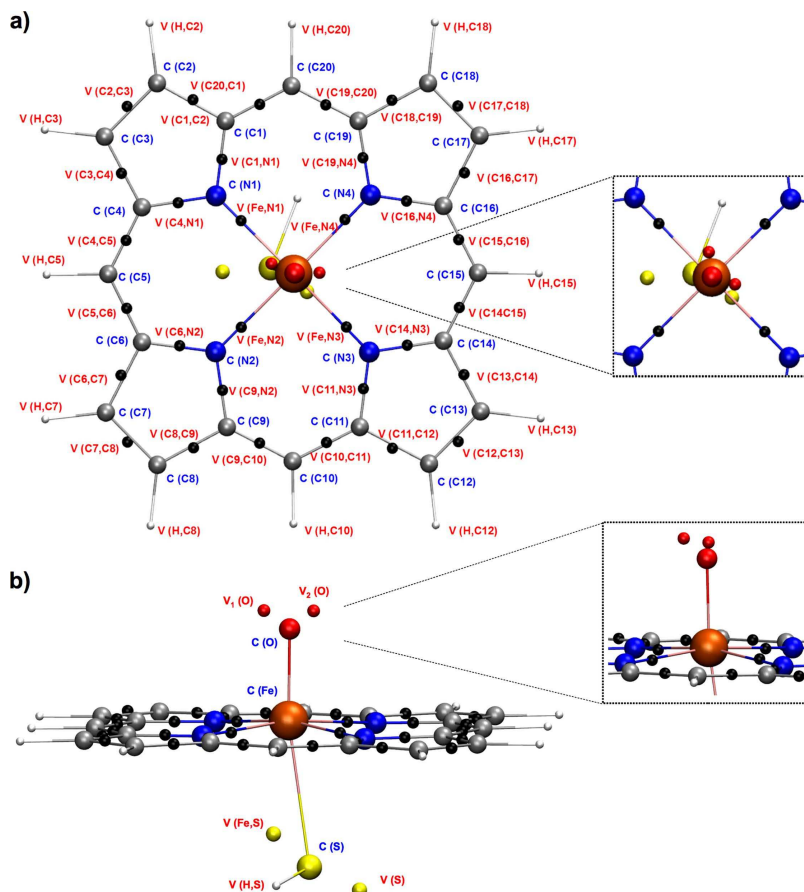


Figure 2. Local maxima attractors localized for the field of electron localization function $\eta(r)$ in Cpd I. Core attractors are depicted in blue and valence attractors in red. a) Attractors in porphyrin ring in doublet state. b) Attractors localized for iron axial ligands in double state. Small boxes represent quartet state. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

$C(N)_{i=1-4}$ attractors in Cpd I (0.64–0.66 Å) are shorter than between the $V(N)$ and $C(N)$ attractors in the pyrrole anion (0.73 Å). This fact may be explained as result of a Pauli repulsion between electrons “contained” in the $C(Fe)$ and $V(Fe,N)_{i=1-4}$ basins and the corresponding “compression” of the electron density in this region.

The electronic structure of isolated Por obtained using topological analysis of $\eta(r)$ is shown in Figure 3c. An interesting topological feature is found for the C2–C3 and C12–C13 bonds of the pyrrole anion fragments, where pairs of the bonding disynaptic attractors $V_{i=1,2}(C,C)$ are localized. The same topological feature has been achieved at B3LYP/6-311++G(d,p) computational level. It is worth emphasizing that in isolated pyrrole anion only a single disynaptic attractor is found (Fig. 3b). The presence of two bonding $V_1(C,C)$ and $V_2(C,C)$ attractors suggests that these carbon–carbon bonds have double bond character, and the degree of electron delocalization is smaller than for other C–C bonds. As showed by

Silvi et al.,^[27] the localization of two $V_{i=1,2}(C,C)$ attractors is typical for compounds with formal C=C bonds and has been described using the synapticity concept for the first time in ethylene, propene, and trans-butadiene.^[46] The double bond character is also supported by a comparison of the bond length in the porphyrin (1.36 Å), pyrrole (1.43 Å), and pyrrole anion (1.42 Å), which shows that the C=C bond in the porphyrin is the shortest one. Since all C–C bonds in Cpd I are represented by single $V(C,C)_{i=1-20}$ attractors, the $Por^+ Fe^IV=O(S-H)$ complex formation would result in an increase of the electron density delocalization over the molecule. Thus, a number of resonance hybrids, based on Lewis formula, have to be considered. The bond order for delocalized bonds should be between 1 and 2, and the values of the basin populations support this expectation, being in the range 2.39e–3.29e, and topological bond orders between 1.2 and 1.6, in both electronic states (see Supporting Information Tables S1 and S2).

Three groups of carbon–carbon bonds with very similar values of the basin population (Fig. 4) are found. These values may be associated with the localized nature of the bonds. Furthermore, the topological analysis of ELF shows that the perturbation exerted by the HS ligand in the Cpd I, which lacks of the C4 symmetry axis present in the porphyrin ring, is quite small providing almost the same basin populations in each of the four Cpd I fragments.

The largest values of \bar{N} , (3.30e) are obtained for the C2,C3; C7,C8; C12,C13; and C17,C18 pairs of atoms (I group) in the pyrrole subunits. In fact, the resonance hybrids which are usually proposed to illustrate the delocalized character of the porphyrin bonds suggest a dominant contribution of the C=C bonding. It is worth noting that these bonds have a clear double bond character in the isolated porphyrin, as long as they are represented by two $V_{i=1,2}(C,C)$ attractors. Conversely, the carbon–carbon bonds of the four methine fragments are slightly less populated, with basin populations in the range 2.88e–3.01e (II group). As result, the contribution of resonance

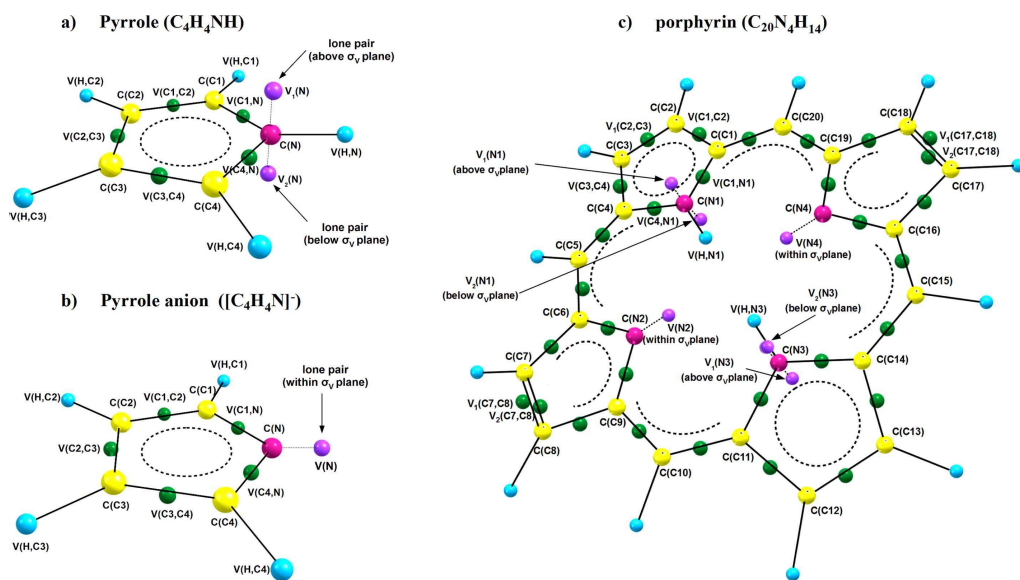


Figure 3. Core and valence attractors localized for pyrrole a), pyrrole anion b), and porphyrin ring c). [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

hybrids with double C=C bonds is smaller than postulated for I group. The smallest values of \bar{N} are found for the carbon-carbon bonds in the pyrrole subunits (III group) connecting the group I bonds with the methine "bridges" (=CH-). Their basin populations are approximately 2.39e, values that can be interpreted in terms of the predominant single C-C character. In the same way, those carbon-nitrogen bonds with the smallest basin populations, equal to 2.15e–2.17e, may be characterized as single type C-N bonds. In spite of the essential contribution of the C=N double bond character suggested by the resonance hybrids, the valence electrons are "concentrated" in the disynaptic $V(\text{Fe,N})_{i=1-4}$ of the Cpd I, equivalent to the pyrrole anion $V(\text{N})$ basins, being their value equal to 3.04e in average, which is much larger than the 2e formal value. This fact can be explained analyzing the covariance matrix of the ELF function. In this case, we observe a large cross term in average of the localization

indexes among the $V(\text{Fe,N})_{i=1-4}$ and the corresponding $V(\text{C,N})_{i=1-8}$ (0.36e each, see Supporting Information Table S3). This would mean that the porphyrin ring would be donating electrons to the Fe-N dative bonds.

The presence of these four $V(\text{Fe,N})_{i=1-4}$ basins localized between Fe and four N atoms (see Figs. 2 and 5) may be interpreted as covalent bonds. Topographical analysis of $\eta(r)$ function (2D map, Fig. 5), performed at the molecular plane of Cpd I ($M = 2$) shows that the valence domains observed between the iron and nitrogen core domains are well separated, and the values of $\eta(r)$ function in the $\text{Fe}\cdots\text{N}_{i=1-4}$ regions approach to zero (blue color at Fig. 5). Besides, the degree of electron localization in those regions is below 0.5e, the value corresponding to the free electron gas. Furthermore, the observed separation between 2D-domains indicates that the region of the N lone pair should be associated with a monosynaptic nonbonding basin $V(\text{N})_{i=1-4}$ and interpreted as a lone pair. In fact, a lack of a bonding disynaptic basin implies that the nature of the iron-nitrogen bond is governed by electrostatic interactions, $\text{Fe}^{\delta+}\cdots\text{N}^{\delta-}$, and is not originated in a covalent bonding. However, a 2D map of the $\eta(r)$ function does not clearly shows the localization basins but only a distribution of the $\eta(r)$ values. Therefore, a more accurate description must be performed by means of localization basins: the four nitrogen cores $\text{C}(\text{N})_{i=1-4}$, the iron core $\text{C}(\text{Fe})$, and the valence basins localized between the Fe and N cores, as

Table 1. Basin populations of the pyrrole C₄H₄NH, pyrrole anion [C₄H₄N]⁻, porphyrin and pyrrole fragment (C₄H₂N) in Cpd I ($M = 2, 4$); as a result of the topological analysis of electron density $\rho(r)$ and electron localization function $\eta(r)$.

Basin ^[a] /molecule	C ₄ H ₄ NH	[C ₄ H ₄ N] ⁻	Porphyrin	Cpd I ($M = 2$)	Cpd I ($M = 4$)
C(N)	2.11e	2.11e	2.11e	2.13e	2.14e
$V_{i=1,2}(\text{N})$	0.54e/0.54e ^[b]	3.03e ^[c]	0.53e ^[b] /3.04e ^[c]	3.04e ^[d]	3.03 ^[d]
$r[V_{i=1,2}(\text{N})\cdots\text{C}(\text{N})]$	0.65 Å/0.65 Å ^[e]	0.73 Å ^[e]	0.65 Å/0.72 Å ^[e]	0.65 Å ^[f]	0.64 Å ^[f]
$V(\text{C1,C2});V(\text{C3,C4})$	3.30e/3.30e	3.34e/3.34e	2.38e/2.51e	2.39e	2.39e
$V(\text{C2,C3})$	2.54e	2.65e	3.11e/1.61e	3.30e	3.29e
$V(\text{C1,N1}); V(\text{C4,N1})$	2.18e/2.16e	2.18e/2.15e	2.23e/2.19e	2.16e	2.16e

[a] Numbering of atoms and basins according to Figures 2 and 3. [b] Nonbonding basin $V(\text{N})$ localized in both above and below the molecular plane. [c] Single nonbonding basin $V(\text{N})$ localized in the molecular plane. [d] Bonding disynaptic basin $V(\text{Fe,N})$. [e] Distance between the core attractor of nitrogen $\text{C}(\text{N})$ and nonbonding monosynaptic attractor $V(\text{N})$. [f] Distance between the core attractor of nitrogen $\text{C}(\text{N})$ and bonding disynaptic attractor $V(\text{Fe,N})$.

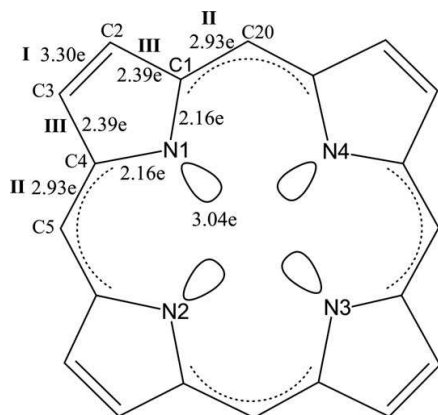


Figure 4. ELF population analysis for Cpd I porphyrin ring. The different carbon–carbon bond groups (I, II, and III), depending on their population values, are also depicted.

presented in Figure 6. Certainly, each valence basin between N and Fe belongs to a disynaptic $V(\text{Fe},\text{N})_{i=1-4}$ type, as long as it has a common surface with both the $C(\text{N})_{i=1-4}$ and $C(\text{Fe})$ core basins (see Fig. 6c). Thus, the analysis of the synaptic order implies that the iron–nitrogen bonding belongs to the covalent-dative type, as long as the distribution of ELF values shows that the electron density of the bond is mainly formed by the nitrogen lone pair.

A very similar topological structure is observed for the localization basin associated with the valence attractor localized between the $C(\text{Fe})$ and $C(\text{S})$ cores (see Fig. 5). The 2D representation of $\eta(r)$ presented at Figure 5b shows that the domain corresponding to the Fe–S bonding is well isolated from the iron core domain. Most probably it is almost entirely formed by the nonbonding electron density of the sulfur. However, its localization basin has a common surface with $C(\text{Fe})$ core basin (Fig. 6b); and therefore, this valence basin has bonding nature also being disynaptic $V(\text{Fe},\text{S})$, and the covariance matrix of the ELF shows a localization index of 0.45e for the $V(\text{S})$ and 0.25e for the $V(\text{H},\text{S})$. Thus, the sulfur–iron bond, similarly to iron–nitrogen one, has a topological feature of covalent dative character.

Additional information on the nature of the sulfur–iron interaction can be obtained from the comparison with simple molecules (H_2S , HS^- anion, and HS^\cdot radical), containing sulfhydryl group. The results of topological analysis of the $\eta(r)$ and $\rho(r)$ functions, performed for B3LYP/6-31G(d) optimized geometrical structures, are shown in Table 2 and Supporting Information Figure S4. In the $[\text{HS}]^-$ and $[\text{HS}]^\cdot$ molecules, the lone pairs of sulfur are represented by single nonbonding basin $V(\text{S})$ and such topology is similar to that observed for Cpd I. The total basin population of the Fe–S bond and $V(\text{S})$ valence basins, $N[V(\text{Fe},\text{S}) + V(\text{S})]$, in the complex is 5.43e for both $M = 2$ and $M = 4$, respectively, which is smaller than the computed for the $[\text{HS}]^-$ anion (6.12e) but slightly larger than for the HS^\cdot radical (5.17e). The population of the H–S bond in Cpd I (1.84e) is similar to the H_2S , $[\text{HS}]^-$ and $[\text{HS}]^\cdot$ molecules. Thus, the H–S bond may be considered as a single bond from a topological

point of view. Anyway, the most interesting finding is obtained when comparing the positions of the $V(\text{Fe},\text{S})$ and $V(\text{S})$ point attractors in Cpd I and in the $[\text{HS}]^-$, $[\text{HS}]^\cdot$, H_2S molecules. The distance between the $C(\text{S})$ and $V(\text{Fe},\text{S})$ attractors is shorter than between the $C(\text{S})$ and $V(\text{S})$ attractors in H_2S , $[\text{HS}]^-$ and $[\text{HS}]^\cdot$. This implies that the disynaptic $V(\text{Fe},\text{S})$ basin does not seem to be a standard Fe–S covalent bond, but rather a $V(\text{S})$ basin lone pair. This result confirms the concept of dative $\text{S} \rightarrow \text{Fe}$ bond. In addition, the effect of the lone pair ‘compression’ is very similar to that one observed for the N atoms. The same kind of conclusion can be derived from the analysis of the localizations basins, which shows the presence of common surfaces among the disynaptic $V(\text{Fe},\text{S})$ and the $C(\text{Fe})$ and $C(\text{S})$ cores (see Fig. 6b).

The oxygen atom, formally bound to iron atom by the double $\text{Fe}=\text{O}$ bond is described by the core $C(\text{O})$ attractor and two monosynaptic nonbonding attractors $V_1(\text{O})$ and $V_2(\text{O})$. These last two attractors reflect the nonbonding electron density of the oxygen, which according to the symmetry of the complex is represented by two local maxima. Such topology of $\eta(r)$ stays in agreement with the Lewis structure (see

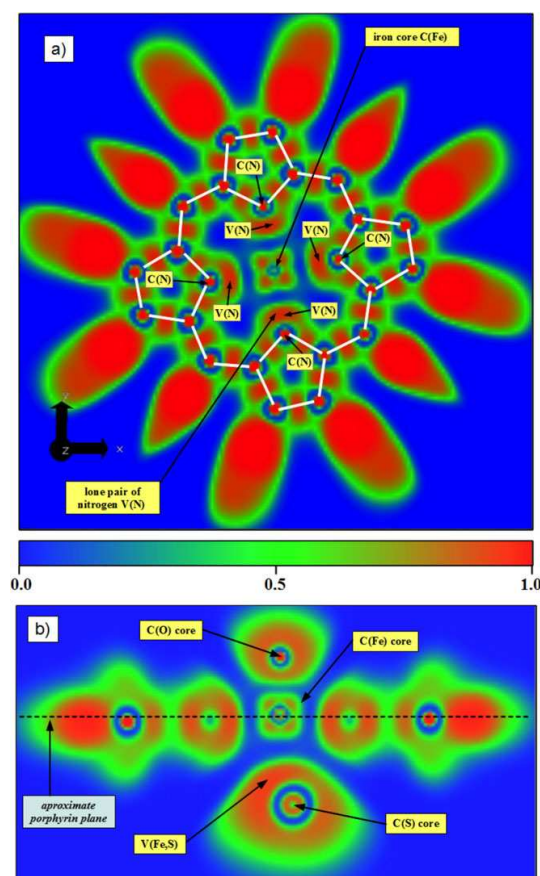


Figure 5. Two different 2D plots of the ELF function of Cpd I at the porphyrin ring plane a) and perpendicular to this plane b).

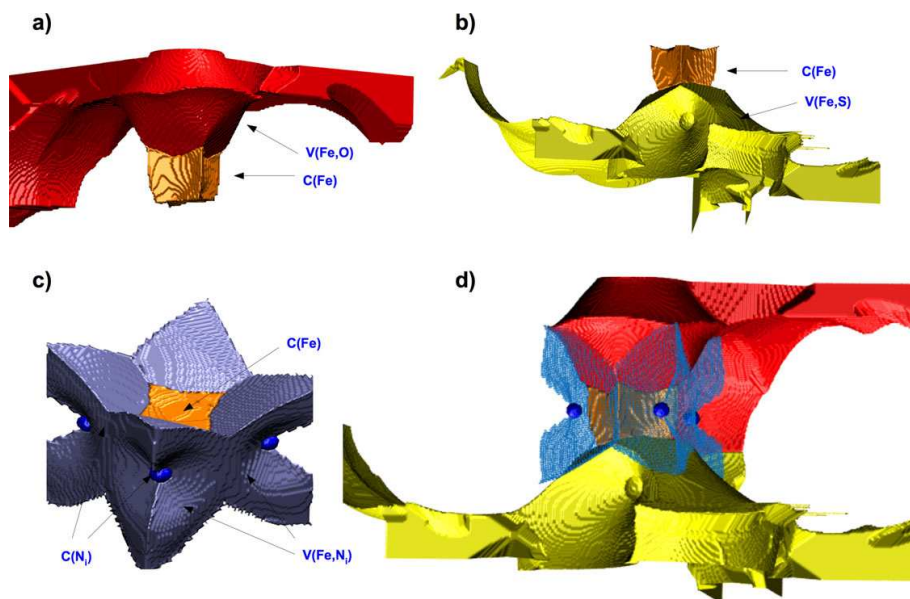


Figure 6. 3D manifolds of the iron core C(Fe) with the different valence basins of the different ligands: a) the oxygen basin V(Fe,O), b) the sulfur basin V(Fe,S), c) the four nitrogen cores C(Ni = 1, 4) and basins V(Fe,Ni), and d) a combined representation of the C(Fe) core and the ligand basins.

Scheme 1) predicting the double Fe=O bond with two lone pairs on oxygen. Furthermore, is very similar to that observed for the simplest carbonyl compound $\text{H}_2\text{C}=\text{O}$ ^[47–49] or in molecules with carbon monoxide ligands.^[50] The positions of the $V_1(\text{O})$ and $V_2(\text{O})$ attractors (see Fig. 2a) suggest that the electron cloud of valence shell of the O atom is polarized to minimize the Pauli repulsion with the electron clouds associated with the $V(\text{N})_{i=1-4}$ localization basins.

It is worth emphasizing that in the Fe...O region, where the double bond Fe=O is expected, no bonding attractor is observed (see Fig. 2b). Therefore, a clear difference between the iron–oxygen bonding and both iron–nitrogen and iron–sulfur bonding can be observed. Therefore, the covalent bond (Fe–O) or covalent-dative bond (O→Fe) between Fe and O is missing and, from the point of view of the ELF topology analysis, the nature of the binding stems mainly from $\text{Fe}^{\delta+}\cdots\text{O}^{\delta-}$ electrostatic interactions. Similar results have been reported for other metal–oxygen bonds in the literature.^[34,35]

The analysis of the synaptic order performed for the valence basins of the oxygen atom shows (see Fig. 6a) that they also have common surfaces with the iron core and might be classified to the disynaptic type V(Fe,O). However, this result does not testify about covalent character of the bonding since the basins (and attractors) are clearly associated with the lone pairs (see Fig. 2b). Nevertheless the analysis of the

covariance matrix of the ELF calculations reveals the presence of 0.43e localized among the iron and oxygen lone pairs (see Supporting Information Table S3). This fact could be an indicator of a large fluctuation of the electronic density between these basins, which in turns could indicate the presence of a charge-shift bond.^[51]

Finally, the mean electron populations calculated for selected core and valence basins are presented in Table 3. The iron core, represented by the C(Fe) basin, contains 23.72e ($M = 2, 4$) and a value of the alpha spin density ($n_\alpha - n_\beta$) of 1.24e ($M = 2, 4$). The analysis of the spin density calculated for the η -basins between the doublet to quartet electronic states, shows different results. The values of the basin population and spin densities summed up for different fragments of the $\text{Por}^+\text{Fe}^{\text{IV}}=\text{O}(\text{S}-\text{H})$ complex are also showed in Table 3. On one hand, there is a lack of spin density change for the iron core, oxygen atom, oxygen–iron bond, and the nitrogen lone pairs (represented by the $V(\text{Fe},\text{N})_{i=1-4}$ basins). Thus, local electronic

Table 2. Basin populations of for the hydrogen sulfide H_2S , hydrosulfide ion $[\text{HS}]^-$, hydrosulfide radical $[\text{HS}]^\cdot$, and HS fragment in Cpd I ($M = 2, 4$); as a result of the topological analysis of electron density $\rho(r)$ and electron localization function $\eta(r)$.

Basin ^[a] /molecule	H_2S	$[\text{HS}]^-$	$[\text{HS}]^\cdot (M = 2)$	Cpd I ($M = 2$)	Cpd I ($M = 4$)
C(S)	10.07e	10.09e	10.06e	10.08e	10.08e
$V_{i=1,2}(\text{S})$	2.15e ^[b] /2.15e ^[b]	6.12e ^[b]	2.59e ^[b] /2.58e ^[b]	2.75e ^[b] /2.68e ^[c]	2.77 ^[b] /2.66e ^[c]
$r[V_{i=1,2}(\text{S})\cdots\text{C}(\text{S})]$	0.99 Å ^[d]	0.98 Å ^[d]	0.99 Å ^[d]	0.97 Å ^[d] /0.95 Å ^[e]	0.96 Å ^[e] /0.96 Å ^[e]
$V_{i=1,2}(\text{H},\text{S})$	1.81e/1.82e	1.80e	1.77e	1.84e	1.83e

[a] Numbering of atoms and basins according to Figure 2. [b] Single nonbonding basin V(S). [c] Disynaptic bonding basin V(Fe,S). [d] Distance between the core attractor of sulfur C(S) and attractor of the basin $V_1(\text{S})$. [e] Distance between the core attractor of nitrogen C(S) and bonding disynaptic attractor V(Fe,S).

Table 3. The approximate total electron populations and spin electron densities [e] calculated for different fragments of the Cpd I ($M = 2, 4$) using the basin populations obtained from topological analysis of ELF function.

Molecular fragment/multiplicity	$M = 2$		$M = 4$	
	$n_x + n_\beta^{[a]}$	$n_x - n_\beta^{[b]}$	$n_x + n_\beta$	$n_x - n_\beta$
Sulfur-hydrogen bond [S—H] δ^- -C	17.35	-0.55	17.34	0.52
(S)∪V(Fe,S)∪V(S)∪V(H,S)	$\delta = -1.35$		$\delta = -1.34$	
Iron atom [Fe] $\delta^{[c]}$	23.72	1.24	23.72	1.18
C(Fe)	$\delta = +2.28$		$\delta = +2.28$	
Oxygen atom [O] δ^-	8.90	0.74	8.88	0.80
C(O)∪V ₁ (O)∪V ₂ (O)	$\delta = -0.90$		$\delta = -0.88$	
Oxygen-iron bond [Fe=O] δ^{+}	32.62	1.98	32.60	1.98
C(Fe)∪C(O)∪V ₁ (O)∪V ₂ (O)	$\delta = +1.38$		$\delta = +1.40$	
Nitrogen N1—N4 cores and four dative N→Fe bonds	20.68	-0.22	20.68	0.20
C(N1)∪V(Fe,N1)∪C(N2)∪V(Fe,N2)∪C(N3)∪V(Fe,N3)∪C(N3)∪V(Fe,N4)	$\delta = -4.68$		$\delta = -4.68$	
Porphyrin fragment	161.44	-0.43	161.52	0.50
[C ₂₀ H ₁₂ N ₄] δ^{--}	$\delta = -1.44$		$\delta = -1.52$	

[a] Total number of electrons. [b] Spin density. [c] Net charge δ calculated regarding to formal population [Ar]3d⁶4s².

structure of these fragments is not affected by the electronic state. However, the essential change of spin density is observed for the porphyrin skeleton and the sulfhydryl group (S—H) since the value of $n_x - n_\beta$ difference changes from $-0.43e$ to $0.50e$ and $-0.55e$ to $0.52e$, respectively, when comparing the doublet with the quartet state. This is an expected result, confirming the radical nature of the porphyrin ring, which leads to the different multiplicities of the Cpd I.

Conclusions

For the first time, an analysis of the electronic structure of the porphyrin-iron complex Cpd I has been performed by means of the topological analysis of the ELF and electron density. Without invoking the concept of molecular orbital, we have been able to describe the nature of the chemical bonds in real space, and depict the main differences between two spin states, doublet, and quartet ($M = 2, 4$). The main conclusions can be summarized as follows:

1. A difference between the doublet and quartet spin states of Cpd I is seen through different positions of the $V_1(O)$ and $V_2(O)$ nonbonding attractors, in respect to the bonds of the porphyrin. Although the same number of valence and core attractors with the same synaptic order are observed.

2. The change of the spin state from doublet to quartet leads to a change of the spin density mainly in the porphyrin fragment and the sulfhydryl group.

3. The delocalized network of carbon-carbon and carbon-nitrogen bonds in the porphyrin macrocycle is reflected by the point attractors $V(C,C)_{i=1-20}$ and $V(C,N)_{i=1-8}$ localized approximately in the plane of the porphyrin. Since the topological analysis of $\eta(r)$ does not show pairs of the $V_{i=1,2}(C,C)$ and $V_{i=1,2}(C,N)$ attractors (above and below the porphyrin plane), the topological signature for the double C=C and C=N bonds are, therefore, not observed.

4. The C=C bonds in the pyrrole anion fragments [C₄H₂N]⁻ of the native porphyrin ring are described by the $V_{i=1,2}(C,C)$ disynaptic attractors, while they are "reduced" to

single attractor $V(C,C)_{i=1-20}$ in the Cpd I.

5. According to the synaptic order, the Fe—N_{*f*=1-4} and Fe—S bonds are covalent-dative bonds but the respective localization basins are "formed" by electron density from lone pairs.

6. The nature of the bonding between Fe and O, formally double Fe=O, predicted by the ELF analysis stems mainly from electrostatic interactions, as long as no bonding attractor


$V(Fe,O)$ between the C(Fe) and C(O) core attractors is observed. Conversely, the localization indexes obtained from the covariance matrix for the iron and oxygen lone-pair basins would indicate the presence of a charge-shift bond.

Acknowledgments

The authors are also grateful to the Servei d'Informàtica from UJI and the Wrocław Centre for Networking and Supercomputing for generous allocation of computer time. Finally, the authors are grateful to P. González-Navarrete for his valuable comments that improved the manuscript.

Keywords: quantum chemical topology · electron localization function · electron density · chemical bond · compound I · cytochrome P450 · porphyrin · pyrrole

How to cite this article: I. Viciano, S. Berski, S. Martí, J. Andrés, J. Comput. Chem. **2013**, *34*, 780–789. DOI: 10.1002/jcc.23201

 Additional Supporting Information may be found in the online version of this article.

- [1] J. S. Lee, R. S. Obach, M. B. Fisher, Drug Metabolizing Enzymes: Cytochrome P450 and Other Enzymes in Drug Discovery and Development; FontisMedia SA; Marcel Dekker: Lausanne, Switzerland, New York, **2003**.
- [2] D. R. Nelson, *Hum. Genom.* **2009**, *4*, 59.
- [3] P. R. Ortiz de Montellano, Cytochrome P450 Structure, Mechanism, and Biochemistry; Kluwer Academic/Plenum Publishers: New York, Boston, MA, **2005**.
- [4] J. T. Groves, Y. Watanabe, *J. Am. Chem. Soc.* **1988**, *110*, 8443.
- [5] J. C. Schoneboom, H. Lin, N. Reuter, W. Thiel, S. Cohen, F. Ogliaro, S. Shaik, *J. Am. Chem. Soc.* **2002**, *124*, 8142.
- [6] M. Newcomb, R. Shen, S.-Y. Choi, P. H. Toy, P. F. Hollenberg, A. D. N. Vaz, M. J. Coon, *J. Am. Chem. Soc.* **2000**, *122*, 2677.
- [7] S. Shaik, D. Kumar, S. P. de Visser, *J. Am. Chem. Soc.* **2008**, *130*, 10128.
- [8] M. Newcomb, R. Zhang, R. E. P. Chandrasena, J. A. Halgrimson, J. H. Horner, T. M. Makris, S. G. Sligar, *J. Am. Chem. Soc.* **2006**, *128*, 4580.

- [9] H. Isobe, S. Yamanaka, M. Okumura, K. Yamaguchi, J. Shimada, *J. Phys. Chem. B* **2011**, *115*, 10730.
- [10] J. Rittle, M. T. Green, *Science* **2010**, *330*, 933.
- [11] B. Meunier, S. P. de Visser, S. Shaik, *Chem. Rev.* **2004**, *104*, 3947.
- [12] S. Shaik, D. Kumar, S. P. de Visser, A. Altun, W. Thiel, *Chem. Rev.* **2005**, *105*, 2279.
- [13] S. Shaik, S. Cohen, Y. Wang, H. Chen, D. Kumar, W. Thiel, *Chem. Rev.* **2010**, *110*, 949.
- [14] M. T. Green, *J. Am. Chem. Soc.* **1999**, *121*, 7939.
- [15] T. Ohta, K. Matsuura, K. Yoshizawa, I. Morishima, *J. Inorg. Biochem.* **2000**, *82*, 141.
- [16] D. L. Harris, *Curr. Opin. Chem. Biol.* **2001**, *5*, 724.
- [17] F. Ogliaro, S. Cohen, S. P. de Visser, S. Shaik, *J. Am. Chem. Soc.* **2000**, *122*, 12892.
- [18] F. Ogliaro, S. Cohen, M. Filatov, N. Harris, S. Shaik, *Angew. Chem. Int. Ed. Engl.* **2000**, *39*, 3851.
- [19] S. Shaik, S. P. de Visser, F. Ogliaro, H. Schwarz, D. Schroder, *Curr. Opin. Chem. Biol.* **2002**, *6*, 556.
- [20] D. Harris, G. Loew, L. Waskell, *J. Am. Chem. Soc.* **1998**, *120*, 4308.
- [21] G. H. Loew, D. L. Harris, *Chem. Rev.* **2000**, *100*, 407.
- [22] D. Harris, G. Loew, L. Waskell, *J. Inorg. Biochem.* **2001**, *83*, 309.
- [23] D. L. Harris, G. H. Loew, *J. Porphy. Phthalocya.* **2001**, *5*, 334.
- [24] P. Hohenberg, W. Kohn, *Phys. Rev. B* **1964**, *136*, B864.
- [25] R. F. W. Bader, *Atoms in Molecules: A Quantum Theory*; Clarendon Press, Oxford University Press: Oxford, New York, **1990**.
- [26] A. D. Becke, K. E. Edgecombe, *J. Chem. Phys.* **1990**, *92*, 5397.
- [27] B. Silvi, A. Savin, *Nature* **1994**, *371*, 683.
- [28] H. Chevreau, F. Fuster, B. Silvi, *Actual. Chimique*, **2001**, *3*, 15.
- [29] B. Silvi, I. Fourre, M. E. Alikhani, *Monatsh. Chem.* **2005**, *136*, 855.
- [30] N. O. J. Malcolm, P. L. A. Popelier, *Faraday Discuss.* **2003**, *124*, 353.
- [31] P. L. A. Popelier, In *Intermolecular Forces and Clusters*; D. J. Wales, Ed.; Springer-Verlag, Germany, **2005**; pp. 1–56.
- [32] I. V. Novozhilova, P. Coppens, J. Lee, G. B. Richter-Addo, K. A. Bagley, *J. Am. Chem. Soc.* **2006**, *128*, 2093.
- [33] N. Xu, J. Yi, G. B. Richter-Addo, *Inorg. Chem.* **2010**, *49*, 6253.
- [34] M. D. Michelini, N. Russo, E. Sicilia, *J. Am. Chem. Soc.* **2007**, *129*, 4229.
- [35] L. Gracia, P. Gonzalez-Navarrete, M. Calatayud, J. Andres, *Catal. Today* **2008**, *139*, 214.
- [36] P. J. Hay, W. R. Wadt, *J. Chem. Phys.* **1985**, *82*, 299.
- [37] M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Rob, J. R. Cheeseman, J. A. Montgomery, Jr., T. Vreven, K. N. Kudin, J. C. Burant, J. M. Millam, S. S. Iyengar, J. Tomasi, V. Barone, B. Mennucci, M. Cossi, G. Scalmani, N. Rega, G. A. Petersson, H. Nakatsuji, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, M. Klene, X. Li, J. E. Knox, H. P. Hratchian, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, P. Y. Ayala, K. Morokuma, G. A. Voth, P. Salvador, J. J. Dannenberg, V. G. Zakrzewski, S. Dapprich, A. D. Daniels, M. C. Strain, O. Farkas, D. K. Malick, A. D. Rabuck, K. Raghavachari, J. B. Foresman, J. V. Ortiz, Q. Cui, A. G. Baboul, S. Clifford, J. Cioslowski, B. B. Stefanov, G. Liu, A. Liashenko, P. Piskorz, I. Komaromi, R. L. Martin, D. J. Fox, T. Keith, M. A. Al-Laham, C. Y. Peng, A. Nanayakkara, M. Challacombe, P. M. W. Gill, B. Johnson, W. Chen, M. W. Wong, C. Gonzalez, J. A. Pople, I. Gaussian, Eds. *Gaussian 03 program (revision d02)*, Wallingford, CT, **2003**.
- [38] B. Silvi, *J. Mol. Struct.* **2002**, *614*, 3.
- [39] M. Kohout, DGrid, Version 4.5, Radebeul, DGrid 4.5 program, **2009**.
- [40] JMol program: an open-source Java viewer for chemical structures in 3D version 12.2.5, 2011.
- [41] W. Humphrey, A. Dalke, K. Schulten, *J. Mol. Graphics* **1996**, *14*, 33.
- [42] E. F. Pettersen, T. D. Goddard, C. C. Huang, G. S. Couch, D. M. Greenblatt, E. C. Meng, T. E. Ferrin, *J. Comput. Chem.* **2004**, *25*, 1605.
- [43] K. L. Stone, R. K. Behan, M. T. Green, *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 16563.
- [44] B. Silvi, E. S. Kryachko, O. Tishchenko, F. Fuster, M. T. Nguyen, *Mol. Phys.* **2002**, *100*, 1659.
- [45] F. Fuster, A. Sevin, B. Silvi, *J. Comput. Chem.* **2000**, *21*, 509.
- [46] A. Savin, B. Silvi, F. Coionna, *Can. J. Chem.* **1996**, *74*, 1088.
- [47] S. Berski, G. Gajewski, Z. Latajka, *J. Mol. Struct.* **2007**, *844*, 278.
- [48] I. Fourre, B. Silvi, P. Chaquin, A. Sevin, *J. Comput. Chem.* **1999**, *20*, 897.
- [49] A. Sanchez-Gonzalez, S. Melchor, J. A. Dobado, B. Silvi, *J. Andres, J. Phys. Chem. A* **2011**, *115*, 8316.
- [50] J. Pilme, B. Silvi, M. E. Alikhani, *J. Phys. Chem. A* **2005**, *109*, 10028.
- [51] S. Shaik, D. Danovich, B. Silvi, D. L. Lauvergnat, P. C. Hiberty, *Chem.—Eur. J.* **2005**, *11*, 6358.

Revised: 6 September 2012
 Accepted: 15 November 2012
 Published online on 12 December 2012

Supporting information for the paper entitled:

New Insight into the Electronic Structure of Iron(IV)-oxo Porphyrin Compound I. A Quantum Chemical Topological Analysis

Ignacio Viciano, Slawomir Berski, Sergio Martí,* and Juan Andrés

Figure 1. Snapshot of the localization domains obtained for Cpd I. Doublet and quartet states provide the same representation. Color code: blue for core basins, purple for disynaptic $V(\text{Fe},\text{N})_{i=1-4}$ basins, red for $V_{i=1,2}(\text{O})$ monosynaptic basins, yellow for both monosynaptic $V(\text{S})$ and disynaptic $V(\text{Fe},\text{S})$, orange for the $V(\text{Fe})$ monosynaptic basins, white for protonated disynaptic basins, and green for $V(\text{C},\text{C})_{i=1-20}$ and $V(\text{C},\text{N})_{i=1-8}$ disynaptic basins.

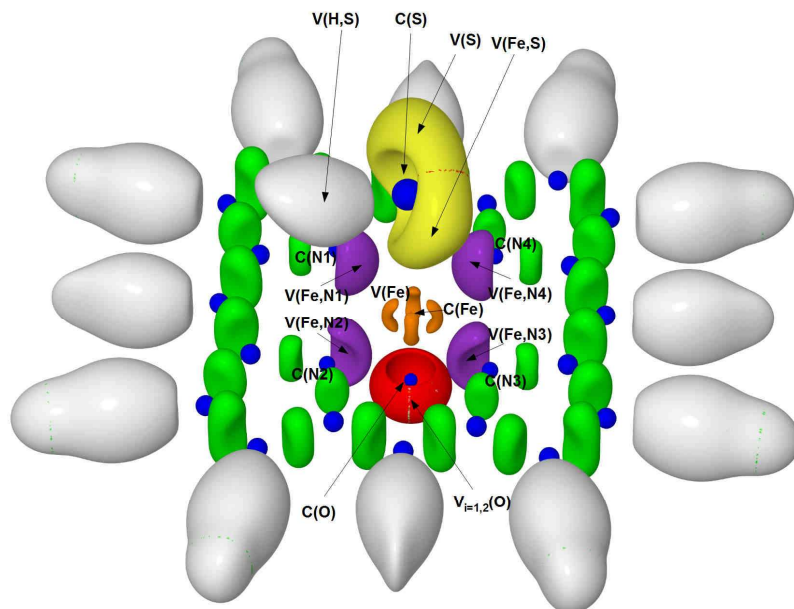


Figure 2. Spin density representation for both (a) doublet and (b) quartet spin states of the $\text{Por}^+ \text{Fe}^{\text{IV}}=\text{O}(\text{S-H})$ model. Blue regions correspond to negative density charge, whereas red ones correspond to positive density charge.

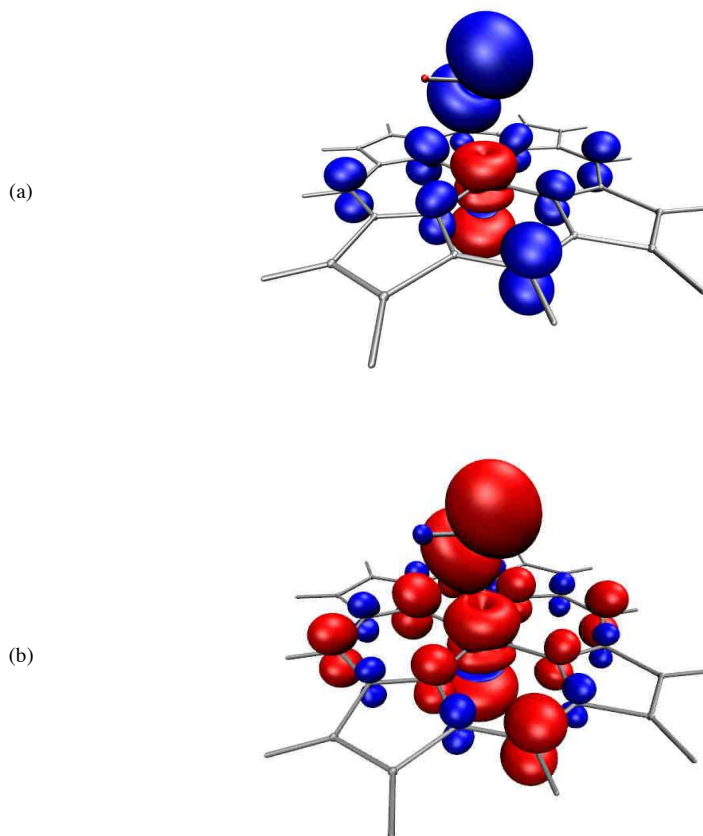


Figure 3. Snapshots of the ELF localization domains for: a) pyrrole, b) pyrrole anion and c) porphyrin ring. Color code: blue for core basins, purple for mponosynaptic basins, white for protonated disynaptic basins, and green for $V(C,C)_{i=1-3}$ and $V(C,N)_{i=1,2}$ disynaptic basins.

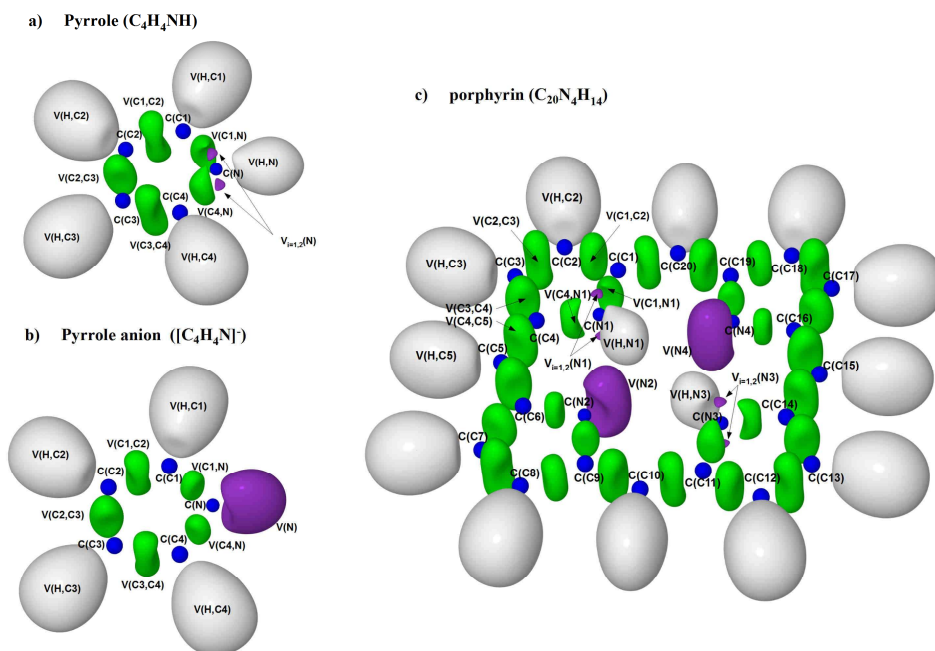


Figure 4. Snapshots of the ELF localization domains for: a) H_2S , b) $[\text{HS}]^-$ anion and c) $[\text{HS}]^\cdot$ radical. Color code: blue for core basins, yellow for mponosynaptic basins, and white for protonated disynaptic basins.

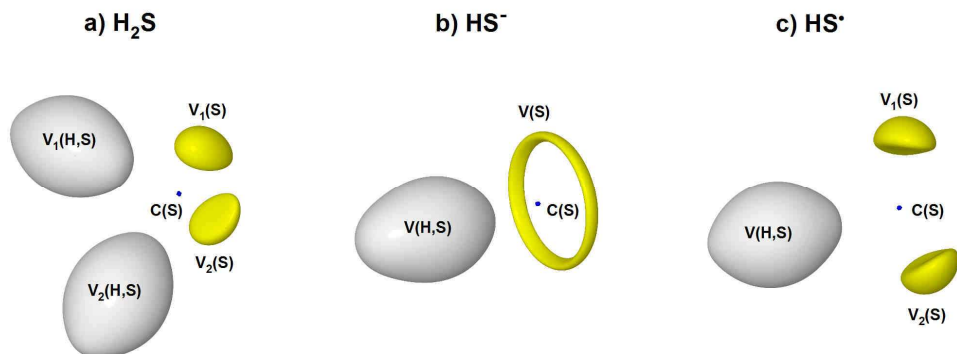


Table 1. Population obtained using the electron localization function $\eta(r)$ for the Cpdl in the doublet state.

BASIN	N _{total}	N _{alpha}	N _{beta}	N _{alpha} - N _{beta}	BASIN	N _{total}	N _{alpha}	N _{beta}	N _{alpha} - N _{beta}
C (Fe)	23.72	12.48	11.24	1.24	V (C5,C6)	2.90	1.44	1.46	-0.02
C (S)	10.08	5.03	5.05	-0.02	V (C6,C7)	2.39	1.20	1.19	0.01
C (O)	2.14	1.09	1.05	0.04	V (C7,C8)	3.30	1.65	1.65	0.00
C (N1)	2.13	1.06	1.07	-0.01	V (C8,C9)	2.38	1.19	1.19	0.00
C (N2)	2.13	1.06	1.07	-0.01	V (C9,C10)	2.97	1.47	1.50	-0.03
C (N3)	2.12	1.06	1.06	0.00	V (C10,C11)	2.89	1.43	1.46	-0.03
C (N4)	2.13	1.06	1.07	-0.01	V (C11,C12)	2.39	1.20	1.19	0.01
V ₁ (S)	2.75	1.25	1.50	-0.25	V (C12,C13)	3.30	1.65	1.65	0.00
V ₂ (S)	2.68	1.22	1.46	-0.24	V (C13,C14)	2.39	1.20	1.19	0.01
V (H,S)	1.84	0.90	0.94	-0.04	V (C14,C15)	2.91	1.44	1.47	-0.03
V ₁ (O)	0.95	0.52	0.43	0.09	V (C15,C16)	2.95	1.46	1.49	-0.03
V ₂ (O)	5.81	3.21	2.60	0.61	V (C16,C17)	2.39	1.20	1.19	0.01
V (N1)	3.02	1.47	1.55	-0.08	V (C17,C18)	3.30	1.65	1.65	0.00
V (N2)	3.04	1.50	1.54	-0.04	V (C18,C19)	2.40	1.20	1.20	0.00
V (N3)	3.05	1.51	1.54	-0.03	V (C19,C20)	2.88	1.43	1.45	-0.02
V (N4)	3.06	1.51	1.55	-0.04	V (C20,C1)	3.01	1.50	1.51	-0.01
V (C1,N1)	2.14	1.06	1.08	-0.02	V (H,C2)	2.14	1.07	1.07	0.00
V (C4,N1)	2.15	1.06	1.09	-0.03	V (H,C3)	2.14	1.07	1.07	0.00
V (C6,N2)	2.17	1.08	1.09	-0.01	V (H,C5)	2.12	1.05	1.07	-0.02
V (C9,N2)	2.15	1.07	1.08	-0.01	V (H,C7)	2.14	1.07	1.07	0.00
V (C11,N3)	2.17	1.08	1.09	-0.01	V (H,C8)	2.14	1.07	1.07	0.00
V (C14,N3)	2.17	1.08	1.09	-0.01	V (H,C10)	2.12	1.05	1.07	-0.02
V (C16,N4)	2.15	1.07	1.08	-0.01	V (H,C12)	2.14	1.07	1.07	0.00
V (C19,N4)	2.17	1.08	1.09	-0.01	V (H,C13)	2.14	1.07	1.07	0.00
V (C1,C2)	2.38	1.19	1.19	0.00	V (H,C15)	2.12	1.05	1.07	-0.02
V (C2,C3)	3.30	1.65	1.65	0.00	V (H,C17)	2.14	1.07	1.07	0.00
V (C3,C4)	2.39	1.20	1.19	0.01	V (H,C18)	2.14	1.07	1.07	0.00
V (C4,C5)	2.98	1.48	1.50	-0.02	V (H,C20)	2.12	1.05	1.07	-0.02

Table 2. Population obtained using the electron localization function $\eta(r)$ for the CpDI in the quartet state.

BASIN	Ntotal	Nalpha	Nbeta	Nalpha - Nbeta	BASIN	Ntotal	Nalpha	Nbeta	Nalpha - Nbeta
C (Fe)	23.72	12.45	11.27	1.18	V (C5,C6)	2.89	1.45	1.44	0.01
C (S)	10.08	5.05	5.03	0.02	V (C6,C7)	2.39	1.19	1.20	-0.01
C (O)	2.13	1.09	1.04	0.05	V (C7,C8)	3.30	1.65	1.65	0.00
C (N1)	2.14	1.07	1.07	0.00	V (C8,C9)	2.39	1.19	1.20	-0.01
C (N2)	2.13	1.07	1.06	0.01	V (C9,C10)	2.98	1.50	1.48	0.02
C (N3)	2.14	1.07	1.07	0.00	V (C10,C11)	2.88	1.45	1.43	0.02
C (N4)	2.14	1.07	1.07	0.00	V (C11,C12)	2.39	1.19	1.20	-0.01
V1 (S)	2.77	1.51	1.26	0.25	V (C12,C13)	3.30	1.65	1.65	0.00
V2 (S)	2.66	1.44	1.22	0.22	V (C13,C14)	2.39	1.19	1.20	-0.01
V (H,S)	1.83	0.93	0.90	0.03	V (C14,C15)	2.91	1.47	1.44	0.03
V1 (O)	4.55	2.53	2.02	0.51	V (C15,C16)	2.95	1.49	1.46	0.03
V2 (O)	2.20	1.22	0.98	0.24	V (C16,C17)	2.39	1.19	1.20	-0.01
V (N1)	3.02	1.55	1.47	0.08	V (C17,C18)	3.30	1.65	1.65	0.00
V (N2)	3.05	1.55	1.50	0.05	V (C18,C19)	2.40	1.20	1.20	0.00
V (N3)	3.03	1.53	1.50	0.03	V (C19,C20)	2.87	1.45	1.42	0.03
V (N4)	3.03	1.53	1.50	0.03	V (C20,C1)	3.02	1.52	1.50	0.02
V (C1,N1)	2.14	1.08	1.06	0.02	V (H,C2)	2.14	1.07	1.07	0.00
V (C4,N1)	2.14	1.08	1.06	0.02	V (H,C3)	2.14	1.07	1.07	0.00
V (C6,N2)	2.17	1.09	1.08	0.01	V (H,C5)	2.12	1.07	1.05	0.02
V (C9,N2)	2.14	1.07	1.07	0.00	V (H,C7)	2.14	1.07	1.07	0.00
V (C11,N3)	2.18	1.09	1.09	0.00	V (H,C8)	2.14	1.07	1.07	0.00
V (C14,N3)	2.17	1.09	1.08	0.01	V (H,C10)	2.12	1.07	1.05	0.02
V (C16,N4)	2.15	1.08	1.07	0.01	V (H,C12)	2.14	1.07	1.07	0.00
V (C19,N4)	2.17	1.09	1.08	0.01	V (H,C13)	2.14	1.07	1.07	0.00
V (C1,C2)	2.39	1.19	1.20	-0.01	V (H,C15)	2.12	1.07	1.05	0.02
V (C2,C3)	3.29	1.64	1.65	-0.01	V (H,C17)	2.14	1.07	1.07	0.00
V (C3,C4)	2.39	1.19	1.20	-0.01	V (H,C18)	2.14	1.07	1.07	0.00
V (C4,C5)	2.99	1.50	1.49	0.01	V (H,C20)	2.12	1.07	1.05	0.02

Table 3. Localization indexes, obtained from the covariance matrix for the ELF analysis for the doublet spin state, expressed as integrated population minus squared sigma.

$V_{i=1,2}(O)$		4.30e (7.14 integrated)
$V_{i=1}(O)$	$V_{i=2}(O)$	0.83e
$V_{i=1,2}(O)$	$V(Fe)$	0.43e
$V(C1,N1)$	$V(Fe,N1)$	0.35e
$V(C4,N1)$	$V(Fe,N1)$	0.35e
$V(C6,N2)$	$V(Fe,N2)$	0.36e
$V(C9,N2)$	$V(Fe,N2)$	0.36e
$V(C11,N3)$	$V(Fe,N3)$	0.36e
$V(V14,N3)$	$V(Fe,N3)$	0.36e
$V(C16,N4)$	$V(Fe,N4)$	0.36e
$V(C19,N4)$	$V(Fe,N4)$	0.36e
$V(Fe,N1)$	$V(Fe)$	0.21e
$V(Fe,N2)$	$V(Fe)$	0.21e
$V(Fe,N3)$	$V(Fe)$	0.19e
$V(Fe,N4)$	$V(Fe)$	0.20e
$V(Fe,S)$	$V(H,S)$	0.25e
$V(Fe,S)$	$V(S)$	0.45e

4.2. First Catalytic Subcycle of the Enzyme Aromatase

4.2.1. QM/MM Modeling of the Hydroxylation of the Androstenedione Substrate Catalyzed by Cytochrome P450 Aromatase (CYP19A1)

Ignacio Viciano, Raquel Castillo, and Sergio Martí

Journal of Computational Chemistry **2015**, 36, 1736–1747

QM/MM Modeling of the Hydroxylation of the Androstenedione Substrate Catalyzed by Cytochrome P450 Aromatase (CYP19A1)

Ignacio Viciano, Raquel Castillo, and Sergio Martí*

CYP19A1 aromatase is a member of the Cytochrome P450 family of heme proteins, and is the enzyme responsible for the final step of the androgens conversion into the corresponding estrogens, via a three-step oxidative process. For this reason, the inhibition of this enzyme plays an important role in the treatment of hormone-dependent breast cancer. The first catalytic subcycle, corresponding to the hydroxylation of androstenedione, has been proposed to occur through a first hydrogen abstraction and a subsequent oxygen rebound step. In present work, we have studied the mechanism of the first catalytic subcycle by means of hybrid quantum mechanics/molecular mechanics methods. The inclusion of the protein flexibility has been achieved by means of Free Energy Perturbation techniques, giving rise to a free energy of activation for the hydrogen abstraction step of 13.5 kcal/mol. The subsequent oxygen

rebound step, characterized by a small free energy barrier (1.5 kcal/mol), leads to the hydroxylated products through a highly exergonic reaction. In addition, an analysis of the primary deuterium kinetic isotopic effects, calculated for the hydrogen abstraction step, reveals values (~ 10) overpassing the semi-classical limit for the C—H, indicating the presence of a substantial tunnel effect. Finally, a decomposition analysis of the interaction energy for the substrate and cofactor in the active site is also discussed. According to our results, the role of the enzymatic environment consists of a transition state stabilization by means of dispersive and polarization effects. © 2015 Wiley Periodicals, Inc.

DOI: 10.1002/jcc.23967

Introduction

Cytochrome P450 (CYP or P450) is a superfamily of heme-thiolate-containing metalloenzymes, composed by a large number of families and subfamilies, which belong to the three domains of life. Currently there are more than 11,500 CYP named, 57 of which are members of the *homo sapiens* species.^[1] These enzymes play an important role in different biological processes in some organisms such as plants, mammalian, bacteria, fungi, archaea, or protists.^[2,3] P450s can metabolize a wide variety of both endogenous and exogenous substrates, being responsible of the metabolism for a large amount of pharmaceuticals^[4] ($\sim 75\%$). The most common reactions catalyzed by CYP enzymes involve oxygen insertions, including hydroxylation of unactivated C—H bonds, alkene epoxidations, N- and O-dealkylations, heteroatom oxidations such as sulfoxidations, and so forth.^[2,5]

Human Aromatase (CYP19A1) is a member of the superfamily of cytochrome P450 enzymes and is located in a number of tissues that include gonads, adrenal glands, ovaries, placenta, testes, adipose tissue, and numerous sites in the brain.^[6,7] This enzyme is responsible for a key step in the biosynthesis of steroid hormones, and more specifically it is involved in the final step of the conversion of androgens (androstenedione [ASD], testosterone and 16- α -hydroxytestosterone) into the corresponding estrogens (estrone, 17- β -estradiol and estriol, respectively).^[8] This conversion consists of three consecutive oxidation steps of the 19-methyl group of the androgens, which is removed as formic acid during the course of the

reaction, resulting in aromatization of ring A of these steroidal substrates (see Scheme 1).^[9] First and second steps are believed to proceed through two sequential hydroxylations of the 19-methyl group of ASD to produce 19-*gem*-diol species that then undergoes dehydration to yield the aldehyde intermediate (19-oxo-ASD). The third step, which is still under discussion, corresponds to the C₁₀—C₁₉ bond cleavage process (lyase reaction).^[10] The entire process of aromatization consumes a total of three moles of NADPH and three moles of oxygen.^[2]

Aromatase is the only known enzyme responsible for the biosynthesis of estrogens from androgens in vertebrates. The estrogens are the primary female sex hormones, being important for sexual and reproductive development in humans, mainly in woman. However, these steroidal hormones also stimulate the growth of hormone-receptor-positive breast cancer cells playing an important role in the development of hormone-dependent breast cancer. The inhibition of this

I. Viciano, R. Castillo, S. Martí

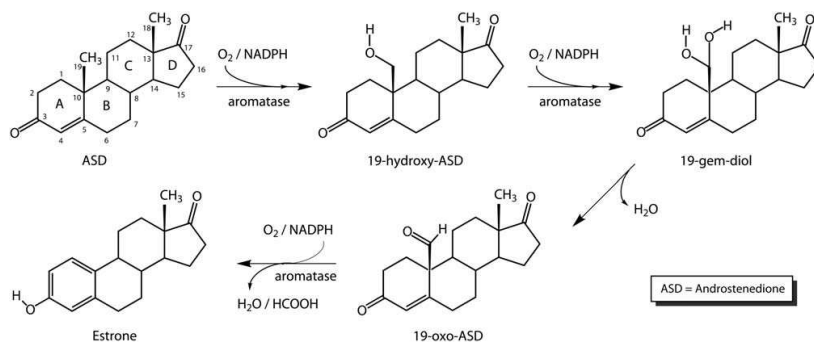
Departament de Química Física i Analítica, Universitat Jaume I, Castelló, 12071, Spain

Fax: +34964728066

E-mail: smarti@uji.es

Contract/grant sponsors: MINECO Project CTQ2012-36253-C03-01, Generalitat Valenciana Prometeo/2009/053 and PrometeoII/2014/022, Universitat Jaume I Project P1.1B2011-23, HPC-EUROPA2 project (project number: 228398) (European Community - Research Infrastructure Action of the FP7); Contract/grant sponsors: Spanish Ministerio de Ciencia e Innovación; Contract/grant number: CTQ2009-14541-C02 (to I.V.)

© 2015 Wiley Periodicals, Inc.



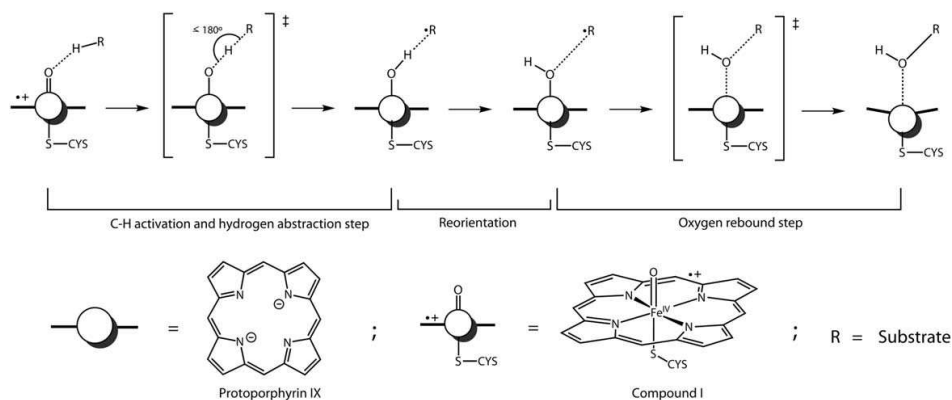
Scheme 1. Catalytic process of conversion of ASD into aromatic estrone catalyzed by cytochrome CYP19A aromatase.

enzyme, and hence cessation of estrogen production, provides one of the first molecular targets for rational drug development in the treatment of breast cancer and of other hormone-responsive cancers.^[11,12] For this reason, the development of potent and selective aromatase inhibitors (AIs) is one of the greatest challenges in the field of endocrinology regarding the breast cancer therapy, as well as a goal to achieve in the pharmaceutical industry.

The knowledge of the structural nature of the enzyme and its mechanism of action provide useful information for designing new AIs and optimal drugs to treat breast cancer. One of the best ways to explore the structure of an enzyme as well as its reaction mechanisms is through the use of computational techniques such as molecular dynamics (MD) simulations and hybrid quantum mechanics/molecular mechanics (QM/MM) methods. Although the cytochrome P450 superfamily of enzymes has been extensively studied using computational methods in some of its most common isoforms,^[13] CYP450 aromatase has not been under theoretical study for a long time, partly because of the lack of a good crystallographic model. This fact prompted a number of homology models based on other P450 isoforms and site-directed mutagenesis data were proposed.^[14] The crystal structure of aromatase purified from human placenta in complex with its natural sub-

strate ASD (protein data bank [PDB] code 3EQM) was determined by Ghosh et al.^[15] in 2009. This crystal structure shows how the natural substrate is accommodated in the active site of this enzyme and provides a good initial molecular structure on which to apply computational tools. In fact, after obtaining this crystal structure, several mechanistic studies on aromatase catalysis using QM/MM methods have been published in recent years.^[16,17] The first oxidation step of the 19-methyl group of ASD has been studied experimentally many times,^[18–26] however, no QM/MM study has been performed for this catalytic subcycle so far, thus remaining the mechanism still unsolved to this day.

This first oxidation step, in which ASD is hydroxylated to 19-hydroxy-ASD, is thought to proceed through the hydrogen abstraction/oxygen rebound mechanism, originally proposed by Groves et al., as occurs in other P450s.^[27] This mechanism, consists of an initial hydrogen atom abstraction of the substrate by means of the reactive species, a high valent ferryl-oxo heme complex (Compound I, Cpd I), leading to the formation of the alkyl radical intermediate and the iron-hydroxo complex. Subsequently, an alkyl (or OH) reorientation is produced to facilitate the final oxygen rebound mechanism, where a recombination of the alkyl radical with the iron-bound hydroxyl radical leads to the formation of the corresponding hydroxylated product (see Scheme 2). The nature of



Scheme 2. Hydroxylation of substrates via cytochromes P450.

the reactive species located in the active site of CYP enzymes, has been extensively discussed in the literature by both theoretical^[13] and experimental^[28] studies. Currently, the most accepted oxidant is the Cpd I, an iron IV oxo-porphyrin cation radical (Por⁺-Fe^{IV}=O), whose ability to activate inert C—H bonds has been amply demonstrated.^[29–33] Even so, due to the elusiveness of this compound and its high reactivity, making it very difficult to detect and capture, other oxidants such as the ferric hydroperoxo complexes (Por-Fe^{III}-OOH⁻, Cpd 0) and perferryl-based species (Por-Fe^V=O) have been proposed.^[28,34–39] However, Rittle and Green^[30] succeeded preparing Cpd I on CYP119 P450 cytochrome in a high yield, being able to spectroscopically and kinetically characterize it, settling the matter on the hydroxylating agent in P450 enzymes. In addition, in a recent study,^[40] Khatri and coworkers concluded that the Cpd I is the reactive intermediate not only in the first two steps of aromatization by aromatase (hydroxylations) but also in the third (and last) aromatization step.

The Compound I has a radicalary nature and possesses three singly occupied orbitals. Two of them are the π_{xz}^* and π_{yz}^* orbitals of the Fe—O moiety, while the third one is delocalized between the a_{2u} orbital of the porphyrin and the π_s of the sulfur atom. A ferromagnetic coupling of the three electrons provides a high-spin quartet state ($S = 3/2$), whereas an antiferromagnetic one leads to a low-spin doublet state ($S = 1/2$). As a result of this dual spin nature of the Cpd I, Shaik et al.^[41] proposed a two-state reactivity mechanism for the hydroxylation and the epoxidation of alkanes, in which product distribution of the reaction is determined by the combination of both spin states.

We herein present a theoretical study for the first oxidation step of androgen substrate ASD via cytochrome CYP19A1 aromatase. In this oxidation step, which involves the hydroxylation of the 19-methyl group of ASD to 19-hydroxy-ASD, the radical rebound mechanism as well as the Cpd I as the oxidizing species have been proposed. The two potential energy surfaces (PESs), corresponding to the two possible spin configurations, doublet, and quartet, have been explored by means of QM/MM methods using the density functional theory (DFT) to describe the QM atoms. In addition, the activation free energies of the biochemical process have been evaluated introducing free energy perturbation (FEP) techniques, thus, taking into account the conformational space of the protein. Also, semiclassical kinetic isotope effects (KIE) of the hydrogen abstraction step from ASD 19-methyl group have been reported. Finally, an exhaustive analysis of the binding nature of both substrate and cofactor has been performed by means of the decomposition of the interaction potential energy along the hydrogen abstraction step.

Computational Methods

System setup and MD simulations

The initial geometry used in our calculations was obtained from the X-ray crystal structure of the human placental aromatase cytochrome P450, in complex with its natural substrate

ASD, solved at 2.9 Å of resolution (PDB code 3EQM).^[15] The original pentacoordinated heme B cofactor found in the PDB file was modeled into an hexacoordinated Iron-Oxo porphyrin Cpd I species, using the atom positions suggested in the literature.^[15] All the hydrogen atoms were added using the NAMD 2.8 program,^[42] according to the empirical pKa predictions rendered by PROPKA web interface.^[43] The standard protonation states in solution were found except for Asp-309, which exhibited a large pKa displacement (7.7). This value indicated a weak acidic character and, therefore, Asp-309 was protonated at pH 7. Histidine residues were protonated as follows: His-171 was doubly protonated, His-(62, 105, 111, 325, 475, and 480) were singly protonated at ϵ position, His-(109, 128, 402, and 459) were singly protonated at δ position. A total of four counter ions (Cl⁻) were placed into optimal electrostatic positions around the enzyme to electroneutralize the excess of positive charge (+4), fulfilling the electroneutrality of the system. The model was placed in a prerelaxed orthorhombic box of water molecules with dimensions of $90 \times 80 \times 80 \text{ \AA}^3$, erasing all those water molecules with an oxygen atom lying less than 2.8 Å from any heavy atom. The resulting model consisted of 452 residues of amino acids, the modeled Cpd I, the substrate ASD, 35 crystallographic water molecules, four counterions, and 16,542 water molecules of the solvation box.

The solvated model was minimized and then equilibrated by means of classical MD at 300 K, using the canonical (NVT) ensemble and the Langevin–Verlet integrator. The MD was run in NAMD for 20 ns with a step size of 1 fs using the OPLS-AA^[44] force field and the TIP3P^[45] water model. The nonbonding interactions were treated in all calculations by periodic boundary conditions with the Particle Mesh Ewald convention, via a force-switch function with a cutoff distance in the range 14–16 Å. No harmonic restraints were applied between atoms and/or fragments during the simulations.

QM and QM/MM calculations

Starting structure used in QM/MM calculations was extracted from one of the last snapshots obtained in MD simulations based on observation criteria. To study the chemical reactivity, the QM region of the model was defined to include the substrate, the Cpd I and the axial Cys-437 ligand (comprising the S_γ , the C_β , and two H_β atoms), giving rise to a total of 124 atoms (see Fig. 1a). The link atom formalism was applied to satisfy the valence of the QM region, due to presence of the classical bond partitioning in the Cys-437 amino acid (C_α — C_β bond). Because of the complexity of the enzymatic model, a truncation scheme was implemented during all hybrid QM/MM calculations. In this scheme, the lists of interaction for the QM atoms (incorporating all atoms within 20 Å from any atom of both the substrate and heme groups) are defined at the very beginning and remain constant during the calculations. In this way, all residues further than 20 Å from either the Cpd I or the substrate were kept frozen, during the simulations (8638 atoms from a total of 57,252).

As the study and characterization of a transition state (TS) usually make use of the Hessian matrix, its evaluation is not

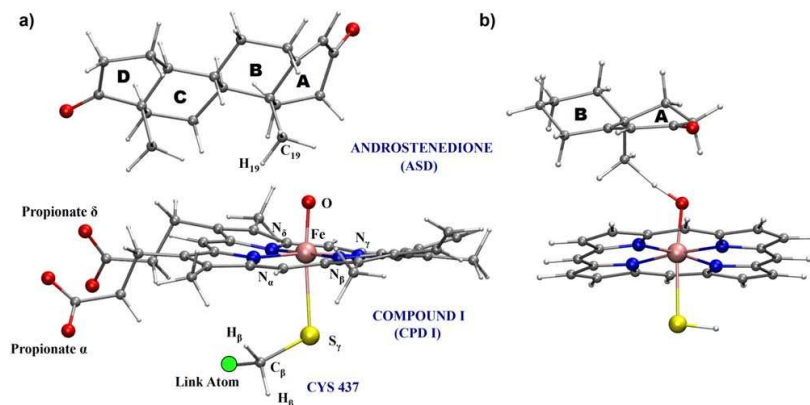


Figure 1. a) Atoms comprising QM model in the enzymatic system (Cpd I, ASD substrate, and residue Cys-437). b) Atoms comprising the QM-only model (Cpd I and ASD models, the latter containing only the steroidal rings A and B).

feasible for enzymatic systems due to the large amount of degrees of freedom. One way to overcome this problem is to divide the full coordinates space into two different subsets: the control space, which incorporates all those atoms or molecules involved in the reaction process (usually the QM atoms); and the complementary space comprising the rest of the system. Therefore, the optimization of the model system can be expressed as the combination of iterations in both subsets: at each step of the control space Hessian guided optimization (based on the Baker^[46,47] algorithm), the rest of the system is kept fully relaxed merely using gradient vectors (using the LBFGS^[48] procedure). This strategy is also known as the micro/macroiteration method, and allows taking advantage of efficient optimization algorithms.^[49] Furthermore, a dual QM:Charge/MM scheme^[49,50] was adopted, where the total QM/MM energy of the system is expressed using different terms, depending on which coordinate space is being optimized. This choice is justified by the fact that a typical complementary space optimization (corresponding to the macroiterations) typically needs a large number of gradient evaluations at each control space movement (or microiteration), and thus some approximations must be introduced to speedup calculations. In this case, the QM atoms are reduced to frozen classical charges during the macroiterations, thereby using different expressions for the energy depending on the current iteration, as shown in following equations:

$$E_{\text{micro}} = E_{\text{MM}} + \langle \Psi | \hat{H}_0 | \Psi \rangle + \sum \langle \Psi | \frac{q_{\text{MM}}}{r} | \Psi \rangle = E_{\text{MM}} + E_{\text{QM,o}} + E_{\text{QM,int}} \quad (1)$$

$$E_{\text{macro}} = E_{\text{MM}} + \sum \frac{q_{\text{QM}}^{\text{fit}} q_{\text{MM}}}{r_{\text{QM,MM}}} \quad (2)$$

The term “ E_{MM} ”, common in both equations, is the energy of the MM region, including the Lennard-Jones interaction between QM and MM atoms. The terms “ $E_{\text{QM,o}}$ ” and “ $E_{\text{QM,int}}$ ” included in eq. (1), correspond to the quantum energy of the QM atoms using a high-level method (such as *ab initio* or den-

sity functional-based methods) under a classical environment described by point charges (q_{MM}). The term “ $E_{\text{QM,o}}$ ” refers to the selected gas-phase Hamiltonian operator acting on the polarized wave function, meanwhile the term “ $E_{\text{QM,int}}$ ” involves the electrostatic interaction among this wave function and the classical point charges. Conversely, in eq. (2) the gas-phase quantum energy calculation has been removed, and the QM/MM electrostatic interaction is replaced by a Coulombic expression, making use of electrostatic potential fitted charges (Merz-Kollman^[51,52]) for the QM atoms.

In this way, the TSs were localized by means of this micro/macroiterations scheme, and were characterized by hessian inspection. Afterward, minimum energy paths were then traced down to their corresponding minima, to ensure that the structures really connect reactants and intermediates. Once the reaction paths were obtained, the Gibbs free energy was evaluated at the stationary points, under the rigid-rotor and the harmonic-oscillator (RRHO) approximations, taking into account both zero point energy (ZPE) and thermal corrections.

Moreover, a small model of the system (see Fig. 1b) was built to study the hydrogen abstraction step, performing gas-phase calculations (QM-only model). This model, consisting of a reduced version of both Cpd I and ASD, was used to contrast some results obtained for the enzymatic model.

The unrestricted Kohn–Sham formalism with the B3LYP^[53,54] hybrid density functional was used to describe the QM region of the enzyme aromatase as well as the quantum atoms of the QM-only model. It has been shown that the DFT and DFT/MM methodologies provide a reasonably good description of the doublet and quartet ground states of the Cpd I.^[55,56] In particular, the B3LYP functional has become the preferred choice for carrying out the modeling of the CYP enzymes as is well-known to give reasonable relative spin state energies as well as to perform well for a number of other properties.^[13,57] The LACVP* basis set (B1),^[58] consisting of the combination of the 6-31G(d) basis for all the atoms except for the iron, which is represented by the Lanl2dz effective core potential (ECP), was used for the exploration of the PES. Additionally, to obtain

more accurate energies and frequency numbers, a larger basis set (B2) was adopted to perform single point calculations on the stationary points previously allocated with the B1 basis set. This B2 basis set corresponds to the combination of the LanL2tz+ ECP for the iron and the 6-311G(d,p) basis for the rest of the atoms. The rest of the enzyme and the water molecules were described by the OPLS-AA and TIP3P force fields, respectively.

All the electronic calculations were performed using the Gaussian03^[59] family of programs, whereas the fDYNAMO^[60] library was used to describe the MM during the hybrid QM/MM simulations.

Free energy perturbation calculations

FEP techniques were introduced to evaluate the free energies of activation of the chemical process. The use of this method, based on statistical mechanics, allows taking into account the conformational space of the enzyme, obtaining accurate free-energy surfaces and solvent effects. For this purpose, different energetically equispaced geometries (windows) were selected from the previously traced reaction paths, being 54 for the doublet spin state and 75 for the quartet state. In each window, the QM region (including the link atom) was kept frozen during the calculations. Following the micro/macro scheme presented before, after each QM calculation performed [see eq. (1)], 100 steps of classical MD were performed with a time step of 1 fs [see eq. (2)], under the same conditions as during the equilibration. This procedure was repeated 200 times, resulting in a total of 20 ps per window. Finally, the coordinates of the QM region were swapped between consecutive windows (for the last 100 structures) to estimate the free energy changes, using the following expression:

$$\Delta F_{i \rightarrow j} = -RT \ln \left\langle e^{-\left(\frac{H_j - H_i}{RT}\right)} \right\rangle_i \quad (3)$$

In eq. (3), H_i corresponds to the QM/MM potential energy obtained during the i th window of MD. The term H_j refers to the QM/MM potential energy, once the coordinates of the atoms for the control space have been exchanged ($j = i + 1$ for forward or $j = i - 1$ for backwards), under the same geometry of the classical environment (i). Finally, the triangular brackets denote an average performed over the i th window. Once the free energy barrier is calculated using eq. (3), the ZPE correction term is included as the difference between the ZPEs of both final and initial states (reactant and TS for the activation energy) by means of QM/MM frequency calculations.

Kinetic isotope effects calculations

Semiclassical KIE were calculated for the hydrogen abstraction step in both QM-only model and the enzymatic system (as an average of 10 window obtained from FEP calculations for both the TS and reactants).

$$\text{KIE}_E = \frac{k_H}{k_D} = \exp \left[\frac{\Delta G_D^\ddagger - \Delta G_H^\ddagger}{RT} \right] \quad (4)$$

Subsequent corrections to the Eyring semiclassical KIE were incorporated by means of the Wigner correction^[61] to estimate the Tunnel effects on these KIEs:

$$\text{KIE}_W = \text{KIE}_E \cdot \left(\frac{Q_{W,H}}{Q_{W,D}} \right); Q_W = 1 + \frac{1}{24} \left(\frac{h\nu^\ddagger}{k_B T} \right)^2 \quad (5)$$

where ν^\ddagger is the imaginary frequency at the saddle point, k_B is Boltzmann's constant and h is the Planck's constant. This choice seems to be appropriate to the tunneling corrections applied to rate constants for which the values of correction coefficients are small to moderate (≤ 2).^[62,63] The CAMVIB/CAMISO^[64,65] programs, based on the RRHO approximations, without scaling the vibrational frequencies, were used to calculate KIEs.

As the analysis of the bond orders (BO) may provide relevant information about the geometrical similarity in the TS structures, which is directly related with the KIE values, the C₁₉—H₁₉ and H₁₉—O formal BO were calculated. For this purpose, natural bond orbital (NBO) calculations were performed using the Gaussian 03 package at the UB3LYP/B1 level to estimate the Wiberg bond indices in the natural atomic orbitals (NAO) basis.^[66]

Results and Discussion

The results presented in this article correspond to the hydroxylation of the ASD 19-methyl group to 19-hydroxy-ASD (see Scheme 2) catalyzed by the cytochrome P450 aromatase. Both the initial hydrogen abstraction step, and the subsequent oxygen rebound step were studied in the active site of the enzyme, in the doublet and the quartet spin states, using the B3LYP[B1]:ESP/MM scheme presented before. Table 1 shows the Gibbs free energies for the QM-only model (using the B1 basis set), and for the enzymatic system (using both B1 and B2 basis set). In addition, the free energies of activation obtained from FEP in the enzymatic model for both basis sets are also presented.

The QM-only model gas-phase calculations, exhibit almost degenerated reactants at both spin states. The Gibbs activation energies, as well as the corresponding imaginary frequencies, for the hydrogen abstraction step show also similar values, leading to the formation of the intermediates through an endergonic process. The comparison of these values with those obtained in the enzymatic environment (using the same B1 basis set), displays a modest breakdown of the reactants degeneracy by 1.4 kcal/mol. The subsequent Gibbs activation free energies for this step are lowered by 3.4 kcal/mol in the doublet and 1.3 kcal/mol in the quartet, when comparing to the gas-phase model. This modest catalytic effect obtained from the PES analysis suggests a mild participation of the enzymatic environment in the chemical process. The intermediates that follow the hydrogen abstraction step give place to the oxygen rebound mechanism by means of a small

Table 1. Relative free energies obtained by means of RRHO (ΔG) and FEP (ΔF) for the hydroxylation of ASD in the doublet and quartet spin states (ZPE included in both cases).

	ΔG_{vac} (B1)	ΔG_{enz} (B1)	ΔF_{enz} (B1)	ΔG_{enz} (B2)	ΔF_{enz} (B2)
^2R	0.0	0.0	0.0	0.0	0.0
^4R	-0.1	1.4	2.9 ± 0.1	1.9	3.0
$^2\text{TS}_{\text{H-abs}}$	25.7 (1601i)	22.3 (1606i)	13.5 ± 0.4	20.0 (1619i)	11.4
$^4\text{TS}_{\text{H-abs}}$	25.8 (1713i)	24.5 (1645i)	13.5 ± 0.3	21.8 (1675i)	12.0
^2I	16.3	17.1	11.5 ± 0.6	13.5	8.7
^4I	16.4	17.2	12.3 ± 0.4	14.7	9.3
$^2\text{TS}_{\text{reb}}$		20.0 (258i)	13.0 ± 0.8	15.5 (230i)	10.0
$^4\text{TS}_{\text{reb}}$		19.3 (174i)	15.1 ± 1.0	17.0 (119i)	12.6
^2P		-32.4		-36.9	
^4P		-42.7		-43.3	

The basis set used in each calculation is also shown in parentheses. The imaginary frequencies associated with the TS obtained in the PES explorations are also reported. All energies are expressed in kcal/mol, and frequencies in cm^{-1} .

barrier (~ 3 kcal/mol), leading to the formation of the hydroxylated products through a highly exergonic process (-32 and -43 kcal/mol for the doublet and quartet states, respectively).

The inclusion of the conformational space effects (enzyme flexibility), by means of the FEP methodology, dramatically reduces the free energy of activation for the hydrogen abstraction step. The obtained value of 13.5 kcal/mol for both spin states appears to be as a more realistic prediction for this biochemical reaction.^[67] At the same time, the free energy difference between reactants in both spin states obtained with the FEP is increased when compared with the PES results, despite that both spin states are accessible during the dynamics. Once the intermediate species have been formed, the oxygen rebound step takes place after a previous reorientation process (see Scheme 2).

For the doublet electronic state, the rotation has a small free energy barrier of 2.1 kcal/mol leading to a rotation intermediate (-3.3 kcal/mol) from which the rebound step takes place with a barrier of 2.3 kcal/mol. This behavior is different in the quartet state, where the rotation has a free energy barrier of 2.2 kcal/mol leading to a rotation intermediate roughly thermoneutral (0.7 kcal/mol) from which the rebound takes place with a barrier of 2.7 kcal/mol. Finally, the inclusion of the ZPE energy term reduces the overall barrier for the rebound step to 1.5 and 2.8 kcal/mol, for the doublet and quartet states, respectively. As can be noted, the free energies of the product species are not reported in Table 1, as we have focused on obtaining only the free energies of activation for the biochemical process with the FEP methodology.

The effect of the basis set has been also inspected by means of the incorporation of a larger basis set (B2), thus, performing single point calculations of the structures obtained in

the former PES (B1 basis set). The new Gibbs free energies obtained depict a decrease in the activation energy of the hydrogen abstraction around 2–3 kcal/mol, while the remaining stationary points show similar values of relative energy. Additionally, the basis set correction has been included in the FEP calculation, but only in those terms that can be factored out of the average, to avoid the large amount of expensive QM(B2) calculations:

$$\Delta F \approx -RT \ln \left(e^{-\frac{\Delta E_{\text{int}}^{\text{B1}}}{RT}} \right) + \Delta E_{\text{vac}}^{\text{B2}} + \Delta \text{ZPE}^{\text{B2}} \quad (6)$$

In eq. (6), the interaction energy term ($\Delta E_{\text{int}}^{\text{B1}}$) has been previously calculated using the B1 basis set, while both the gas-phase energy ($\Delta E_{\text{vac}}^{\text{B2}}$) and the ZPE ($\Delta \text{ZPE}^{\text{B2}}$) terms correspond to the expensive calculations with the large basis set. The obtained results show a free energy of activation for the hydrogen abstraction lower than the corresponding to the B1 basis set (11.4 and 12.0 kcal/mol for the doublet and quartet states, respectively). This trend, associated to the use of the B2 basis set, is not observed in the rebound step, where the activation barriers remain virtually unchanged (1.3 and 3.3 kcal/mol for the doublet and quartet state, respectively).

Primary deuterium (^2H) KIE values of the H_{19} abstraction step for the two models (QM-only and the enzymatic system) were determined using the frequency data obtained with the B1 basis set. This basis set was used to compare the results for the QM-only model and the enzymatic system (obtained as an average of the FEP structures), due to the complexity of the latter.

Inspection of the computed KIEs presented in Table 2 show values around 7 in both models and in both spin states

Table 2. Primary deuterium (^2H) KIE for the hydrogen transfer process for both gas-phase model and enzymatic system.

Spin state	Gas-phase model		Enzyme aromatase	
	$^2\text{H}_{19}$ primary KIE	$^2\text{H}_{19}$ corrected KIE	$^2\text{H}_{19}$ primary KIE	$^2\text{H}_{19}$ corrected KIE
Doublet	7.027	10.225	7.281 ± 0.004	10.315 ± 0.006
Quartet	7.013	10.445	7.074 ± 0.003	10.456 ± 0.007

Corrected KIEs incorporate the tunneling effects by means of the Wigner approximation.

Table 3. Atomic spin densities of selected atoms or fragments included in the QM atoms of the enzymatic model with the B2 basis set.

Fragment	Hydrogen abstraction			Oxygen rebound		
	Reactant	TS (H-Abs)	Intermediate	Intermediate	TS (rebound)	Product
SCH ₃ ⁻	-0.20 [0.17]	-0.21 [0.14]	-0.14 [0.10]	-0.10 [-0.08]	-0.19 [-0.06]	-0.17 [0.28]
Fe	1.30 [1.22]	1.06 [0.96]	1.02 [0.92]	1.02 [2.06]	1.16 [2.24]	1.34 [2.70]
O	0.83 [0.83]	0.49 [0.52]	0.13 [0.15]	0.09 [0.13]	-0.06 [0.04]	-0.00 [0.01]
Porphyrin	-0.93 [0.78]	-0.87 [0.82]	-0.95 [0.87]	-0.98 [-0.09]	-0.62 [-0.10]	-0.17 [0.01]
C ₁₉	0.00 [0.00]	0.60 [0.62]	1.03 [1.04]	1.04 [1.05]	0.76 [0.94]	0.00 [0.00]
H ₁₉	-0.00 [-0.00]	-0.05 [-0.05]	-0.01 [0.00]	0.00 [0.01]	0.00 [0.02]	0.00 [0.00]
Substrate ^[a]	0.00 [0.00]	-0.02 [-0.02]	-0.07 [-0.07]	-0.07 [-0.07]	-0.05 [-0.06]	0.00 [0.00]

The values are presented for both spin states: doublet and quartet (in brackets). [a] The substrate does not include the C₁₉ and H₁₉ atoms.

(doublet and quartet). The fact that both models provide almost identical values (very close to semiclassical Eyring limit, $k_H/k_D \approx 7$) suggests that this KIE value is intrinsic to the reaction and is not affected by the enzymatic environment. Regarding the slight difference in values between different spin states, we can invoke the effect of the symmetry of the TS as it is well-known that the KIE value depends on the geometry adopted in this stationary point. Thus, the analysis of the distances among atoms involved in the H-transfer coordinate ($d[C_{19}-H_{19}]$ and $d[H_{19}-O]$) for the TS might provide a picture of the relationship between the symmetry and the KIE values. The geometrical parameters obtained for the TS species along this coordinate in the enzymatic system are: $d(C_{19}-H_{19}) = 1.399 \text{ \AA}$, $d(H_{19}-O) = 1.133 \text{ \AA}$ and $C_{19}-H_{19}-O = 171.2^\circ$ for the doublet, and $d(C_{19}-H_{19}) = 1.394 \text{ \AA}$, $d(H_{19}-O) = 1.134 \text{ \AA}$ and $C_{19}-H_{19}-O = 169.6^\circ$ for the quartet. These distances and angles obtained along the H-transfer coordinate are practically the same in both spin states, and hence denote a great geometrical similarity between the TS species obtained in both states. In fact, this behavior of geometrical similarity is also borne out by the results obtained from NBO analysis performed on TS structures. The values provided by Wiberg BO indices on the atoms involved in the H-transfer coordinate are: $BO(C_{19}-H_{19}) = 0.343$ and $BO(H_{19}-O) = 0.471$ for the doublet and $BO(C_{19}-H_{19}) = 0.344$ and $BO(H_{19}-O) = 0.470$ for the quartet state.

In this way, due to the strong parallel drawn between TS structures, we can find similar imaginary frequencies for the hydrogen abstraction step in doublet (1606*i*) and quartet (1645*i*) spin states, and even when the deuterium atom replaces the transferred hydrogen atom (1209*i* and 1236*i* for doublet and quartet, respectively).

This analogous behavior between doublet and quartet spin states along the H-transfer coordinate, can also be found in the case of the QM-only model. In this model, the geometrical parameters obtained for the TS species along this coordinate are: $d(C_{19}-H_{19}) = 1.409 \text{ \AA}$, $d(H_{19}-O) = 1.141 \text{ \AA}$, and $C_{19}-H_{19}-O = 166.4^\circ$ for the doublet, and $d(C_{19}-H_{19}) = 1.413$

\AA , $d(H_{19}-O) = 1.143 \text{ \AA}$, and $C_{19}-H_{19}-O = 164.1^\circ$ for the quartet. The Wiberg BO indices obtained are: $BO(C_{19}-H_{19}) = 0.350$ and $BO(H_{19}-O) = 0.468$ for the doublet, $BO(C_{19}-H_{19}) = 0.348$ and $BO(H_{19}-O) = 0.469$ for the quartet. Finally, the imaginary frequencies for the hydrogen abstraction step are: 1601*i* in doublet and 1713*i* in quartet spin states, and when the deuterium atom replaces the transferred hydrogen atom are 1200*i* and 1277*i* for doublet and quartet, respectively. It is noteworthy that both the geometrical parameters and the BO indices as well as imaginary frequency numbers obtained are very similar between the QM-only model and the enzymatic system, and therefore, it is not surprising that the KIE results are similar in both models.

Corrections for barrier tunneling were determined by Wigner approximation and provided values in the range 1.4–1.5 in both models and spin states. As the Wigner correction yields values significantly greater than unity, it suggests the involvement of a significant amount of tunneling in this hydrogen abstraction step. Indeed, the corrected KIE values increase up to values around 10 in both systems and spin states, thus, exceeding the semiclassical Eyring limit (see Table 2). It is worth pointing out that tunneling effect has been reported in previous theoretical^[68–73] and experimental^[30,38,67,74–77] studies for this kind of hydrogen atom abstractions.

The Mulliken atomic spin densities obtained for the hydroxylation step using the B2 basis set are reported in Table 3. These results reveal the presence of three unpaired electrons in the reactant species: two electrons with parallel spin are located on the FeO moiety of the Cpd I forming a local triplet ($\rho_{FeO} \sim 2.10$) and a third electron is located between the π_s orbital at the sulfur atom of the Cys-437 and the a_{2u} orbital of the porphyrin ring. The sign of the latter (beta or alpha) determines either the doublet ($\rho_{Por+SCH_3} = -1.13$) or quartet ($\rho_{Por+SCH_3} = 0.95$) nature of the system, respectively. The results in Table 3 show that the porphyrin π cation radical is predominantly of A_{2u} character ($\sim 80\%$) in both doublet and quartet spin states, with a significant contribution of $I\pi_s$ state, leading to a “green” Cpd I instead of the “red” one. As long as

Table 4. Main geometrical parameters (distances in Å and angles in degrees) found in the active site of the enzyme at the stationary points (see Scheme 1 for atom labeling).

	Hydrogen abstraction step			Oxygen rebound step		
	Reactant	TS (H-Abs)	Intermediate	Intermediate	TS (rebound)	Product
Fe—S	2.681 [2.659]	2.588 [2.564]	2.527 [2.488]	2.520 [2.503]	2.540 [2.574]	2.395 [2.586]
Fe—O	1.625 [1.626]	1.761 [1.761]	1.801 [1.802]	1.806 [1.792]	1.875 [1.831]	2.874 [2.816]
O—C ₁₉	2.907 [2.906]	2.525 [2.518]	2.895 [2.871]	2.937 [2.960]	2.508 [2.602]	1.423 [1.422]
C ₁₉ —H ₁₉	1.091 [1.092]	1.399 [1.394]	2.044 [2.182]	2.457 [3.374]	2.500 [2.612]	1.969 [1.964]
O—H ₁₉	2.487 [2.383]	1.133 [1.134]	0.976 [0.974]	0.973 [0.974]	0.971 [0.974]	0.971 [0.970]
O'—H _{D2, Asp-309}	1.646 [1.665]	1.695 [1.692]	1.668 [1.667]	3.157 [3.185]	3.176 [3.207]	3.128 [3.149]
O'—H _{N, Met-374}	1.768 [1.767]	1.823 [1.821]	1.787 [1.793]	1.784 [1.771]	1.808 [1.794]	1.824 [1.815]
S—Fe—O	173.9 [173.6]	170.2 [170.2]	172.4 [172.8]	173.1 [171.9]	176.2 [172.8]	171.3 [171.6]
C ₁₉ —H ₁₉ —O	101.5 [107.6]	171.2 [169.6]	144.5 [126.6]	110.1 [57.1]	79.3 [78.7]	43.1 [43.2]
H ₁₉ —O—Fe—N _x	12.3 [11.6]	23.1 [20.2]	17.5 [−5.1]	−17.7 [−111.8]	−49.0 [−64.7]	−57.6 [−60.2]

The values are presented for both spin states: doublet and quartet (in brackets).

the first reaction step takes place, the hydrogen H₁₉ is transferred to the oxygen of the Cpd I from the C₁₉, leading to the formation of an alkyl radical on this carbon and a hydroxyl group linked to the iron atom. These results can be observed based on the values of the spin density population obtained for intermediate species, $\rho_{C_{19}} \sim 1.00$ and $\rho_{O-H_{19}} \leq 0.15$. Therefore, the unpaired electron of the hydrogen H₁₉ coming from the homolytic C—H cleavage is transferred to the oxygen atom, thus increasing the negative atomic charge on this atom (see Supporting Information Table S1).

It is worth noting that the iron-hydroxo complex formed in intermediate species was found in two different electromeric configurations for the quartet spin state (⁴Fe^{III}Por⁺ and ⁴Fe^{IV}Por) and only one for the doublet state (²Fe^{III}Por⁺). Both electromers found in the quartet state share the same geometry, but differ in the distribution of electrons on their atoms.^[13,78] The ^{2,4}Fe^{III}Por⁺ electromers contain one unpaired electron on the iron center ($\rho_{Fe} = 0.92$ (⁴Fe^{III}Por⁺), 1.02 (²Fe^{III}Por⁺)) and the other one distributed on both the a_{2u} orbital and the sulfur atom ($\rho_{Por+SCH_3} = 0.97$ (⁴Fe^{III}Por⁺), −1.09 (²Fe^{III}Por⁺)). The ⁴Fe^{IV}Por electromer has the two unpaired electrons located on the iron center ($\rho_{Fe} = 2.06$ and $\rho_{Por+SCH_3} = -0.17$). However, all the electromeric situations retain the radical character on the C₁₉ atom with a singly occupied orbital ($\rho_{C_{19}} = 1.03$ (²Fe^{III}Por⁺), 1.04 (⁴Fe^{III}Por⁺), and 1.05 (⁴Fe^{IV}Por)).

The oxygen rebound step is characterized by the formation of a new C—O bond, and involves the elimination of the radical character in the C₁₉ carbon atom and the formation of a closed shell porphyrin in the product species. As a result, two different electronic distributions for Fe^{III} are obtained in products defining the multiplicity state of the molecular system: the doublet state, which have an uncoupled electron on the iron atom ($\rho_{Fe} = 1.34$ and $\rho_{Por+SCH_3} = -0.34$) and the quartet state which have three uncoupled electrons on the iron (and partially in the sulfur axial ligand) ($\rho_{Fe} = 2.70$ and $\rho_{Por+SCH_3} = 0.29$).

Table 4 illustrates some of the most relevant geometrical parameters of the substrate and cofactor in the active site (for more details see Supporting Information Tables S2 and S3). The values obtained for both the Fe—S and Fe—O distances in

the reactant species, are consistent with those observed experimentally for the Cpd I^[79] (2.48 and 1.65 Å, respectively, using EXAFS spectroscopy). Concerning the evolution of these distances as the hydrogen abstraction step proceeds, it can be observed that the Fe—O bond is stretched due to the formation of the new hydroxyl group linked to the iron atom (from ~1.63 to ~1.80 Å), being this process accompanied by a shortening of the Fe—S distance (from ~2.70 to ~2.50 Å).

Furthermore, it can be seen that the donor–acceptor distance (C₁₉—O) remains practically invariant in the TSs of both doublet and quartet spin states (~2.52 Å). This fact can also be observed in the QM-only model (see Supporting Information Table S3), where the C₁₉—O bond distance in the TSs (~2.53 Å) is similar to that found in the enzymatic system. However, we can observe that for the reactant species obtained for the enzymatic model, the C₁₉—O distance (~2.91 Å) is shorter than the one found in the QM-only model (~3.51 Å). This observation suggests that one of the possible roles of the aromatase active site consists of the preorganization of the reactant state toward the TS along the reaction coordinate.

Regarding the C₁₉—H₁₉—O angle, it shows its maximum value in the geometries of the TSs corresponding to the hydrogen abstraction step. This angle takes values of ~170° in the TS of the enzymatic system and ~165° in the TS of the QM-model. The fact that this angle is virtually linear in the TS structures is what is expected for the hydrogen abstraction step, being this value closer to the linearity in the enzymatic system than in gas-phase model.

Previous to the oxygen rebound step, a reorientation of the hydroxyl group takes place to prepare the formation of the new C₁₉—O bond. This rotation can be observed by following the change that occurs in the H₁₉—O—Fe—N_x dihedral angle between intermediates and TS of the rebound step (around 31° for doublet and 47° for quartet), together with the shortening of the C₁₉—O bond (from ~2.94 to ~2.60 Å). Subsequently, this distance decreases dramatically from TS to products (from ~2.60 to ~1.42 Å) leading to the formation of a new C₁₉—O bond as well as a penta-coordinated iron complex.

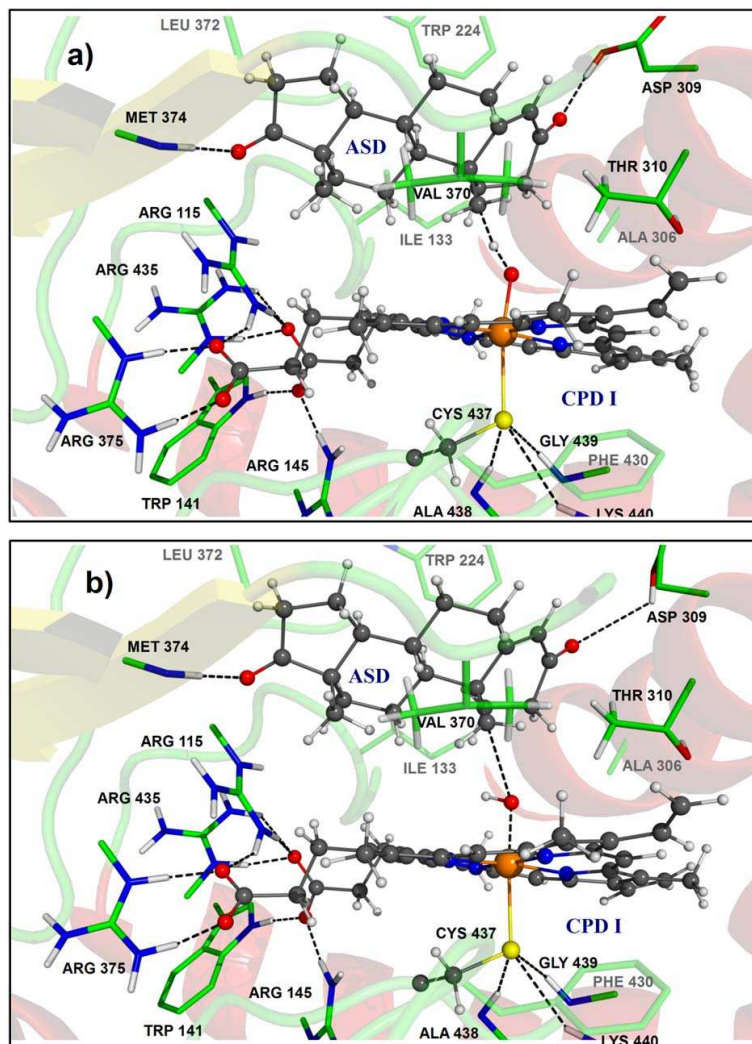


Figure 2. Snapshot of the enzymatic active site showing the TS structure for a) the hydrogen abstraction and b) the oxygen rebound steps. The most relevant interactions with the propionate chains of the Cpd I, the Cys-437, and the substrate are included.

Regarding the interactions established among the substrate/cofactor and the residues present in the active site, the following observations can be made: (i) there is a large network of hydrogen bonds at the binding site of the heme propionate chains involving four charged Arg (115, 145, 375, and 435) and a Trp-141 residues (see Fig. 2). These interactions balance the large repulsion of the carboxylate side chains of the porphyrin ring, facilitating thus an efficient binding site for the heme group in the active site. (ii) The partially charged sulfur atom of the Cys-437 is stabilized by the interaction with three polar peptidic hydrogens (Ala-438, Gly-439, and Lys-440). In fact, hydrogen bonds have been reported to stabilize the heme thiolate coordination and regulate the redox potential of the heme iron,^[80] preventing the elongation of the Fe–S bond

and avoiding the formation of a radical on the sulfur. (iii) The binding of the substrate is accomplished via hydrophobic interactions at the active site with the residues Ile-133, Trp-224, Ala-306, Thr-310, Val-370, Leu-372, Val-373, Phe-430, and Leu-477. (iv) Two hydrogen bonds are established between the polar moieties of the substrate ASD (carbonyl groups) with the Met-374 and protonated Asp-309 residues. Previous research suggests that the Asp-309 residue can participate in the enolization reaction of ASD as a proton donor in the third and last catalytic subcycle of aromatase enzyme.^[15]

The interaction energy between the substrate/cofactor and the enzymatic environment for the hydrogen atom abstraction step was calculated in the enzymatic system, using the windows corresponding to the reactants and the TSs obtained

Table 5. Relative QM/MM energy decomposition terms (values in kcal/mol) and gas-phase energies (ΔE_{vac}) for the hydrogen abstraction rate-limiting step.

	ΔE_{elec}	E_{pol}	E_{LJ}	$\Delta E_{\text{LJ}}(\text{ASD})$	$\Delta E_{\text{LJ}}(\text{Cys} + \text{heme})$	ΔE_{vac}
² R	[−630.4 ± 7.9]	0.0	[79.0 ± 1.5]	0.0	[−316.9 ± 7.1]	0.0
² TS	[−629.0 ± 8.7]	1.4	[76.1 ± 2.3]	−2.9	[−321.0 ± 9.3]	−4.1
⁴ R	[−621.4 ± 6.9]	9.0	[80.4 ± 1.8]	1.4	[−324.4 ± 7.8]	−7.5
⁴ TS	[−620.4 ± 8.0]	10.0	[74.5 ± 1.5]	−4.5	[−328.9 ± 8.6]	−12.0

Absolute values with the corresponding errors are shown in brackets. Lennard-Jones term has been broken down into substrate (ASD) and cofactor (Cys + heme).

from the FEP. This energy has been in turn decomposed into different energetic contributions as follows: (i) the electrostatic term (E_{elec}) which accounts for the interaction between the MM charges of the enzymatic environment and the polarized wave function of the substrate; (ii) the polarization energy (E_{pol}) of the wave function due to its changes on-going from gas-phase to the enzymatic active site; (iii) the Lennard-Jones term (E_{LJ}) representing the dispersive interactions between the MM and the QM atoms and (iv) the gas-phase energy (E_{vac}) of each species. Each of the interaction energy terms was obtained as an average of both doublet and quartet FEP calculations.

As can be seen in Table 5, the electrostatic interaction term (E_{elec}) for the doublet TS species shows a weaker value when compared with the corresponding reactant, leading to an unfavorable contribution to the activation energy (1.4 kcal/mol). This effect is similar in the quartet state, but less adverse than in the doublet (1 kcal/mol). Conversely, it is worth pointing that both polarization (E_{pol}) and Lennard-Jones (E_{LJ}) terms stabilize the TSs with respect to the corresponding reactant species, being different the weight of each term depending on the spin state. In this way, the polarization term in the quartet (−5.9 kcal/mol) is higher than in the doublet (−2.9 kcal/mol); however, a different behavior occurs for the Lennard-Jones where both the doublet and quartet TSs are stabilized in a similar way (−4.1 and −4.5 kcal/mol, respectively). Furthermore, the E_{LJ} term can in turn be broken down into the contributions of the substrate ($\Delta E_{\text{LJ}}(\text{ASD})$) and cofactor ($\Delta E_{\text{LJ}}(\text{Cys} + \text{heme})$), showing that the main contribution to this stabilization comes from the substrate (ASD) when comparing with the cofactor (Cys + heme), giving rise to values of −2.6 and −4.9 kcal/mol for the doublet and quartet states, respectively.

All in all, both the gas-phase barrier and the unfavorable electrostatic TS stabilization are offset by those contributions from the polarization and Lennard-Jones interaction, leading to lower activation free energies for both spin states. Despite the fact that the free energy of activation is lower for the quartet state, the reactant is destabilized compared with the doublet, making the overall process to be less affordable.

Conclusions

The above study investigates the hydroxylation of the substrate ASD to 19-hydroxy-ASD, occurring during the first catalytic subcycle of the enzyme aromatase, by means of MD simulations and hybrid QM/MM calculations. With this study,

we attempt to improve the understanding of the conversion mechanism of androgens into estrogens via the enzyme aromatase from the standpoint of theoretical chemistry.

In this oxidation step, the radical rebound mechanism (consisting in a first hydrogen abstraction step followed by an oxygen rebound step) as well as the Cpd I as the oxidizing species have been proposed.

According to the results obtained from the analysis performed on the PES, the rate-limiting step corresponds to the hydrogen abstraction process. The Gibbs free energy barriers for this step (obtained by means of the RRHO approximation) were found to be slightly lower for the enzymatic system (22.3 kcal/mol) than for the QM-only model (25.7 kcal/mol), showing a catalytic effect of the enzyme on the biochemical reaction. An analysis of the ²H-KIE calculated for this step show almost identical values in both models studied (QM-only model and enzymatic system), indicating that KIE value is intrinsic to the reaction and is not heavily affected by the enzymatic environment. Moreover, the results show a slight difference in KIE values between different spin states, which is explained by the great geometrical similarity found between the TS species obtained in both states. Once the tunneling correction is included, the KIE values exceed the semiclassical limit thus denoting a substantial tunnel effect (as has already been reported in some studies involving C—H abstractions).

Furthermore, we have incorporated (for the first time in this P450 isoform) the enzymatic conformational diversity in the calculations by means of the FEP methodology. As this methodology takes into account the enzyme flexibility, it provides a more realistic view of the role played by the enzyme in the biochemical process. According to the results obtained with this methodology, the free energy barriers for the hydroxylation step are lower than those calculated by way of the RRHO approximation, being the values obtained within the range expected for this kind of biological processes. The hydrogen abstraction step takes place through an endergonic process with an activation barrier of 13.5 kcal/mol, leading to the formation of an alkyl radical on the substrate and a hydroxyl group linked to the heme cofactor. From these intermediates, a reorientation of the hydroxyl group occurs previous to the oxygen rebound step. The energetics of this reorientation process depends on the spin state of the substrate, being the doublet state the more favored. Thereby, the rebinding of the oxygen befalls via the hydroxyl rotation and the formation of the new C—O bond. The coupled reorientation-rebound process is characterized by a remarkably low free activation

barrier of 1.5 kcal/mol, giving rise to the formation of the products through a highly exergonic reaction. Based in these energetic results and the KIEs obtained, we can conclude that the hydroxylation of substrates via the enzyme aromatase is compatible with the conventional hydrogen abstraction-oxygen rebound mechanism observed in other P450 isoforms.

Finally, the analysis of the decomposition of the free energy barrier (obtained by FEP calculations) in its different contributions, suggests that the electric field produced by the enzyme during the biochemical reaction is not directly involved in the stabilization of the TS when compared with the reactant species. In fact, it is noteworthy that both the polarization of the wave function and the Lennard-Jones term are the only ones responsible for such stabilization. These findings are consistent with the kind of chemical process and/or substrate. Therefore, we can conclude that the main role of the enzyme aromatase during the hydroxylation of ASD consists of a stabilization of the TS by means of the participation of both dispersive and polarization effects.

Acknowledgments

We acknowledge the Servei d'Informàtica of the Universitat Jaume I, GENCI-CINES, and BSC-Marenostrum for providing us with computer capabilities. The authors thank V. Moliner for valuable comments and discussion.

Keywords: cytochrome P450 · compound 1 · aromatase · hydrogen abstraction · androstenedione · CYP19A1 · quantum mechanics/molecular mechanics

How to cite this article: I. Viciano, R. Castillo, S. Martí. *J. Comput. Chem.* **2015**, *36*, 1736–1747. DOI: 10.1002/jcc.23967



Additional Supporting Information may be found in the online version of this article.

- [1] D. R. Nelson, *Hum. Genomics* **2009**, *4*, 59.
- [2] P. R. Ortiz de Montellano, Ed. *Cytochrome P450 Structure, Mechanism, and Biochemistry*, 3rd ed.; Kluwer Academic/Plenum Publishers: New York, **2005**.
- [3] J. S. Lee, R. S. Obach, M. B. Fisher, Eds. *Drug Metabolizing Enzymes: Cytochrome P450 and Other Enzymes in Drug Discovery and Development*; FontisMedia SA and Taylor & Francis Group: Boca Raton, FL 33487-2742, **2003**.
- [4] F. P. Guengerich, *Chem. Res. Toxicol.* **2008**, *21*, 70.
- [5] F. P. Guengerich, *Chem. Res. Toxicol.* **2001**, *14*, 611.
- [6] E. R. Simpson, C. Clyne, G. Rubin, W. C. Boon, K. Robertson, K. Britt, C. Speed, M. Jones, *Annu. Rev. Physiol.* **2002**, *64*, 93.
- [7] E. R. Simpson, S. R. Davis, *Endocrinology* **2001**, *142*, 4589.
- [8] E. R. Simpson, M. S. Mahendroo, G. D. Means, M. W. Kilgore, M. M. Hinshelwood, S. Grahmlorance, B. Amarneh, Y. J. Ito, C. R. Fisher, M. D. Michael, C. R. Mendelson, S. E. Bulun, *Endocrinol. Rev.* **1994**, *15*, 342.
- [9] M. Akhtar, J. N. Wright, P. Lee-Robichaud, *J. Steroid Biochem. Mol. Biol.* **2011**, *125*, 2.
- [10] M. Akhtar, M. R. Calder, D. L. Corina, J. N. Wright, *Biochem. J.* **1982**, *201*, 569.
- [11] R. J. Santen, H. Brodie, E. R. Simpson, P. K. Siiteri, A. Brodie, *Endocrinol. Rev.* **2009**, *30*, 343.
- [12] R. W. Brueggemeier, J. C. Hackett, E. S. Diaz-Cruz, *Endocrinol. Rev.* **2005**, *26*, 331.
- [13] S. Shaik, S. Cohen, Y. Wang, H. Chen, D. Kumar, W. Thiel, *Chem. Rev.* **2010**, *110*, 949.
- [14] D. Ghosh, J. Griswold, M. Erman, W. Pangborn, *J. Steroid Biochem. Mol. Biol.* **2010**, *118*, 197.
- [15] D. Ghosh, J. Griswold, M. Erman, W. Pangborn, *Nature* **2009**, *457*, 219.
- [16] K. Sen, J. C. Hackett, *Biochemistry* **2012**, *51*, 3039.
- [17] B. Krámos, J. Oláh, *J. Phys. Chem. B* **2013**, *118*, 390.
- [18] M. Akhtar, S. J. Skinner, *Biochem. J.* **1968**, *109*, 318.
- [19] D. Arigoni, R. Battaglia, M. Akhtar, T. Smith, *J. Chem. Soc. Chem. Commun.* **1975**, 185.
- [20] E. Caspi, T. Arunachalam, P. A. Nelson, *J. Am. Chem. Soc.* **1983**, *105*, 6987.
- [21] E. Caspi, T. Arunachalam, P. A. Nelson, *J. Am. Chem. Soc.* **1986**, *108*, 1847.
- [22] J. Fishman, J. Goto, *J. Biol. Chem.* **1981**, *256*, 4466.
- [23] S. L. Gantt, I. G. Denisov, Y. V. Grinkova, S. G. Sligar, *Biochem. Biophys. Res. Commun.* **2009**, *387*, 169.
- [24] M. Numazawa, M. Nagaoka, N. Sohtome, *Biochemistry* **2005**, *44*, 10839.
- [25] S. J. Skinner, M. Akhtar, *Biochem. J.* **1969**, *114*, 75.
- [26] Y. Khatri, A. Luthra, R. Duggal, S. G. Sligar, *FEBS Lett.* **2014**, *588*, 3117.
- [27] J. T. Groves, G. A. McCluskey, *J. Am. Chem. Soc.* **1976**, *98*, 859.
- [28] P. R. O. de Montellano, *Chem. Rev.* **2010**, *110*, 932.
- [29] T. M. Makris, K. von Koenig, I. Schlichting, S. G. Sligar, *Inorg. Biochem.* **2006**, *100*, 507.
- [30] J. Rittle, M. T. Green, *Science* **2010**, *330*, 933.
- [31] I. Schlichting, J. Berendzen, K. Chu, A. M. Stock, S. A. Maves, D. E. Benson, B. M. Sweet, D. Ringe, G. A. Petsko, S. G. Sligar, *Science* **2000**, *287*, 1615.
- [32] S. G. Sligar, *Science* **2010**, *330*, 924.
- [33] S. G. Sligar, T. M. Makris, I. G. Denisov, *Biochem. Biophys. Res. Commun.* **2005**, *338*, 346.
- [34] M. Newcomb, R. Shen, S.-Y. Choi, P. H. Toy, P. F. Hollenberg, A. D. N. Vaz, M. J. Coon, *J. Am. Chem. Soc.* **2000**, *122*, 2677.
- [35] M. Newcomb, R. Zhang, R. E. P. Chandrasena, J. A. Halgrimson, J. H. Horner, T. M. Makris, S. G. Sligar, *J. Am. Chem. Soc.* **2006**, *128*, 4580.
- [36] H. Isobe, S. Yamanaka, M. Okumura, K. Yamaguchi, J. Shimada, *J. Phys. Chem. B* **2011**, *115*, 10730.
- [37] I. G. Denisov, T. M. Makris, S. G. Sligar, I. Schlichting, *Chem. Rev.* **2005**, *105*, 2253.
- [38] R. E. P. Chandrasena, K. P. Vatsis, M. J. Coon, P. F. Hollenberg, M. Newcomb, *J. Am. Chem. Soc.* **2003**, *126*, 115.
- [39] M. Newcomb, D. Aebischer, R. Shen, R. E. P. Chandrasena, P. F. Hollenberg, M. J. Coon, *J. Am. Chem. Soc.* **2003**, *125*, 6064.
- [40] M. C. Gregory, I. G. Denisov, Y. V. Grinkova, Y. Khatri, S. G. Sligar, *J. Am. Chem. Soc.* **2013**, *135*, 16245.
- [41] S. Shaik, S. P. de Visser, F. Ogliaro, H. Schwarz, D. Schroder, *Curr. Opin. Chem. Biol.* **2002**, *6*, 556.
- [42] J. C. Phillips, R. Braun, W. Wang, J. Gumbart, E. Tajkhorshid, E. Villa, C. Chipot, R. D. Skeel, L. Kalé, K. Schulten, *J. Comput. Chem.* **2005**, *26*, 1781.
- [43] H. Li, A. D. Robertson, J. H. Jensen, *Proteins: Struct. Funct. Bioinf.* **2005**, *61*, 704.
- [44] W. L. Jorgensen, D. S. Maxwell, J. Tirado-Rives, *J. Am. Chem. Soc.* **1996**, *118*, 11225.
- [45] W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, M. L. Klein, *J. Chem. Phys.* **1983**, *79*, 926.
- [46] J. Baker, *J. Comput. Chem.* **1997**, *18*, 1079.
- [47] J. Baker, A. Kessi, B. Delley, *J. Chem. Phys.* **1996**, *105*, 192.
- [48] R. H. Byrd, P. Lu, J. Nocedal, C. Zhu, *SIAM J. Sci. Comput.* **1995**, *16*, 1190.
- [49] S. Martí, V. Moliner, I. Tunon, *J. Chem. Theory Comput.* **2005**, *1*, 1008.
- [50] Y. K. Zhang, H. Y. Liu, W. T. Yang, *J. Chem. Phys.* **2000**, *112*, 3483.
- [51] B. H. Besler, K. M. Merz, P. A. Kollman, *J. Comput. Chem.* **1990**, *11*, 431.
- [52] U. C. Singh, P. A. Kollman, *J. Comput. Chem.* **1984**, *5*, 129.
- [53] A. D. Becke, *J. Chem. Phys.* **1993**, *98*, 5648.
- [54] B. Miehlisch, A. Savin, H. Stoll, H. Preuss, *Chem. Phys. Lett.* **1989**, *157*, 200.
- [55] A. Altun, D. Kumar, F. Neese, W. Thiel, *J. Phys. Chem. A* **2008**, *112*, 12904.
- [56] H. Chen, J. S. Song, W. Z. Lai, W. Wu, S. Shaik, *J. Chem. Theory Comput.* **2010**, *6*, 940.

- [57] S. Shaik, D. Kumar, S. P. de Visser, A. Altun, W. Thiel, *Chem. Rev.* **2005**, *105*, 2279.
- [58] P. J. Hay, W. R. Wadt, *J. Chem. Phys.* **1985**, *82*, 299.
- [59] M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Rob, J. R. Cheeseman, J. A. Montgomery, Jr., T. Vreven, K. N. Kudin, J. C. Burant, J. M. Millam, S. S. Iyengar, J. Tomasi, V. Barone, B. Mennucci, M. Cossi, G. Scalmani, N. Rega, G. A. Petersson, H. Nakatsuji, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, M. Klene, X. Li, J. E. Knox, H. P. Hratchian, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, P. Y. Ayala, K. Morokuma, G. A. Voth, P. Salvador, J. J. Dannenberg, V. G. Zakrzewski, S. Dapprich, A. D. Daniels, M. C. Strain, O. Farkas, D. K. Malick, A. D. Rabuck, K. Raghavachari, J. B. Foresman, J. V. Ortiz, Q. Cui, A. G. Baboul, S. Clifford, J. Cioslowski, B. B. Stefanov, G. Liu, A. Liashenko, P. Piskorz, I. Komaromi, R. L. Martin, D. J. Fox, T. Keith, M. A. Al-Laham, C. Y. Peng, A. Nanayakkara, M. Challacombe, P. M. W. Gill, B. Johnson, W. Chen, M. W. Wong, C. Gonzalez, J. A. Pople, Revision D, 2nd ed.; Gaussian: Wallingford, CT, **2004**.
- [60] M. J. Field, M. Albe, C. Bret, F. Proust-De Martin, A. Thomas, *J. Comput. Chem.* **2000**, *21*, 1088.
- [61] E. Wigner, *Z. Phys. Chem. Abt. B.* **1932**, *19*, 203.
- [62] R. P. Bell, *The Tunnel Effect in Chemistry*; University Press: Cambridge, **1980**.
- [63] F. Louis, C. A. Gonzalez, J.-P. Sawersyn, *J. Phys. Chem. A* **2004**, *108*, 1058.
- [64] I. H. Williams, *Chem. Phys. Lett.* **1982**, *88*, 462.
- [65] I. H. Williams, *J. Mol. Struct. Theochem.* **1983**, *94*, 275.
- [66] K. B. Wiberg, *Tetrahedron* **1968**, *24*, 1083.
- [67] Q. Wang, X. Sheng, J. H. Horner, M. Newcomb, *J. Am. Chem. Soc.* **2009**, *131*, 10629.
- [68] S. P. de Visser, F. Ogliaro, P. K. Sharma, S. Shaik, *J. Am. Chem. Soc.* **2002**, *124*, 11809.
- [69] E. J. Klinker, S. Shaik, H. Hirao, L. Que, *Angew. Chem. Int. Ed.* **2009**, *48*, 129.
- [70] F. Ogliaro, M. Filatov, S. Shaik, *Eur. J. Inorg. Chem.* **2000**, 2455.
- [71] X. Zhang, Y. Liu, Y. Wang, *Theor. Chem. Acc.* **2014**, *133*, 1.
- [72] S. P. de Visser, *J. Am. Chem. Soc.* **2006**, *128*, 9813.
- [73] Z. Guo, X. Guan, J.-S. Huang, W.-M. Tsui, Z. Lin, C.-M. Che, *Chem. Eur. J.* **2013**, *19*, 11320.
- [74] X. Sheng, J. H. Horner, M. Newcomb, *J. Am. Chem. Soc.* **2008**, *130*, 13310.
- [75] P. H. Toy, M. Newcomb, M. J. Coon, A. D. N. Vaz, *J. Am. Chem. Soc.* **1998**, *120*, 9718.
- [76] C. V. Sastri, J. Lee, K. Oh, Y. J. Lee, T. A. Jackson, K. Ray, H. Hirao, W. Shin, J. A. Halfen, J. Kim, L. Que, S. Shaik, W. Nam, *Proc. Natl. Acad. Sci. USA* **2007**, *104*, 19181.
- [77] J. I. Manchester, J. P. Dinnocenzo, L. Higgins, J. P. Jones, *J. Am. Chem. Soc.* **1997**, *119*, 5069.
- [78] M. Filatov, N. Harris, S. Shaik, *Angew. Chem. Int. Ed.* **1999**, *38*, 3510.
- [79] K. L. Stone, R. K. Behan, M. T. Green, *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 16563.
- [80] D. Fishelovitch, C. Hazan, H. Hirao, H. J. Wolfson, R. Nussinov, S. Shaik, *J. Phys. Chem. B* **2007**, *111*, 13822.

Received: 10 December 2014

Revised: 7 April 2015

Accepted: 16 May 2015

Published online on 19 June 2015

Supporting Information

QM/MM modelling of the hydroxylation of the androstenedione substrate catalyzed by Cytochrome P450 Aromatase (CYP19A1)

*Ignacio Viciano, Raquel Castillo and Sergio Martí**

Departament de Química Física i Analítica, Universitat Jaume I, 12071 Castelló, Spain.

* to whom correspondence should be addressed

smarti@uji.es

Table S1: Mulliken Atomic Charges of different atoms or fragments included in the QM atoms of the enzymatic system obtained with the B2 basis set. The values are presented for both spin states: doublet and quartet (in brackets).

Fragment	Hydrogen Abstraction			Oxygen Rebound		
	Reactant	TS (H-Abs)	Intermediate	Intermediate	TS (rebound)	Product
SCH ₃ ⁻	-0.65 [-0.64]	-0.62 [-0.64]	-0.62 [-0.61]	-0.61 [-0.47]	-0.56 [-0.53]	-0.44 [-0.58]
Fe	1.07 [1.06]	1.06 [1.06]	0.92 [0.92]	0.94 [1.03]	0.97 [1.05]	0.96 [1.07]
O	-0.42 [-0.42]	-0.52 [-0.52]	-0.52 [-0.52]	-0.52 [-0.51]	-0.50 [-0.50]	-0.38 [-0.38]
N _α +N _β +N _γ +N _δ	-2.05 [-2.05]	-2.00 [-2.00]	-1.99 [-1.98]	-1.99 [-2.19]	-2.04 [-2.19]	-2.27 [-2.38]
Porphyrin*	1.88 [1.87]	1.70 [1.73]	1.74 [1.77]	1.76 [1.68]	1.55 [1.64]	1.51 [1.64]
Propionate _δ	-0.91 [-0.91]	-0.91 [-0.91]	-0.91 [-0.91]	-0.91 [-0.91]	-0.91 [-0.91]	-0.91 [-0.91]
Propionate _α	-0.89 [-0.89]	-0.90 [-0.90]	-0.90 [-0.90]	-0.89 [-0.89]	-0.90 [-0.89]	-0.90 [-0.89]
C	-0.21 [-0.22]	-0.29 [-0.29]	-0.24 [-0.23]	-0.21 [-0.16]	-0.08 [-0.12]	0.07 [0.07]
H	0.14 [0.15]	0.32 [0.32]	0.31 [0.27]	0.24 [0.26]	0.23 [0.25]	0.27 [0.27]
Substrate**	0.05 [0.04]	0.16 [0.15]	0.21 [0.19]	0.18 [0.15]	0.24 [0.19]	0.09 [0.10]

Table S2: Main distances (in Å) found in the active site of the aromatase enzyme during both steps, the hydrogen abstraction and the oxygen rebound. Values correspond to both the doublet and the quartet (in brackets) spin states. See Figure 1a for atom labeling.

Distance	Hydrogen Abstraction			Oxygen Rebound		
	Reactant	TS (H-Abs)	Intermediate	Intermediate	TS (rebound)	Product
Fe - N $_{\alpha}$	2.007 [2.006]	2.003 [2.011]	2.018 [2.020]	2.013 [1.984]	1.995 [1.997]	1.990 [1.983]
Fe - N $_{\beta}$	2.013 [2.017]	2.007 [2.011]	2.013 [2.014]	2.019 [2.002]	2.015 [2.009]	2.001 [2.007]
Fe - N $_{\gamma}$	2.012 [2.009]	2.026 [2.017]	2.017 [2.017]	2.020 [2.031]	2.021 [2.012]	2.015 [2.005]
Fe - N $_{\delta}$	2.015 [2.016]	2.012 [2.016]	2.011 [2.013]	2.024 [2.006]	2.026 [2.006]	2.012 [2.016]
Fe - S	2.681 [2.659]	2.588 [2.564]	2.527 [2.488]	2.520 [2.503]	2.540 [2.574]	2.395 [2.586]
Fe - O	1.625 [1.626]	1.761 [1.761]	1.801 [1.802]	1.806 [1.792]	1.875 [1.831]	2.874 [2.816]
O - C $_{19}$	2.907 [2.906]	2.525 [2.518]	2.895 [2.871]	2.937 [2.960]	2.508 [2.602]	1.423 [1.422]
C $_{19}$ - H $_{19}$	1.091 [1.092]	1.399 [1.394]	2.044 [2.182]	2.457 [3.374]	2.500 [2.612]	1.969 [1.964]
O - H $_{19}$	2.487 [2.383]	1.133 [1.134]	0.976 [0.974]	0.973 [0.974]	0.971 [0.974]	0.971 [0.970]
S - H $_{N,Ala-438}$	1.970 [1.958]	2.002 [1.994]	1.982 [1.976]	1.966 [1.975]	1.996 [1.976]	2.007 [1.964]
S - H $_{N,Gly-439}$	2.372 [2.378]	2.366 [2.355]	2.359 [2.361]	2.416 [2.409]	2.425 [2.406]	2.453 [2.429]
S - H $_{N,Lys-440}$	2.733 [2.707]	2.902 [2.935]	2.837 [2.796]	2.732 [2.776]	2.801 [2.764]	2.776 [2.653]
Prop-O $_{1\alpha}$ - H $_{E,Arg-115}$	4.999 [5.005]	5.007 [5.007]	5.004 [5.001]	5.050 [5.037]	5.051 [5.036]	5.050 [5.040]
Prop-O $_{1\alpha}$ - H $_{H11,Arg-115}$	3.458 [3.457]	3.453 [3.454]	3.452 [3.452]	3.430 [3.432]	3.428 [3.435]	3.430 [3.435]
Prop-O $_{1\alpha}$ - H $_{H12,Arg-115}$	1.689	1.682	1.681	1.665	1.662	1.664

	[1.688]	[1.682]	[1.682]	[1.668]	[1.670]	[1.671]
Prop-O _{1α} - H _{H21,Arg-115}	3.492 [3.497]	3.495 [3.495]	3.495 [3.492]	3.548 [3.550]	3.549 [3.549]	3.548 [3.548]
Prop-O _{1α} - H _{H22,Arg-115}	1.704 [1.710]	1.705 [1.703]	1.706 [1.705]	1.773 [1.772]	1.777 [1.768]	1.774 [1.771]
Prop-O _{1α} - H _{E,Arg-375}	1.587 [1.581]	1.584 [1.585]	1.586 [1.579]	1.623 [1.623]	1.617 [1.629]	1.624 [1.630]
Prop-O _{1α} - H _{H11,Arg-375}	5.123 [5.113]	5.134 [5.135]	5.135 [5.126]	5.187 [5.203]	5.183 [5.211]	5.189 [5.208]
Prop-O _{1α} - H _{H12,Arg-375}	5.479 [5.471]	5.481 [5.481]	5.483 [5.476]	5.568 [5.587]	5.564 [5.595]	5.569 [5.590]
Prop-O _{1α} - H _{H21,Arg-375}	4.160 [4.150]	4.170 [4.171]	4.171 [4.165]	4.341 [4.366]	4.342 [4.370]	4.342 [4.367]
Prop-O _{1α} - H _{H22,Arg-375}	2.719 [2.718]	2.705 [2.705]	2.709 [2.709]	2.806 [2.809]	2.804 [2.812]	2.806 [2.810]
Prop-O _{2α} - H _{E,Arg-115}	7.162 [7.165]	7.171 [7.171]	7.168 [7.165]	7.236 [7.225]	7.238 [7.225]	7.236 [7.229]
Prop-O _{2α} - H _{H11,Arg-115}	5.345 [5.335]	5.391 [5.393]	5.383 [5.375]	5.294 [5.328]	5.299 [5.330]	5.301 [5.332]
Prop-O _{2α} - H _{H12,Arg-115}	3.660 [3.653]	3.695 [3.696]	3.689 [3.682]	3.636 [3.656]	3.639 [3.657]	3.641 [3.660]
Prop-O _{2α} - H _{H21,Arg-115}	5.439 [5.445]	5.419 [5.418]	5.421 [5.421]	5.534 [5.519]	5.532 [5.521]	5.531 [5.521]
Prop-O _{2α} - H _{H22,Arg-115}	3.711 [3.715]	3.695 [3.692]	3.698 [3.697]	3.797 [3.786]	3.797 [3.786]	3.796 [3.789]
Prop-O _{2α} - H _{E,Arg-375}	2.648 [2.644]	2.656 [2.655]	2.655 [2.648]	2.635 [2.637]	2.633 [2.637]	2.638 [2.641]
Prop-O _{2α} - H _{H11,Arg-375}	5.466 [5.462]	5.477 [5.477]	5.477 [5.471]	5.466 [5.474]	5.459 [5.475]	5.467 [5.474]
Prop-O _{2α} - H _{H12,Arg-375}	5.110 [5.108]	5.105 [5.105]	5.107 [5.103]	5.125 [5.136]	5.117 [5.138]	5.124 [5.134]
Prop-O _{2α} - H _{H21,Arg-375}	3.194 [3.191]	3.186 [3.186]	3.190 [3.188]	3.248 [3.252]	3.241 [3.252]	3.243 [3.248]
Prop-O _{2α} - H _{H22,Arg-375}	1.553 [1.550]	1.549 [1.550]	1.552 [1.548]	1.587 [1.592]	1.579 [1.594]	1.584 [1.590]

Prop-O _{1δ} - H _{E,Arg-115}	5.293 [5.305]	5.243 [5.228]	5.248 [5.251]	5.399 [5.378]	5.396 [5.376]	5.389 [5.374]
Prop-O _{1δ} - H _{H11,Arg-115}	7.597 [7.593]	7.572 [7.563]	7.576 [7.576]	7.664 [7.663]	7.666 [7.663]	7.662 [7.661]
Prop-O _{1δ} - H _{H12,Arg-115}	6.650 [6.640]	6.644 [6.637]	6.645 [6.644]	6.690 [6.705]	6.692 [6.704]	6.689 [6.702]
Prop-O _{1δ} - H _{H21,Arg-115}	3.598 [3.594]	3.595 [3.585]	3.592 [3.592]	3.643 [3.642]	3.645 [3.638]	3.634 [3.636]
Prop-O _{1δ} - H _{H22,Arg-115}	4.762 [4.747]	4.785 [4.780]	4.781 [4.778]	4.780 [4.790]	4.782 [4.789]	4.777 [4.785]
Prop-O _{1δ} - H _{E1,Trp-141}	1.639 [1.631]	1.652 [1.663]	1.650 [1.642]	1.636 [1.645]	1.641 [1.655]	1.664 [1.666]
Prop-O _{1δ} - H _{E,Arg-145}	4.926 [4.915]	4.929 [4.944]	4.929 [4.928]	5.045 [5.060]	5.048 [5.063]	5.062 [5.068]
Prop-O _{1δ} - H _{H11,Arg-145}	1.674 [1.670]	1.669 [1.674]	1.672 [1.676]	1.682 [1.677]	1.682 [1.674]	1.685 [1.677]
Prop-O _{1δ} - H _{H12,Arg-145}	3.045 [3.042]	3.061 [3.061]	3.060 [3.063]	3.044 [3.045]	3.043 [3.039]	3.038 [3.038]
Prop-O _{1δ} - H _{H21,Arg-145}	5.228 [5.219]	5.223 [5.227]	5.224 [5.228]	5.215 [5.217]	5.214 [5.214]	5.213 [5.213]
Prop-O _{1δ} - H _{H22,Arg-145}	5.733 [5.724]	5.726 [5.735]	5.728 [5.729]	5.798 [5.802]	5.799 [5.801]	5.806 [5.804]
Prop-O _{1δ} - H _{E,Arg-435}	2.672 [2.663]	2.646 [2.641]	2.648 [2.645]	2.721 [2.713]	2.722 [2.716]	2.730 [2.720]
Prop-O _{1δ} - H _{H11,Arg-435}	6.131 [6.128]	6.101 [6.100]	6.104 [6.101]	6.309 [6.306]	6.312 [6.313]	6.324 [6.319]
Prop-O _{1δ} - H _{H12,Arg-435}	5.727 [5.718]	5.688 [5.690]	5.693 [5.689]	5.845 [5.847]	5.850 [5.855]	5.867 [5.863]
Prop-O _{1δ} - H _{H21,Arg-435}	5.133 [5.140]	5.128 [5.124]	5.127 [5.124]	5.304 [5.301]	5.304 [5.305]	5.308 [5.307]
Prop-O _{1δ} - H _{H22,Arg-435}	3.665 [3.675]	3.664 [3.656]	3.662 [3.659]	3.821 [3.814]	3.820 [3.816]	3.818 [3.816]
Prop-O _{2δ} - H _{E,Arg-115}	3.076 [3.091]	3.031 [3.014]	3.037 [3.040]	3.181 [3.164]	3.177 [3.162]	3.170 [3.157]
Prop-O _{2δ} - H _{H11,Arg-115}	5.652	5.624	5.631	5.699	5.696	5.691

	[5.661]	[5.615]	[5.636]	[5.698]	[5.697]	[5.689]
Prop-O _{2δ} - H _{H12,Arg-115}	4.988 [4.994]	4.973 [4.969]	4.976 [4.982]	5.004 [5.015]	5.002 [5.012]	4.998 [5.006]
Prop-O _{2δ} - H _{H21,Arg-115}	1.621 [1.623]	1.620 [1.618]	1.619 [1.623]	1.638 [1.640]	1.640 [1.635]	1.631 [1.632]
Prop-O _{2δ} - H _{H22,Arg-115}	3.275 [3.273]	3.289 [3.292]	3.287 [3.290]	3.260 [3.268]	3.260 [3.266]	3.256 [3.260]
Prop-O _{2δ} - H _{E1,Trp-141}	3.123 [3.122]	3.138 [3.145]	3.140 [3.139]	3.139 [3.147]	3.136 [3.150]	3.149 [3.151]
Prop-O _{2δ} - H _{E,Arg-145}	6.961 [6.959]	6.967 [6.983]	6.967 [6.968]	7.081 [7.097]	7.083 [7.098]	7.092 [7.100]
Prop-O _{2δ} - H _{H11,Arg-145}	3.609 [3.606]	3.593 [3.607]	3.596 [3.600]	3.604 [3.595]	3.607 [3.594]	3.614 [3.602]
Prop-O _{2δ} - H _{H12,Arg-145}	4.249 [4.248]	4.262 [4.279]	4.259 [4.260]	4.217 [4.221]	4.222 [4.219]	4.226 [4.226]
Prop-O _{2δ} - H _{H21,Arg-145}	6.618 [6.610]	6.611 [6.628]	6.610 [6.614]	6.589 [6.592]	6.593 [6.591]	6.597 [6.596]
Prop-O _{2δ} - H _{H22,Arg-145}	7.410 [7.408]	7.404 [7.421]	7.405 [7.408]	7.459 [7.462]	7.461 [7.461]	7.469 [7.465]
Prop-O _{2δ} - H _{E,Arg-435}	1.589 [1.599]	1.574 [1.572]	1.582 [1.586]	1.616 [1.621]	1.607 [1.622]	1.618 [1.615]
Prop-O _{2δ} - H _{H11,Arg-435}	5.294 [5.303]	5.293 [5.282]	5.299 [5.299]	5.370 [5.377]	5.361 [5.380]	5.367 [5.371]
Prop-O _{2δ} - H _{H12,Arg-435}	5.155 [5.164]	5.143 [5.138]	5.151 [5.154]	5.195 [5.205]	5.187 [5.208]	5.195 [5.199]
Prop-O _{2δ} - H _{H21,Arg-435}	4.057 [4.064]	4.090 [4.080]	4.089 [4.084]	4.148 [4.159]	4.142 [4.162]	4.144 [4.155]
Prop-O _{2δ} - H _{H22,Arg-435}	2.376 [2.382]	2.421 [2.410]	2.417 [2.410]	2.476 [2.486]	2.471 [2.490]	2.471 [2.483]
O' - H _{D2,Asp-309}	1.646 [1.665]	1.695 [1.692]	1.668 [1.667]	3.157 [3.185]	3.176 [3.207]	3.128 [3.149]
O'' - H _{N,Met-374}	1.768 [1.767]	1.823 [1.821]	1.787 [1.793]	1.784 [1.771]	1.808 [1.794]	1.824 [1.815]
N _α - H ₁₉	3.085 [3.021]	2.870 [2.857]	2.549 [2.464]	2.490 [3.370]	2.648 [2.811]	2.782 [2.807]

$N_8 - H_{19}$	4.295 [4.198]	3.408 [3.382]	3.171 [2.926]	2.793 [2.347]	2.416 [2.381]	2.603 [2.543]
----------------	------------------	------------------	------------------	------------------	------------------	------------------

Table S3: Main geometrical parameters (distances in Å and angles in degrees) obtained in the QM-only model during the hydrogen abstraction step for both doublet and quartet (in brackets) spin states. See Figure 1b for atom labeling.

	Hydrogen Abstraction		
	Reactant	TS (H-Abs)	Intermediate
Fe - S	2.639 [2.636]	2.542 [2.462]	2.482 [2.412]
Fe - O	1.625 [1.626]	1.752 [1.748]	1.807 [1.799]
O -C ₁₉	3.510 [3.532]	2.532 [2.532]	3.372 [3.358]
C ₁₉ -H ₁₉	1.094 [1.094]	1.409 [1.413]	3.274 [3.288]
O -H ₁₉	2.494 [2.506]	1.141 [1.143]	0.973 [0.974]
S-Fe-O	169.9 [171.3]	174.0 [174.7]	171.3 [171.7]
C ₁₉ -H ₁₉ -O	153.9 [155.8]	166.4 [164.1]	87.3 [85.7]
H ₁₉ -O-Fe-N _α	80.0 [79.5]	45.0 [53.8]	43.7 [74.6]

Table S4. Relative zero point energy (ZPE) and Gibbs contributions to the free energy (in kcal/mol) for the hydroxylation of ASD in the doublet and quartet spin states, using the B1 basis set. The Gibbs contribution includes both the ZPE and the thermal corrections, based on the rigid-rotor and the harmonic-oscillator (RRHO) approximations.

	ZPE	Gibbs
² R	0.0	0.0
⁴ R	0.0	-0.4
² TS _{H-abs}	-5.8	-5.6
⁴ TS _{H-abs}	-6.1	-5.4
² I	-3.3	-3.1
⁴ I	-2.9	-4.0
² TS _{reb}	-5.3	-2.9
⁴ TS _{reb}	-3.5	-3.0
² P	-0.7	1.3
⁴ P	-0.9	1.3

Details for the FEP error calculation:

The uncertainties reported for FEP energies in present work are based on the first-order expansion of the free energy, expressed as $\pm 2\sigma$:

$$\Delta F_{i \rightarrow j} = -RT \ln \left(\frac{1}{N} \sum_k e^{-\frac{\Delta U_k}{RT}} \right) \pm RT \frac{2\delta}{\frac{1}{N} \sum_k e^{-\frac{\Delta U_k}{RT}}}$$

$$\delta^2 = \frac{1 + 2\tau}{N} \left[\left(\frac{1}{N} \sum_k e^{-2\frac{\Delta U_k}{RT}} \right) - \left(\frac{1}{N} \sum_k e^{-\frac{\Delta U_k}{RT}} \right)^2 \right]$$

where $1+2\tau$ is the sampling ratio, defined as:

$$1 + 2\tau = \frac{1 + r_1}{1 - r_1} \quad r_1 = \frac{\sum_{i=2}^N (x_i - \bar{x})(x_{i-1} - \bar{x})}{\sum (x_i - \bar{x})^2}$$

For further details, please consider the following references:

[1] Chipot, C.

Free Energy Calculations in Biological Systems. How Useful Are They in Practice?

New Algorithms for Macromolecular Simulation, Springer Berlin Heidelberg, **2006**, 49, 185-211

[2] T. P. Straatsma, H. J. C. Berendsen, A. J. Stam, *Molecular Physics* **1986**, 57, 89-95

[3] E. B. Smith, B. H. Wells, *Molecular Physics* **1984**, 52, 701-704

4.2.2. Joint Use of Bonding Evolution Theory and QM/MM Hybrid Method for Understanding the Hydrogen Abstraction Mechanism via Cytochrome P450 Aromatase

Ignacio Viciano, Patricio González-Navarrete, Juan Andrés, and Sergio Martí

Journal of Chemical Theory and Computation **2015**, 11, 1470–1480

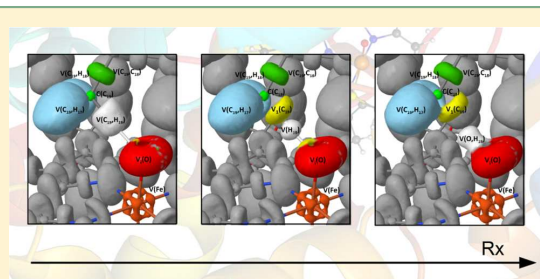
Joint Use of Bonding Evolution Theory and QM/MM Hybrid Method for Understanding the Hydrogen Abstraction Mechanism via Cytochrome P450 Aromatase

Ignacio Viciano, Patricio González-Navarrete, Juan Andrés,* and Sergio Martí

Departament de Química Física i Analítica, Universitat Jaume I, 12071, Castellón, Spain

Supporting Information

ABSTRACT: Bonding evolution theory (BET), as a combination of the electron localization function (ELF) and Thom's catastrophe theory (CT), has been coupled with quantum mechanics/molecular mechanics (QM/MM) method in order to study biochemical reaction paths. The evolution of the bond breaking/forming processes and electron pair rearrangements in an inhomogeneous dynamic environment provided by the enzyme has been elucidated. The proposed methodology is applied in an enzymatic system in order to clarify the reaction mechanism for the hydrogen abstraction of the androstenedione (ASD) substrate catalyzed by the cytochrome P450 aromatase enzyme. The use of a QM/MM Hamiltonian allows inclusion of the polarization of the charges derived from the amino acid residues in the wave function, providing a more accurate and realistic description of the chemical process. The hydrogen abstraction step is found to have five different ELF structural stability domains, whereas the C–H breaking and O–H forming bond process rearrangements are taking place in an asynchronous way.



1. INTRODUCTION

In most of the theoretical studies on chemical reactivity, relevant insights into the electronic structure are based on the localization of minimum energy reaction pathways between stable configurations on a high-dimensional potential energy surface (PES). This energy-based representation provides a wealth of information on the chemical reactivity; however, energy as a global quantity misses mechanistic details such as bond breaking/forming processes that are the essence of chemical reactivity.

Since the chemical reaction can be described as a sequence of events of a complex process in space and time, an adequate representation of these chemical events should be given by a physical observable. The electron density, $\rho(r)$, meets these requirements which can be achieved by Bader's quantum theory of atoms in molecules (QTAIM).¹ This analysis affords a rigorous and exact definition of bonding within an atomic ensemble in terms of topological properties of the $\rho(r)$.¹ Likewise, QTAIM provides a partition of the geometrical space occupied by the chemical system into adjacent nonoverlapping volumes called "basins", which can be thought as chemical entities such as atoms in molecules. Thus, determining the electron densities of all atoms can lead to establishing the evolution of the connectivity among them, while an energy-based representation associated with an accurate description of the electronic density along the reaction coordinate can provide a valuable notion about the reaction mechanism. In consequence, the methods which involve the analysis of $\rho(r)$

should have a particular appeal for chemists contributing to understanding the electronic structures of molecular processes and the chemical reactivity.¹

One of the most popular alternative partitions of the molecular space is provided by the topological analysis of the electron localization function (ELF), originally introduced by Becke and Edgecombe.² ELF can be understood as a local measure of the Pauli repulsion between electrons due to the exclusion principle, which allows defining regions of space that are associated with different electron pairs in a molecule.³ In this sense, the topological analysis of the ELF can also be treated as a continuous and differentiable scalar field in three-dimensional space which provides information on the structure of atomic shells, location, size of bonding, and lone electron pairs.⁴ Popelier proposed that these type of studies, based on the topological analysis of $\rho(r)$ and its related scalar field, form a unified theoretical framework, named "quantum chemical topology" (QCT), inspired by the seminal work of Bader.⁵ QCT provides a bridge between the traditional scheme of the chemical bond derived from the Lewis theory and first-principles quantum-mechanical methodologies.

Hybrid schemes based on quantum mechanics/molecular mechanics (QM/MM) are well-established computational techniques to describe theoretically the electronic structures in biochemical reactions taking into account the effects of

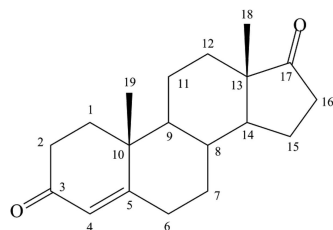
Received: November 19, 2014

Published: February 20, 2015

environment.⁶ The use of the topological analysis of the ELF has been successfully applied to large biological systems treated at the QM/MM level of theory, demonstrating its applicability in this kind of macromolecular entity.⁷ A further step is to take into account René Thom's catastrophe theory (CT)⁸ coupled with ELF, i.e., the use of bonding evolution theory (BET).^{8a} The mathematical foundations, as well as a number of cases where BET has been successfully applied, have been reviewed comprehensively in recent times.⁹ Since the topological behavior of the ELF gradient field can be studied within the framework provided by CT, BET allows finding a direct relationship between the ELF topology and the evolution of the bond breaking/forming processes and electron pair rearrangements as the chemical reaction proceeds.

In the present work, the joint use of the BET methodology with the QM/MM hybrid scheme has been applied in order to study the hydrogen abstraction process catalyzed by the cytochrome P450 aromatase enzyme. This biomolecule is a member of the superfamily of cytochrome P450 metalloenzymes (CYP19A1) and is uniquely known to be able to catalyze the biosynthesis of estrogens from androgens in humans.¹⁰ Since this enzyme is responsible for the final (and rate-limiting) step of estrogen formation, it has aroused considerable interest in the field of endocrinology due to its success as a therapeutic target in the treatment of hormone-dependent breast cancer.¹¹ The most important reactions catalyzed by aromatase are the conversion of androstenedione (ASD) (Scheme 1) and testosterone to estrone and estradiol¹²

Scheme 1. Structure of the Androstenedione Substrate



respectively via a three-step oxidative transformation (see Scheme 2). The first and second steps are accepted to be sequential hydroxylation of the C₁₉-methyl via the classical oxygen rebound mechanism that leads to the formation of the C₁₉-aldehyde (or C₁₉-oxo-ASD). The third "lyase" step (which is still controversial) culminates in the cleavage of the C₁₀-C₁₉ bond of the C₁₉-oxo-ASD, leading to the aromatization of the steroid A-ring of androgens and expelling the C₁₉ as formic acid.¹³

During the first oxidation step of ASD via the enzyme aromatase, the P450-mediated hydrogen atom abstraction-oxygen rebound mechanism originally proposed by Groves and co-workers¹⁴ takes place. Along this step, three substages can be distinguished as can be seen in Scheme 3:

(i) → (ii) An inert C-H bond of the substrate is activated by means of the high valent ferryl-oxo species commonly referred to as Compound I (Cpd I), leading to the formation of the alkyl radical intermediate and the iron-hydroxo complex.

(ii) → (iii) A reorientation (rotation) of the alkyl group and the OH radical takes place in order to prepare the oxygen rebound mechanism.

(iii) → (iv) A recombination of radical intermediate with the iron-bound hydroxyl radical leads to the formation of a new C-O bond (hydroxylated substrate).

Recently, the electronic structure of Cpd I (Fe(IV)-oxo porphyrin π -cation radical) was studied in our research group¹⁵ by means of the analysis of the ELF and the electron density. Cpd I has a radical nature due to three single occupied orbitals (two orbitals forming a local triplet focused on the Fe-O moiety, and one delocalized between the A_{2u} orbital of the porphyrin and the π_s of the sulfur atom). Depending on the coupling of these three electrons, Cpd I can adopt two different close-lying spin configurations: doublet (low spin) or quartet (high spin). For this reason, Shaik and co-workers¹⁶ have proposed a two-state reactivity mechanism (TSR)¹⁷ for the oxidation reactions carried out by cytochrome P450 enzymes. In our previous work, the interaction of the central iron atom with its coordinated ligands, as well as the structure of the carbon-carbon bonds in porphyrin moiety, was studied, providing deeper insights into the nature of the bonds and the electronic distribution of Cpd I.

Herein, we present an alternative representation of the electron density redistribution for the hydrogen abstraction process of the ASD substrate via the enzyme aromatase, at two competing doublet and quartet spin state surfaces. We provide the first study where the BET^{8a} has been used in a large biological system, employing electronic wave functions obtained from QM/MM calculations, thus incorporating the polarization effect of the charges derived from the enzymatic environment. The strategy adopted in this research provides a new insight into the electronic structure and related properties of the stationary points along the PES, enabling a nice guide to elucidate the mechanism of this biochemical process. Using this procedure, we are able to answer the following questions: (i) How could the electronic reorganization proceed along the reaction path? (ii) Is the electron density flowing synchronously, and in which direction? (iii) When and how do the bond formation/breaking processes along the reaction path take place? (iv) Does the formation/rupture of bonds take place before or after the transition structure (TS) is reached?

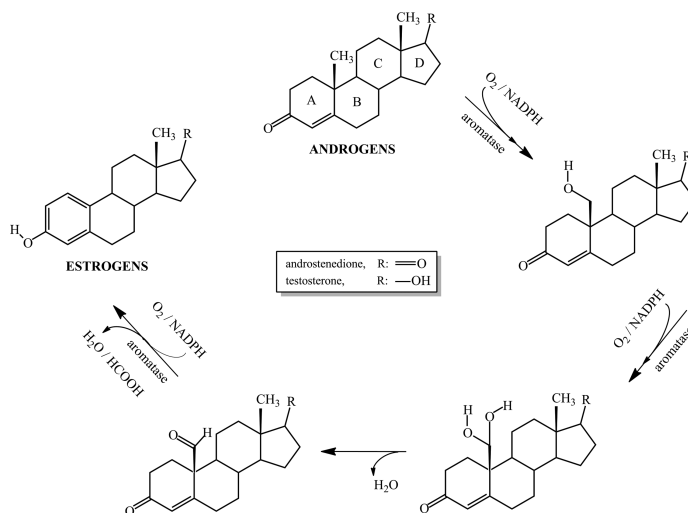
2. COMPUTATIONAL DETAILS

2.1. Obtaining QM/MM Potential Energy Surfaces.

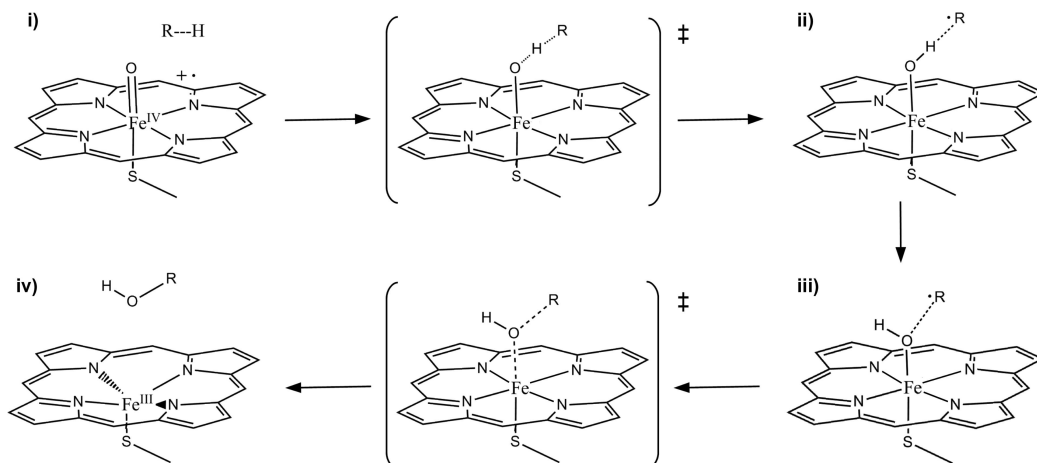
The initial geometry of the human placental aromatase (CYP19A1) in complex with its natural substrate androstenedione was obtained from the X-ray crystal structure (PDB code 3EQM).¹⁸ The iron pentacoordinated heme species was manually modeled into a hexacoordinated iron-oxo porphyrin, Cpd I. The protonation states of the residues were carefully defined according to the pK_a values obtained with the PROPKA web interface.¹⁹ All the hydrogen atoms were added using the fDYNAMO²⁰ library, and a total of four counterions (Cl⁻) were placed into electrostatically favorable positions around the enzyme in order to electroneutralize the excess of positive charge. Subsequently, the system was placed in a prerelaxed orthorhombic box of water molecules with dimensions of 90 Å × 80 Å × 80 Å. The resulting model consisted of 452 residues of amino acids, the modeled Cpd I, the substrate ASD, 4 counterions, 35 crystallographic water molecules, and 16 542 water molecules of the solvation box.

A semiempirical QM/MM relaxation of this model was carried out using the AM1²¹ Hamiltonian for the ASD substrate, and the OPLS-AA²²/TIP3P²³ force field. Then, the resulting structure was equilibrated by means of hybrid

Scheme 2. Catalytic Process of Conversion of Androgens into Aromatic Estrogens Catalyzed by Aromatase



Scheme 3. Hydroxylation of Substrates via Cytochrome P450



molecular dynamics (MD) at 300 K, using the NVT ensemble and the Langevin–Verlet integrator. The MD was run for 200 ps with a step size of 1 fs, keeping frozen any atom 20 Å apart from the Cpd I or the substrate during all the simulations.

Once the system was equilibrated, the PES was explored for the hydrogen abstraction reaction in both doublet and quartet spin states. The QM region of the model consisted of the substrate androstenedione, Cpd I, and the axial Cys-437 ligand (comprising the S_γ , the C_β , and two H_β atoms), yielding a total amount of 124 atoms, as shown in Figure 1. The link atom^{6b} procedure was applied to satisfy the valence of the QM region, due to the presence of the classical bond partitioning in the Cys-437 amino acid (C_α – C_β bond).

The transition structures corresponding to doublet and quartet spin states were localized and characterized as saddle points of order 1 by means of a micro–macro iteration optimization scheme.²⁴ In this method, the system is divided into two parts: the core or control space, and the environment

or complementary space. The Hessian matrix is calculated explicitly only for those atoms belonging to the core. Thus, before each step of the control space Hessian guided optimization (based on the Baker algorithm^{25a,b}), the degrees of freedom of the environment are optimized in order to maintain an approximately zero gradient and to minimize the potential energy (using the L-BFGS procedure^{25c}). In our case, the core region matches up with the QM region described above. Once the TSs were located and characterized, subsequent minimum energy paths (MEPs) were traced down toward the corresponding minima, to ensure that the structures really connect reactants and intermediates. The reaction paths were obtained for both doublet and quartet spin state PESs from the corresponding TSs using a reaction coordinate (Rx) in a mass-weighted step of 0.01 amu^{1/2} bohr. The calculations were performed at UB3LYP^{26a,b}/LACVP*^{26c} (consisting of the combination of the 6-31G(d) basis for all the atoms except for the iron one, which was represented by the

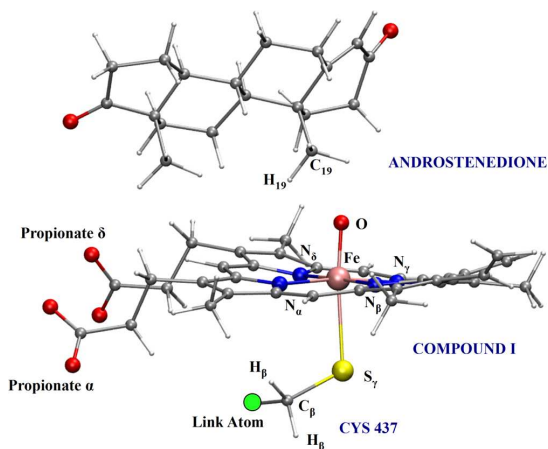


Figure 1. Atoms comprising the QM model: Cys-437, Compound I, and the substrate androstenedione.

LANL2DZ effective core potential) level. In addition, single point energy calculations for selected points along the MEP (50 points toward reactants and 50 toward intermediates) were carried out at the UB3LYP/6-31G(d) level, thus avoiding the use of pseudopotentials on the iron atom.

Finally, in order to check the effect of the enzymatic environment, calculations on an unpolarized gas-phase model were carried out. For this purpose, the same geometrical structures of the reaction paths were selected, and single point energy calculations were performed using the UB3LYP/6-31G(d) level, but without incorporating the enzymatic environment.

All the electronic calculations were performed using the Gaussian 03²⁷ family of programs, whereas the fDYNAMO²⁰ library was used for the molecular mechanics and MD simulations.

2.2. Evaluation of the Topological Analysis of the ELF and Thom's CT (BET Method). ELF and electron density analyses were performed at each point selected from the reaction path (100 points) by means of Dgrid-4.6 package²⁸ considering a rectangular parallelepiped grid with a step size of 0.1 bohr. The ELF basins were visualized using the Jmol²⁹ and VMD programs.³⁰

The ELF function is a mathematical function that measures the likelihood of finding an electron in the neighborhood space of a reference electron located at a given point and with the same spin. This function takes numerical values mapped on the interval (0,1). Values close to 1 provide regions of the space where the probability of finding an electron pair in the molecular space is high. The topological analysis of the ELF gradient field³¹ represents an alternative to the valence bond theory and molecular orbital approaches and provides the partition of the molecular space into basins of attractors that correspond to atomic cores, bonds, and lone pairs. The atomic cores coincide with the atomic nuclei and are labeled as C(A), being A the atomic symbol of the element. The concept of synpacticity, introduced by Silvi et al.,^{31b,32} allows the classification of η -localization basins in the ELF domain. In this way, it is possible to find monosynaptic basins, labeled as V(A), which correspond to lone pairs of the Lewis model, as well as disynaptic basins labeled as V(X,Y) which correspond

with the ELF-localization basins with common surfaces between two C(X) and C(Y) core basins and are related to the chemical bonds. The mathematical branch that provides an appropriate framework to achieve the partition of the molecular space in regions related to the chemical properties is the theory of dynamical systems.^{8b,33} A dynamical system is a mathematical model that describes the temporal evolution of the state of a physical system. The simplest model is the gradient dynamical system in which the gradient field of a scalar function provides the vector field.³⁴ For example, the differential equation of motion $\dot{x} = (dx/dt) = v(x)$ involves the time derivative of a trajectory on the phase space.

The elementary CT studies how the equilibria of a gradient system (or dynamical system) change as the control parameters change, by considering the behavior of the Hessian matrix (H_{ij}). According to this theory, the set of points at which the determinant of the Hessian matrix is nonzero, $\det |H_{ij}| \neq 0$, are hyperbolic critical points; in contrast, the set of points at which the $\det |H_{ij}| = 0$ are nonhyperbolic critical points or bifurcation points. The set of values for which the determinant of the Hessian matrix of a given critical point is nonzero defines the domain of stability of the critical point also known as structural stability domains (SSDs). When a small perturbation of the function (by means of changes in the control parameters) leads to a zero value of the determinant of the Hessian matrix (nonhyperbolic point), the system changes from one SSD to another through a bifurcation point (catastrophic point).

In a chemical reaction, the changes on the control parameters defining the reaction pathway (such as the nuclear coordinates and the electronic state) can lead to different topologies of the ELF. Along the reaction pathway, the chemical system goes from a given ELF-SSD to another by means of bifurcation catastrophes occurring at the turning points (TPs). In this way, a chemical reaction can be viewed as a dynamical system where a sequence of elementary chemical processes occur, being characterized by catastrophes. The bifurcation catastrophes occurring at these turning points may be identified according to Thom's classification. Only three types of bifurcation catastrophes have been found in chemical reactivity: (i) the fold catastrophe, corresponding to the creation or to the annihilation of two critical points of different parities; (ii) the cusp catastrophe, which transforms one critical point into three (and vice versa) such as in the formation or the breaking of a covalent bond; (iii) the elliptic umbilic, in which the index of a critical point changes by two. The characterization of the TPs connecting the ELF-SSDs along the reaction pathway allows us to catalog the different chemical events occurring during the course of the reaction as a sequence of electron pair rearrangements taking place during a chemical transformation.

3. RESULTS AND DISCUSSION

3.1. Energetic and Geometrical Parameters. The results presented in this paper correspond to the hydrogen abstraction from the substrate ASD by Cpd I in the active site of enzyme aromatase. In this reaction, a hydrogen abstraction process, that is also viewed as a case of hydrogen-atom transfer,³⁵ takes place from the C₁₉ methyl of ASD to the ferryl (Fe=O) group of Cpd I, leading to the formation of the ASD alkyl radical intermediate as well as the iron-hydroxo complex.

The relative electrostatic potential energies (eq 1) obtained for the stationary points are shown in Table 1:

$$E = E_{\text{elec}} = \langle \Psi | \hat{H}_0 | \Psi \rangle + \sum_i^{\text{electrons}} \sum_m^{\text{MM atoms}} \left\langle \Psi \left| \frac{q_m}{r_{i,m}} \right| \Psi \right\rangle \quad (1)$$

Table 1. Relative Electrostatic Potential Energies (ΔE , kcal/mol) for the Hydrogen Abstraction Process of Substrate ASD via Aromatase Enzyme Calculated at the UB3LYP/6-31G(d) Level^{a,b}

spin	electromer	ΔE (kcal/mol)		
		R	TS	I
doublet	$^2\text{Fe}^{\text{III}}\text{Por}^{\bullet+}$	0.0	32.8	25.4
quartet	$^4\text{Fe}^{\text{III}}\text{Por}^{\bullet+}$	0.5	35.0	25.0
	$^4\text{Fe}^{\text{IV}}\text{Por}$		36.3	23.2

^aThe values correspond to the doublet and quartet spin state surfaces. The latter includes the energy of both electromeric situations found in the intermediate ($^4\text{Fe}^{\text{III}}\text{Por}^{\bullet+}$ and $^4\text{Fe}^{\text{IV}}\text{Por}$). The origin of energies has been set at the lowest energy reactant (doublet). ^bR, reactant; TS, transition state; I, intermediate.

The term $\langle \Psi | \hat{H}_0 | \Psi \rangle$ of eq 1 accounts for the selected gas-phase Hamiltonian (\hat{H}_0) operating on the polarized wave function (Ψ), while the second term corresponds to the electrostatic interaction of Ψ with the charges representing the atoms of the enzymatic environment (q_m) at a distance $r_{i,m}$. The MM charges that have been considered are those that are within 20 Å of Cpd I or the ASD substrate (a total of 8512 charges). It is worth noting that those terms corresponding to the MM potential of the total QM/MM energy have been neglected, since we are only interested in the electronic properties of the QM part. Thus, although the reaction paths obtained depend upon the Lennard-Jones interaction between the QM and MM regions, this classical contribution does not affect directly the wave function, and therefore it is important to remark that the results obtained from subsequent ELF based analysis are maintained.

As can be observed in Table 1 and Figure 2, the activation energy barriers for the enzymatic process are predicted to be 32.8 and 35.0 kcal/mol for the doublet and quartet, respectively. The slight difference between both values suggests that both states might be accessible during the reaction.

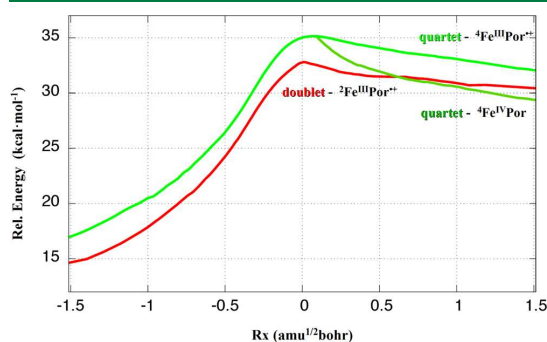


Figure 2. Potential energy profile corresponding to the hydrogen abstraction reaction from ASD in doublet and quartet spin states (reactant and intermediate species are not included). A bifurcation point on the quartet PES can be observed after the transition state, leading to the formation of two different electromers for the quartet state ($^4\text{Fe}^{\text{III}}\text{Por}^{\bullet+}$ and $^4\text{Fe}^{\text{IV}}\text{Por}$).

An analysis of Figure 2 shows the presence of a post-transition state bifurcation on the quartet state PES leading toward two different close-lying electromer intermediates that vary in the oxidation state of the iron center and the porphyrin ligand.³⁶

According to the $^{2S+1}\text{I}(\text{Fe}^{\text{OS}}\text{Por})$ nomenclature, four low-lying electronic states may be found ($^2\text{Fe}^{\text{III}}\text{Por}^{\bullet+}$ and $^2\text{Fe}^{\text{IV}}\text{Por}$) depending on the spin quantum number ($S = 1/2, 3/2$) and the oxidation states ($\text{OS} = \text{III}, \text{IV}$) on the iron center and the ligands.³⁷ In particular, only one electromeric situation was found for the doublet spin state ($^2\text{Fe}^{\text{III}}\text{Por}^{\bullet+}$) and two for the quartet state ($^4\text{Fe}^{\text{III}}\text{Por}^{\bullet+}$, $^4\text{Fe}^{\text{IV}}\text{Por}$) in the enzymatic system. In fact, this bifurcation corresponds to the crossing point of two different adiabatic surfaces for each electromeric distribution in the quartet state. In the present work we have eluded the initial part of the $^4\text{Fe}^{\text{IV}}\text{Por}$ reaction path previous to the crossing point, as long as it corresponds to a higher energy mechanistic profile.

On the other hand, the presence of electromers is not observed for gas-phase models. This can be attributed to the absence of specific interactions with the enzymatic environment³⁸ (mostly related to the sulfur axial ligand of the heme group). Therefore, a single energy profile has been obtained for each electronic state (doublet and quartet) when single point gas-phase calculations are performed along the enzymatic reaction paths (see Table S1 and Figure S1 of the Supporting Information).

The hydrogen abstraction process has been found to be endothermic with ΔE values of 25.4 kcal/mol for the doublet spin state, and 25.0 and 23.2 kcal/mol for the $^4\text{Fe}^{\text{III}}\text{Por}^{\bullet+}$ and $^4\text{Fe}^{\text{IV}}\text{Por}$ electromers, respectively. Likewise, after the system overcomes the transition state, the energy of the doublet lies between both quartet electromers, with $^4\text{Fe}^{\text{IV}}\text{Por}$ being about 2 kcal/mol more stable than $^4\text{Fe}^{\text{III}}\text{Por}^{\bullet+}$. Thus, the intermediate formed in this step brings about an alkyl radical on the C₁₉ methyl of ASD as well as a hydroxyl group linked to the iron atom, and later gives rise to the oxygen rebound mechanism step.

The fact that the reported energies are too large for a chemical process catalyzed by an enzyme can be explained taking into account the following considerations: (i) These energies solely arise from the electronic terms of the hybrid QM/MM potential (see eq 1). (ii) The results are based merely on the PES analysis. In fact, the estimation of the Gibbs free energy of activation calculated with the 6-31G(d) basis set, by means of the total QM/MM potential energy and the rigid-rotor and harmonic-oscillator (RRHO) approximations applied to the ASD substrate, provides a lower activation barrier (25.5 kcal/mol) for the doublet process.

Furthermore, we have recently found that the free energy of activation for the hydrogen abstraction process, evaluated using free energy perturbation techniques, is about 13 kcal/mol, using the LACVP* basis set.³⁹ In the same study we determined that the major contributions to reducing the activation energy come from the polarization of the wave function and the Lennard-Jones potential; the electrostatic interaction is nonfavorable for the reaction in this particular chemical system.

All in all, it is not our intention to discuss in the present paper the role of cytochrome P450 aromatase as a biocatalyst, but to shed light on the different electronic rearrangements that take place during the hydrogen abstraction process carried out by this enzyme.

Table 2. Atomic Spin Densities for Atoms or Fragments Included in the QM Model^a

fragment	spin	R	TS		I	
			Fe ^{III} Por ^{••}	Fe ^{IV} Por	Fe ^{III} Por ^{••}	Fe ^{IV} Por
SCH ₃ ⁻	doublet	-0.19	-0.19		-0.16	
	quartet	0.17	0.13	-0.07	0.08	-0.07
Fe	doublet	1.39	1.13		1.03	
	quartet	1.32	1.06	2.03	0.96	2.10
O	doublet	0.73	0.46		0.10	
	quartet	0.74	0.50	0.46	0.12	0.11
Fe=O	doublet	2.12	1.59		1.14	
	quartet	2.06	1.56	2.48	1.09	2.21
C ₁₉	doublet	0.00	0.57		1.01	
	quartet	0.00	0.59	0.68	1.03	1.02
H ₁₉	doublet	0.00	-0.07		-0.00	
	quartet	0.00	-0.08	-0.04	-0.00	0.02
porphyrin	doublet	-0.93	-0.87		-0.95	
	quartet	0.76	0.81	-0.05	0.87	-0.11
substrate ^b	doublet	0.00	-0.02		-0.07	
	quartet	0.00	-0.02	-0.01	-0.06	-0.06

^aThe values are presented for both doublet and quartet spin states, with the latter including the two electromeric states found in the transition states and intermediate species (Fe^{III}Por^{••} and Fe^{IV}Por). ^bDoes not include the C₁₉ and H₁₉ atoms.

The Mulliken atomic spin densities obtained from single point calculations are reported in Table 2 (as well as in Table S2 of the Supporting Information for gas-phase calculations). These results reveal that two spin-up electrons are located on the FeO moiety ($\rho(\text{FeO}) \sim 2.10$), forming a local triplet. A third electron is localized between the a_{2u} orbital of the porphyrin ring that leads to the A_{2u} electronic state and the π_s orbital of the Cys-437 sulfur atom leading to a Π_s electronic state. The sign of this unpaired spin density population (beta or alpha) determines either the doublet or quartet nature of the system, respectively. The results in Table 2 show a predominant A_{2u} state versus the Π_s state (with a ratio of $\sim 80\%$), in both doublet and quartet spin states, corresponding to a “green” Cpd I instead of the “red” one. As the reaction proceeds, the hydrogen H₁₉ is transferred to the oxygen of Cpd I from the C₁₉, leading to the formation of a radical on this carbon ($\rho(\text{C}_{19}) \sim 1.00$) and a hydroxyl group ($\rho(\text{OH}) \sim 0.10$) linked to the iron atom. As discussed above, the iron–hydroxo complex has been found in two different electromeric configurations in the quartet spin state and only one in the doublet electronic state. The first electromer (Fe^{III}Por^{••}), containing one unpaired electron on the iron center and the other one distributed over the a_{2u} orbital and the sulfur atom, was found in both doublet ($\rho(\text{Fe}) = 1.03$, $\rho(\text{Por}+\text{SCH}_3) = -1.08$) and quartet states ($\rho(\text{Fe}) = 0.96$, $\rho(\text{Por}+\text{SCH}_3) = 1.05$). The second electromeric state (Fe^{IV}Por) containing the two unpaired electrons on the iron center ($\rho(\text{Fe}) = 2.10$ and $\rho(\text{Por}+\text{SCH}_3) = -0.19$) was only found in the quartet state, being 2 kcal/mol lower in energy than the other configuration. However, it is worth noting that all electromeric situations found retain the radical character on the C₁₉ atom with a singly occupied orbital ($\rho(\text{C}_{19}) = 1.01$ (²Fe^{III}Por^{••}), 1.03 (⁴Fe^{III}Por^{••}), and 1.02 (⁴Fe^{IV}Por)).

The most important geometrical parameters of the substrate and cofactor obtained during the hydrogen abstraction step are presented in Table 3. As can be noticed, the values obtained for both the Fe–S and Fe–O distances in the reactant species (R) are consistent with those observed experimentally for Cpd I⁴⁰ (2.48 and 1.65 Å respectively, using EXAFS spectroscopy). As the reaction proceeds, the Fe–O bond is stretched and there is

Table 3. Selected Geometrical Parameters^a

	doublet			quartet		
	R	TS	I	R	TS	I
Fe–S	2.681	2.588	2.527	2.659	2.564	2.488
Fe–O	1.625	1.761	1.801	1.626	1.761	1.802
O–C ₁₉	2.907	2.525	2.895	2.906	2.518	2.871
C ₁₉ –H ₁₉	1.091	1.399	2.044	1.092	1.394	2.182
O–H ₁₉	2.487	1.133	0.976	2.383	1.134	0.974
S–Fe–O	173.9	170.2	172.4	173.6	170.2	172.8
C ₁₉ –H ₁₉ –O	101.5	171.2	144.5	107.6	169.6	126.6

^aDistances are in angstroms, and angles are in degrees.

a shortening in the bond length between the iron atom and the sulfur axial ligand due to the formation of the new hydroxyl group ($d(\text{O}–\text{H}_{19}) = 0.976$ Å). Moreover, the C₁₉–H₁₉–O angle shows its maximum value close to the linearity ($\sim 170^\circ$) at the transition structure of the hydrogen abstraction step, as expected for this kind of reaction.

3.2. Topological Analysis. The basin populations along the MEP pathway are summarized in Tables 4, 5, and 6. For convenience, only basin populations directly related to the rearrangements of the C₁₉–H₁₉ and O–H₁₉ bonds have been considered for discussion. Moreover, due to the valence shell compactness of the O atom, the overall population of its monosynaptic basins ($V_u(\text{O})$) has been considered instead of their internal reorganizations, thus focusing on those ELF topological changes which are more relevant for the hydrogen abstraction. Likewise, the specific monosynaptic basin found surrounding the C(O) core basin region on the line joining the H₁₉ and O nuclei, labeled $V_s(\text{O})$, has been treated apart from the rest of $V_u(\text{O})$ monosynaptic basins, as long as it would be directly related to the radical character of the ferryl moiety (FeO). Finally, the reaction coordinates have been traced from -9.8 to 6.9 amu^{1/2} bohr for the doublet and from -10.7 to 8.9 amu^{1/2} bohr for the quartet.

The MEP of the hydrogen abstraction step is reported in Figure 3 together with the different ELF topologies along the reaction coordinate (also, a video animation is available in the Supporting Information). In terms of the ELF topological

Table 4. Population of the Most Significant Basins Involved in the Hydrogen Abstraction Reaction in Doublet Spin State^a

basin	SSD-I		SSD-II		SSD-III			SSD-IV		SSD-V	
	reactant	last point	first point	last point	first point	TS	last point	first point	last point	first point	intermediate
V(C ₁₈ ,C ₁₉)	1.85	1.86	1.86	1.85	1.85	1.82	1.78	1.78	1.75	1.75	1.66
V(C ₁₉ ,H ₁₇)	2.01	2.03	2.03	2.04	2.04	2.06	2.07	2.07	2.01	2.01	1.84
V(C ₁₉ ,H ₁₈)	2.00	2.02	2.02	2.03	2.03	2.05	2.09	2.10	2.22	2.02	1.80
V(C ₁₉ ,H ₁₉)	2.01	2.01	2.01	2.01							
V ₁ (C ₁₉)					1.29	1.60	1.69	1.69	1.69	1.69	1.41
V ₂ (C ₁₉)										0.20	1.07
V(H ₁₉)					0.71	0.51	0.59				
V _a (O)	7.38	7.29	6.92	6.95	6.98	6.34	6.14	6.13	6.07	6.07	5.97
V _x (O)			0.38	0.34	0.29	0.87	0.94				
V(O,H ₁₉)								1.53	1.57	1.57	1.60
Rx (amu ^{1/2} bohr)	-9.841	-0.527	-0.489	-0.323	-0.304	0.000	0.262	0.280	0.947	0.966	6.947
energy (kcal/mol)	0.00	23.78	24.44	27.90	28.34	32.77	31.88	31.83	31.01	30.97	25.39

^aThe anterior and posterior points to the turning points as well as the stationary points are also included.

Table 5. Population of the Most Significant Basins Involved in the Hydrogen Abstraction Reaction in Quartet Spin State, for the ⁴Fe^{III}Por^{•+} Electromer^a

basin	SSD-I		SSD-II		SSD-III			SSD-IV		SSD-V	
	reactant	last point	first point	last point	first point	TS	last point	first point	last point	first point	intermediate
V(C ₁₈ ,C ₁₉)	1.85	1.86	1.84	1.86	1.84	1.85	1.77	1.76	1.74	1.75	1.68
V(C ₁₉ ,H ₁₇)	2.01	2.04	2.04	2.04	2.04	2.06	2.06	2.06	1.99	1.99	1.85
V(C ₁₉ ,H ₁₈)	2.01	2.02	2.02	2.02	2.02	2.04	2.09	2.09	2.23	2.03	1.81
V(C ₁₉ ,H ₁₉)	2.00	1.96	2.06	1.95							
V ₁ (C ₁₉)					1.42	1.43	1.76	1.75	1.73	1.73	1.30
V ₂ (C ₁₉)										0.20	1.09
V(H ₁₉)					0.66	0.54	0.59				
V _a (O)	7.40	7.03	6.68	6.66	6.66	6.53	6.05	6.04	5.99	5.99	5.82
V _x (O)			0.63	0.60	0.65	0.61	1.00				
V(O,H ₁₉)								1.59	1.61	1.61	1.61
Rx (amu ^{1/2} bohr)	-10.787	-0.321	-0.302	-0.284	-0.265	0.000	0.246	0.265	0.631	0.652	8.935
energy (kcal/mol)	0.54	30.13	30.56	30.97	31.39	35.02	34.75	34.69	33.73	33.69	25.03

^aThe anterior and posterior points to the turning points as well as the stationary points are also included.

Table 6. Population of the Most Significant Basins Involved in the Hydrogen Abstraction Reaction in Quartet Spin State, for the ⁴Fe^{IV}Por Electromer^a

basin	SSD-I		SSD-II		SSD-III			SSD-IV		SSD-V	
	reactant	last point	first point	last point	first point	TS	last point	first point	last point	first point	intermediate
V(C ₁₈ ,C ₁₉)	1.85	1.86	1.84	1.86	1.84	1.85	1.82	1.82	1.78	1.77	1.73
V(C ₁₉ ,H ₁₇)	2.01	2.04	2.04	2.04	2.04	2.06	2.07	2.07	2.04	2.02	1.92
V(C ₁₉ ,H ₁₈)	2.01	2.02	2.02	2.02	2.02	2.04	2.07	2.07	2.22	2.05	1.87
V(C ₁₉ ,H ₁₉)	2.00	1.96	2.06	1.95							
V ₁ (C ₁₉)					1.42	1.43	1.53	1.53	1.53	1.53	1.14
V ₂ (C ₁₉)										0.20	0.95
V(H ₁₉)					0.66	0.54	0.60				
V _a (O)	7.40	7.03	6.68	6.66	6.66	6.53	6.26	6.24	6.06	6.06	5.92
V _x (O)			0.63	0.60	0.65	0.61	0.78				
V(O,H ₁₉)								1.39	1.53	1.54	1.60
Rx (amu ^{1/2} bohr)	-10.787	-0.321	-0.302	-0.284	-0.265	0.000	0.189	0.208	1.120	1.174	8.935
energy (kcal/mol)	0.54	30.13	30.56	30.97	31.39	36.33	33.90	33.69	30.26	30.12	23.15

^aThe anterior and posterior points to the turning points as well as the stationary points are also included.

description, the C–H breaking and the subsequent O–H bond formation has been found to take place in five ELF-SSDs, which are connected by their respective turning points. This topological response has been found both in low and high spin configurations. The ELF evolution along the MEP displays in both spin configurations a certain synchronicity: (i) SSD-I

and SSD-II have been localized before the respective transition states; (ii) the TSs are localized on the SSD-III; (iii) SSD-IV and SSD-V have been found after the TS.

In Figure 4, the evolution of valence attractors obtained from the ELF gradient field along the reaction path is schematically displayed. At the beginning of the process the V(C₁₉,H₁₉) basin

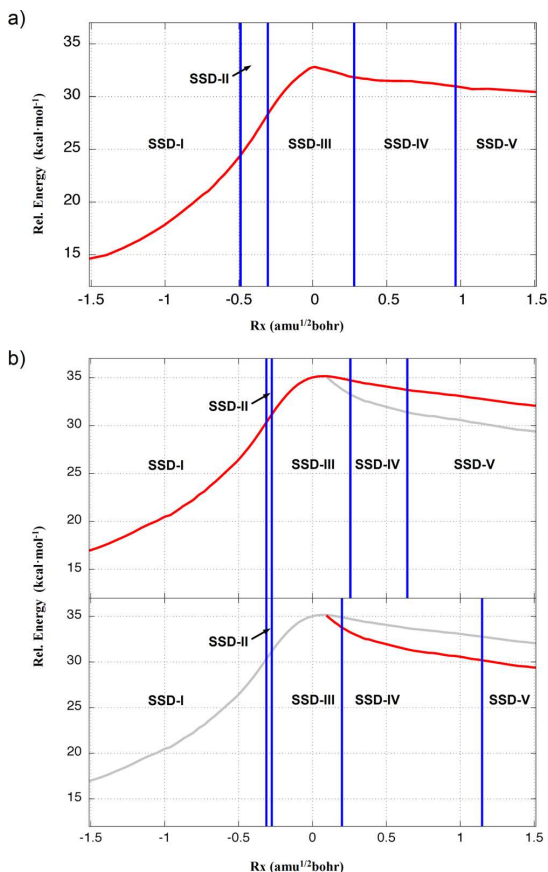


Figure 3. MEP of the hydrogen abstraction reaction in doublet and quartet spin states with marked SSDs obtained from BET analysis. (a) MEP path for the doublet spin. (b) MEP path for both electronic states obtained for the quartet spin. The top image shows the ${}^4\text{Fe}^{\text{III}}\text{Por}^{\bullet+}$ electromer, while the bottom image shows ${}^4\text{Fe}^{\text{IV}}\text{Por}$ electromer.

populations are calculated to be 2.00 e for the doublet and the quartet, while the overall $V_x(\text{O})$ basin populations are predicted to be 7.38 and 7.40, respectively. In this way, the first topological change is predicted to be at $\approx -0.49 \text{ amu}^{1/2} \text{ bohr}$ (doublet) and $\approx -0.30 \text{ amu}^{1/2} \text{ bohr}$ (quartet) connecting the SSD-I and SSD-II, which is associated with the creation of the

$V_x(\text{O})$ monosynaptic basin and a fold-type of catastrophe. The $V_x(\text{O})$ basin population has been predicted to be 0.38 e and 0.63 e for the doublet and the quartet, respectively. It is worth noting that the energetic cost associated with SSD-I and SSD-II not only comprises the energetic cost associated with electronic density rearrangement anticipating the breaking process of the $\text{C}_{19}\text{--H}_{19}$ bond, it also involves the rotation of the methyl group allowing the closest orientation between the H_{19} and O atoms. Later, the second topological change connecting SSD-II and SSD-III has been predicted at $\approx -0.30 \text{ amu}^{1/2} \text{ bohr}$ (doublet) and $\approx -0.26 \text{ amu}^{1/2} \text{ bohr}$ (quartet). This topological change represents thus a cusp type of catastrophe associated with the breaking process of the $\text{C}_{19}\text{--H}_{19}$ bond, where the $V(\text{C}_{19}, \text{H}_{19})$ disynaptic basin splits into two monosynaptic basins, namely, $V(\text{H}_{19})$ and $V(\text{C}_{19})$; see Figure 4. The respective basin populations for these monosynaptic basins are predicted to be 0.71 e and 1.29 e for the doublet and 0.66 e and 1.42 e for the quartet, respectively. Despite this topological change, the nearby $V(\text{C}_{19}, \text{C}_{18})$, $V(\text{C}_{19}, \text{H}_{18})$, and $V(\text{C}_{19}, \text{H}_{17})$ basin populations which belong to the ASD moiety do not seem to be affected by the sudden topological change of the ELF field (see Tables 4–6). However, when the system reaches the transition state, the radical character of the C_{19} and O atoms enhances; their respective basin populations increase up to 1.60 e and 1.43 e for $V(\text{C}_{19})$ and up to 0.87 e and 0.61 e for the $V_x(\text{O})$ in low and high spin configurations, respectively. In contrast the $V(\text{H}_{19})$ basin population decreases up to 0.51 e and 0.54 e, for low and high spin configurations, respectively. Interestingly, from an ELF-topological point of view, when the system reaches the transition state there is no evidence of the formation of the O--H_{19} bond. Subsequently, when the system overcomes the transition state, and then reaches the end of the SSD-III, a considerable increase in the $V_x(\text{O})$ basin population is observed in both spin configurations (around ≈ 1.0 e). Likewise, the $V(\text{H}_{19})$ basin population increases its value up to 0.60 e in both electronic states. The role of the SSD-III is therefore to prepare the system for the imminent formation of the O--H_{19} bond, while the energy increases by $\approx 4.4/3.6 \text{ kcal/mol}$ (doublet and quartet) until the TS is reached.

Afterward, a new turning point connecting SSD-III and SSD-IV is found. For the low spin configuration it is localized at $\approx 0.28 \text{ amu}^{1/2} \text{ bohr}$, whereas for the quartet state the corresponding turning points of each electromer are localized at 0.26 and 0.21 $\text{amu}^{1/2} \text{ bohr}$. The nonbonding monosynaptic attractors $V(\text{H}_{19})$ and $V_x(\text{O})$ are replaced by a single bonding disynaptic attractor $V(\text{O}, \text{H}_{19})$, which corresponds to a cusp type of catastrophe. Thus, from an ELF-topological point of view, this is the first evidence of the formation of the O--H

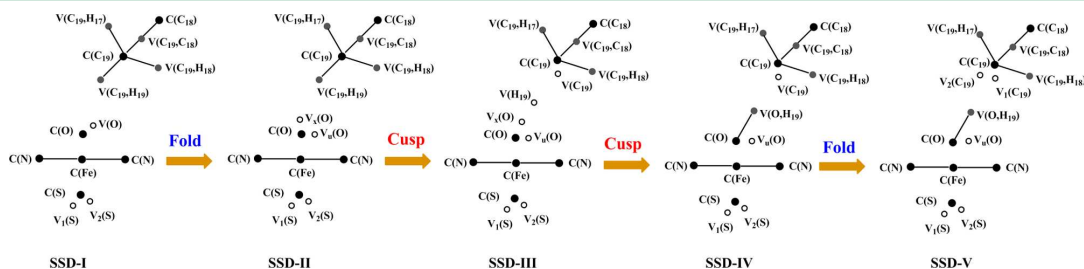


Figure 4. Evolution of valence attractors obtained from ELF gradient field along the reaction path. The respective types of catastrophes are indicated. Black balls correspond to core basins, and gray balls correspond to valence basins.

bond, which in turn takes place after overcoming the transition state. On the first point of the SSD-IV, the $V(\text{O},\text{H}_{19})$ basin populations are predicted to be 1.53 e for the doublet, and 1.59 e and 1.39 e for the respective quartet electromers. Likewise, the $V(\text{C}_{19})$ basin populations are calculated to be 1.69 e for the doublet, and 1.75 e and 1.53 e for the quartet. Subsequently, during the progress of the SSD-IV, the $V(\text{C}_{19})$ monosynaptic basin populations remain practically constant while the populations of the $V(\text{C}_{19},\text{H}_{18})$ disynaptic basin increases. This fact indicates that along the SSD-IV a partial charge density is concentrated on the $V(\text{C}_{19},\text{H}_{18})$ disynaptic basin rather than $V(\text{C}_{19})$. As shown in previous articles,⁴¹ a basin population larger than ≈ 2.00 e for a C–H bond suggests that it can hide a partial charge density of the monosynaptic basin $V(\text{C})$, corresponding to an unpaired electron on the C atom, overlapped by the hydrogenated basin. Therefore, during the progress of the SSD-IV the electronic flux is directed toward the C_{19} atom, increasing the radical character of this atom.

Finally, when the system reaches the last turning point connecting SSD-IV and SSD-V, a fold-type of catastrophe is localized at $\approx 0.97 \text{ amu}^{1/2}$ bohr for the doublet, and 0.65 and $1.17 \text{ amu}^{1/2}$ bohr for the respective quartet electromers. It is worth noting that the SSD-V found in the ${}^4\text{Fe}^{\text{IV}}\text{Por}$ electromer is delayed along the reaction path with respect to the ${}^4\text{Fe}^{\text{III}}\text{Por}^{*+}$ electromer; this fact is consistent with the different electronic structures of both electromers. In this turning point a new $V_2(\text{C}_{19})$ monosynaptic attractor is created due to an excess of the charge density around the C_{19} atom indicating a permanent electronic flow toward this atom from the $V_1(\text{C}_{19})$ monosynaptic basin as well as from the $V(\text{C}_{19},\text{H}_{17})$ and $V(\text{C}_{19},\text{H}_{18})$ disynaptic basins. Thus, when the system reaches the radical intermediate, the basin population for $V_2(\text{C}_{19})$ has increased up to 1.07 e at low spin configuration, whereas for the quartet state these values have been predicted to be 1.09 e and 0.95 e for the respective ${}^4\text{Fe}^{\text{III}}\text{Por}^{*+}$ and ${}^4\text{Fe}^{\text{IV}}\text{Por}$ electromers (see Tables 4, 5 and 6).

On the other hand, the results obtained once the BET analysis was performed in the gas phase (single point energy calculations) show results similar to those in the enzymatic system. In this way, the same number and type of stability domains (SSD-I–SSD-V) are found in the gas-phase model, being characterized by the same type of catastrophes. Furthermore, the populations for the different basins obtained with the ELF analysis in both spin states give rise to virtually the same values for either the gas-phase model or the enzymatic system (see Tables S3 and S4 and Figure S1 of the Supporting Information). The only difference between these models lies in that these catastrophes appear in different values of the reaction coordinate, showing that electronic rearrangement occurs slightly differently in the enzyme and gas phase (see Figures S2 and S3 in the Supporting Information). In particular, it can be shown that the last turning point that connects the SSD-IV and SSD-V through a fold-type catastrophe is found at lower values along the reaction coordinate for the gas-phase model than in the enzymatic system (0.76 and $0.54 \text{ amu}^{1/2}$ bohr for the doublet and the quartet respectively in the gas-phase model). These findings also show that the quartet state in the gas-phase model is more similar to the ${}^4\text{Fe}^{\text{III}}\text{Por}^{*+}$ electromer ($0.65 \text{ amu}^{1/2}$ bohr) than to the ${}^4\text{Fe}^{\text{IV}}\text{Por}$ one ($1.17 \text{ amu}^{1/2}$ bohr). However, the two first turning points that give rise to the stability domains from SSD-I to SSD-III take place in similar values of the reaction coordinate in both the enzymatic and gas-phase models. Since those SSDs are involved in the activation

of the C–H bond, this suggests that the electronic process that takes place during the C–H bond activation is minimally affected by the electrostatic environment of the enzyme.

4. CONCLUSIONS

In summary, from the results of the present work, we can draw the following conclusions:

(i) In this article we have introduced a novel approach to analyze the progress of a biochemical reaction taking into account an inhomogeneous enzymatic environment. We have demonstrated the applicability of BET coupled with QM/MM methods as a useful and powerful tool to study complex enzymatic reactions involving large biological systems. This is the first study where the combination of both methods has been applied to study the hydrogen abstraction process catalyzed by P450. We are able to describe the electronic reorganization, how the electrons flow along the reaction progress, and when and how the bond breaking/forming processes and/or electron pair rearrangements take place.

(ii) Chemical events provided by BET allows us to characterize five different SSDs for the hydrogen abstraction process. Thus, the hydrogen abstraction step is found to be a nonconcerted process, inasmuch as the C–H bond breaking and the O–H bond formation processes do not take place at the same time. As a result, the reaction pathway associated with the hydrogen abstraction process first involves a methyl rotation, then the breaking process of the C–H bond, and finally the formation of the O–H bond, localized after the transition state.

(iii) The presence of electromers in the high spin (quartet) configuration obtained in the enzymatic system (not in the gas-phase model) has been found; however, these electromers do not affect the evolution of the ELF field along the reaction coordinate. Both electromers give rise to the same radical intermediate; nevertheless the SSD-V appears slightly delayed for the ${}^4\text{Fe}^{\text{IV}}\text{Por}$ electromer.

(iv) When comparing the results provided by the BET in the enzyme system and in the gas-phase model, it has been observed that the same domains (SSD-I – SSD-V) are found in both models, which are characterized by the same type of catastrophes. The only difference between these models lies in the different values of the reaction coordinate at which these catastrophes take place. This difference, which is not large, shows that electronic rearrangement occurs slightly differently in the enzyme and in the gas phase, suggesting that the chemical reaction is almost electronically equivalent in both models. Based on these findings, we can conclude that this cytochrome may provide different sources of stabilization than the electrostatic contribution, and may increase the degree of electronic diversity of the heme group, i.e., the presence of electromers.

We believe that the general approach presented here can be further extended to study an even wider range of biological systems. Moving forward, a number of extensions of the present method should be explored, in particular in enzyme catalysis that are currently underway within our group.

■ ASSOCIATED CONTENT

Supporting Information

Energies, spin densities, and populations of the most significant basins obtained from single point energy calculations in gas-phase are reported. Video animations of the different ELF topologies along the reaction coordinate, showing the different

SSDs, are available for both the doublet and quartet states. This material is available free of charge via the Internet at <http://pubs.acs.org>.

AUTHOR INFORMATION

Corresponding Author

*E-mail: andres@uji.es.

Funding

This work was supported by MINECO Projects CTQ2012-36253-C03-01 and CTQ2012-36253-C03-02, by Prometeo/2009/053 and Prometeo/2014/022 (Generalitat Valenciana), and by BANCAIXA Project P1.1B2011-23. I.V. thanks the Spanish Ministerio de Ciencia e Innovación for a doctoral grant (CTQ2009-14541-C02).

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

We acknowledge the Servei d'Informàtica of the Universitat Jaume I.

REFERENCES

- (1) Bader, R. F. W. *Atoms in Molecules: A Quantum Theory*; Oxford University Press: Oxford, 1990; p xviii, 438 pp.
- (2) Becke, A. D.; Edgecombe, K. E. A simple measure of electron localization in atomic and molecular systems. *J. Chem. Phys.* **1990**, *92* (9), 5397–5403.
- (3) Chesnut, D. B. An Electron Localization Function Study of the Lone Pair. *J. Phys. Chem. A* **2000**, *104* (49), 11644–11650.
- (4) (a) Nesper, R. Bonding Patterns in Intermetallic Compounds. *Angew. Chem., Int. Ed. Engl.* **1991**, *30* (7), 789–817. (b) Savin, A.; Nesper, R.; Wengert, S.; Fässler, T. F. ELF: The Electron Localization Function. *Angew. Chem., Int. Ed. Engl.* **1997**, *36* (17), 1808–1832.
- (5) (a) Popelier, P. L. A. Quantum Chemical Topology: on Bonds and Potentials. In *Intermolecular Forces and Clusters*; Wales, D. J., Ed.; Springer-Verlag: Berlin, 2005; Vol. 115, pp 1–56. (b) Popelier, P. L. A.; Aicken, F. M.; O'Brien, S. E. Atoms in molecules. In *Chemical Modelling: Applications and Theory*, Hinchliffe, A., Ed.; The Royal Society of Chemistry: Cambridge, U.K., 2000; Vol. 1, pp 143–198. (c) Popelier, P. L. A.; Smith, P. J. Quantum topological atoms. In *Chemical Modelling: Applications and Theory*, Hinchliffe, A., Ed.; The Royal Society of Chemistry: Cambridge, U.K., 2002; Vol. 2, pp 391–448. (d) Malcolm, N. O. J.; Popelier, P. L. A. The full topology of the Laplacian of the electron density: scrutinising a physical basis for the VSEPR model. *Faraday Discuss.* **2003**, *124*, 353–363.
- (6) (a) Warshel, A.; Levitt, M. Theoretical studies of enzymic reactions: Dielectric, electrostatic and steric stabilization of the carbonium ion in the reaction of lysozyme. *J. Mol. Biol.* **1976**, *103* (2), 227–249. (b) Singh, U. C.; Kollman, P. A. A combined ab initio quantum mechanical and molecular mechanical method for carrying out simulations on complex molecular systems: Applications to the CH₃Cl + Cl⁻ exchange reaction and gas phase protonation of polyethers. *J. Comput. Chem.* **1986**, *7* (6), 718–730. (c) Field, M. J.; Bash, P. A.; Karplus, M. A combined quantum mechanical and molecular mechanical potential for molecular dynamics simulations. *J. Comput. Chem.* **1990**, *11* (6), 700–733. (d) Senn, H. M.; Thiel, W. QM/MM Methods for Biomolecular Systems. *Angew. Chem., Int. Ed.* **2009**, *48* (7), 1198–1229. (e) Lonsdale, R.; Ranaghan, K. E.; Mulholland, A. J. Computational enzymology. *Chem. Commun.* **2010**, *46* (14), 2354–2372. (f) Ferrer, S.; Ruiz-Pernía, J.; Martí, S.; Moliner, V.; Tuñón, I.; Bertrán, J.; Andrés, J. Hybrid Schemes Based on Quantum Mechanics/Molecular Mechanics Simulations: Goals to Success, Problems, and Perspectives. In *Advances in Protein Chemistry and Structural Biology*; Christo, C., Ed.; Academic Press: San Diego, CA, USA, 2011; Vol. 85, pp 81–142. (g) Sousa, S. F.; Fernandes, P. A.; Ramos, M. J. Computational enzymatic catalysis—clarifying

enzymatic mechanisms with the help of computers. *Phys. Chem. Chem. Phys.* **2012**, *14* (36), 12431–12441.

(7) (a) Fang, D.; Chaudret, R.; Piquemal, J.-P.; Cisneros, G. A. Toward a Deeper Understanding of Enzyme Reactions Using the Coupled ELF/NCI Analysis: Application to DNA Repair Enzymes. *J. Chem. Theory Comput.* **2013**, *9* (5), 2156–2160. (b) Chaudret, R.; Piquemal, J.-P.; Cisneros, G. A. Correlation between electron localization and metal ion mutagenicity in DNA synthesis from QM/MM calculations. *Phys. Chem. Chem. Phys.* **2011**, *13* (23), 11239–11247. (c) Fang, D.; Lord, R. L.; Cisneros, G. A. Ab Initio QM/MM Calculations Show an Intersystem Crossing in the Hydrogen Abstraction Step in Dealkylation Catalyzed by AlkB. *J. Phys. Chem. B* **2013**, *117* (21), 6410–6420.

(8) (a) Krokidis, X.; Noury, S.; Silvi, B. Characterization of Elementary Chemical Processes by Catastrophe Theory. *J. Phys. Chem. A* **1997**, *101* (39), 7277–7282. (b) Thom, R. *Structural Stability and Morphogenesis; An Outline of a General Theory of Models*; W. A. Benjamin: Reading, MA, USA, 1975.

(9) (a) Andres, J.; Berski, S.; Domingo, L. R.; Polo, V.; Silvi, B. Describing the Molecular Mechanism of Organic Reactions by Using Topological Analysis of Electronic Localization Function. *Curr. Org. Chem.* **2011**, *15* (20), 3566–3575. (b) Andrés, J.; González-Navarrete, P.; Safont, V. S. Unraveling reaction mechanisms by means of Quantum Chemical Topology Analysis. *Int. J. Quantum Chem.* **2014**, *114* (19), 1239–1252. (c) Polo, V.; Andres, J.; Berski, S.; Domingo, L. R.; Silvi, B. Understanding Reaction Mechanisms in Organic Chemistry from Catastrophe Theory Applied to the Electron Localization Function Topology. *J. Phys. Chem. A* **2008**, *112* (31), 7128–7136. (d) Gillet, N.; Chaudret, R.; Contreras-García, J.; Yang, W.; Silvi, B.; Piquemal, J.-P. Coupling Quantum Interpretative Techniques: Another Look at Chemical Mechanisms in Organic Reactions. *J. Chem. Theory Comput.* **2012**, *8* (11), 3993–3997. (e) Andrés, J.; Gracia, L.; González-Navarrete, P.; Safont, V. S. Chemical structure and reactivity by means of quantum chemical topology analysis. *Comput. Theor. Chem.* **2015**, *1053*, 15–17.

(10) (a) Simpson, E. R.; Mahendroo, M. S.; Means, G. D.; Kilgore, M. W.; Hinshelwood, M. M.; Graham-Lorence, S.; Amarnah, B.; Ito, Y. J.; Fisher, C. R.; Michael, M. D.; Mendelson, C. R.; Bulun, S. E. Aromatase cytochrome P450, the enzyme responsible for estrogen biosynthesis. *Endocr. Rev.* **1994**, *15* (3), 342–355. (b) Ghosh, D.; Griswold, J.; Erman, M.; Pangborn, W. X-ray structure of human aromatase reveals an androgen-specific active site. *J. Steroid Biochem. Mol. Biol.* **2010**, *118* (4–5), 197–202.

(11) (a) Santen, R. J.; Brodie, H.; Simpson, E. R.; Siiteri, P. K.; Brodie, A. History of Aromatase: Saga of an Important Biological Mediator and Therapeutic Target. *Endocr. Rev.* **2009**, *30* (4), 343–375. (b) Bruggemeier, R. W.; Hackett, J. C.; Diaz-Cruz, E. S. Aromatase inhibitors in the treatment of breast cancer. *Endocr. Rev.* **2005**, *26* (3), 331–345.

(12) Hong, Y.; Cho, M.; Yuan, Y. C.; Chen, S. Molecular basis for the interaction of four different classes of substrates and inhibitors with human aromatase. *Biochem. Pharmacol.* **2008**, *75* (5), 1161–1169.

(13) Ortiz de Montellano, P. R. *Cytochrome P450 Structure, Mechanism, and Biochemistry*, 3rd ed.; Kluwer Academic/Plenum Publishers: New York, NY, USA, 2005.

(14) (a) Groves, J. T.; McCluskey, G. A. Aliphatic hydroxylation via oxygen rebound. Oxygen transfer catalyzed by iron. *J. Am. Chem. Soc.* **1976**, *98* (3), 859–861. (b) Groves, J. T. Key elements of the chemistry of cytochrome P450: The oxygen rebound mechanism. *J. Chem. Educ.* **1985**, *62* (11), 928–931. (c) Filatov, M.; Harris, N.; Shaik, S. On the “rebound” mechanism of alkane hydroxylation by cytochrome P450: Electronic structure of the intermediate and the electron transfer character in the rebound step. *Angew. Chem., Int. Ed.* **1999**, *38* (23), 3510–3512.

(15) Viciano, I.; Berski, S.; Martí, S.; Andrés, J. New insight into the electronic structure of iron(IV)-oxo porphyrin compound I. A quantum chemical topological analysis. *J. Comput. Chem.* **2013**, *34* (9), 780–789.

- (16) Shaik, S.; de Visser, S. P.; Ogliaro, F.; Schwarz, H.; Schroder, D. Two-state reactivity mechanisms of hydroxylation and epoxidation by cytochrome P-450 revealed by theory. *Curr. Opin. Chem. Biol.* **2002**, *6* (5), 556–567.
- (17) Schröder, D.; Shaik, S.; Schwarz, H. Two-State Reactivity as a New Concept in Organometallic Chemistry. *Acc. Chem. Res.* **2000**, *33* (3), 139–145.
- (18) Ghosh, D.; Griswold, J.; Erman, M.; Pangborn, W. Structural basis for androgen specificity and oestrogen synthesis in human aromatase. *Nature* **2009**, *457* (7226), 219–223.
- (19) Li, H.; Robertson, A. D.; Jensen, J. H. Very fast empirical prediction and rationalization of protein pKa values. *Proteins: Struct., Funct., Bioinf.* **2005**, *61* (4), 704–721.
- (20) Field, M. J.; Albe, M.; Bret, C.; Proust-De Martin, F.; Thomas, A. The dynamo library for molecular simulations using hybrid quantum mechanical and molecular mechanical potentials. *J. Comput. Chem.* **2000**, *21* (12), 1088–1100.
- (21) Dewar, M. J. S.; Zuebis, E. G.; Healy, E. F.; Stewart, J. J. P. Development and use of quantum mechanical molecular models. 76. AM1: a new general purpose quantum mechanical molecular model. *J. Am. Chem. Soc.* **1985**, *107* (13), 3902–3909.
- (22) Jorgensen, W. L.; Maxwell, D. S.; Tirado-Rives, J. Development and Testing of the OPLS All-Atom Force Field on Conformational Energetics and Properties of Organic Liquids. *J. Am. Chem. Soc.* **1996**, *118* (45), 11225–11236.
- (23) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **1983**, *79* (2), 926–935.
- (24) (a) Marti, S.; Moliner, V.; Tunon, I. Improving the QM/MM description of chemical processes: A dual level strategy to explore the potential energy surface in very large systems. *J. Chem. Theory Comput.* **2005**, *1* (5), 1008–1016. (b) Zhang, Y. K.; Liu, H. Y.; Yang, W. T. Free energy calculation on enzyme reactions with an efficient iterative procedure to determine minimum energy paths on a combined ab initio QM/MM potential energy surface. *J. Chem. Phys.* **2000**, *112* (8), 3483–3492.
- (25) (a) Baker, J. Constrained optimization in delocalized internal coordinates. *J. Comput. Chem.* **1997**, *18* (8), 1079–1095. (b) Baker, J.; Kessi, A.; Delley, B. The generation and use of delocalized internal coordinates in geometry optimization. *J. Chem. Phys.* **1996**, *105* (1), 192–212. (c) Byrd, R. H.; Lu, P.; Nocedal, J.; Zhu, C. A Limited Memory Algorithm for Bound Constrained Optimization. *SIAM J. Sci. Comput.* **1995**, *16* (5), 1190–1208.
- (26) (a) Lee, C.; Yang, W.; Parr, R. G. Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density. *Phys. Rev. B* **1988**, *37* (2), 785–789. (b) Becke, A. D. Density-functional thermochemistry. III. The role of exact exchange. *J. Chem. Phys.* **1993**, *98* (7), 5648–5652. (c) Hay, P. J.; Wadt, W. R. Ab initio effective core potentials for molecular calculations. Potentials for K to Au including the outermost core orbitals. *J. Chem. Phys.* **1985**, *82* (1), 299–310.
- (27) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, J. A., Jr.; Vreven, T.; Kudin, K. N.; Burant, J. C.; Millam, J. M.; Iyengar, S. S.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; Petersson, G. A.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Klene, M.; Li, X.; Knox, J. E.; Hratchian, H. P.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Ayala, P. Y.; Morokuma, K.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Zakrzewski, V. G.; Dapprich, S.; Daniels, A. D.; Strain, M. C.; Farkas, O.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Ortiz, J. V.; Cui, Q.; Baboul, A. G.; Clifford, S.; Cioslowski, J.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Gonzalez, C.; Pople, J. A. *Gaussian 03*, revision D.02; Gaussian, Inc.: Wallingford, CT, USA, 2004.
- (28) Kohout, M. *DGrid*, version 4.6; Radebeul, Germany, 2011.
- (29) Jmol: an open-source Java viewer for chemical structures in 3D. *Jmol*, version 14.2.7; SourceForge, Inc.: Mountain View, CA, USA, 2014. <http://www.jmol.org/>.
- (30) Humphrey, W.; Dalke, A.; Schulten, K. VMD: Visual molecular dynamics. *J. Mol. Graphics* **1996**, *14* (1), 33–38.
- (31) (a) Häussermann, U.; Wengert, S.; Hofmann, P.; Savin, A.; Jepsen, O.; Nesper, R. Localization of Electrons in Intermetallic Phases Containing Aluminum. *Angew. Chem., Int. Ed. Engl.* **1994**, *33* (20), 2069–2073. (b) Silvi, B.; Savin, A. Classification of chemical bonds based on topological analysis of electron localization functions. *Nature* **1994**, *371* (6499), 683–686.
- (32) Silvi, B. The synaptic order: a key concept to understand multicenter bonding. *J. Mol. Struct.* **2002**, *614* (1–3), 3–10.
- (33) (a) Abraham, R.; Marsden, J. E. *Foundations of Mechanics*; American Mathematical Society: New York, NY, USA, 1978. (b) Abraham, R.; Shaw, C. D. *Dynamics—the Geometry of Behavior: Global Behavior*. Aerial Press: Santa Cruz, CA, USA, 1985.
- (34) Silvi, B.; Pilme, J.; Fuster, F.; Alikhani, M. E. What can Tell Topological Approaches on the Bonding in Transition Metal Compounds. In *Metal-Ligand Interactions*; Russo, N.; Salahub, D.; Witko, M., Eds.; Kluwer Academic Publishers: Dordrecht, Netherlands, 2003; Vol. 116, pp 241–284.
- (35) Dietl, N.; Schlangen, M.; Schwarz, H. Thermal Hydrogen-Atom Transfer from Methane: The Role of Radicals and Spin States in Oxo-Cluster Chemistry. *Angew. Chem., Int. Ed.* **2012**, *51* (23), 5544–5555.
- (36) (a) Shaik, S.; Kumar, D.; de Visser, S. P.; Altun, A.; Thiel, W. Theoretical perspective on the structure and mechanism of cytochrome P450 enzymes. *Chem. Rev.* **2005**, *105* (6), 2279–2328. (b) Shaik, S.; Cohen, S.; Wang, Y.; Chen, H.; Kumar, D.; Thiel, W. P450 Enzymes: Their Structure, Reactivity, and Selectivity-Modeled by QM/MM Calculations. *Chem. Rev.* **2010**, *110* (2), 949–1017.
- (37) Lai, W.; Chen, H.; Cohen, S.; Shaik, S. Will P450cam Hydroxylate or Desaturate Alkanes? QM and QM/MM Studies. *J. Phys. Chem. Lett.* **2011**, *2* (17), 2229–2235.
- (38) Shaik, S.; Hirao, H.; Kumar, D. Reactivity of high-valent iron-oxo species in enzymes and synthetic reagents: A tale of many states. *Acc. Chem. Res.* **2007**, *40* (7), 532–542.
- (39) Viciano, I.; Castillo, R.; Marti, S. QM/MM modelling of the hydroxylation of the androstenedione substrate catalyzed by Cytochrome P450 Aromatase (CYP19A1). Submitted for publication in *J. Comput. Chem.*
- (40) Stone, K. L.; Behan, R. K.; Green, M. T. X-ray absorption spectroscopy of chloroperoxidase compound I: Insight into the reactive intermediate of P450 chemistry. *Proc. Natl. Acad. Sci. U. S. A.* **2005**, *102* (46), 16563–16565.
- (41) (a) Melin, J.; Fuentealba, P. Application of the electron localization function to radical systems. *Int. J. Quantum Chem.* **2003**, *92* (4), 381–390. (b) Polo, V.; Andrés, J. A joint study based on the electron localization function and catastrophe theory of the chameleonic and centauric models for the Cope rearrangement of 1,5-hexadiene and its cyano derivatives. *J. Comput. Chem.* **2005**, *26* (14), 1427–1437. (c) Polo, V.; Andrés, J.; Castillo, R.; Berski, S.; Silvi, B. Understanding the Molecular Mechanism of the 1,3-Dipolar Cycloaddition between Fulminic Acid and Acetylene in Terms of the Electron Localization Function and Catastrophe Theory. *Chem.—Eur. J.* **2004**, *10* (20), 5165–5172.

Supporting Information

Joint Use of Bonding Evolution Theory and QM/MM Hybrid Method for Understanding the Hydrogen Abstraction Mechanisms via Cytochrome P450 Aromatase

Ignacio Viciano, Patricio González-Navarrete, Juan Andrés and Sergio Martí*

Departament de Química Física i Analítica, Universitat Jaume I, 12071, Castellón, Spain

*** to whom correspondence should be addressed**

andres@uji.es

Table S1. Relative potential energies (ΔE , kcal/mol) for the hydrogen abstraction process of substrate ASD via Aromatase enzyme calculated at the UB3LYP/6-31G(d) level. The values correspond to the single energy point calculations in gas phase for both doublet and quartet spin state surfaces. The origin of energies has been set at the lowest energy reactant (doublet).

ΔE (kcal/mol)			
Spin	Reactant	TS	Intermediate
doublet	0.0	28.7	24.0
quartet	0.2	29.0	23.5

Table S2. Atomic spin densities for atoms or fragments included in the QM model, obtained from single energy point calculations for both doublet and quartet spin states surfaces.

fragment	spin	R	TS	I
SCH ₃ ⁻	doublet	-0.30	-0.29	-0.24
	quartet	0.25	0.19	0.21
Fe	doublet	1.41	1.22	1.02
	quartet	1.39	1.20	0.99
O	doublet	0.71	0.48	0.13
	quartet	0.71	0.48	0.11
Fe=O	doublet	2.12	1.70	1.15
	quartet	2.09	1.68	1.10
C ₁₉	doublet	0.00	0.51	1.01
	quartet	0.00	0.51	1.02
H ₁₉	doublet	0.00	-0.08	-0.01
	quartet	0.00	-0.08	-0.01
Porphyrin	doublet	-0.83	-0.83	-0.84
	quartet	0.65	0.72	0.74
Substrate*	doublet	0.00	-0.02	-0.06
	quartet	0.00	-0.02	-0.06

* Does not include the C₁₉ and H₁₉ atoms.

Table S3. Population of the most significant basins involved in the hydrogen abstraction reaction in the single energy point calculations from the doublet spin state surface in gas-phase. The anterior and posterior points to the turning points as well as the stationary points are also included.

BASIN	SSD I		SSD II		SSD III			SSD IV		SSD V	
	Reactant	last point	1st point	last point	1st point	TS	last point	1st point	last point	1st point	Product
V(C ₁₈ ,C ₁₉)	1.86	1.85	1.85	1.85	1.85	1.81	1.77	1.77	1.76	1.75	1.65
V(C ₁₉ ,H ₁₇)	2.01	2.03	2.03	2.04	2.04	2.06	2.05	2.04	2.03	1.98	1.82
V(C ₁₉ ,H ₁₈)	2.00	2.02	2.02	2.02	2.03	2.05	2.12	2.13	2.17	2.00	1.79
V(C ₁₉ ,H ₁₉)	2.01	2.06	2.06	2.06							
V ₁ (C ₁₉)					1.32	1.70	1.76	1.76	1.69	1.75	1.42
V ₂ (C ₁₉)										0.24	1.12
V(H ₁₉)					0.75	0.49	0.57				
V _u (O)	7.27	7.16	6.48	6.37	6.37	5.76	5.56	5.56	5.61	5.50	5.40
V _x (O)			0.68	0.76	0.76	1.11	1.03				
V(O,H ₁₉)								1.60	1.56	1.61	1.57
Rx (amu ^{1/2} bohr)	-9.845	-0.489	-0.452	-0.379	-0.360	0.000	0.245	0.262	0.745	0.759	6.947
Energy (kcal/mol)	0.00	18.38	19.01	20.85	21.14	28.75	28.64	28.64	28.43	28.42	24.04

Table S4. Population of the most significant basins involved in the hydrogen abstraction reaction in the single energy point calculations from the quartet spin state surface in gas-phase. The anterior and posterior points to the turning points as well as the stationary points are also included.

BASIN	SSD I		SSD II		SSD III			SSD IV		SSD V	
	Reactant	last point	1st point	last point	1st point	TS	last point	1st point	last point	1st point	Product
V(C ₁₈ ,C ₁₉)	1.85	1.86	1.86	1.86	1.86	1.82	1.78	1.78	1.75	1.75	1.65
V(C ₁₉ ,H ₁₇)	2.01	2.03	2.03	2.04	2.04	2.06	2.06	2.06	1.99	1.99	1.83
V(C ₁₉ ,H ₁₈)	2.01	2.02	2.02	2.02	2.03	2.05	2.10	2.10	2.23	2.00	1.78
V(C ₁₉ ,H ₁₉)	2.00	2.02	2.02	2.02							
V ₁ (C ₁₉)					1.30	1.62	1.72	1.72	1.72	1.72	1.34
V ₂ (C ₁₉)										0.24	1.19
V(H ₁₉)					0.71	0.50	0.58				
V _u (O)	7.17	7.17	6.69	6.70	6.69	5.88	5.61	5.61	5.52	5.51	5.39
V _x (O)			0.46	0.41	0.42	0.98	1.01				
V(O,H ₁₉)								1.59	1.61	1.61	1.59
Rx (amu ^{1/2} bohr)	-10.787	-0.473	-0.454	-0.321	-0.302	0.000	0.208	0.227	0.507	0.540	8.935
Energy (kcal/mol)	0.19	19.28	19.66	22.66	23.14	29.01	29.85	29.84	29.41	29.44	23.47

Figure S1. The MEP of the hydrogen abstraction reaction obtained from single energy point calculations in gas phase for doublet (a) and quartet (b) spin states. The different SSDs obtained from BET analysis are also depicted.

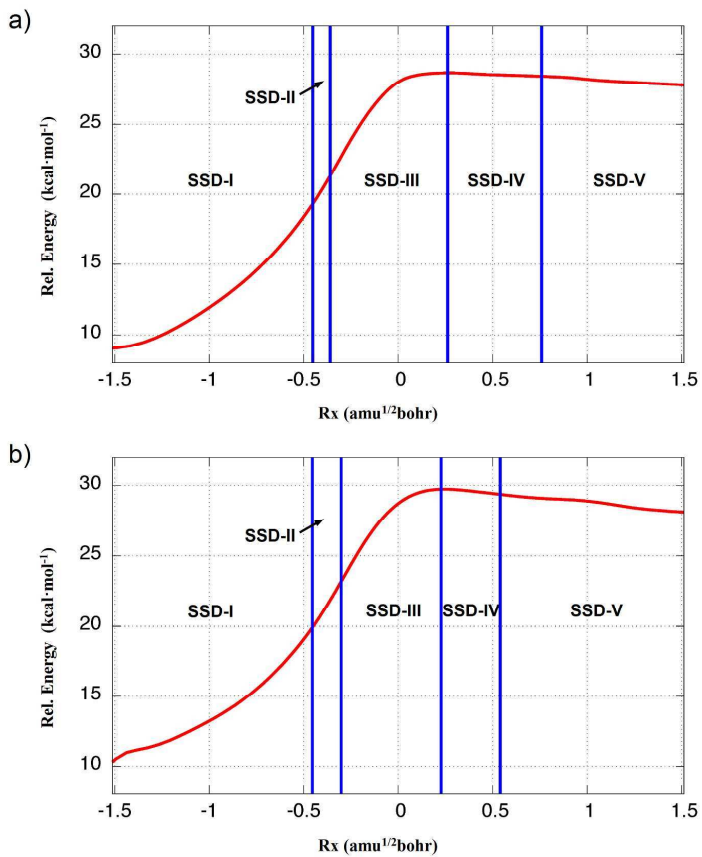


Figure S2. Comparison between the MEP obtained for the doublet spin state catalyzed by the enzyme and the corresponding obtained from single point energy calculations in gas phase. The different SSDs obtained from the BET analysis are also depicted.

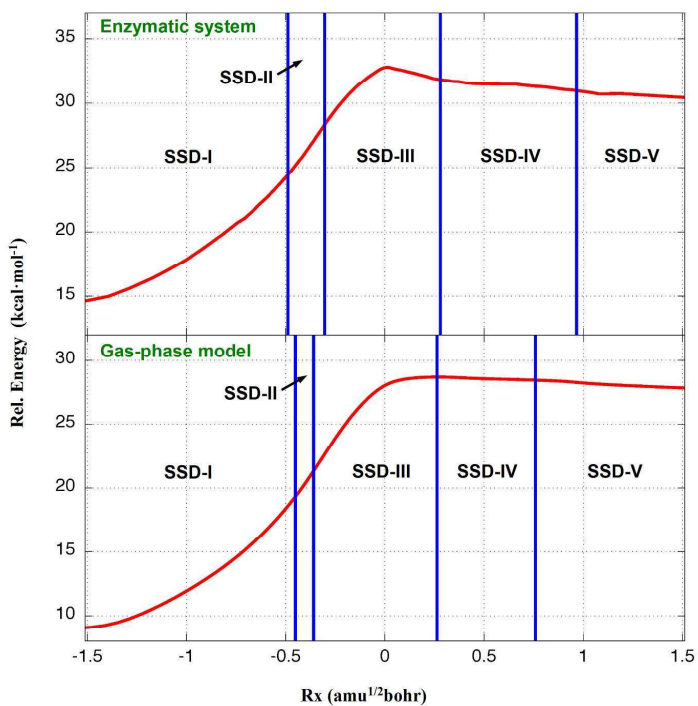
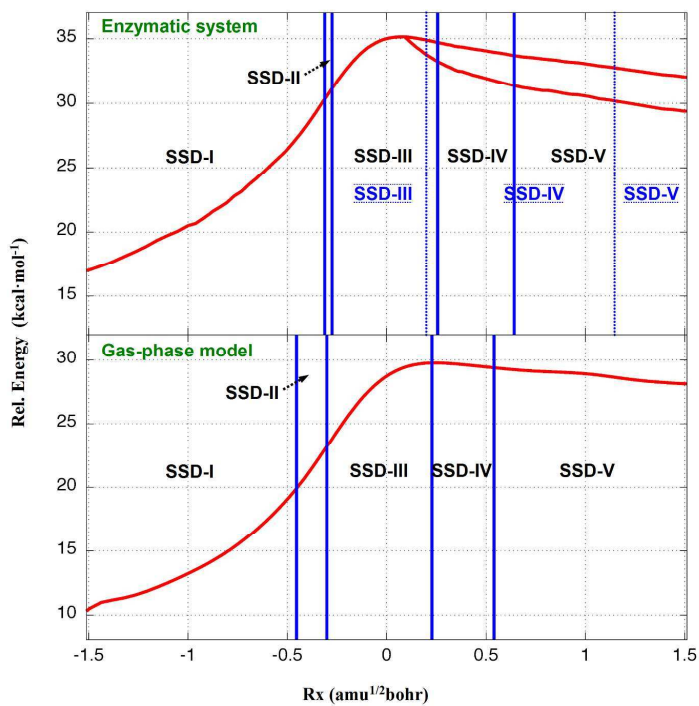


Figure S3. Comparison between the MEP for the doublet spin state catalyzed by the enzyme and the corresponding obtained from single energy point calculations in gas phase. The different SSDs obtained from the BET analysis are also depicted. Blue dashed lines corresponds to SSDs for the ${}^4\text{Fe}^{\text{IV}}\text{Por}$ electromer.



4.3. Study of the Hydroxylation of an Aromatase Inhibitor: Exemestane

4.3.1. Theoretical Study of the Mechanism of Exemestane Hydroxylation Catalyzed by Human Aromatase Enzyme

Ignacio Viciano, and Sergio Martí

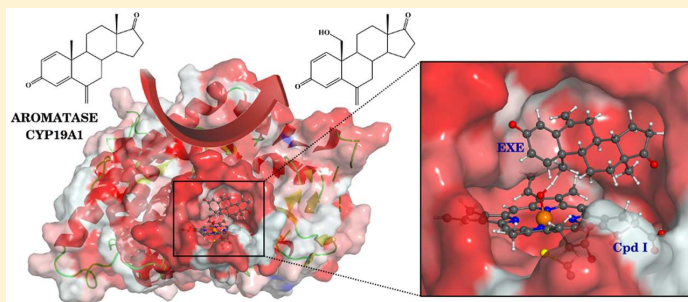
The Journal of Physical Chemistry B **2016**, 120, 3331–3343

Theoretical Study of the Mechanism of Exemestane Hydroxylation Catalyzed by Human Aromatase Enzyme

Ignacio Viciano and Sergio Martí*

Departament de Química Física i Analítica, Universitat Jaume I, 12071 Castelló, Spain

S Supporting Information



ABSTRACT: Human aromatase (CYP19A1) aromatizes the androgens to form estrogens via a three-step oxidative process. The estrogens are necessary in humans, mainly in women, because of the role they play in sexual and reproductive development. However, these also are involved in the development and growth of hormone-dependent breast cancer. Therefore, inhibition of the enzyme aromatase, by means of drugs known as aromatase inhibitors, is the frontline therapy for these types of cancers. Exemestane is a suicidal third-generation inhibitor of aromatase, currently used in breast cancer treatment. In this study, the hydroxylation of exemestane catalyzed by aromatase has been studied by means of hybrid QM/MM methods. The Free Energy Perturbation calculations provided a free energy of activation for the hydrogen abstraction step (rate-limiting step) of 17 kcal/mol. The results reveal that the hydroxylation of exemestane is not the inhibition stage, suggesting a possible competitive mechanism between the inhibitor and the natural substrate androstenedione in the first catalytic subcycle of the enzyme. Furthermore, the analysis of the interaction energy for the substrate and the cofactor in the active site shows that the role of the enzymatic environment during this reaction consists of a transition state stabilization by means of electrostatic effects.

1. INTRODUCTION

The cytochromes P450 (CYP or P450) form a superfamily of heme-enzymes, which are made up of a large number of isoforms, and they are responsible for the oxidative metabolism of a broad variety of substrates, both endogenous and exogenous.^{1,2} These enzymes have the ability to activate dioxygen by means of their prosthetic group, an iron–heme porphyrin, to catalyze the oxidation of unactivated hydrocarbons. In this way, P450 uses molecular oxygen to insert an oxygen atom into the C–H bond of their substrates, usually as a hydroxyl group, while the other oxygen is reduced to a water molecule.^{3–6} This process is known as monooxygenase reaction (that is the reason why P450s are called monooxygenases), and requires both molecular oxygen and NADPH as cosubstrates.

Human aromatase (CYP19A1) is an isoform of this superfamily and is located in different tissues such as gonads, adrenal glands, ovaries, testes, placenta, ovaries, adipose tissue, and in the brain.^{7,8} This enzyme, also known as estrogen biosynthetase, is responsible for the last (and key) step of the biosynthesis of steroid hormones from cholesterol. Specifically, aromatase is involved in the formation of C₁₈-estrogens, estrone

(E1), 17 β -estradiol (E2), and 17 β ,16 α -estradiol (E3), from the C₁₉-androgens, androstenedione (ASD), testosterone (TST), and 16 α -hydroxytestosterone (HTST), respectively.^{9,10} In fact, this is the only known enzyme involved in the biosynthesis of estrogens from androgens in vertebrates.¹¹ This conversion consists of the aromatization of the A ring of the androgens, which occurs through a process of three consecutive oxidations of the angular C₁₉-methyl group of the androgens.^{12–16} In this catalytic process, each oxidation step consumes 1 mol of NADPH, 1 mol of molecular oxygen, and requires the presence of the cytochrome P450 reductase (CPR) as a source of electrons.^{10,17–22} The overall process of aromatization of androgens via the enzyme aromatase has been depicted in Scheme 1.

As can be observed in this scheme, the first and the second oxidation steps occur through two consecutive hydroxylations of the C₁₉-methyl group of the androgen substrates. The first

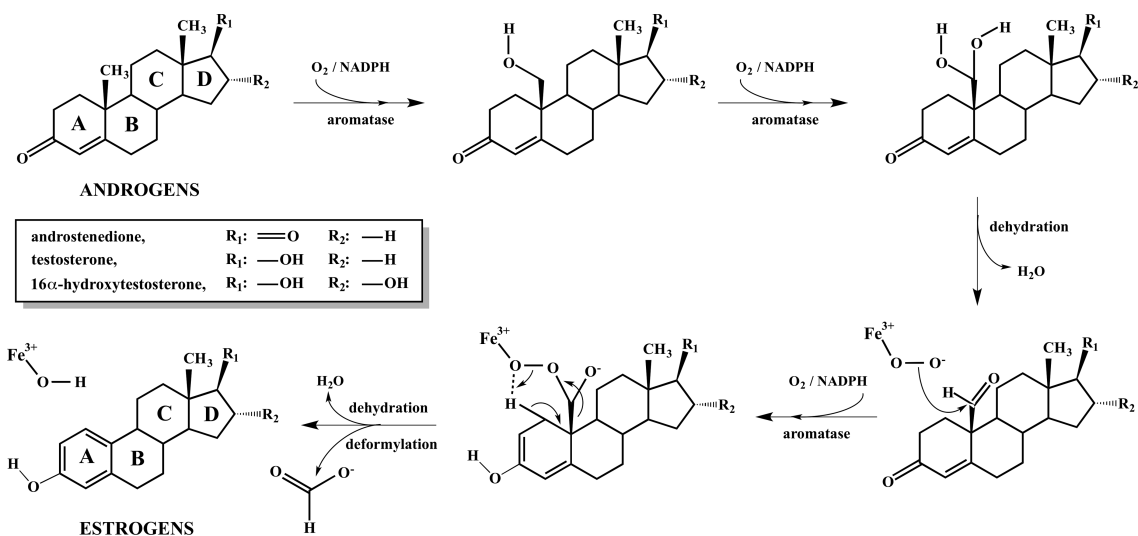
Received: January 29, 2016

Revised: March 11, 2016

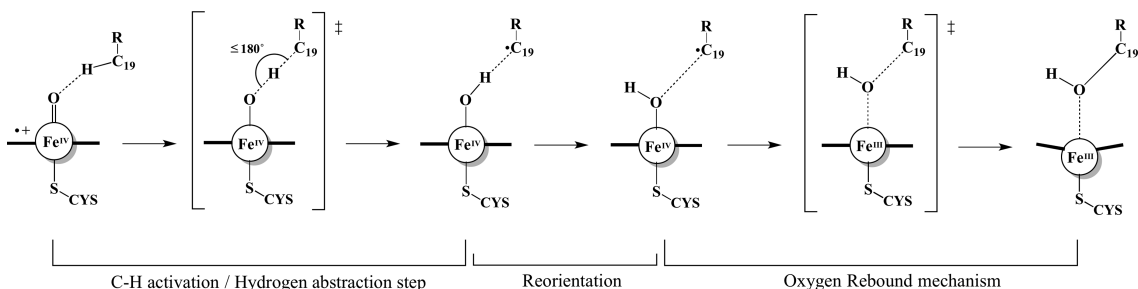
Published: March 14, 2016



Scheme 1. Proposed Catalytic Cycle for the Aromatization of Androgens into Estrogens via the Enzyme Aromatase (CYP19A1)



Scheme 2. Mechanism of Hydroxylation by Cytochrome P450



step produces 19-hydroxy-androgens, while the second one leads to the formation of the 19-gem-diol species; the latter is then dehydrated to the aldehyde intermediate (19-oxo-androgen). Finally, the third step consists of a lyase reaction, in which the C₁₀–C₁₉ bond of the androgens is cleaved, resulting in the aromatization of the phenolic A ring of the androgens, and expelling formic acid and a water molecule. The literature on this mechanism is very extensive because it has been studied by means of both experimental^{12–15,23–31} and theoretical techniques.^{32–39} However, the third oxidation step is still under discussion, and has not yet reached a consensus. It must be noted that the purpose of this Article is out of participating in this debate.

As mentioned above, the first of the three consecutive oxidation steps of estrogens formation consists of the hydroxylation of the androgens to form 19-hydroxy-androgens. As can be seen in Scheme 2, this process proceeds through two distinct stages: Initially, a C–H bond of the C₁₉-methyl group of the androgens is activated by means of the reactive species, an iron(IV) oxo-porphyrin cation radical also known as “Compound I” (Cpd I), which subsequently abstracts the hydrogen atom to the substrate. In this way, an alkyl radical is generated on the C₁₉-methyl group, and the iron–hydroxo porphyrin complex is formed. Later, a reorientation of both the alkyl and the OH groups is produced to facilitate the

subsequent oxygen rebound step in which a new C–O bond is formed. In this step, the alkyl radical is recombined with the iron-bound hydroxyl radical to form the corresponding hydroxylated substrate and the iron(III) porphyrin complex.

Recently, the hydroxylation of the androgen ASD to 19-hydroxy-ASD via the enzyme aromatase was studied by our research group, using hybrid quantum mechanics/molecular mechanics (QM/MM) techniques.⁴⁰ According to our findings, the mechanism discussed above, which is common for all cytochromes P450, is fully compatible with the androgen hydroxylation by aromatase enzyme. In this work, apart from calculating the free energy of the whole hydroxylation mechanism, an analysis of the decomposition of the interaction energy for substrate and cofactor was performed. This analysis revealed that the role of the enzyme aromatase during this catalytic subcycle consists of the TS stabilization achieved through both dispersive and polarization effects.

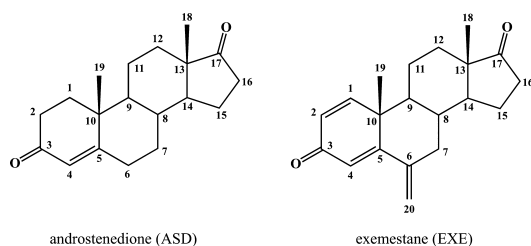
The estrogens are important in humans because of the role they play in sexual and reproductive development. Particularly the estrogens are essential for women, because these are the primary female sex hormones. In premenopausal women, the main source of estrogens is the ovary, which produces mostly E1. However, in postmenopausal women, the source of estrogen production is attributed to the aromatization of adrenal and ovarian androgens (ASD and TST) to estrogens

(E1 and E2) via the enzyme aromatase in peripheral tissues.⁴¹ These steroid hormones, apart from being essential in female sexual development, also play a significant role in the growth and proliferation of hormone-dependent breast cancer. This is because the estrogens bind to the estrogen receptor (ER), which activates the transcription of its target genes, thereby stimulating proliferation of human mammary cells. The complex estrogen-ER interacts with the DNA and may cause mutations due to the resulting increase in cell division, thus encouraging the development of breast cancer cells.^{42,43} Around two-thirds of all cases of breast cancer are related to the type of hormone-dependent breast cancer;^{44–46} however, this ratio increased up to 75% when we refer to postmenopausal breast cancer.⁴⁷ For this reason, blocking the estrogen biosynthesis has proved to be a good strategy to combat estrogen-dependent breast cancers, and has been a prime goal in the field of endocrinology. Therefore, inhibition of the enzyme aromatase, by means of drugs known as aromatase inhibitors (AIs), is the frontline therapy for these types of cancers.^{10,44,48}

The AIs can be classified according to their structures into Type I (steroidal) and Type II (nonsteroidal) inhibitors.^{49,50} The former, such as exemestane (EXE) or formestane, have structure similar to natural substrate ASD, and hence they act as pseudo substrates of aromatase. In such a way, these inhibitors irreversibly bind the enzyme active site and thus are considered as mechanism-based inactivators or suicidal inhibitors. The latter are mostly azole-based compounds, such as anastrozole or letrozole, which are triazole derivatives. These inhibitors compete reversibly with the natural substrate for binding the enzyme, thus interfering with the heme–iron moiety of the cofactor.^{41,43,51,52} Currently, the first-line therapy for advanced hormone dependent breast cancer is based on the third generation of AIs. This family of inhibitors consists of three compounds, anastrozole, letrozole, and exemestane, which are marketed under the brand names ARIMIDEX, FEMARA, and AROMASIN, respectively.^{41,48}

In recent times, the aromatase inhibition has been studied by means of different computational techniques, where several AIs have been tested.^{10,43,51,53–57} In this Article, we have focused in particular on the EXE compound (6-methylene-androsta-1,4-diene-3,17-dione) because of the resemblance it has with the natural substrate ASD (androst-4-ene-3,17-dione). EXE shares the steroidal backbone with ASD; however, the main differences that exist between them are (i) a C₆-substituted methyldiene group located at the B-ring of EXE, and (ii) a double bond between C₁ and C₂ carbons, which is present at the A-ring of the EXE but not in the ASD (see Scheme 3 for comparison).

Scheme 3. Structure of the Androstenedione and Exemestane Substrates



On one hand, the ability to block enzyme activity by the exemestane has been attributed to the presence of this C₁–C₂ double bond.^{41,58} On the other hand, it has been reported that the presence of the C₆-methylidene improves the binding of the EXE in the active site, by means of a “hydrophobic clamp” comprised of the residues Thr-310, Val-370, and Ser-478.¹⁰ In this way, the tight interaction between the C₆-methylidene and the hydrophobic crevice might reduce the mobility of the Thr-310, thus interfering with the mechanism of formation of the Cpd I (in the absence of CPR), with the EXE not being hydroxylated.^{10,11}

Given the structural characteristics of EXE, it seems reasonable to assume that it shares the catalytic cycle with the natural substrate (ASD), at least in its early stages. Therefore, we have performed a theoretical QM/MM study of the hydroxylation of EXE substrate to 19-hydroxy-EXE catalyzed by the enzyme aromatase, with the aim of comparing this mechanism with that which occurs during the first catalytic subcycle of the enzyme. With the present work, we attempt to address the following questions: (i) Is the hydroxylation of EXE compatible with that observed for the ASD, in terms of activation energy? (ii) Does the suicidal inhibition take place in this mechanistic step? (iii) How does the hydrophobic cleft affect the activation energy of the process?

2. COMPUTATIONAL METHODS

System Setup. The X-ray crystal structure of the human placental aromatase cytochrome P450, in complex with its natural substrate ASD (PDB code 3EQM),¹¹ was used as the startup geometry in our simulations. The heme cofactor existing in the original PDB file was altered to obtain the iron–oxo porphyrin Cpd I species, according to the atomic positions suggested in the literature.¹¹ In addition, the substrate ASD accommodated in the active site of aromatase was modeled to EXE, the latter being retained in the same position as the crystallized ASD. To this end, a methylidene group ($\text{C}=\text{CH}_2$) was substituted in the C₆ carbon, and the single bond between positions C₁ and C₂ was modified to a double bond. All of the hydrogen atoms were added using the fDYNAMO library,⁵⁹ according to the pK_a values rendered by the empirical propKa program.⁶⁰ Standard protonation states in solution were found with the exception of the residue Asp-309, which was protonated at pH 7, given that it exhibited a large pK_a displacement (7.7). Histidine amino acids were protonated as follows: His-(62, 105, 111, 325, and 475) were singly protonated at ϵ position, His-(109, 128, 402, 459, and 480) were singly protonated at δ position, and His-171 was doubly protonated. A total of three counterions (Cl^-) were accommodated into optimal electrostatic positions with the aim to electroneutralize the system. The model was placed in an orthorhombic box of water molecules of $90 \times 80 \times 80 \text{ \AA}$, erasing any water molecule whose oxygen atom lies less than 2.8 \AA from any heteroatom. The resulting model consisted of 452 residues of amino acids, the Cpd I, the substrate EXE, as well as 35 crystallographic water molecules, 3 counterions, and 16 553 water molecules of the solvation box.

A QM/MM relaxation of the resulting model was accomplished, using the L-BFGS⁶¹ algorithm. With this purpose, the substrate molecule was described using the AM1⁶² semiempirical Hamiltonian, while the Cpd I, the enzyme, and the counterions were described classically in terms of the OPLS-AA⁶³ force field. All of the water molecules, including the crystallographic ones, were treated using the

TIP3P⁶⁴ water potential. Finally, the equilibration of the resulting structure was performed by means of hybrid molecular dynamics (MD) at 300 K, using the Langevin–Verlet integrator and under the NVT ensemble. The MD was run for 200 ps with a step size of 1 fs. The nonbonding interactions were applied in the calculations by means of periodic boundary conditions with the minimum image convention, using a force-switch function with a cutoff distance in the range 14.5–16 Å.

QM/MM Calculations. To study the reactivity of the enzyme aromatase with the substrate EXE, the potential energy surfaces (PES) were explored by means of QM/MM calculations. The strategy used to localize the TSs during the QM/MM calculations was the micro-/macroiteration method.⁶⁵ In this method, the full system is divided into two different subgroups: the control space, consisting of the atoms participating in the reaction, which are treated by means of quantum methods, and the complementary space that includes the rest of the system. In this way, the optimization of the system is performed as a combination of iterations in both subgroups, which allows taking advantage of efficient optimization algorithms. For each optimization step of the control space (based on the Baker^{66,67} algorithm), the complementary space remains completely relaxed (using the L-BFGS algorithm). Because the complementary space optimization (macroiterations) usually demands a large number of gradient calculations for each optimization step of the control space (microiteration), a dual QM:Charge/MM scheme^{65,68} was adopted. In this procedure, the QM/MM potential energy of the system is evaluated using different expressions, depending on the coordinate space being optimized. In this way, the QM atoms are represented as frozen classical charges throughout the macroiterations, which accelerates the calculations. In addition, due to the complexity of the system, a truncation scheme was carried out during all of the QM/MM simulations. Therefore, a list of interactions (comprising all atoms within 20 Å from both the Cpd I and the substrate EXE) is defined at the beginning of the simulations and is held constant during all of them. In such a way, all of the residues further than this distance remained frozen during the simulations, while all of the residues included in the list were allowed to move.

Once the TSs were obtained and characterized, minimum energy paths (MEP) were mapped out to obtain all of the stationary points involved in the reaction. The Gibbs free energy then was computed on these stationary points, under the rigid-rotor and the harmonic-oscillator (RRHO) approximations, taking into account both thermal corrections and zero point energy (ZPE).

The QM region of the model consisted of 123 atoms distributed among the EXE substrate, the Cpd I, and the axial Cys-437 ligand (comprising the S_γ, the C_β, and two H_β atoms), as can be seen in Figure 1.

The quantum atoms belonging to the QM region were described by means of the LACVP* basis set⁶⁹ (named as B1), where the 6-31G(d) basis is used to describe all of the atoms except the iron, which is represented by the Lan12dz effective core potential (ECP). Additionally, with the aim of calculating more precise energies and frequency numbers, single-point calculations were performed using a larger basis set (B2) on the stationary points previously located with the B1 basis set. The new B2 basis set consists of the combination of the Lan12tz+ ECP for the iron and the 6-311G(d,p) basis for the rest of the

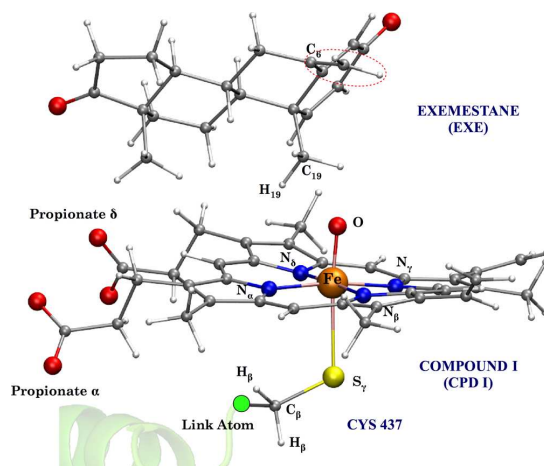


Figure 1. Atoms belonging to the QM model: Cys-437 residue, the substrate EXE, and the cofactor Cpd I. The C₆-methylidene of the exemestane has been highlighted with a red dashed line.

atoms. The level of theory employed during the QM calculations was the unrestricted Kohn–Sham formalism with the B3LYP^{70,71} density functional. The DFT and DFT/MM methodologies have been reported to adequately describe the ground states of both the doublet and the quartet for the Cpd I.^{72,73} In fact, B3LYP has become the widely accepted choice to treat the quantum calculations of the CYP enzymes, because it is well-known that this functional provides reasonable relative spin state energies as well as a number of other properties.^{74,75} As in the system setup, the rest of the enzyme as well as the water molecules (MM region) were described classically by means of the OPLS-AA and TIP3P force fields, respectively. In addition, the link atom formalism was adopted to satisfy the valence of the QM region, due to the existence of a classical bond partitioning in the Cys-437 amino acid (C_α–C_β bond). All of the DFT calculations were carried out using the Gaussian 03⁷⁶ program, while the MD simulations and the MM calculations were performed with the fDYNAMO library.

Free Energy Perturbation Calculations. To compute the free energies of activation of this enzymatic reaction, free energy perturbation (FEP) calculations were performed on the system. Because this method is based on statistical mechanics, the conformational space of the enzyme is taken into account in the calculations, thereby allowing us to obtain more accurate free energy surfaces. With this purpose, a series of windows (geometries with equally spaced energy) were chosen from the formerly mapped out reaction paths. Specifically, 67 windows were selected for the doublet spin state and 76 for the quartet state. During the FEP calculations, the QM region (including the link atom) of each window remained frozen. Every time a QM energy calculation was performed, 100 steps of classical MD were carried out (following the micro/macro scheme), under the same conditions used during the equilibration. This scheme was applied 200 times with a time step of 1 fs, giving rise to a total of 20 ps per window. Finally, to estimate the variation of free energy change, the coordinates of the QM region were shifted between consecutive windows (for the last 100 structures), using the following expression:

$$\Delta F_{i \rightarrow j} = -RT \ln \langle e^{-(H_j - H_i/RT)} \rangle_i \quad (1)$$

This equation is expressed as an average of potential QM/MM energy differences as denoted by the triangular brackets. The term H_i refers to the potential energy of the i th window of MD; meanwhile, H_j corresponds to the potential energy calculated after swapping the coordinates of the atoms of the control space ($j = i + 1$ for forward or $j = i - 1$ for backward). The average was carried out under the same geometry of the classical environment corresponding to the i th window. Once the free energy barrier is computed using this equation, the ZPE correction term is added by means of QM/MM frequency calculations, performed on both the reactant and the TS structures.

3. RESULTS AND DISCUSSION

The results presented in this Article correspond to the hydroxylation of EXE substrate to 19-hydroxy-EXE catalyzed via the enzyme aromatase. This process is divided in two stages: an initial hydrogen abstraction process and a subsequent oxygen rebound step. At first, the reaction was studied at the active site of the enzyme in the doublet spin state, using the B3LYP[B1]:ESP/MM scheme discussed above. Then, the search of the transition states in the quartet state was performed from the optimized structures of the doublet, from which the entire PES was obtained.

Table 1 shows the Gibbs free energies (ΔG) obtained with the B2 basis set for the entire hydroxylation process, as well as

Table 1. Relative Potential Energies (ΔE_{pot}) and Free Energies Obtained by Means of RRHO (ΔG) and Free Energy Perturbation (ΔF) for the Hydroxylation of EXE in Doublet and Quartet Spin States^a

	ΔE_{pot} (B2)	ΔG (B2)	ΔF (B1)	ΔF (B2)
² R	1.5	1.0	0.4 ± 0.0	0.0
⁴ R	0.0	0.0	0.0 ± 0.1	0.2
² TS _{H-abs}	24.1	18.7 (1764i)	17.3 ± 0.1	15.2
⁴ TS _{H-abs}	24.0	19.1 (1878i)	19.2 ± 0.2	17.8
² I	14.7	11.7	14.5 ± 0.2	10.8
⁴ I	13.4	10.5	13.3 ± 0.2	9.7
² TS _{reb}	15.6	13.4 (171i)		
⁴ TS _{reb}	18.4	16.0 (328i)		
² P	-43.2	-41.7		
⁴ P	-47.6	-46.7		

^aThe FEP energy has been calculated only for the hydrogen abstraction step, and the ZPE has been included in both ΔG and ΔF . The basis set used (B1 or B2) is shown in parentheses. The energies are expressed in kcal/mol, and the imaginary frequencies (cm^{-1}) associated with each TS obtained are also reported.

the activation free energies obtained from FEP (ΔF) only for the first hydrogen abstraction step. The FEP calculations were carried out with the B1 basis set and afterward were corrected with the larger basis set B2. However, this correction was only applied to those energy terms that can be extracted from the average. With this purpose, eq 1 was rewritten to reduce the large amount of time-consuming QM(B2) calculations:

$$\Delta F \approx -RT \ln \langle e^{-\Delta E_{\text{int}}^{\text{B1}}/RT} \rangle + E_{\text{vac}}^{\text{B2}} + \Delta \text{ZPE}^{\text{B2}} \quad (2)$$

In eq 2, the interaction energy term ($\Delta E_{\text{int}}^{\text{B1}}$) corresponds to that previously calculated with the B1 basis set, while both the gas-

phase energy ($E_{\text{vac}}^{\text{B2}}$) and the ZPE ($\Delta \text{ZPE}^{\text{B2}}$) terms were estimated with the B2 basis set.

As can be observed in Table 1 and Figure 2, the reactant species show Gibbs free energy values quite similar in both spin states, the quartet being 1.0 kcal/mol lower than the doublet state. The ΔG of activation for the hydrogen abstraction step (the rate-limiting step of the reaction) falls in the range of 18–19 kcal/mol, in this case being lower for the doublet than for the quartet. The intermediates generated in this step correspond with the iron–hydroxo complex and alkyl species and give place to the oxygen rebound process. The ΔG of activation for this step shows values of 1.7 kcal/mol for the doublet and 5.5 kcal/mol for the quartet, leading to the formation of the 19-hydroxy-EXE through a highly exergonic process (−42 and −47 kcal/mol for the doublet and quartet, respectively).

The free energies were also calculated by means of FEP techniques, to include the flexibility of the enzymatic environment. The values calculated with this methodology are reported in Table 1 and are depicted in Figure 3. As can be seen in Table 1, the ΔF energies obtained for the reactant species with the B1 basis set are almost degenerate (<0.5 kcal/mol), being lower in the case of the quartet state. Indeed, when the B2 basis set is considered, this difference is even lower, with both spin states being virtually degenerated. The ΔF of activation for the hydrogen abstraction step obtained with the B1 basis falls in the range of 17–19 kcal/mol, with the barrier of the doublet state being lower than for the quartet. However, these barriers decrease to values in the range of 15–18 kcal/mol once the B2 basis has been considered. Interestingly, when the values calculated from PESs (ΔG) are compared to those calculated through FEP techniques (ΔF), it can be noted that the inclusion of the conformational space of the enzyme does not have a significant impact on the activation barriers. Surprisingly, this trend is in clear contrast with that observed in a previous investigation including the ASD substrate,⁴⁰ where the barriers calculated for the H-abstraction by means of FEP were about 9 kcal/mol smaller on average than those calculated from the PES. In addition, our results indicate that the hydroxylation step is slower for the EXE than for the natural substrate ASD, because the barrier for the rate-limiting step of the reaction is 3.4 kcal/mol higher for the EXE (16.9 kcal/mol) than for the ASD (13.5 kcal/mol). It is worth mentioning that the ΔF values used in the comparison between both studies are those obtained for the doublet state with the B1 basis set. Indeed, this observation is in good agreement with the values derived from experimental results, specifically when compared to the data obtained from kinetic studies for the hydroxylation of both substrates. According to the kinetic parameters obtained experimentally for the EXE (global rate of $0.83 \times 10^{-3} \text{ s}^{-1}$),⁷⁷ the free energy barrier for the entire inhibition process at 310 K is around 22.5 kcal/mol. Moreover, the whole aromatization process of the ASD presents a free energy barrier around 19.8 kcal/mol at the same temperature (experimental global rates of 0.06 s^{-1} and 5.69 min^{-1}).^{27,29} The hydroxylation of ASD is in turn the rate-limiting step of the aromatization, and it is characterized by a free energy barrier of 18.2 kcal/mol (kcat of 1.0 s^{-1}).²⁷ Under the assumption that the hydroxylation of the EXE is also the rate-limiting step of the catalytic cycle, as is the case of the ASD, the free energy of activation for EXE would be around 2.4 kcal/mol higher than for the ASD. Although our reported activation barriers are underestimated (around 5 kcal/mol) when compared to the

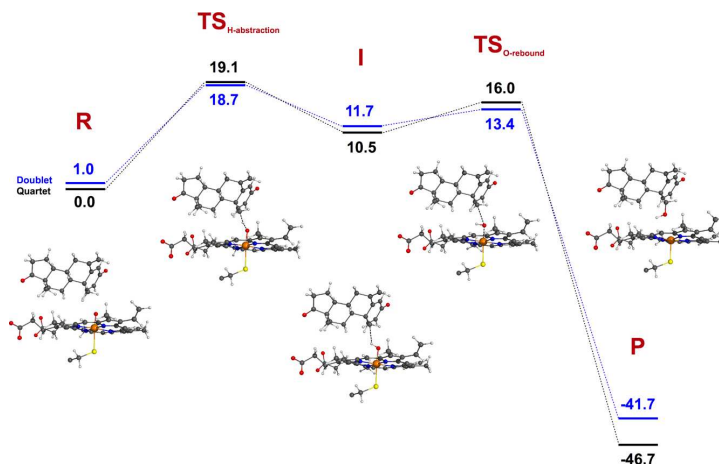


Figure 2. Gibbs free energy profile (kcal/mol) for the hydroxylation of the EXE substrate. The structures corresponding to the QM part for each stationary point have been depicted.

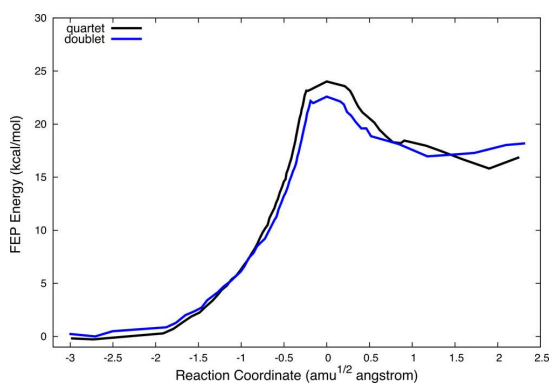


Figure 3. Free energy profile (kcal/mol) of the hydrogen abstraction step from the EXE substrate. This profile has been obtained by means of FEP techniques using the B1 basis set. The ZPE energy has not been included in the profile.

experimental values, the difference in the activation barrier that we have obtained between the EXE and ASD falls in the range experimentally observed.

Furthermore, the hydroxylation of EXE is more exergonic than the hydroxylation of ASD (4.8 kcal/mol for the doublet and 3.3 kcal/mol for the quartet state, respectively). This observation suggests that EXE could act as a competitive inhibitor in the first catalytic subcycle of human aromatase.

The reaction process can be monitored through the values of spin density and the geometrical parameters of the structures obtained. The geometrical data obtained for the hydroxylation of EXE and the corresponding Mulliken atomic spin densities using the B2 basis set are reported in Tables S1 and S2, respectively. Moreover, the evolution of the molecular orbitals occupancy along the hydroxylation process, for doublet and quartet states, has been depicted in Figure 4.

As can be observed in Figure 4 and in Table S2, the reactant species concentrate the entire spin density on the FeO moiety (local triplet) and on the a_{2u} orbital of the porphyrin ring: $\rho(\text{FeO}) = 2.12$, $\rho(\text{Por}+\text{SCH}_3^-) = -1.12$ for doublet and

$\rho(\text{FeO}) = 2.04$, $\rho(\text{Por}+\text{SCH}_3^-) = 0.96$ for quartet. Inasmuch the hydrogen abstraction takes place, the H_{19} atom along with the unpaired electron coming from the homolytic C–H cleavage are transferred from the C_{19} to the oxygen of the Cpd I. During this process, the Fe–O bond is stretched (from ~ 1.6 to ~ 1.8 Å), while the Fe–S distance is slightly contracted (from ~ 2.6 to ~ 2.5 Å). At the same time, the distance between the donor (C_{19}) and the acceptor (O) atoms becomes shorter when going from the reactant species to the transition structures (from 2.9 to 2.5 Å). It is noteworthy that the TS shows the maximum value of the angle $\text{C}_{19}\text{--H}_{19}\text{--O}$ ($\sim 168^\circ$), which is a value close to linearity, as expected for this type of hydrogen transfer mechanisms. However, the O– C_{19} distance again elongates as the formation of intermediate species occurs, thus taking values similar to those of the reactants (~ 2.9 Å). This step leads to the formation of an alkyl radical on this carbon ($\rho(\text{C}_{19}) \approx 1.00$) and of the iron–hydroxo complex ($\rho(\text{O–H}_{19}) < 0.20$) in the intermediate species. Interestingly, the iron–hydroxo complex was found in two different electronic configurations for the quartet spin state as can be seen in Figure 4: ${}^4\text{Fe}^{\text{III}}\text{Por}^{\bullet+}$ ($\rho(\text{Fe}) = 0.90$, $\rho(\text{Por}+\text{SCH}_3^-) = 0.96$) and ${}^4\text{Fe}^{\text{IV}}\text{Por}$ ($\rho(\text{Fe}) = 2.00$, $\rho(\text{Por}+\text{SCH}_3^-) = -0.17$). Both electromers share the same geometry, but differ in the occupancy of their molecular orbitals.^{75,78} However, in the case of the doublet state, only the configuration corresponding to ${}^2\text{Fe}^{\text{III}}\text{Por}^{\bullet+}$ was found ($\rho(\text{Fe}) = 1.00$ and $\rho(\text{Por}+\text{SCH}_3^-) = -1.08$). It is worth pointing out that despite the fact that different electromers were found in the iron–hydroxo complex, all of the configurations present in the intermediate species retain the radical character on the C_{19} atom with a singly occupied orbital Φ_c ($\rho(\text{C}_{19}) = 1.00$ (${}^2\text{Fe}^{\text{III}}\text{Por}^{\bullet+}$), 1.03 (${}^4\text{Fe}^{\text{III}}\text{Por}^{\bullet+}$), and 1.03 (${}^4\text{Fe}^{\text{IV}}\text{Por}$)). The subsequent rebound step is preceded by a reorientation of the hydroxyl group, which is being prepared to form a new $\text{C}_{19}\text{--O}$ bond. This rotation process can be monitored through the change in the $\text{H}_{19}\text{--O--Fe--N}_\alpha$ dihedral angle between the intermediate and the TS of the rebound step (around 55° for doublet and 45° for quartet). The rebinding of the hydroxyl group involves the shortening of the $\text{C}_{19}\text{--O}$ bond (from ~ 3.0 to ~ 2.6 Å in doublet and from ~ 3.0 to ~ 2.5 Å in quartet). This distance, in turn, decreases drastically from the TS to product species (~ 1.4 Å), giving

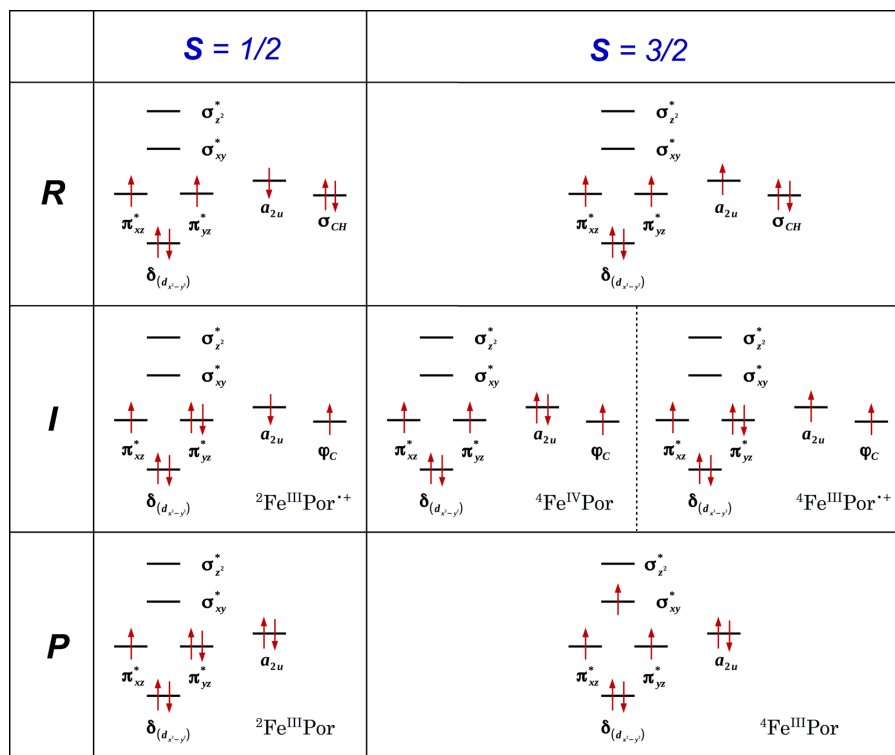


Figure 4. Evolution of the molecular orbitals occupancy along the hydroxylation of the EXE substrate, for doublet ($S = 1/2$) and quartet ($S = 3/2$) spin states.

place to the formation of the new C_{19} -O bond as well as the penta-coordinated iron complex. During the rebound step, the radical character on C_{19} disappears, leading to product species showing a closed-shell porphyrin as well as unpaired electrons on the Fe(III) center: three in the quartet state and one in the doublet state (see Figure 4). All of the spin density found in the products is concentrated on the iron atom ($\rho(\text{Fe}) = 2.69$ and 1.30 for quartet and doublet, respectively) and partly in the porphyrin and the thiolate ligand ($\rho(\text{Por}+\text{SCH}_3^-) = 0.31$ and -0.33 for quartet and doublet, respectively).

The main interactions established among the EXE and the Cpd I with the surrounding amino acids of the active site are presented in Figures 5 and 6, and Table S1. As can be observed in Figure 5, there are several hydrogen bonds established among the propionate side chains attached to the heme group and the following residues: Arg (115, 145, 375, and 435) and Trp-141. These interactions are responsible for the efficient binding of the heme group in the active site, because they neutralize the excess of negative charge present in the heme-propionate side chains. Similarly, the residues Ala-438, Gly-439, and Lys-440 stabilize the sulfur atom of the Cys-437, which is partially charged, by means of the interaction with three polar peptidic hydrogens. It has been reported that this interaction stabilizes the coordination of the thiolate with the heme group, and at the same time regulates the redox potential of the heme group, which prevents the formation of a radical on the sulfur atom.⁷⁹

Regarding the interactions of the EXE substrate, it is noteworthy that it is enclosed in a mainly hydrophobic cleft

formed by the residues: Ile-133 and Phe-134 from B-C loop, Phe-221 and Trp-224 from F-helix, Ile-305, Ala-306, Asp-309, and Thr-310 from I-helix, Val-369, Val-370, Leu-372, and Val-373 from the K- β loop, Met-374 from β 3, and Leu-477 and Ser-478 from the β 8- β 9 loop. EXE in turn forms two hydrogen bonds between its polar moieties (3-keto and 17-keto oxygens) and the residues Asp-309 (which is protonated) and Met-374, respectively. Some of the interactions discussed above can be seen in Figure 5.

When comparing the interactions between the EXE and ASD substrates with the active site of aromatase, we see that the main difference lies in the interactions between the substituted C_6 carbon and its nearest surrounding amino acids. It should be remembered that both substrates differ in the functional group substituted on C_6 , which is located in a quite exposed position, due to its proximity to the access channel to the aromatase active site. The EXE substrate has a methylene group substituted in this position that makes it bulkier than at the same position in the ASD (which is not substituted), and hence it has more surface accessible to fit in the available space. A study based on docking techniques¹⁰ has proposed the existence of a “hydrophobic clamp” in the active site, which comprise the residues Thr-310, Val-370, and Ser-478, and that it would be responsible for anchoring the C_6 -methylene of EXE, thus improving the binding of the inhibitor.

In our simulations, we have observed that the C_6 -methylene interacts with the residues Phe-221, Thr-310, Ser-478, Val-369, Val-370, and a water molecule (WM) (see Figure 6). The distances among the main atoms of the residues

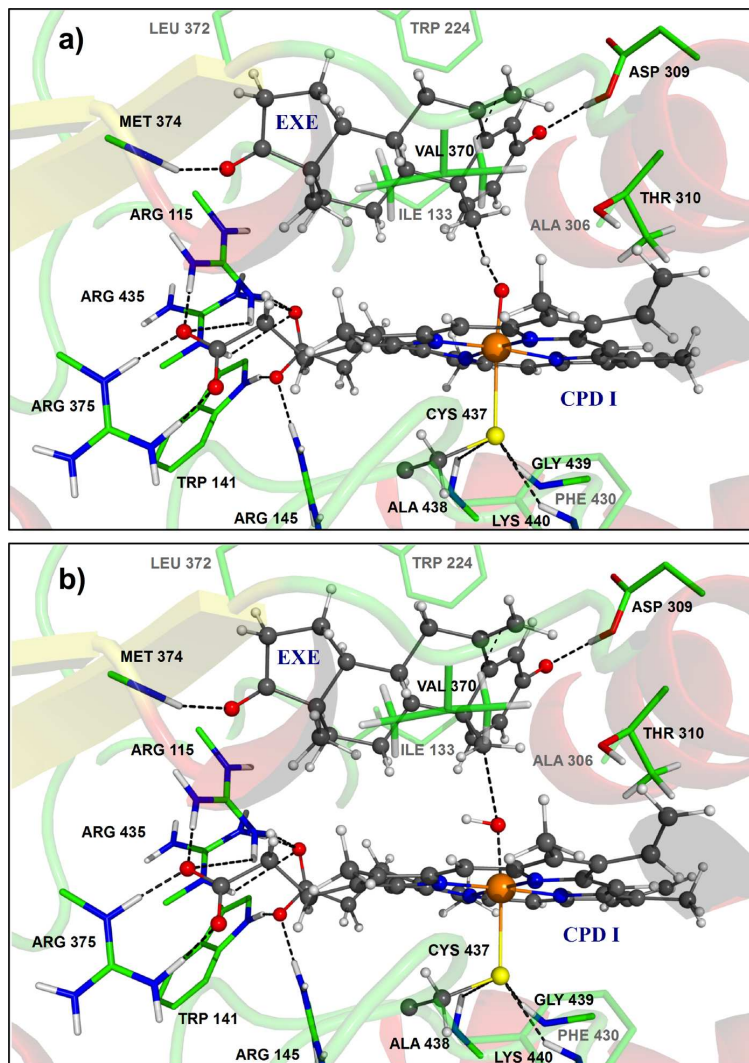


Figure 5. Snapshots corresponding to the TS structures obtained during the hydroxylation reaction of the EXE substrate: (a) for the hydrogen abstraction and (b) for the oxygen rebound step. The most relevant interactions among the substrate and the cofactor with the enzymatic environment have been depicted.

described above and the C_{20} atom of the EXE range from 3.2 to 4.9 Å (see Table S1). Interestingly, we have observed a bridge of polar nature formed by the peptide carbonyl group of Asp-309, the WM, and the Hy of the Ser-478, which is located in the vicinity of the access channel. This polar bridge is aligned with the methylene group and would provide a binding region of electrostatic nature. It should be pointed out that we have observed the presence of a conformer of the Thr-310 residue, in which the $O_{\gamma 1}$ hydroxyl moiety is located between the residue EXE and the heme group. The orientation of this residue also reveals an interaction of the $O_{\gamma 1}$ atom with a C_{19} -methyl hydrogen of the EXE substrate, whose average value is 2.7 Å (see Figure 5). On the basis of the short distance between both atoms (<2.9 Å), one could think that this interaction assists the orientation of the C_{19} atom, thus facilitating the

hydrogen abstraction process. It is worth pointing out that this disposition differs from that found in the original X-ray structure, in which the $C_{\gamma 2}$ is located between the substrate and the heme group. However, it is necessary to mention that during the calculations that we performed with the ASD substrate, the orientation of Thr-310 was retained in the same position as in the original structure from X-ray. To check how stable the Thr-310 rotamer is, classical MD simulations were carried out for two different conformations: (a) the one obtained from the analysis of the PES, and (b) another one in which the Thr-310 was modified with the aim of reproducing the orientation found in the X-ray structure. Each MD was run for 30 ns with the program NAMD⁸⁰ with a step size of 1 fs using the TIP3P water model and the OPLS-AA force field. The results of the MD calculations showed that the orientation

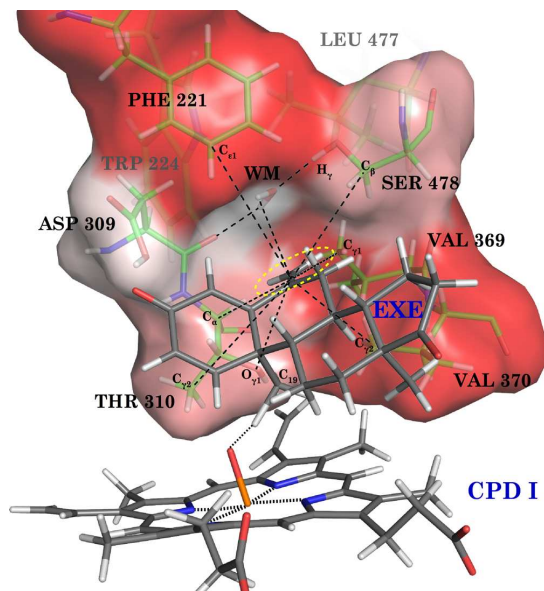


Figure 6. Representation of the main interactions of the C_6 -methylidene group (highlighted in yellow dashed lines) with the surrounding amino acids. Notice that WM means water molecule. The solvent-accessible surface in the vicinity of the methylidene group has been depicted, and it has been colored according to the hydrophobicity (white for the polar interactions and red for hydrophobic interactions).

of the Thr-310 remains unaltered in each of the simulations, which suggests that the two rotamers are representative of the conformational space of the protein. Indeed, the distance between the $O_{\gamma 1}$ atom and the C_{10} -methyl hydrogen of the EXE substrate shows an average value of 3.1 ± 0.4 Å for conformer (a), while this value is 4.9 ± 0.3 Å for conformer (b). On another note, conformer (a) shows an average potential energy lower than that of conformer (b); specifically, this energy is ~ 146 kcal/mol more negative in (a) than in (b) for the last 1000 structures analyzed (see Figure S1 for more detail). Moreover, the activation energy for the hydrogen abstraction step was also calculated (with the B1 basis set) in the case of conformer (b), giving place to a free energy barrier 3.5 kcal/mol higher than for conformer (a). Further details regarding the energies and structures obtained in these calculations can be seen in Table S3 and Figure S2.

Furthermore, the stability of the polar bridge, which was discussed above, was also checked from the MD simulations. On the basis of the analysis of 30 000 structures, we observed that this bridge is only present in around 7% of the time of

simulation. This finding can be partly explained by the fact that the residues involved in the bridge are found in a region quite exposed to solvent, and, therefore, the water molecules are free to come in and go out through the access channel. Indeed, the $H_{\gamma 1}$ atom belonging to the Ser-478 residue establishes hydrogen bonds with the peptide carbonyl group of the residues, Ser-478, Val-369, and Asp-309, and with the N_{ϵ} atom of the residue His-480. In any case, this polar bridge would introduce certain electrostatic character to the interaction between the methylene group and the active site.

Finally, the interaction energy between the enzymatic environment and the substrate/cofactor was calculated for the hydrogen abstraction step (rate-limiting step) in both doublet and quartet spin states. This energy was obtained from the FEP windows corresponding to reactant and TS structures, and thus is expressed as an average value. Moreover, to unravel the source of the catalytic activity of aromatase during the chemical process, the interaction energy was further broken down into its constituent parts. Table 2 shows the results obtained for the different contributions of the interaction energy, which are as follows: (i) the electrostatic term (ΔE_{elec}), which corresponds to the interaction between the polarized wave function obtained for the substrate/cofactor and the MM charges representing the enzymatic environment; (ii) the polarization energy (ΔE_{pol}), consisting of the energy associated with the deformation of the wave function obtained in the gas phase, when it is perturbed by the effect of including the enzyme environment; and (iii) the Lennard-Jones term (ΔE_{LJ}), representing the classical dispersive interaction between the QM and MM atoms. This interaction has been in turn broken down into the contributions coming from the atoms of the substrate (EXE) and from the cofactor (Cys+heme). Additionally, the gas-phase energy (ΔE_{vac}) obtained for each species is also presented in Table 2.

Our results reveal that the electrostatic interaction term is the only responsible for the TS stabilization with respect to the reactant species, being $\Delta E_{\text{elec}} = -14.5$ and -3.2 kcal/mol for the doublet and quartet, respectively. In contrast, it is observed that both the polarization and the Lennard-Jones terms lead to an unfavorable contribution to the activation energy. However, these contributions have different weights depending on the spin state being monitored, in such a way that ΔE_{pol} and ΔE_{LJ} show values of 4.0 and 7.9 kcal/mol in the case of the doublet, which are higher than those obtained in the case of the quartet (0.8 and 0.2 kcal/mol, respectively). In addition, when the ΔE_{LJ} energy is decomposed in terms of substrate (ΔE_{LJ} (EXE)) and cofactor (ΔE_{LJ} (Cys+heme)), a different trend is also noted depending on the spin state. Interestingly, the latter gives rise to a destabilizing nature in the case of the doublet (6.5 kcal/mol), but a slightly stabilizer nature in the case of the quartet state (-0.8 kcal/mol). Nevertheless, despite the fact that the different contributions to the activation energy for the doublet

Table 2. Relative QM/MM Energy Decomposition Terms for the Hydrogen Abstraction (Rate-Limiting Step)^a

	ΔE_{elec}	ΔE_{pol}	ΔE_{LJ}	ΔE_{LJ} (EXE)	ΔE_{LJ} (Cys+heme)	ΔE_{vac}
² R	$[-577.8 \pm 9.6]$	0.0	$[60.8 \pm 1.6]$	0.0	$[346.5 \pm 10.4]$	0.0
² TS	$[-592.3 \pm 10.5]$	-14.5	$[64.8 \pm 2.2]$	4.0	$[338.6 \pm 11.4]$	25.4
⁴ R	$[-576.1 \pm 9.0]$	1.7	$[60.5 \pm 1.6]$	-0.4	$[348.8 \pm 10.4]$	-0.6
⁴ TS	$[-579.3 \pm 11.1]$	-1.5	$[61.2 \pm 2.0]$	0.4	$[348.7 \pm 11.7]$	25.9

^aThe absolute values, along with the corresponding errors, are shown in brackets. The Lennard-Jones term has been in turn broken down into substrate (EXE) and cofactor (Cys+heme), and the gas-phase energies (ΔE_{vac}) have been also included. All of the values are expressed in kcal/mol.

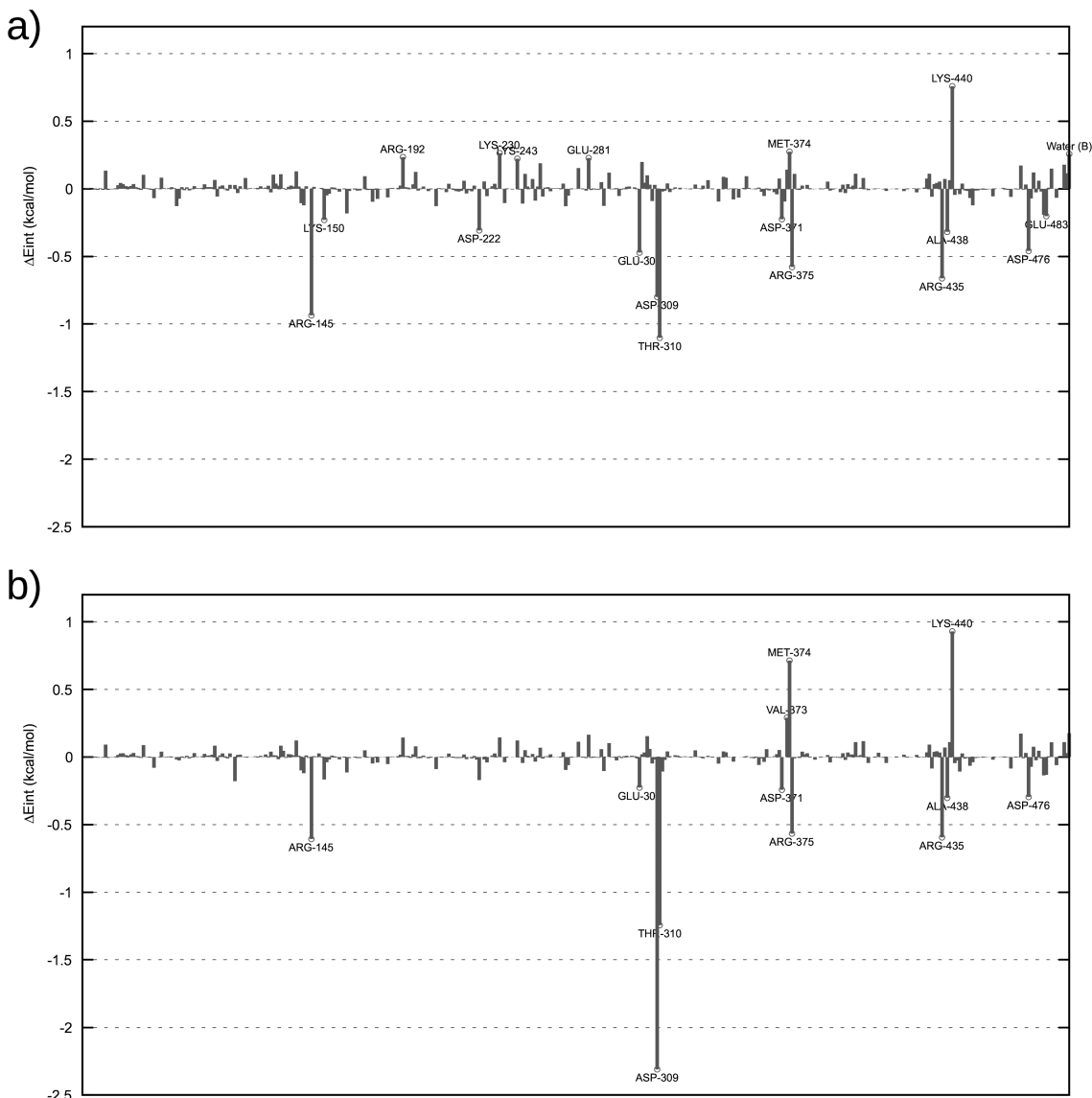


Figure 7. Decomposition “residue by residue” of the activation energy for the electrostatic interaction term (ΔE_{elec}) obtained for the doublet (a) and the quartet (b) spin states. Notice that only those residues contributing with an absolute value greater than 0.2 kcal/mol are labeled in the charts.

state are higher in magnitude than for the quartet, the global interaction energy of the activation process (the sum of all of the different contributions) falls into the same range for both the doublet (−2.6 kcal/mol) and the quartet (−2.3 kcal/mol).

Surprisingly, in a previous study that we performed on the same enzyme but with the natural substrate (ASD),⁴⁰ we showed that the stabilization of the transition state for the hydrogen abstraction step is achieved by means of Lennard-Jones and polarization effects. Because both substrates share the same steroidal backbone, it would be expected a similar behavior in this regard; however, in the present study we observe that the electrostatic term is the only one that facilitates the binding of the transition state in the EXE substrate. To

better clarify the source of this preferential electrostatic stabilization, we performed a decomposition “residue by residue” of the electrostatic interaction term (ΔE_{elec}) involved in the activation energy. The results obtained are plotted in the Figure 7 and reveal that the residue Thr-310 participates in the stabilization of the TS. These findings suggest that the changing nature of the TS stabilization for the hydrogen abstraction step between both substrates is due to (i) the presence of a methylenedioxy group substituted in the C_6 of the EXE substrate and (ii) the different orientation of the residue Thr-310 during the chemical reaction, which forms a hydrogen bond ($O\gamma_1-H_{10}$) with the EXE substrate.

4. CONCLUSIONS

In the present work, the hydroxylation of the EXE substrate to 19-hydroxy-EXE, catalyzed by the enzyme aromatase, has been studied using hybrid QM/MM calculations and MD simulations. The aim of this study has been to compare the hydroxylation mechanism of this aromatase inhibitor with that which occurs during the first catalytic subcycle of the natural substrate (ASD). To this end, the hydrogen abstraction–oxygen rebound mechanism that takes place during hydroxylation of the ASD has been also proposed as the mechanism for the hydroxylation of the EXE. With the results obtained, we have attempted to address different questions:

(i) Is the hydroxylation of EXE compatible with that observed for the ASD, in terms of activation energy?

According to the results derived from the analysis performed on the PES, the rate-limiting step during the hydroxylation reaction of EXE corresponds to the hydrogen abstraction process, as in the case of the hydroxylation of the ASD. In this step, the Cpd I oxidant species abstracts a hydrogen atom from the C₁₉-methyl group of the substrate, leading to the formation of an alkyl radical on the C₁₉ carbon as well as of the iron–hydroxo complex. The activation barriers for this step, obtained through FEP techniques, show that it takes place through an endergonic process with activation barriers in the range from 16.9 kcal/mol (doublet) to 19.2 kcal/mol (quartet). When comparing these results to those obtained for the natural substrate, it is found that the activation barriers obtained for the ASD (13.5 kcal/mol for both the doublet and the quartet) are around 3–5 kcal/mol lower than those for the EXE. According to these findings, the hydroxylation reaction for the inhibitor proceeds through a slower process than for the natural substrate. These results are consistent with the experimental studies, from which it may be estimated an activation barrier for hydroxylation of the EXE 2.4 kcal/mol higher than that for the hydroxylation of the ASD. Therefore, we can conclude that, although this process is slower for the inhibitor than for the natural substrate, the hydroxylation of the EXE proceeds via the hydrogen abstraction–oxygen rebound mechanism as in the case of ASD.

(ii) Does the suicidal inhibition take place in this mechanistic step?

Because the hydroxylation of the EXE substrate is compatible with the mechanism proposed above, and the activation barriers are in the range of values expected, everything suggests that this is not the stage where the inhibition of the aromatase takes place. It has been reported that the EXE is a mechanism-based inactivator or suicidal inhibitor, and thus it should irreversibly bind the active site of the enzyme at some stage. However, this behavior could not be observed throughout the hydroxylation process. Therefore, we dismiss the first catalytic subcycle of aromatase as the stage of the inhibition, and thus it probably will take place in subsequent catalytic subcycles.

Furthermore, the fact that the products from the EXE hydroxylation show lower free energies than those obtained for the ASD suggests a competitive inhibition in the first subcycle of human aromatase.

(iii) How does the hydrophobic cleft affect the activation energy of the process?

The examination of the different terms in which the activation energy can be decomposed suggests that the electrostatic term is the only one involved in the stabilization of the transition state when compared to the reactant species.

This finding obtained for the EXE substrate is not in accordance with that obtained for ASD, where the opposite behavior was observed. In the case of the latter, both the Lennard-Jones term and the polarization of the wave function are responsible for such stabilization. Therefore, the role of the catalytic activity of aromatase during the hydroxylation of the EXE substrate lies in the stabilization of the TS by means of the electric field produced by the enzyme, while the role of this enzyme in the hydroxylation of the ASD consists of a stabilization of the TS by means of both polarization and dispersive effects.

Further analysis has revealed that the difference in the source of the catalytic activity between both substrates lies in two reasons: (i) the presence of a methylene group substituted in the C₆ of the EXE substrate, which mostly interacts with the surrounding amino acids of hydrophobic nature, and partly with a labile polar bridge that provides a binding region of electrostatic character; and (ii) the different orientation of the Thr-310 residue in both substrates. In the case of the EXE, the O_{γ1} of this residue forms a hydrogen bond with a C₁₉-methyl hydrogen, which could set the position of this group throughout the reaction, thereby facilitating the hydrogen abstraction by the Cpd I.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.jpbc.6b01014.

Selection of the main geometrical parameters obtained from PES calculations throughout the hydroxylation reaction; atomic spin densities of different atoms or fragments included in the QM model; and a further analysis of the two different conformers observed for the Thr-310 residue, which comprises the representation of the potential energy obtained from classical MD for each conformer, the decomposition of the Gibbs free energy into its different contributions for each conformer, as well as the snapshots corresponding to the reactant and TS that were found for each rotamer (PDF).

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: smarti@uji.es.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

This work was supported by MINECO Project CTQ2012-36253-C03-01, by Generalitat Valenciana Prometeo/2009/053 and PrometeoII/2014/022, and by Universitat Jaume I Project P1.1B2011-23. We acknowledge the Servei d'Informàtica of the Universitat Jaume I and BSC-Marenostrum (application id: QCM-2013-2-0044) for providing us with computer capabilities. I.V. thanks Spanish Ministerio de Ciencia e Innovación for a doctoral grant (CTQ2009-14541-C02).

■ REFERENCES

- (1) Meunier, B.; de Visser, S. P.; Shaik, S. Mechanism of oxidation reactions catalyzed by cytochrome P450 enzymes. *Chem. Rev.* **2004**, *104* (9), 3947–3980.

- (2) Ortiz de Montellano, P. R. *Cytochrome P450 Structure, Mechanism, and Biochemistry*, 3rd ed.; Kluwer Academic/Plenum Publishers: New York, 2005.
- (3) Hrycay, E. G.; Bandiera, S. M. *Monoxygenase, Peroxidase and Peroxygenase Properties and Mechanisms of Cytochrome*, 1st ed.; Springer International Publishing: New York, 2015.
- (4) de Montellano, P. R. O. Hydrocarbon hydroxylation by cytochrome P450 enzymes. *Chem. Rev.* **2010**, *110* (2), 932–948.
- (5) Hrycay, E. G.; Bandiera, S. M. The monooxygenase, peroxidase, and peroxygenase properties of cytochrome P450. *Arch. Biochem. Biophys.* **2012**, *522* (2), 71–89.
- (6) Sligar, S. G.; Makris, T. M.; Denisov, I. G. Thirty years of microbial P450 monooxygenase research: Peroxo-heme intermediates - The central bus station in heme oxygenase catalysis. *Biochem. Biophys. Res. Commun.* **2005**, *338* (1), 346–354.
- (7) Simpson, E. R.; Clyne, C.; Rubin, G.; Boon, W. C.; Robertson, K.; Britt, K.; Speed, C.; Jones, M. Aromatase - A brief overview. *Annu. Rev. Physiol.* **2002**, *64* (1), 93–127.
- (8) Simpson, E. R.; Davis, S. R. Minireview: Aromatase and the regulation of estrogen biosynthesis - Some new perspectives. *Endocrinology* **2001**, *142* (11), 4589–4594.
- (9) Hong, Y.; Cho, M.; Yuan, Y. C.; Chen, S. Molecular basis for the interaction of four different classes of substrates and inhibitors with human aromatase. *Biochem. Pharmacol.* **2008**, *75* (5), 1161–1169.
- (10) Ghosh, D.; Lo, J.; Morton, D.; Valette, D.; Xi, J.; Griswold, J.; Hubbell, S.; Egbuta, C.; Jiang, W.; An, J.; et al. Novel aromatase inhibitors by structure-guided design. *J. Med. Chem.* **2012**, *55* (19), 8464–8476.
- (11) Ghosh, D.; Griswold, J.; Erman, M.; Pangborn, W. Structural basis for androgen specificity and oestrogen synthesis in human aromatase. *Nature* **2009**, *457* (7226), 219–223.
- (12) Akhtar, M.; Skinner, S. J. The intermediary role of a 19-oxoandrogen in the biosynthesis of oestrogen. *Biochem. J.* **1968**, *109* (2), 318–321.
- (13) Skinner, S. J.; Akhtar, M. The stereospecific removal of a C-19 hydrogen atom in oestrogen biosynthesis. *Biochem. J.* **1969**, *114* (1), 75–81.
- (14) Arigoni, D.; Battaglia, R.; Akhtar, M.; Smith, T. Stereospecificity of oxidation at C-19 in oestrogen biosynthesis. *J. Chem. Soc., Chem. Commun.* **1975**, *0* (6), 185–186.
- (15) Akhtar, M.; Calder, M. R.; Corina, D. L.; Wright, J. N. Mechanistic studies on C-19 demethylation in estrogen biosynthesis. *Biochem. J.* **1982**, *201* (3), 569–580.
- (16) Akhtar, M.; Wright, J. N.; Lee-Robichaud, P. A review of mechanistic studies on aromatase (CYP19) and 17 alpha-hydroxylase-17,20-lyase (CYP17). *J. Steroid Biochem. Mol. Biol.* **2011**, *125* (1–2), 2–12.
- (17) Lu, A. Y. H.; Junk, K. W.; Coon, M. J. Resolution of the cytochrome P-450-containing ω -hydroxylation system of liver microsomes into three components. *J. Biol. Chem.* **1969**, *244* (13), 3714–3721.
- (18) Iyanagi, T. Structure and function of NADPH-cytochrome P450 reductase and nitric oxide synthase reductase domain. *Biochem. Biophys. Res. Commun.* **2005**, *338* (1), 520–528.
- (19) Hong, Y. Y.; Li, H. Z.; Yuan, Y. C.; Chen, S. A. Sequence-function correlation of aromatase and its interaction with reductase. *J. Steroid Biochem. Mol. Biol.* **2010**, *118* (4–5), 203–206.
- (20) Pandey, A. V.; Kempna, P.; Hofer, G.; Mullis, P. E.; Fluck, C. E. Modulation of human CYP19A1 activity by mutant NADPH P450 oxidoreductase. *Mol. Endocrinol.* **2007**, *21* (10), 2579–2595.
- (21) Surawatanawong, P.; Tye, J. W.; Hall, M. B. Density functional theory applied to a difference in pathways taken by the enzymes cytochrome P450 and superoxide reductase: Spin states of ferric hydroperoxy intermediates and hydrogen bonds from water. *Inorg. Chem.* **2010**, *49* (1), 188–198.
- (22) Vatsis, K. P.; Coon, M. J. Oxidative aldehyde deformylation catalyzed by NADPH-cytochrome P450 reductase and the flavoprotein domain of neuronal nitric oxide synthase. *Biochem. Biophys. Res. Commun.* **2005**, *337* (4), 1107–1111.
- (23) Akhtar, M.; Njar, V. C. O.; Neville Wright, J. Mechanistic studies on aromatase and related C-C bond cleaving P-450 enzymes. *J. Steroid Biochem. Mol. Biol.* **1993**, *44* (4–6), 375–387.
- (24) Akhtar, M.; Corina, D.; Miller, S.; Shyadehi, A. Z.; Wright, J. N. Mechanism of the acyl-carbon cleavage and related reactions catalyzed by multifunctional P-450s - studies on cytochrome-P-450(17-alpha). *Biochemistry* **1994**, *33* (14), 4410–4418.
- (25) Grinkova, Y. V.; Denisov, I. G.; Waterman, M. R.; Arase, M.; Kagawa, N.; Sligar, S. G. The ferrous-oxy complex of human aromatase. *Biochem. Biophys. Res. Commun.* **2008**, *372* (2), 379–382.
- (26) Gantt, S. L.; Denisov, I. G.; Grinkova, Y. V.; Sligar, S. G. The critical iron-oxygen intermediate in human aromatase. *Biochem. Biophys. Res. Commun.* **2009**, *387* (1), 169–173.
- (27) Sohl, C. D.; Guengerich, F. P. Kinetic analysis of the three-step steroid aromatase reaction of human cytochrome P450 19A1. *J. Biol. Chem.* **2010**, *285* (23), 17734–17743.
- (28) Guengerich, F. P.; Sohl, C. D.; Chowdhury, G. Multi-step oxidations catalyzed by cytochrome P450 enzymes: Processive vs. distributive kinetics and the issue of carbonyl oxidation in chemical mechanisms. *Arch. Biochem. Biophys.* **2011**, *507* (1), 126–134.
- (29) Khatri, Y.; Luthra, A.; Duggal, R.; Sligar, S. G. Kinetic solvent isotope effect in steady-state turnover by CYP19A1 suggests involvement of Compound I for both hydroxylation and aromatization steps. *FEBS Lett.* **2014**, *588* (17), 3117–3122.
- (30) Mak, P. J.; Luthra, A.; Sligar, S. G.; Kincaid, J. R. Resonance raman spectroscopy of the oxygenated intermediates of human CYP19A1 implicates a compound I intermediate in the final lyase step. *J. Am. Chem. Soc.* **2014**, *136* (13), 4825–4828.
- (31) Yoshimoto, F. K.; Guengerich, F. P. Mechanism of the third oxidative step in the conversion of androgens to estrogens by cytochrome P450 19A1 steroid aromatase. *J. Am. Chem. Soc.* **2014**, *136* (42), 15016–15025.
- (32) Hackett, J. C.; Brueggemeier, R. W.; Hadad, C. M. The final catalytic step of cytochrome P450 aromatase: A density functional theory study. *J. Am. Chem. Soc.* **2005**, *127* (14), 5224–5237.
- (33) Sen, K.; Hackett, J. C. Peroxo-iron mediated deformylation in sterol 14 alpha-demethylase catalysis. *J. Am. Chem. Soc.* **2010**, *132* (30), 10293–10305.
- (34) Lonsdale, R.; Oláh, J.; Mulholland, A. J.; Harvey, J. N. Does compound I vary significantly between isoforms of cytochrome P450? *J. Am. Chem. Soc.* **2011**, *133* (39), 15464–15474.
- (35) Elenewski, J. E.; Hackett, J. C. Cytochrome P450 compound I in the plane-wave pseudopotential framework: GGA electronic and geometric structure of thiolate-ligated iron(IV)-oxo porphyrin. *J. Comput. Chem.* **2013**, *34*, 1647–1660.
- (36) Krámos, B.; Oláh, J. Enolization as an alternative proton delivery pathway in human aromatase (P450 19A1). *J. Phys. Chem. B* **2014**, *118* (2), 390–405.
- (37) Krámos, B.; Oláh, J. The mechanism of human aromatase (CYP 19A1) revisited: DFT and QM/MM calculations support a compound I-mediated pathway for the aromatization process. *Struct. Chem.* **2015**, *26* (1), 279–300.
- (38) Sgrignani, J.; Iannuzzi, M.; Magistrato, A. Role of water in the puzzling mechanism of the final aromatization step promoted by the human aromatase enzyme. Insights from QM/MM MD simulations. *J. Chem. Inf. Model.* **2015**, *55* (10), 2218–2226.
- (39) Xu, K.; Wang, Y.; Hirao, H. Estrogen formation via H-abstraction from the O–H bond of gem-diol by compound I in the reaction of CYP19A1: Mechanistic scenario derived from multiscale QM/MM calculations. *ACS Catal.* **2015**, *5* (7), 4175–4179.
- (40) Viciano, L.; Castillo, R.; Martí, S. QM/MM modeling of the hydroxylation of the androstenedione substrate catalyzed by cytochrome P450 aromatase (CYP19A1). *J. Comput. Chem.* **2015**, *36* (23), 1736–1747.
- (41) Lombardi, P. Exemestane, a new steroidal aromatase inhibitor of clinical relevance. *Biochim. Biophys. Acta, Mol. Basis Dis.* **2002**, *1587* (2–3), 326–337.

- (42) Ciocca, D. R.; Fanelli, M. A. Estrogen receptors and cell proliferation in breast cancer. *Trends Endocrinol. Metab.* **1997**, *8* (8), 313–321.
- (43) Galeazzi, R.; Massaccesi, L. Insight into the binding interactions of CYP450 aromatase inhibitors with their target enzyme: A combined molecular docking and molecular dynamics study. *J. Mol. Model.* **2012**, *18*, 1153–1166.
- (44) Brueggemeier, R. W.; Hackett, J. C.; Diaz-Cruz, E. S. Aromatase inhibitors in the treatment of breast cancer. *Endocr. Rev.* **2005**, *26* (3), 331–345.
- (45) Brueggemeier, R. W.; Richards, J. A.; Joomprabutra, S.; Bhat, A. S.; Whetstone, J. L. Molecular pharmacology of aromatase and its regulation by endogenous and exogenous agents. *J. Steroid Biochem. Mol. Biol.* **2001**, *79* (1–5), 75–84.
- (46) Robinson, A. A review of the use of exemestane in early breast cancer. *Ther. Clin. Risk Manage.* **2009**, *5*, 91–98.
- (47) Chen, S. Aromatase and breast cancer. *Front. Biosci.* **1998**, *3*, d922–d933.
- (48) Santen, R. J.; Brodie, H.; Simpson, E. R.; Siiteri, P. K.; Brodie, A. History of aromatase: Saga of an important biological mediator and therapeutic target. *Endocr. Rev.* **2009**, *30* (4), 343–375.
- (49) Ahmad, I. Shagufita Recent developments in steroidal and nonsteroidal aromatase inhibitors for the chemoprevention of estrogen-dependent breast cancer. *Eur. J. Med. Chem.* **2015**, *102*, 375–386.
- (50) Dowsett, M. Drug and hormone interactions of aromatase inhibitors. *Endocr.-Relat. Cancer* **1999**, *6* (2), 181–5.
- (51) Hong, Y. Y.; Rashid, R.; Chen, S. A. Binding features of steroidal and nonsteroidal inhibitors. *Steroids* **2011**, *76* (8), 802–806.
- (52) Cepa, M.; da Silva, E. J. T.; Correia-da-Silva, G.; Roleira, F. M. F.; Teixeira, N. A. A. Synthesis and biochemical studies of 17-substituted androst-3-enes and 3,4-epoxyandrostanes as aromatase inhibitors. *Steroids* **2008**, *73* (14), 1409–1415.
- (53) Caporuscio, F.; Rastelli, G.; Imbriano, C.; Del Rio, A. Structure-based design of potent aromatase inhibitors by high-throughput docking. *J. Med. Chem.* **2011**, *54* (12), 4006–4017.
- (54) Ghosh, D.; Lo, J.; Egbuta, C. Recent progress in the discovery of next generation inhibitors of aromatase from the structure–function perspective. *J. Med. Chem.* **2015**, DOI: 10.1021/acs.jmedchem.Sb01281.
- (55) Neves, M. A. C.; Dinis, T. C. P.; Colombo, G.; Melo, M. Fast three dimensional pharmacophore virtual screening of new potent non-steroid aromatase inhibitors. *J. Med. Chem.* **2009**, *52* (1), 143–150.
- (56) Roy, P. P.; Roy, K. Docking and 3D-QSAR studies of diverse classes of human aromatase (CYP19) inhibitors. *J. Mol. Model.* **2010**, *16* (10), 1597–1616.
- (57) Sgrignani, J.; Bon, M.; Colombo, G.; Magistrato, A. Computational approaches elucidate the allosteric mechanism of human aromatase inhibition: A novel possible route to small-molecule regulation of CYP450s activities? *J. Chem. Inf. Model.* **2014**, *54* (10), 2856–2868.
- (58) Brueggemeier, R. W. Overview of the pharmacology of the aromatase inactivator exemestane. *Breast Cancer Res. Treat.* **2002**, *74* (2), 177–185.
- (59) Field, M. J.; Albe, M.; Bret, C.; Proust-De Martin, F.; Thomas, A. The dynamo library for molecular simulations using hybrid quantum mechanical and molecular mechanical potentials. *J. Comput. Chem.* **2000**, *21* (12), 1088–1100.
- (60) Li, H.; Robertson, A. D.; Jensen, J. H. Very fast empirical prediction and rationalization of protein pKa values. *Proteins: Struct., Funct., Bioinf.* **2005**, *61* (4), 704–721.
- (61) Byrd, R. H.; Lu, P.; Nocedal, J.; Zhu, C. A limited memory algorithm for bound constrained optimization. *SIAM J. Sci. Comput.* **1995**, *16* (5), 1190–1208.
- (62) Dewar, M. J. S.; Zoebisch, E. G.; Healy, E. F.; Stewart, J. J. P. Development and use of quantum mechanical molecular models. 76. AM1: A new general purpose quantum mechanical molecular model. *J. Am. Chem. Soc.* **1985**, *107* (13), 3902–3909.
- (63) Jorgensen, W. L.; Maxwell, D. S.; Tirado-Rives, J. Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids. *J. Am. Chem. Soc.* **1996**, *118* (45), 11225–11236.
- (64) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **1983**, *79* (2), 926–935.
- (65) Marti, S.; Moliner, V.; Tunon, I. Improving the QM/MM description of chemical processes: A dual level strategy to explore the potential energy surface in very large systems. *J. Chem. Theory Comput.* **2005**, *1* (5), 1008–1016.
- (66) Baker, J. Constrained optimization in delocalized internal coordinates. *J. Comput. Chem.* **1997**, *18* (8), 1079–1095.
- (67) Baker, J.; Kessi, A.; Delley, B. The generation and use of delocalized internal coordinates in geometry optimization. *J. Chem. Phys.* **1996**, *105* (1), 192–212.
- (68) Zhang, Y. K.; Liu, H. Y.; Yang, W. T. Free energy calculation on enzyme reactions with an efficient iterative procedure to determine minimum energy paths on a combined ab initio QM/MM potential energy surface. *J. Chem. Phys.* **2000**, *112* (8), 3483–3492.
- (69) Hay, P. J.; Wadt, W. R. Abinitio effective core potentials for molecular calculations - potentials for K to Au including the outermost core orbitals. *J. Chem. Phys.* **1985**, *82* (1), 299–310.
- (70) Becke, A. D. Density-functional thermochemistry III. The role of exact exchange. *J. Chem. Phys.* **1993**, *98* (7), 5648–5652.
- (71) Miehlich, B.; Savin, A.; Stoll, H.; Preuss, H. Results obtained with the correlation energy density functionals of Becke and Lee, Yang and Parr. *Chem. Phys. Lett.* **1989**, *157* (3), 200–206.
- (72) Altun, A.; Kumar, D.; Neese, F.; Thiel, W. Multireference ab initio quantum mechanics/molecular mechanics study on intermediates in the catalytic cycle of cytochrome P450(cam). *J. Phys. Chem. A* **2008**, *112* (50), 12904–12910.
- (73) Chen, H.; Song, J. S.; Lai, W. Z.; Wu, W.; Shaik, S. Multiple low-lying states for compound I of P450(cam) and chloroperoxidase revealed from multireference ab initio QM/MM calculations. *J. Chem. Theory Comput.* **2010**, *6* (3), 940–953.
- (74) Shaik, S.; Kumar, D.; de Visser, S. P.; Altun, A.; Thiel, W. Theoretical perspective on the structure and mechanism of cytochrome P450 enzymes. *Chem. Rev.* **2005**, *105* (6), 2279–2328.
- (75) Shaik, S.; Cohen, S.; Wang, Y.; Chen, H.; Kumar, D.; Thiel, W. P450 enzymes: Their structure, reactivity, and selectivity-modeled by QM/MM calculations. *Chem. Rev.* **2010**, *110* (2), 949–1017.
- (76) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Rob, M. A.; Cheeseman, J. R.; Montgomery, J. A., Jr.; Vreven, T.; Kudin, K. N.; Burant, J. C.; et al. *Gaussian 03*, revision D.02; Gaussian, Inc.: Wallingford, CT, 2004.
- (77) Giudici, D.; Ornati, G.; Briatico, G.; Buzzetti, F.; Lombardi, P.; di Salle, E. 6-Methylenandrosta-1,4-diene-3,17-dione (FCE 24304): A new irreversible aromatase inhibitor. *J. Steroid Biochem.* **1988**, *30*, 391–394.
- (78) Filatov, M.; Harris, N.; Shaik, S. On the “rebound” mechanism of alkane hydroxylation by cytochrome P450: Electronic structure of the intermediate and the electron transfer character in the rebound step. *Angew. Chem., Int. Ed.* **1999**, *38* (23), 3510–3512.
- (79) Fishelovitch, D.; Hazan, C.; Hirao, H.; Wolfson, H. J.; Nussinov, R.; Shaik, S. QM/MM study of the active species of the human cytochrome P450 3A4, and the influence thereof of the multiple substrate binding. *J. Phys. Chem. B* **2007**, *111* (49), 13822–13832.
- (80) Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Kalé, L.; Schulten, K. Scalable molecular dynamics with NAMD. *J. Comput. Chem.* **2005**, *26* (16), 1781–1802.

Supporting Information

Theoretical Study of the Mechanism of Exemestane Hydroxylation Catalyzed by Human Aromatase Enzyme

*Ignacio Viciano and Sergio Martí**

Departament de Química Física i Analítica, Universitat Jaume I, 12071 Castelló, Spain.

* to whom correspondence should be addressed

smart@uji.es

Table S1: Some selected geometrical data for the stationary points obtained from PES calculations throughout the hydroxylation reaction (see Figure 1 of the paper to clarify the atom labeling). This Table also includes selected geometrical parameters for the interactions established among the thiolate residue and the heme-propionate side chains, with their surrounding amino acids. Distances are expressed in Å and angles in degrees. The numbers in brackets correspond to the quartet state, while those without brackets correspond to the doublet state.

Distance	Hydrogen Abstraction			Oxygen Rebound		
	Reactant	TS (H-Abs)	Intermediate	Intermediate	TS (rebound)	Product
Fe - S	2.599 [2.600]	2.578 [2.586]	2.514 [2.502]	2.492 [2.486]	2.497 [2.559]	2.364 [2.528]
Fe - O	1.625 [1.625]	1.752 [1.753]	1.791 [1.789]	1.792 [1.775]	1.853 [1.865]	2.911 [2.906]
O - C ₁₉	2.852 [2.900]	2.539 [2.547]	2.915 [2.929]	3.017 [3.017]	2.611 [2.505]	1.418 [1.418]
C ₁₉ - H ₁₉	1.091 [1.090]	1.398 [1.401]	2.006 [2.039]	2.264 [2.166]	2.558 [2.271]	1.977 [1.990]
O - H ₁₉	2.385 [2.418]	1.153 [1.159]	0.977 [0.976]	0.975 [0.977]	0.970 [0.974]	0.970 [0.969]
O' - H _{δ₂,Asp-309}	1.864 [1.937]	1.800 [1.786]	1.834 [1.870]	1.740 [1.747]	1.747 [1.746]	1.695 [1.714]
O'' - H _{N,Met-374}	1.780 [1.741]	1.804 [1.805]	1.811 [1.768]	1.768 [1.779]	1.792 [1.792]	1.810 [1.789]
C ₂₀ - C _{ε₁,Phe-221}	4.867 [4.605]	4.936 [4.950]	4.915 [4.796]	4.718 [4.717]	4.863 [4.726]	4.894 [4.807]
C ₂₀ - C _{α,Thr-310}	4.483 [4.612]	4.471 [4.440]	4.456 [4.511]	4.648 [4.647]	4.475 [4.557]	4.405 [4.499]
C ₂₀ - C _{γ₂,Thr-310}	5.757 [5.917]	5.730 [5.714]	5.737 [5.808]	5.983 [5.984]	5.778 [5.868]	5.699 [5.783]
C ₂₀ - O _{γ₁,Thr-310}	4.057 [4.278]	4.010 [3.984]	4.018 [4.116]	4.279 [4.280]	4.061 [4.186]	3.974 [4.091]
C ₂₀ - C _{γ₁,Val-369}	4.668 [4.805]	4.653 [4.621]	4.616 [4.680]	4.668 [4.664]	4.622 [4.705]	4.681 [4.705]
C ₂₀ - C _{γ₂,Val-370}	3.761 [4.015]	3.696 [3.685]	3.711 [3.827]	3.869 [3.870]	3.757 [3.887]	3.755 [3.824]
C ₂₀ - C _{β,Ser-478}	4.215 [4.149]	4.239 [4.236]	4.204 [4.175]	4.019 [4.014]	4.170 [4.134]	4.282 [4.213]
C ₂₀ - H _{2,WM}	3.228 [3.079]	3.278 [3.229]	3.231 [3.143]	3.078 [3.074]	3.144 [3.093]	2.932 [3.154]

H ₁₉ - O _{γ1,Thr-310}	2.877 [2.821]	2.728 [2.729]	2.723 [2.674]	2.769 [2.684]	2.733 [2.694]	2.762 [2.740]
S-Fe-O	174.5 [176.0]	171.8 [173.6]	172.1 [172.9]	174.9 [177.1]	176.7 [174.3]	162.5 [160.9]
C ₁₉ -H ₁₉ -O	104.0 [105.1]	168.4 [168.4]	153.9 [150.6]	133.2 [144.8]	82.2 [92.2]	42.3 [41.5]
H ₁₉ -O-Fe-N _α	16.9 [20.7]	28.1 [28.2]	32.8 [28.2]	5.2 [35.8]	-49.5 [-9.5]	-21.6 [-3.7]
Prop-O _{2δ} - H _{H21,Arg-115}	1.641 [1.646]	1.641 [1.647]	1.642 [1.645]	1.654 [1.653]	1.650 [1.646]	1.643 [1.648]
Prop-O _{2δ} - H _{H22,Arg-115}	3.185 [3.181]	3.180 [3.179]	3.186 [3.182]	3.185 [3.185]	3.182 [3.179]	3.185 [3.181]
Prop-O _{2δ} - H _{E1,Trp-141}	3.289 [3.282]	3.275 [3.279]	3.295 [3.291]	3.288 [3.285]	3.282 [3.276]	3.283 [3.281]
Prop-O _{2δ} - H _{E,Arg-435}	2.572 [2.560]	2.554 [2.557]	2.579 [2.570]	2.581 [2.576]	2.570 [2.559]	2.565 [2.567]
Prop-O _{2δ} - H _{H21,Arg-435}	3.239 [3.241]	3.236 [3.247]	3.243 [3.244]	3.263 [3.258]	3.254 [3.243]	3.236 [3.249]
Prop-O _{2δ} - H _{H22,Arg-435}	1.584 [1.584]	1.577 [1.588]	1.589 [1.588]	1.607 [1.602]	1.597 [1.584]	1.579 [1.591]

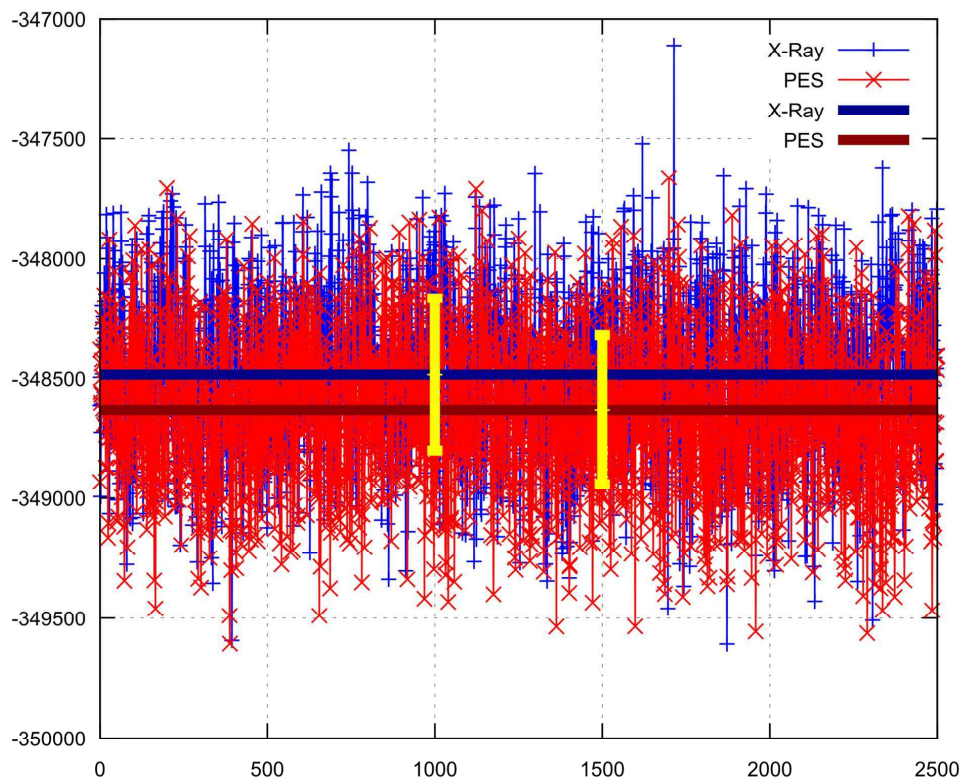
Table S2. Atomic Spin Densities of different atoms or fragments included in the QM model. These values have been obtained with the B2 basis set for the doublet and the quartet (in brackets) spin states.

Fragment	Hydrogen Abstraction			Oxygen Rebound		
	Reactant	TS (H-Abs)	Intermediate	Intermediate	TS (rebound)	Product
SCH ₃ ⁻	-0.235 [0.207]	-0.252 [0.202]	-0.184 [0.144]	-0.136 [-0.068]	-0.182 [0.024]	-0.158 [0.309]
Fe	1.271 [1.169]	1.048 [0.925]	1.009 [0.905]	0.982 [2.001]	1.117 [2.330]	1.303 [2.689]
O	0.850 [0.870]	0.506 [0.570]	0.140 [0.180]	0.125 [0.192]	-0.065 [-0.037]	0.000 [0.004]
Porphyrin* (Por)	-0,889 [0.751]	-0,803 [0.767]	-0,892 [0.820]	-0,936 [-0.100]	-0,571 [-0.124]	-0,146 [-0.001]
C ₁₉	0.001 [0.001]	0.555 [0.595]	1.007 [1.029]	1.027 [1.030]	0.742 [0.831]	0.000 [0.001]
H ₁₉	0.000 [-0.001]	-0.039 [-0.043]	-0.011 [-0.009]	0.002 [0.014]	0.005 [0.021]	0.000 [-0.004]
Substrate**	0.001 [0.002]	-0.015 [-0.016]	-0.068 [-0.068]	-0.065 [-0.069]	-0.046 [-0.045]	-0.000 [0.001]

* The Porphyrin fragment includes the propionate side chains.

** The Substrate does not include the C₁₉ and H₁₉ atoms

Figure S1: Representation of the potential energy (kcal/mol) obtained from classical molecular dynamics (only the last nanosecond) carried out for the two possible conformers of the residue Thr-310. The rotamer which establishes an interaction with the C₁₉-methyl hydrogen (O γ 1-H) is labelled as "PES" and it is depicted in red. The rotamer with the C γ 2 methyl group oriented towards the methylene group of the EXE substrate is labelled as "X-Ray" and it is depicted in blue. The corresponding average values of the potential energy for each conformer are depicted in dark red and dark blue, respectively. The RMS values for both averages are also depicted in yellow color.



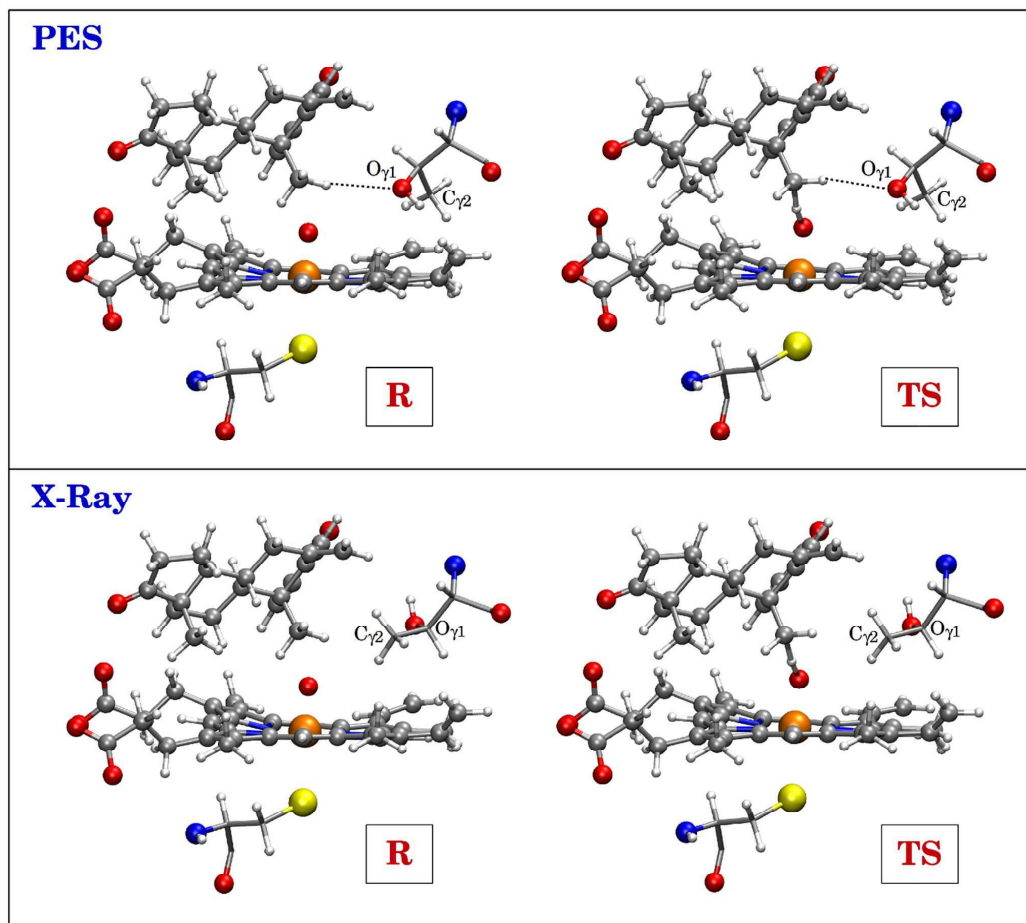
$$\langle E_{Pot} \rangle_{PES} \sim -348629 \pm 312 \text{ kcal/mol}$$

$$\langle E_{Pot} \rangle_{X-Ray} \sim -348483 \pm 318 \text{ kcal/mol}$$

Table S3: Decomposition of the Gibbs free energy into its different contributions for the two conformers of the Thr-310 residue. The rotamer which establishes an interaction with the C₁₉-methyl hydrogen (O γ 1-H) is labelled as "PES". The rotamer with the C γ 2 methyl group oriented towards the methylenide group of the EXE substrate is labeled as "X-Ray". The energies (kcal/mol) have been obtained for the doublet spin state using the B1 basis set.

	² R (PES)	² Ts (PES)	² R (X-Ray)	² Ts (X-Ray)
QM Electrostatic	-1888816.6	-1888793.0	-1888819.2	-1888791.5
MM Bond	619.2	619.7	620.2	620.2
MM Angle	1060.0	1060.1	1060.7	1060.6
MM Dihedral	938.8	939.0	951.4	951.4
MM Improper	31.7	31.7	31.7	31.8
MM Electrostatic	-32980.1	-32980.4	-32999.1	-32999.7
QM+MM Lennard-Jones	-587.7	-587.0	-580.2	-579.4
QM/MM Potential Energy	-1919734.7	-1919709.9	-1919734.5	-1919706.7
Gibbs (RRHO) + ZPE	586.7	581.0	587.0	581.3
ΔG	0.0	19.1	0.5	22.7

Figure S2: Snapshots for the Reactant and TS resulting from the exploration of the two conformers of the Thr-310 residue. The rotamer which establishes an interaction with the C₁₉-methyl hydrogen (O_{γ1}-H) is labelled as "PES". The rotamer with the C_{γ2} methyl group oriented towards the methylene group of the EXE substrate is labelled as "X-Ray". All the geometries have been obtained for the doublet spin state.



Chapter 5

Discussion



The results of this thesis, which were presented in the previous chapter, are divided into three main sections: (1) Study of the reactive species of the enzyme aromatase: Compound I; (2) First catalytic subcycle of the enzyme aromatase; and (3) Study of the hydroxylation of an aromatase inhibitor: Exemestane. The first section comprises a single subsection and investigates the electronic structure of the cytochrome P450 reactant species, Cpd I. The second section, which includes two subsections, investigates the process of hydroxylation of the natural substrate androstenedione (ASD) taking place during the first catalytic subcycle of the enzyme aromatase. The third section is composed of a single subsection and investigates the hydroxylation of the suicidal inhibitor exemestane (EXE) catalyzed by the enzyme aromatase. Each subsection corresponds to a scientific article published in an international peer-reviewed journal. This section presents a brief summary of each subsection.

5.1. Study of the Reactive Species of the Enzyme Aromatase: Compound I

5.1.1. New Insight into the Electronic Structure of Iron(IV)-oxo Porphyrin Compound I. A Quantum Chemical Topological Analysis

The electronic structure of iron-oxo porphyrin π -cation radical complex, known as Compound I (Cpd I), $Por^{\bullet+}Fe^{IV} = O(S-H)$, has been studied for doublet and quartet electronic states by means of two methods of quantum chemical topology analysis: the electron localization function (ELF) $\eta(\vec{r})$, and electron density $\rho(\vec{r})$. According to the ELF approach, the electronic structure of Cpd I, in both doublet and quartet electronic states, is characterized by 76 local maxima (attractors) localized for the ELF field $\eta(\vec{r})$. The core electron density is characterized by a total of 27 core attractors: $C(Fe)$, $C(S)$, $C(O)$, $C(N)_{i=1-4}$, and $C(C)_{i=1-20}$ which coincide with positions of atomic nuclei. Conversely, in the valence shell, the electrons are associated with three monosynaptic non-bonding basins (attractors): $V(S)$ and $V_{i=1,2}(O)$, corresponding to the lone pairs of sulfur and oxygen atoms, respectively. Furthermore, a total of 33 valence disynaptic bonding attractors can be observed: $V(C, C)_{i=1-20}$, $V(C, N)_{i=1-8}$, $V(Fe, S)$, and $V(Fe, N)_{i=1-4}$, corresponding to the $C-N_{i=1-8}$, $Fe-S$, $Fe-N_{i=1-4}$ chemical bonds, respectively. The $C-H$ and $S-H$ bonds are reflected by 12 protonated disynaptic attractors $V(H, C)_{i=1-12}$ and one $V(H, S)$ attractor. It is worth noting that the same number and type of attractors are localized for the Cpd I system in both the doublet and quartet electronic states. The only difference between the two states is the slightly different spatial orientation of the $V_{i=1,2}(O)$ attractors with respect to the $Fe-N$ bonds and the porphyrin plane, which can be associated with the occupation of different orbitals in both electronic states. The core attractors $C(Fe)$, $C(N)_{i=1-4}$, $C(C)_{i=1-20}$ and the valence attractors $V(H, C)_{i=1-12}$, $V(C, C)_{i=1-20}$, $V(C, N)_{i=1-8}$, $V(Fe, N)_{i=1-4}$ reflect the electronic structure of the iron-porphyrin complex: the planar geometry of the Cpd I porphyrin ring is associated with the aromatic nature originated by the conjugation of the double $C=C$ and $C=N$ bonds. However, the lack of the $V_{i=1,2}(C, C)$ or $V_{i=1,2}(C, N)$ attractors, which is the topological signature for the localized double $C=C$ and $C=N$ bonds, above and below the macrocycle ring plane denotes a large delocalization of the electron density. In addition, since the porphyrin molecule, Por , is composed of two isolated pyrrole subunits, C_4H_4NH , and two pyrrole anions, $[C_4H_4N]^-$, the ELF topology of Por , C_4H_4NH , and $[C_4H_4N]^-$ was compared with the porphyrin ring present in Cpd I. The results show that, considering a number of attractors and their synapcticity, the pyrrole anion is the most similar to that found in Cpd I. This fact supports

the observation that the $V(Fe, N)_{i=1-4}$ basin consists mainly of electrons from the lone pair of the nitrogen atom. However, the distance between the $V(Fe, N)_{i=1-4}$ and $C(N)_{i=1-4}$ attractors in Cpd I are shorter than between the $V(N)$ and $C(N)$ attractors in the pyrrole anion. This fact may be explained as a result of a Pauli repulsion between electrons “contained” in the $C(Fe)$ and $V(Fe, N)_{i=1-4}$ basins and the corresponding “compression” of the electron density in this region. The presence of these four $V(Fe, N)_{i=1-4}$ basins localized between Fe and four N atoms may be interpreted as covalent bonds. However, the topographical analysis of the $\eta(\vec{r})$ function performed on the molecular plane of Cpd I shows that the valence domains observed between the iron and nitrogen core domains are well separated, and the values of the $\eta(\vec{r})$ function in the $Fe \cdots N_{i=1-4}$ regions approach zero. This would indicate that the region of the N lone pair should be associated with a monosynaptic non-bonding basin $V(N)_{i=1-4}$ and interpreted as a lone pair. In fact, the lack of a bonding disynaptic basin implies that the nature of the $Fe - N$ bond is governed by electrostatic interactions, $Fe^{\delta+} \cdots N^{\delta-}$, and is not originated in a covalent bonding. Since these interpretations seem to be contradictory, further analysis was performed. In this way, the examination of the covariance matrix of the ELF function shows a large cross term in the average of the localization indexes among the $V(Fe, N)_{i=1-4}$ and the corresponding $V(C, N)_{i=1-8}$. This would mean that the porphyrin ring would be donating electrons to the $Fe - N$ dative bonds. In addition, it was observed that each valence basin between N and Fe belongs to a disynaptic $V(Fe, N)_{i=1-4}$ type, as it has a common surface with both the $C(N)_{i=1-4}$ and $C(Fe)$ core basins. Thus, the analysis of the synaptic order implies that the $Fe - N$ bonding belongs to the covalent-dative type ($N \rightarrow Fe$), as the distribution of ELF values shows that the electron density of the bond is mainly formed by the nitrogen lone pair. The same pattern was observed as regards $Fe - S$ when the disynaptic basin $V(Fe, S)$ was studied. In this case, the topographical analysis of the $\eta(\vec{r})$ function also shows that the valence domains between the iron and sulfur core domains are well separated. However, an analysis based on the indexes obtained from the covariance matrix, and the observation of common surfaces among the disynaptic basin $V(Fe, S)$ and the $C(Fe)$ and $C(S)$ cores, confirms a covalent dative bond ($S \rightarrow Fe$), which is formed mainly by the sulfur lone pair. The oxygen atom, formally bound to the iron atom by the double $Fe = O$ bond is described by the core $C(O)$ attractor and two monosynaptic non-bonding attractors $V_1(O)$ and $V_2(O)$. These last two attractors reflect the non-bonding electron density of the oxygen, which, according to the symmetry of the complex, is represented by two local maxima. Such a topology of $\eta(\vec{r})$ remains in agreement with the Lewis structure predicting the double $Fe = O$ bond with two lone pairs on oxygen. The positions of the $V_1(O)$ and $V_2(O)$ attractors suggest that the electron cloud of valence shell of the O atom is polarized to minimize the Pauli repulsion with the electron clouds associated with the $V(N)_{i=1-4}$ localization

basins. It is worth emphasizing that in the $Fe \cdots O$ region, where the double bond $Fe = O$ is expected, no bonding attractor is observed. Therefore, a clear difference between the iron–oxygen bonding and both iron–nitrogen and iron–sulfur bonding can be observed. Therefore, the covalent bond ($Fe - O$) or covalent-dative bond ($O \rightarrow Fe$) between Fe and O is missing and, from the point of view of the ELF topology analysis, the nature of the binding stems mainly from $Fe^{\delta+} \cdots O^{\delta-}$ electrostatic interactions. The analysis of the synaptic order performed for the valence basins of the oxygen atom shows that they also have common surfaces with the iron core and might be classified as the disynaptic type $V(Fe, O)$. However, this result does not corroborate the covalent nature of the bonding, since the basins (and attractors) are clearly associated with the lone pairs. Nevertheless, the analysis of the covariance matrix of the ELF calculations reveals the presence of 0.43 electrons localized among the Fe and O lone pairs. This fact could be an indicator of a large fluctuation of the electronic density between these basins, which in turn could indicate the presence of a charge-shift bond. Finally, the analysis of the spin density calculated for the η -basins between the doublet to quartet electronic states confirms the radical nature of the porphyrin ring and the sulfhydryl group ($S - H$), which leads to the different multiplicities of Cpd I.

5.2. First Catalytic Subcycle of the Enzyme Aromatase

5.2.1. QM/MM Modeling of the Hydroxylation of the Androstenedione Substrate Catalyzed by Cytochrome P450 Aromatase (CYP19A1)

The hydroxylation of the natural substrate ASD to 19-hydroxy-ASD, occurring during the first catalytic subcycle of the enzyme aromatase, was studied by means of MD simulations and hybrid QM/MM calculations. The starting point of the study was the X-ray crystal structure of the human placenta aromatase in complex with the natural substrate, PDB code 3EQM. Since the cofactor found in the PDB file was the pentacoordinated *heme B* group, this was modeled into the hexacoordinated Compound I (Cpd I) oxidant species. Once the system setup had been completed, a molecular dynamics simulation of 20 ns was carried out in order to balance the enzymatic system. To investigate this hydroxylation step, the oxygen rebound mechanism, originally proposed by Groves et al., was adopted. According to this mechanism, Cpd I abstracts a hydrogen atom from the substrate to give a carbon radical intermediate ($R\bullet$), which recombines with the formal equivalent of an iron-bound hydroxyl radical [$Fe(IV) - OH$] to give the final alcohol product. The hydroxylation of ASD was studied in two different spin states: doublet and

quartet, which lead to two different potential energy surfaces. To explore these PES, the QM subsystem was described by the B3LYP density functional, while the MM subsystem was treated by means of the OPLS and TIP3P force fields. According to the results obtained from the analysis performed on the PES, the rate-limiting step corresponds to the hydrogen abstraction process. The Gibbs free energy barriers for this step (obtained by means of the RRHO approximation) were found to be slightly lower for the enzymatic system (22.3 kcal/mol) than for the QM-only model (25.7 kcal/mol), showing a mild catalytic effect of the enzyme on the biochemical reaction. Further study of this step, based on the analysis of the 2H -KIE, shows almost identical values in both models studied (QM-only model and enzymatic system), indicating that the KIE value is intrinsic to the reaction and is not heavily affected by the enzymatic environment. Moreover, the results show a slight difference in KIE values between different spin states, which is explained by the great geometrical similarity found between the *TS* species obtained in both states. Once the tunneling correction is included, the KIE values exceed the semi-classical limit thus denoting a substantial tunnel effect (as has already been reported in some studies involving *C-H* abstractions). Furthermore, we have incorporated (for the first time in this P450 isoform) the enzymatic conformational diversity in the calculations by means of the FEP methodology. As this methodology takes into account enzyme flexibility, it provides a more realistic view of the role played by the enzyme in the biochemical process. According to the results obtained with this methodology, the free energy barriers for the hydroxylation step are lower than those calculated by way of the RRHO approximation, the values obtained being within the range expected for this kind of biological processes. The hydrogen abstraction step takes place through an endergonic process with an activation barrier of 13.5 kcal/mol, leading to the formation of an alkyl radical on the substrate and a hydroxyl group linked to the heme cofactor. It is worth noting that the iron-hydroxo complex formed in the intermediate species was found in two different electromeric configurations for the quartet spin state ($^4Fe^{III}Por^{\cdot+}$ and $^4Fe^{IV}Por$) and only one for the doublet state ($^2Fe^{III}Por^{\cdot+}$). Both electromers found in the quartet state share the same geometry, but differ in the distribution of electrons in their atoms. From these intermediates, a reorientation of the hydroxyl group occurs prior to the oxygen rebound step. The energetics of this reorientation process depend on the spin state of the substrate, the doublet state being the most favored. Thereby, the rebinding of the oxygen occurs via the hydroxyl rotation and the formation of the new *C-O* bond. The coupled reorientation-rebound process is characterized by a remarkably low free activation barrier of 1.5 kcal/mol, giving rise to the formation of the products through a highly exergonic reaction. Finally, the analysis of the decomposition of the free energy barrier (obtained by FEP calculations) in its different contributions, suggests that the electric field produced by the enzyme during the biochemical reaction is not directly involved in the stabiliza-

tion of the TS when compared with the reactant species. In fact, it is noteworthy that both the polarization of the wave function and the Lennard-Jones term are the only ones responsible for such stabilization. These findings are consistent with the kind of chemical process and/or substrate. Therefore, the main role of the enzyme aromatase during the hydroxylation of ASD consists in a stabilization of the TS by means of the participation of both dispersive and polarization effects.

5.2.2. Joint Use of Bonding Evolution Theory and QM/MM Hybrid Method for Understanding the Hydrogen Abstraction Mechanism via Cytochrome P450 Aromatase

In order to better understand the process occurring in the hydroxylation of ASD, and more specifically the mechanistic step corresponding to the hydrogen abstraction, a new methodology consisting in the joint use of the bonding evolution theory and the QM/MM method was applied. BET combines the electron localization function with Thom's catastrophe theory or, in other words, the topological behavior of the ELF gradient field can be studied within the framework provided by CT. In this way, BET allows a direct relationship to be found between the ELF topology and the evolution of the bond breaking/forming processes and electron pair rearrangements as the chemical reaction proceeds. Accordingly, we offer an alternative representation of the electron density redistribution for the hydrogen abstraction process of the ASD substrate via the enzyme aromatase, at two competing doublet and quartet spin state surfaces. Indeed, this is the first study where the BET has been used in a large biological system, where electronic wave functions obtained from QM/MM calculations are employed, thereby incorporating the polarization effect of the charges derived from the enzymatic environment.

For convenience, only basin populations directly related to the rearrangements of the $C_{19} - H_{19}$ and $O - H_{19}$ bonds have been considered for discussion. In terms of the ELF topological description, the $C_{19} - H_{19}$ breaking and the subsequent $O - H_{19}$ bond formation have been found to take place in five *ELF-SSDs*, which are connected by their respective turning points (see Fig. 5.1). This topological response has been found in both low- and high-spin configurations. The ELF evolution along the MEP displays a certain synchronicity in both spin configurations: (i) *SSD-I* and *SSD-II* have been localized before the respective transition states; (ii) the TS s are localized on the *SSD-III*; and (iii) *SSD-IV* and *SSD-V* have been found after the TS .

The first SSD, *SSD-I*, is characterized by the presence of two basins involving the C_{19} , H_{19} and O atoms, which are the disynaptic basin $V(C_{19}, H_{19})$ and the monosynaptic basin $V(O)$. The first topological change (catastrophe) connecting the *SSD-I* and *SSD-II* takes place through a *fold-type* of catastrophe and is associ-

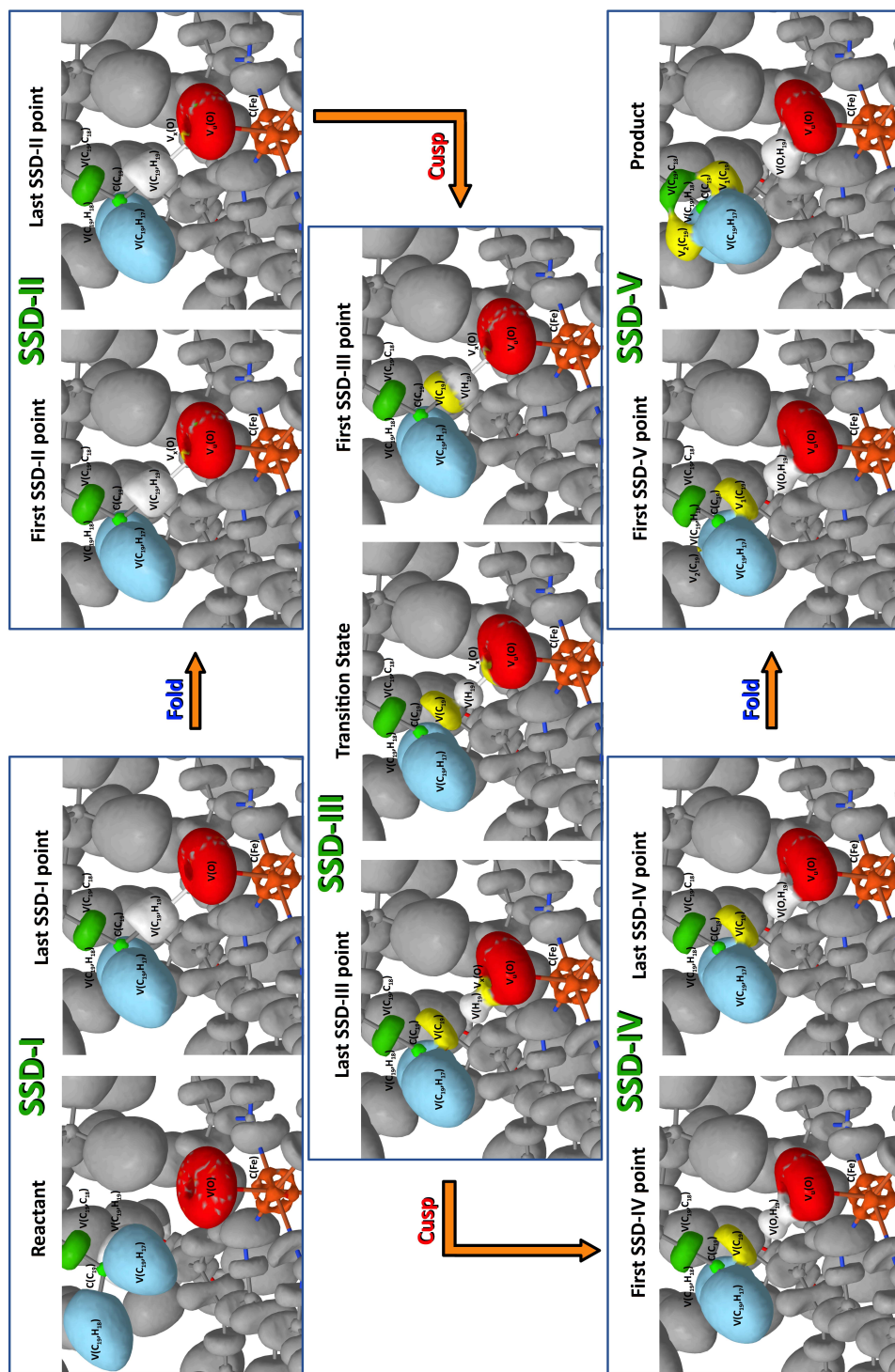


Figure 5.1.: Different SSDs along the hydrogen abstraction from the ASD substrate via aromatase enzyme, and the turning points (catastrophes) connecting them.

ated with the creation of the $V_x(O)$ monosynaptic basin. It is worth noting that the energetic cost associated with *SSD-I* and *SSD-II* not only comprises the energetic cost associated with electronic density rearrangement anticipating the breaking process of the $C_{19} - H_{19}$ bond, but also involves the rotation of the methyl group to allow the closest orientation between the H_{19} and O atoms. Later, the second topological change connecting *SSD-II* and *SSD-III* therefore represents a *cusp-type* of catastrophe associated with the breaking process of the $C_{19} - H_{19}$ bond, where the $V(C_{19}, H_{19})$ disynaptic basin splits into two monosynaptic basins, namely, $V(H_{19})$ and $V(C_{19})$. It is worth mentioning that the sudden topological change of the ELF field produced by this catastrophe does not seem to affect the nearby $V(C_{19}, C_{18})$, $V(C_{19}, H_{18})$, and $V(C_{19}, H_{17})$ basins belonging to ASD. Within the *SSD-III*, the system reaches the *TS* and, at this point, the radical nature of the C_{19} and O atoms enhances. Interestingly, from an ELF-topological point of view, when the system reaches the *TS* there is no evidence of the formation of the $O - H_{19}$ bond. The role of the *SSD-III* is therefore to prepare the system for the imminent formation of this bond. Afterwards, a new turning point connecting *SSD-III* and *SSD-IV* is found, which corresponds to a *cusp-type* of catastrophe. At this turning point, the non-bonding monosynaptic attractors $V(H_{19})$ and $V_x(O)$ are replaced by a single bonding disynaptic attractor $V(O, H_{19})$. Thus, from an ELF-topological point of view, this is the first evidence of the formation of the $O - H_{19}$ bond, which in turn takes place after overcoming the *TS*. Subsequently, during the progress of *SSD-IV*, the $V(C_{19})$ monosynaptic basin populations remain practically constant while the populations of the $V(C_{19}, H_{18})$ disynaptic basin increase. This fact indicates that along the *SSD-IV* a partial charge density is concentrated on the $V(C_{19}, H_{18})$ disynaptic basin rather than $V(C_{19})$. Therefore, during the progress of the *SSD-IV* the electronic flux is directed toward the C_{19} atom, increasing the radical nature of this atom. Finally, when the system reaches the last turning point connecting *SSD-IV* and *SSD-V*, a *fold-type* of catastrophe is found. At this turning point a new $V_2(C_{19})$ monosynaptic attractor is created due to an excess of the charge density around the C_{19} atom indicating a permanent electronic flow toward this atom from the $V_1(C_{19})$ monosynaptic basin as well as from the $V(C_{19}, H_{17})$ and $V(C_{19}, H_{18})$ disynaptic basins.

It is worth mentioning that a post-transition state bifurcation occurs on the quartet state PES, which leads toward two different close-lying electromer intermediates. These electromers, ${}^4Fe^{III}Por^{\bullet+}$ and ${}^4Fe^{IV}Por$, vary in the oxidation state of the iron center and the porphyrin ligand, and are observed in the enzymatic system but not in the gas-phase model. The main difference between these electromers from the point of view of BET is that the *SSD-V* found in the ${}^4Fe^{IV}Por$ electromer is delayed along the reaction path with respect to the ${}^4Fe^{III}Por^{\bullet+}$ electromer; this fact is consistent with the different electronic structures of both electromers. On the other hand, the findings obtained once the BET analysis has

been performed in the gas phase show results similar to those in the enzymatic system. In this way, the same number and type of stability domains (*SSD-I–SSD-V*) are found in the gas-phase model, and are characterized by the same type of catastrophes. The only difference between these models lies in the fact that these catastrophes appear in different values of the reaction coordinate, thereby showing that electronic rearrangement occurs slightly differently in the enzyme and gas phase. In particular, it can be shown that the last turning point that connects the *SSD-IV* and *SSD-V* through a *fold-type* catastrophe is found at lower values along the reaction coordinate for the gas-phase model than in the enzymatic system. These findings also show that the quartet state in the gas-phase model is more similar to the ${}^4Fe^{III}Por^{*+}$ electromer than to the ${}^4Fe^{IV}Por$ one. However, the first two turning points that give rise to the stability domains from *SSD-I* to *SSD-III* take place at similar values of the reaction coordinate in both the enzymatic and gas-phase models. Since those *SSDs* are involved in the activation of the $C_{19} - H_{19}$ bond, this suggests that the electronic process that takes place during the $C_{19} - H_{19}$ bond activation is minimally affected by the electrostatic environment of the enzyme.

5.3. Study of the Hydroxylation of an Aromatase Inhibitor: Exemestane

5.3.1. Theoretical Study of the Mechanism of Exemestane Hydroxylation Catalyzed by Human Aromatase Enzyme

The hydroxylation of the substrate EXE to 19-hydroxy-EXE catalyzed by the enzyme aromatase was studied by means of MD simulations and hybrid QM/MM calculations. The aim of this study was to compare the hydroxylation mechanism of this aromatase inhibitor with that which occurs during the first catalytic subcycle of the natural substrate (ASD). To this end, the hydrogen abstraction–oxygen rebound mechanism that takes place during hydroxylation of ASD was also proposed as the mechanism for the hydroxylation of EXE. As in the study concerning ASD, PDB with code 3EQM was used as the startup geometry in the simulations. The *heme B* cofactor existing in the original PDB file was altered to obtain the iron–oxo porphyrin Cpd I species, and the substrate ASD accommodated in the active site of aromatase was modeled to EXE. It is noteworthy that EXE was retained in the same position as the crystallized ASD. The rest of the system setup process was performed in a similar way to that conducted in the study involving the ASD substrate. The PES were calculated by describing the QM subsystem with the B3LYP density functional, while the MM subsystem was treated by means of the OPLS and TIP3P force fields. According to the results obtained from the

analysis performed on the PES, the ΔG of activation for the hydrogen abstraction step (the rate-limiting step of the reaction) falls in the range of 18–19 kcal/mol. The intermediates generated in this step correspond to the iron–hydroxo complex and alkyl species. As in the study with ASD, the former was found in two different electromeric configurations for the quartet spin state (${}^4Fe^{III}Por\cdot^+$ and ${}^4Fe^{IV}Por$) and only one for the doublet state (${}^2Fe^{III}Por\cdot^+$). The ΔG of activation for the oxygen rebound step shows values of 1.7 kcal/mol for the doublet and 5.5 kcal/mol for the quartet, leading to the formation of 19-hydroxy-EXE through a highly exergonic process (-42 and -47 kcal/mol for the doublet and quartet, respectively). The free energies were also calculated by means of FEP techniques for the hydrogen abstraction step, providing values in the range of 17–19 kcal/mol. Interestingly, when the values calculated from PESs (ΔG) are compared to those calculated through FEP techniques (ΔF), it can be noted that the inclusion of the conformational space of the enzyme does not have a significant impact on the activation barriers. Surprisingly, this trend is in clear contrast to that observed in the investigation including the ASD substrate, where the barriers calculated for the H_{19} -abstraction by means of FEP were about 9 kcal/mol smaller on average than those calculated from the PES. In addition, our results indicate that the hydroxylation step is slower for EXE than for the natural substrate ASD, because the barrier for the rate-limiting step of the reaction is 3.4 kcal/mol higher for EXE (16.9 kcal/mol) than for ASD (13.5 kcal/mol). Indeed, this observation is in good agreement with the values derived from experimental results, specifically when compared to the data obtained from kinetic studies for the hydroxylation of both substrates. According to this experimental data, the free energy of activation for EXE would be around 2.4 kcal/mol higher than for the ASD. Although our reported activation barriers are underestimated (around 5 kcal/mol) when compared to the experimental values, the difference in the activation barrier that we have obtained between EXE and ASD falls inside the experimentally observed range. Furthermore, the hydroxylation of EXE is more exergonic than the hydroxylation of ASD (4.8 kcal/mol for the doublet and 3.3 kcal/mol for the quartet state, respectively). This observation suggests that EXE could act as a competitive inhibitor in the first catalytic subcycle of human aromatase. When comparing the interactions between the EXE and ASD substrates with the active site of aromatase, we see that the main difference lies in the interactions between the substituted C_6 carbon and its nearest surrounding amino acids. In our simulations, we have noted that the C_6 methylene of EXE interacts with the residues Phe-221, Thr-310, Ser-478, Val-369, Val-370, and a water molecule. Interestingly, we have observed a bridge of a polar nature formed by the peptide carbonyl group of Asp-309, the water molecule, and the H_γ of the Ser-478, which is located in the vicinity of the access channel. This polar bridge is aligned with the methylene group and would provide a binding region of an electrostatic nature. In addition, it would introduce a certain electro-

static nature into the interaction between the methyldene group and the active site. It should be pointed out that we have observed the presence of a conformer of the Thr-310 residue, in which the O_{γ_1} hydroxyl moiety is located between the residue EXE and the heme group. The spatial arrangement of this residue also reveals an interaction of the O_{γ_1} atom with a C_{19} -methyl hydrogen of the EXE substrate which might assist the orientation of the C_{19} atom, thus facilitating the hydrogen abstraction process. It is worth pointing out that this disposition differs from that found in the calculations that we performed with the ASD substrate. The stability of this conformer was verified by performing molecular dynamics simulations of 30 ns. Finally, the examination of the different terms into which the activation energy can be decomposed suggests that the electrostatic term is the only one involved in the stabilization of the transition state when compared to the reactant species. This finding obtained for the EXE substrate is not in accordance with that obtained for ASD, where the opposite behavior was observed. In the case of the latter, both the Lennard-Jones term and the polarization of the wave function are responsible for such stabilization. Therefore, the role of the catalytic activity of aromatase during the hydroxylation of the EXE substrate lies in the stabilization of the *TS* by means of the electric field produced by the enzyme, while the role of this enzyme in the hydroxylation of the ASD consists in a stabilization of the *TS* by means of both polarization and dispersive effects. Further analysis has revealed that the difference in the source of the catalytic activity between both substrates is due to two reasons: (i) the presence of a methyldene group substituted in the C_6 of the EXE substrate, which mostly interacts with the surrounding amino acids of a hydrophobic nature, and partly with a labile polar bridge that provides a binding region of an electrostatic nature; and (ii) the different orientation of the Thr-310 residue in both substrates. In the case of EXE, the O_{γ_1} of this residue forms a hydrogen bond with a C_{19} -methyl hydrogen, which could set the position of this group throughout the reaction, thereby facilitating hydrogen abstraction by Cpd I.

Chapter 6

Conclusions



In accordance with the original objectives, four different studies have been presented in this thesis. The main conclusions drawn from this research can be summarized as follows:

- The electronic structure and the bonding nature of Cpd I has been studied for the first time from the standpoint of the ELF. According to this study, the only difference existing between the doublet and the quartet spin states of Cpd I lies in the different spatial positions of the non-bonding attractors of the oxygen atom. The porphyrin macrocycle of Cpd I is composed of a delocalized network of carbon-carbon and carbon-nitrogen bonds, since there is no evidence of localized double bonds in the porphyrin plane. On the other hand, the $Fe-N$ and $Fe-S$ bonds are covalent-dative bonds, although the respective localization basins are “formed” by electron density from lone pairs. Finally, the nature of the bonding between Fe and O predicted by the ELF analysis is primarily attributed to the electrostatic interaction, as no bonding attractor between the Fe and O core attractors is observed. Conversely, the localization indexes obtained from the covariance matrix for the iron and oxygen lone-pair basins indicate the presence of a charge-shift bond.

- The hydroxylation of the substrate ASD to 19-hydroxy-ASD, which occurs during the first catalytic subcycle of the enzyme aromatase, is compatible with the oxygen rebound mechanism proposed by Groves et al. in 1976. Therefore, this mechanism begins with the abstraction of a hydrogen atom from ASD by means of the oxidant species, Cpd I, which takes place through an endergonic process with an activation barrier of 13.5 kcal/mol. This chemical step, which is the rate-limiting step, leads to the formation of an alkyl radical on the substrate and a hydroxyl group linked to the heme cofactor. An analysis of the 2H -KIE calculated for this step shows almost identical values not only in both models studied (QM-only model and enzymatic system), but also in both spin states. This indicates that the KIE value is intrinsic to the reaction and is not heavily affected by the enzymatic environment or by the electronic structure of Cpd I. Once the tunneling correction is included, the KIE values exceed the semi-classical limit, thus denoting a substantial tunnel effect. From the intermediates of the hydrogen abstraction, a reorientation of the hydroxyl group occurs prior to the oxygen rebound step. The energetics of this reorientation process depend on the spin state of the substrate, the doublet state being the most favored. Thereby, the rebinding of the oxygen occurs via the hydroxyl rotation and the formation of the new $C-O$ bond. The coupled reorientation-rebound process is characterized by a remarkably low free activation barrier of 1.5 kcal/mol, giving rise to the formation of the products (19-hydroxy-ASD) through a highly exergonic reaction. Finally, the analysis of the decomposition of the free energy barrier in its different contributions suggests that the electric field produced by the enzyme during the biochemical reaction is not directly involved in the stabilization of the TS when compared with the reactant

species. In fact, it is noteworthy that both the polarization of the wave function and the Lennard-Jones term are the only ones responsible for such stabilization. Therefore, the main role of the enzyme aromatase during the hydroxylation of ASD consists in a stabilization of the *TS* by means of the participation of both dispersive and polarization effects.

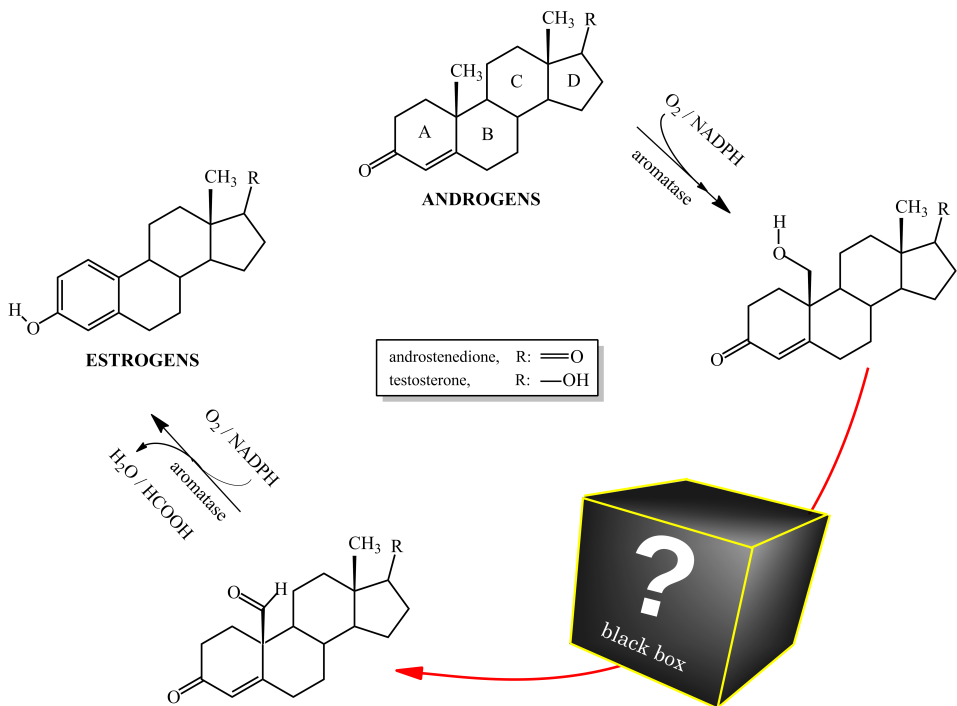
- We have introduced a novel approach to analyze the progress of a biochemical reaction based on the joint use of BET and the QM/MM hybrid method, and we have demonstrated its applicability in reactions involving large biological systems. According to this methodology, the abstraction of a hydrogen atom from the ASD substrate catalyzed by the enzyme aromatase is a non-concerted process, inasmuch as the *C–H* bond breaking and the *O–H* bond formation processes do not take place at the same time. The chemical events provided by BET allows us to characterize five different *SSDs* for this process, which are connected by four catastrophes, two of them fold-type and the other two cusp-type. As a result, the reaction pathway associated with the hydrogen abstraction process first involves a methyl rotation, followed by the breaking process of the *C–H* bond, and finally the formation of the *O–H* bond; the latter occurring after the transition state. In addition, two different electromeric configurations have been found in the high spin state (quartet) of the enzymatic system (not in the gas-phase model); however, these electromers do not affect the evolution of the ELF field along the reaction coordinate. Both electromers give rise to the same radical intermediate; nevertheless, the *SSD-V* appears slightly delayed for the ${}^4Fe^{IV}Por$ electromer. When comparing the results provided by BET in the enzyme system and in the gas-phase model, it has been observed that the same domains (*SSD-I* – *SSD-V*) are found in both models, which are characterized by the same type of catastrophes. The only difference between these models lies in the different values of the reaction coordinate at which these catastrophes take place. This difference, which is not large, shows that the electronic rearrangement occurs slightly differently in the enzyme and in the gas phase, thus suggesting that the chemical reaction is electronically almost equivalent in both models.

- The hydroxylation of the EXE inhibitor to 19-hydroxy-EXE, catalyzed by the enzyme aromatase, has been proposed to proceed via the oxygen rebound mechanism, as in the hydroxylation of ASD. According to the results derived from the analysis performed on the PES, the rate-limiting step during the hydroxylation reaction of EXE also corresponds to the hydrogen abstraction process. The activation barriers for this step reveal that it occurs through an endergonic process with activation barriers in the range from 16.9 kcal/mol (doublet) to 19.2 kcal/mol (quartet), which are around 3–5 kcal/mol higher than those found in the same step for ASD. According to these findings, the hydroxylation reaction for the inhibitor proceeds through a slower process than for the natural substrate. These results are

consistent with the experimental studies, from which an activation barrier for the hydroxylation of EXE may be estimated as being 2.4 kcal/mol higher than that for the hydroxylation of ASD. Because the hydroxylation of the EXE substrate is compatible with the mechanism proposed above, and the activation barriers are within the expected range of values, everything suggests that this is not the stage where the inhibition of aromatase takes place. Besides, EXE is a suicidal inhibitor, and thus it should irreversibly bind the active site of the enzyme at some stage. Since this behavior could not be observed throughout the hydroxylation process, we dismiss the first catalytic subcycle of aromatase as the stage of the inhibition. Furthermore, the fact that the products from the EXE hydroxylation show lower free energies than those obtained for ASD suggests a competitive inhibition in the first subcycle of human aromatase. Finally, the examination of the different terms in which the activation energy can be decomposed suggests that the electrostatic term is the only one involved in the stabilization of the transition state when compared to the reactant species. This finding obtained for the EXE substrate is not in accordance with that obtained for ASD, where the opposite behavior was observed. Therefore, the role of the catalytic activity of aromatase during the hydroxylation of the EXE substrate lies in the stabilization of *TS* by means of the electric field produced by the enzyme. Further analysis has revealed that the difference in the source of the catalytic activity between both substrates is due to two reasons: (i) the presence of a methyldene group substituted in the C_6 of the EXE substrate, which mostly interacts with the surrounding amino acids of a hydrophobic nature, and partly with a labile polar bridge that provides a binding region of an electrostatic nature; and (ii) the different orientation of the Thr-310 residue in the two substrates. In the case of EXE, the O_{γ_1} of this residue forms a hydrogen bond with a C_{19} -methyl hydrogen, which could set the position of this group throughout the reaction, thereby facilitating hydrogen abstraction by Cpd I.

Chapter 7

Future Perspectives



The first catalytic subcycle of the enzyme aromatase consists in the oxidation of the C_{19} -methyl group of the androgen substrates. In this thesis, we have studied this catalytic stage amply, specifically for the androgen androstenedione, and we have corroborated that this mechanism takes place through the oxygen rebound mechanism, proposed by Groves et al. The third catalytic subcycle consists of a lyase reaction, in which the $C_{10} - C_{19}$ bond of the androgens is cleaved, resulting in the aromatization of the phenolic A ring of the androgens, and expelling formic acid and a water molecule. The literature on this mechanism is very extensive because it has been studied by means of both experimental and theoretical techniques. In fact, several mechanisms have been proposed, including 1β - and 2β -hydroxylation, 4,5-epoxidation, the Baeyer–Villiger oxidation of C_{19} and 10β -hydroxyestr-4-ene-3,17-dione formation, or with the intermediation of a C_{19} -peroxide. In spite of this, the third oxidation step is still under discussion, and no consensus has yet been reached on the issue.

However, the second catalytic subcycle of the enzyme aromatase has remained unknown until now, partly due to the lack of theoretical or practical experiments on this subject. In this catalytic subcycle, the product from the first catalytic subcycle, 19-hydroxy-androgen, is oxidized to the 19-gem-diol species (see Fig. 7.1). The mechanism of this oxidation is thought to proceed through the same mechanism from the first catalytic subcycle, i.e., the oxygen rebound mechanism with Cpd I as the oxidizing agent. In a second step, the 19-gem-diol species would be dehydrated to the aldehyde intermediate, 19-oxo-androgen, which is the reactant species of the third catalytic subcycle.

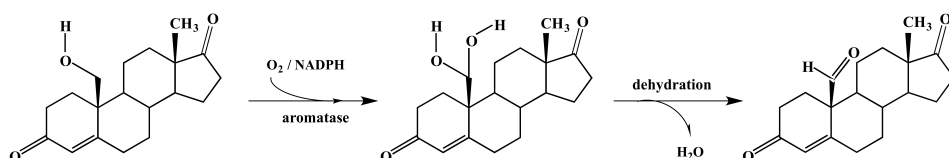


Figure 7.1.: Second catalytic subcycle of the enzyme aromatase.

Because there is no evidence that the above mechanism is the one that occurs in the formation of the 19-oxo-androgens, and since we are not absolutely certain that this is the only possible mechanism, we are currently studying the second catalytic subcycle of the enzyme aromatase. In this way, we have proposed several possible mechanisms that might occur throughout this subcycle, with the aim of shedding light upon this hitherto unknown stage. Specifically, we have proposed three possible mechanistic paths for the oxidation of 19-oxo-ASD, which can be seen in Fig. 7.2.

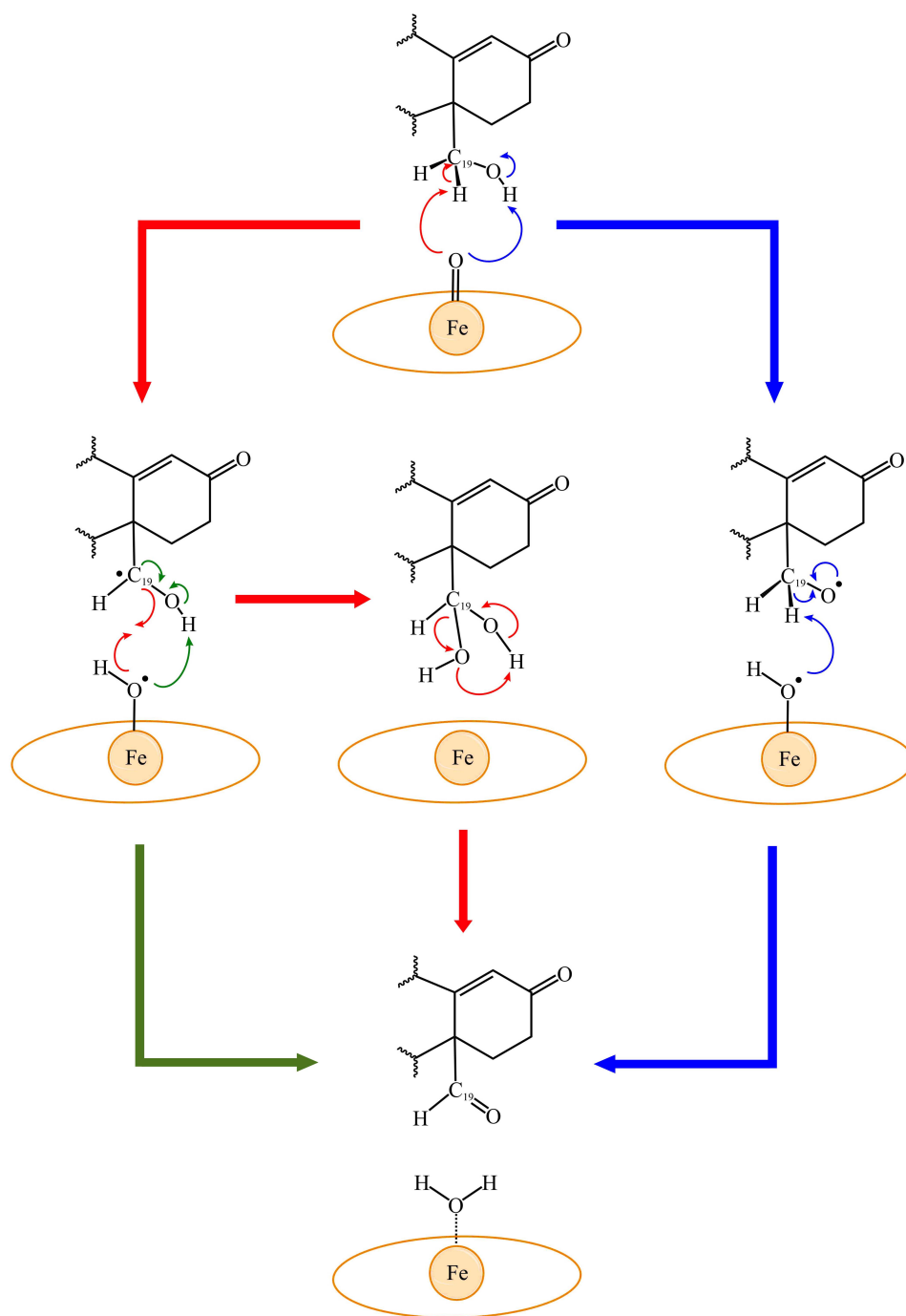


Figure 7.2.: Different mechanisms proposed for the second catalytic subcycle of the enzyme aromatase.

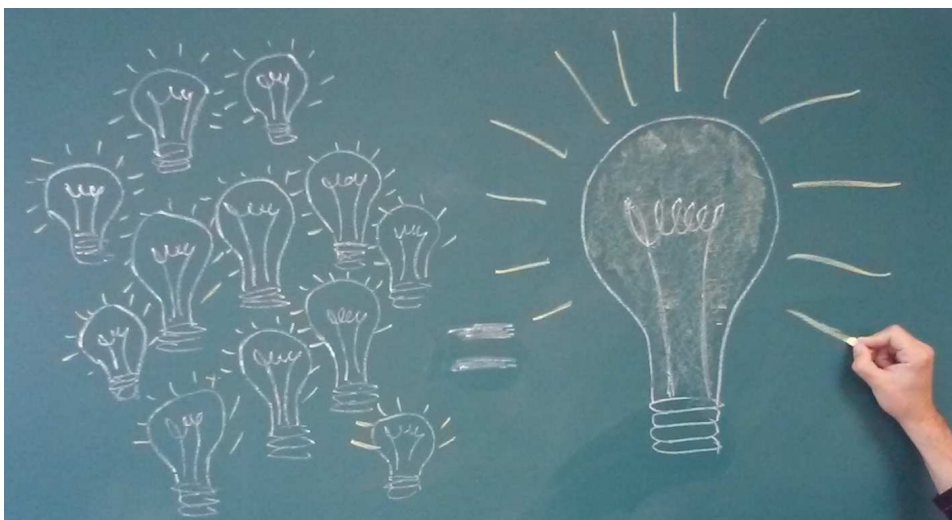
The red path is the one corresponding to the oxygen rebound mechanism. As can be seen, Cpd I abstracts a hydrogen atom from the 19-hydroxy-ASD to give a carbon radical intermediate on the C_{19} and on the iron-bound hydroxyl radical [$Fe(IV) - OH$]. Then a recombination of both radicals leads to the formation of the 19-gem-diol-ASD and a pentacoordinated heme group. In a second step, an oxygen of the OH group abstracts a hydrogen atom of the other hydroxyl group, thereby dehydrating the gem-diol and giving rise to the 19-oxo-ASD and a water molecule; the latter would be linked to the iron atom.

The green path begins at the end of the first step of the red path, i.e., when the carbon radical is formed on the C_{19} and the hydroxyl group is bound to the iron porphyrin. However, at this point, instead of the recombination of both radicals, the iron-bound hydroxyl radical abstracts the hydrogen atom from the OH group. In this way, after an electronic rearrangement the 19-oxo-ASD and a water molecule are formed, the latter being coordinated to the iron atom.

Finally, the blue path proceeds through a hydrogen abstraction from the hydroxyl group of the 19-hydroxy-ASD, by means of Cpd I. This path provides an alternative to the red path, since the hydrogen abstracted comes from the hydroxyl group instead of from the $C - H$ bond. The intermediates of this step are the iron-bound hydroxyl radical, as in the red path, and a radical on the oxygen of the hydroxyl group (alkoxy radical). It should be pointed out that this step could also be viewed as an enolization reaction. Subsequently, a hydrogen abstraction from the $C - H$ bond takes place via the iron-bound hydroxyl radical, leading to the 19-oxo-ASD and a water molecule linked to the iron atom.

In addition, we intend to study the role of the amino acid Thr-310 in the catalysis during this catalytic step. We believe that the orientation of this residue may play a major mediating role, as we could observe when we studied the hydroxylation of the EXE substrate. Another factor to be taken into account is the presence of water molecules in the active site of aromatase, as these might participate in the catalytic process, for example, in the dehydration of the 19-gem-diol species. For this reason, we intend to study the above three mechanisms, but with an extra water molecule involved in the oxidation mechanism.

Resumen



Introducción

Los citocromos P450 (CYP o P450) forman una superfamilia de hemo metaenzimas constituida por varias familias y subfamilias que pertenecen a los tres dominios de la vida. Esta superfamilia incluye una gran cantidad de enzimas isoformas, actualmente se han contabilizado 21039, las cuales juegan un papel muy importante en los procesos biológicos en diversos bioorganismos tales como plantas, mamíferos, hongos, arqueas, bacterias o protistas. La principal función que desempeñan estas enzimas en los seres humanos es la de metabolizar una amplia variedad de sustratos tanto endógenos como exógenos, así como participar activamente en procesos de oxidación y reducción durante el metabolismo de compuestos xenobióticos. De hecho, las enzimas P450 están involucradas en el metabolismo de aproximadamente el 75 % del total de los medicamentos que hay en circulación. Los P450 tienen la habilidad de activar el oxígeno molecular mediante su grupo prostético, una hemo porfirina que contiene un átomo de hierro, para catalizar la oxidación de hidrocarburos inactivados. Esta oxidación consiste en la inserción de uno de los dos átomos de oxígeno en el enlace $C-H$ de sus sustratos, generalmente en forma de hidroxilo, mientras que el otro oxígeno se reduce a una molécula de agua. Esta reacción precisa tanto de oxígeno molecular como NADPH como cosustratos y se conoce como reacción monooxigenasa, lo que le confiere a estas enzimas el sobrenombre de monooxigenasas. De entre todas las enzimas isoformas de la familia de citocromos P450, únicamente 57 de ellas pertenecen a la especie humana, una de estas, el CYP19A1, se conoce como enzima aromatasa humana.

La aromatasa, también conocida como estrógeno sintetasa, es la enzima responsable de la última etapa (paso clave) de la biosíntesis de hormonas esteroideas a partir del colesterol. Concretamente, la enzima aromatasa está involucrada en la formación de los C_{18} -estrógenos: estrona, 17β -estradiol y $17\beta, 16\alpha$ -estriol, a partir de los C_{19} -andrógenos: androstenediona (ASD), testosterona (TST) y 16α -hidroxitestosterona, respectivamente. De hecho, ésta es la única enzima responsable de la biosíntesis de los estrógenos a partir de los andrógenos en los vertebrados. Esta conversión consiste en la aromatización del anillo A de los andrógenos (de ahí procede el nombre aromatasa), que tiene lugar a través de un proceso de tres oxidaciones consecutivas (subciclos catalíticos) del grupo metilo del carbono C_{19} de los andrógenos. Este proceso catalítico requiere la actividad de la enzima citocromo P450 reductasa (CPR), que ejerce como fuente de electrones, y además, cada paso de oxidación consume 1 mol de NADPH y 1 mol de oxígeno molecular.

Según el ciclo catalítico propuesto para la enzima aromatasa, las dos primeras etapas de oxidación proceden a través de dos hidroxilaciones secuenciales del carbono C_{19} de los andrógenos. En la primera etapa se forman los C_{19} -hidroxiandrógenos, mientras que la segunda conduce a la formación de la especie C_{19} -gemdiol; este último se deshidrata posteriormente para formar el intermedio aldehído

(C_{19} -*oxo*-andrógenos). Finalmente, en la tercera etapa de oxidación se produce una reacción *liasa*, en la cual se rompe el enlace formado por los carbonos C_{10} y C_{19} de los andrógenos, dando lugar a la aromatización del anillo fenólico A de los andrógenos y generando como subproductos ácido fórmico y una molécula de agua.

Los estrógenos son muy importantes en los seres humanos debido al papel que juegan en el desarrollo sexual y reproductivo. Particularmente, los estrógenos son esenciales para las mujeres ya que son las principales hormonas sexuales femeninas. En las mujeres premenopáusicas, la principal fuente de estrógenos son los ovarios, que producen mayoritariamente estrona. Por el contrario, la principal fuente de producción de estrógenos en las mujeres posmenopáusicas se atribuye a la aromatización de andrógenos ováricos y suprarrenales (ASD y TST) a estrógenos (estrona y 17β -estradiol), por medio de la enzima aromatasa en los tejidos periféricos. Estas hormonas, a parte de ser esenciales en el desarrollo sexual de las mujeres, también juegan un papel significativo en el desarrollo y proliferación del cáncer de mama posmenopáusico. Ésto es debido a que los estrógenos se unen al receptor de estrógeno (ER), el cual activa la transcripción de sus genes diana, estimulando de este modo la proliferación de las células mamarias humanas. El complejo estrógeno-ER interactúa con el ADN y puede causar mutaciones debido al creciente aumento de la división celular, favoreciendo así el desarrollo de células cancerígenas. Alrededor de dos tercios de todos los casos de cáncer de mama están relacionados con el tipo de cáncer dependiente de hormonas; sin embargo, esta proporción aumenta hasta el 75 % cuando nos referimos al cáncer de mama posmenopáusico. Por esta razón, el bloqueo de la biosíntesis de estrógenos ha demostrado ser una buena estrategia para combatir los cánceres de mama dependientes de estrógenos, además de ser un objetivo primordial en el campo de la endocrinología. Por lo tanto, la inhibición de la enzima aromatasa, a través de fármacos conocidos como inhibidores de la aromatasa (AI), constituye la terapia de primera línea para estos tipos de cánceres. Los AIs se pueden clasificar en función de sus estructuras en inhibidores Tipo I (esteroideos) o Tipo II (no esteroideos). Los inhibidores de Tipo I, como el exemestano (EXE) o el formestano, tienen una estructura similar al sustrato natural ASD, y por lo tanto actúan como pseudo sustratos de la aromatasa. De este modo, estos inhibidores se unen irreversiblemente al centro activo de la enzima y por lo tanto se consideran como inactivadores basados en el mecanismo o inhibidores suicidas. Los inhibidores de Tipo II son en su mayoría compuestos basados en azol, como el anastrozol o letrozol, que son derivados del triazol. Estos inhibidores compiten de forma reversible con el sustrato natural para unirse a la enzima, interfiriendo así con el cofactor del grupo *hemo*. En la actualidad, el tratamiento de primera línea para el cáncer de mama hormonodependiente localmente avanzado está basado en la tercera generación de inhibidores de la aromatasa. Esta familia de inhibidores consta de tres compuestos, anastrozol, letrozol y exemestano, que se comercializan bajo las marcas comerciales ARIMIDEX, FEMARA y AROMASIN, respectivamente.

La primera de las tres etapas de oxidación, que tiene lugar durante la formación de estrógenos, consiste en la hidroxilación de los andrógenos para formar C_{19} -hidroxi-andrógenos. Se ha propuesto que esta hidroxilación procede a través del mecanismo de rebote de oxígeno (oxygen rebound), que fue originalmente propuesto por Groves y otros en 1976. Este mecanismo consiste en dos etapas: abstracción de hidrógeno y rebote de oxígeno. Inicialmente, un enlace $C - H$ del sustrato se activa por medio de las especies reactivas, una oxo-porfirina de hierro (IV) π -cation radical también conocido como "Compound I" (Cpd I), el cual abstrae un átomo de hidrógeno del enlace activado. De esta forma se genera un radical alquilo en el átomo de carbono y un complejo de hierro-hidroxo porfirina. Seguidamente, tiene lugar una reorientación tanto del grupo OH como del alquilo, que facilita la posterior etapa de rebote de oxígeno en el cual se forma un nuevo enlace $C - O$. En este paso de rebote de oxígeno, el radical alquilo se recombina con el radical hidroxilo unido al átomo de hierro para formar el sustrato hidroxilado correspondiente, así como un complejo pentacoordinado de porfirina de hierro (III).

El mecanismo de rebote de oxígeno ha sido probado de forma teórica y experimental como mecanismo de hidroxilación de sustratos por diversas enzimas isoformas del P450. Sin embargo, en el caso de la aromatasa, a pesar de que hay bastantes estudios experimentales sobre este primer subciclo catalítico, no hay ningún estudio teórico al respecto que avale que el mecanismo de Groves y otros sea el que procede durante la etapa de hidroxilación del grupo C_{19} -metilo de los andrógenos. La falta de estudios teóricos sobre la enzima aromatasa se debe a que hasta hace relativamente poco tiempo (2009) no se había conseguido cristalizar la estructura de rayos-X debido a la gran dificultad que entrañaba. De esta forma, solo se disponía de algunos modelos basados en la homología de otras isoformas del P450 o de datos de mutagénesis de sitio dirigido. Finalmente, Ghosh y otros obtuvieron la estructura cristalográfica de la aromatasa en complejo con el sustrato natural ASD en el centro activo. Esta estructura cristalográfica permite realizar cálculos teóricos que puedan apoyar a las observaciones experimentales en las etapas mecanísticas de la enzima aromatasa. Además, como se ha comentado anteriormente, la inhibición de la aromatasa es uno de los objetivos principales en el campo de la endocrinología en lo que respecta al cáncer de mama dependiente de hormonas. De este modo, el estudio del mecanismo de acción de esta enzima y de los mecanismos de unión al centro activo mediante técnicas computacionales, nos puede ayudar a desarrollar nuevos fármacos para el tratamiento de dicha enfermedad.

Objetivos

La enzima citocromo P450 aromatasa juega un papel esencial en la biosíntesis de estrógenos y su inhibición es un objetivo primordial para el desarrollo de fármacos para el tratamiento de cáncer de mama. El objetivo principal de la presente tesis

es mejorar la comprensión del mecanismo catalítico y la bioquímica de esta enzima desde el punto de vista de la química teórica y computacional. Con este objetivo, se ha llevado a cabo un estudio teórico por medio de técnicas computacionales del primer subciclo catalítico de la enzima aromatasasa, así como un análisis de la estructura electrónica de la especie reactiva de esta enzima. En concreto, los objetivos que se han planteado y abordado durante esta tesis son los siguientes:

- **Estudio de la estructura electrónica de las especies reactivas de la enzima aromatasasa (Compound I).**

El objetivo de este estudio es arrojar luz sobre la estructura electrónica y la naturaleza de los enlaces que conforman la especie oxidante Cpd I. Puesto que el Cpd I es la especie reactiva de la enzima aromatasasa, la comprensión de su estructura electrónica proporcionará un conocimiento más amplio de la bioquímica de esta enzima. Con este propósito, el estudio de la estructura electrónica se basa en dos métodos de análisis de la topología químico-cuántica: la función de localización electrónica (ELF) y la densidad electrónica.

- **Estudio del primer subciclo catalítico de la enzima aromatasasa.**

Estudio del mecanismo de reacción de la primera etapa de oxidación del sustrato natural ASD por medio de la metodología híbrida QM/MM. Este subciclo catalítico consiste en la hidroxilación del grupo C_{19} -metilo del sustrato ASD a través de un mecanismo de dos pasos. Una primera etapa en la que un átomo de hidrógeno presente en el grupo C_{19} -metilo de la ASD es abstraído por mediación del Cpd I, seguido por una segunda etapa de rebote de oxígeno, en la cual se forma el 19-hidroxi-ASD. Adicionalmente, se realizará un estudio más detallado sobre el mecanismo de la etapa de abstracción de hidrógeno mediante la utilización conjunta de la teoría de la evolución de enlace (BET) y de la metodología híbrida QM/MM. Este estudio proporcionará una nueva perspectiva sobre esta etapa mecanística tan común en los citocromos P450.

- **Estudio de la hidroxilación del exemestano, un inhibidor suicida de la enzima aromatasasa, durante el primer subciclo catalítico de la enzima.**

El sustrato ASD presente en el centro activo de la aromatasasa será reemplazado por el sustrato EXE, un inhibidor de la aromatasasa de tercera generación que se utiliza en la terapia de cáncer de mama. De este modo, la hidroxilación del grupo C_{19} -metilo de este sustrato se estudiará proponiendo para la hidroxilación el mecanismo que tiene lugar durante el primer subciclo catalítico de la aromatasasa. Puesto que ambos sustratos, ASD y EXE, son derivados esteroideos y dado que sus geometrías moleculares son muy similares, se realizará una comparación entre los procesos de hidroxilación de ambos sustratos. El estudio de la reactividad del EXE

y de su mecanismo de unión (*binding*) al centro activo de la aromatasas, permitirá conocer mejor el mecanismo de acción de los inhibidores esteroideos en la enzima aromatasas.

Resultados y Discusión

- El primer estudio que se presenta tiene un título que se traduce por: “**Nuevas Perspectivas sobre la Estructura Electrónica del Compuesto Oxo-Porfirina de Hierro (IV) (Compound I). Un Análisis Topológico Químico-Cuántico**”.

En este trabajo se estudió la estructura electrónica del complejo oxo-ferrilo con una porfirina π -catiónica radical ($Por^{\cdot+}Fe^{IV} = O(S-H)$), conocido como Cpd I, para los estados electrónicos doblete y cuadruplete, por medio de dos métodos de análisis de la topología químico-cuántica: la función de localización electrónica, $\eta(\vec{r})$, y la densidad electrónica, $\rho(\vec{r})$. De acuerdo con el planteamiento de la ELF, la estructura electrónica del Compound I se caracteriza por 76 máximos locales (atractores), tanto en el estado electrónico doblete como en el cuadruplete, que están localizados en el campo de la ELF $\eta(\vec{r})$. La densidad electrónica de los núcleos se caracteriza por un total de 27 atractores de core: $C(Fe)$, $C(S)$, $C(O)$, $C(N)_{i=1-4}$ y $C(C)_{i=1-20}$, que coinciden con las posiciones de los núcleos atómicos. Por otro lado, los electrones en la capa de valencia están asociados con tres cuencas monosinápticas no enlazantes (atractores): $V(S)$ and $V_{i=1,2}(O)$ correspondientes a los pares solitarios de los átomos de azufre y oxígeno, respectivamente. Además, se puede observar un total de 33 atractores de valencia disinápticos enlazantes: $V(C, C)_{i=1-20}$, $V(C, N)_{i=1-8}$, $V(Fe, S)$, and $V(Fe, N)_{i=1-4}$, correspondientes a los enlaces químicos $C - N_{i=1-8}$, $Fe - S$, $Fe - N_{i=1-4}$ respectivamente. Los enlaces $C - H$ y $S - H$ aparecen reflejados por un total de 12 atractores disinápticos protonados $V(H, C)_{i=1-12}$ y por un atractor $V(H, S)$. Cabe destacar que para el Cpd I se localizaron el mismo número y el mismo tipo de atractores, tanto para el estado doblete como para el cuadruplete. La única diferencia existente entre ambos estados de espín es la diferente orientación espacial que tienen los atractores $V_{i=1,2}(O)$ con respecto a los enlaces $Fe - N$ situados en el plano de la porfirina. Esta diferencia espacial podría estar asociada con la diferente ocupación de los orbitales en ambos estados electrónicos. Los atractores de core $C(Fe)$, $C(N)_{i=1-4}$, $C(C)_{i=1-20}$ y los atractores de valencia $V(H, C)_{i=1-12}$, $V(C, C)_{i=1-20}$, $V(C, N)_{i=1-8}$, $V(Fe, N)_{i=1-4}$ reflejan la estructura electrónica del complejo de hierro-porfirina: la geometría plana del anillo porfirínico del Cpd I está asociado con el carácter aromático originado por la conjugación de los dobles enlaces $C = C$ y $C = N$. Sin embargo, la ausencia de los atractores $V_{i=1,2}(C, C)$ o $V_{i=1,2}(C, N)$, que corresponderían a la firma topológica de los enlaces dobles localizados $C = C$ y $C = N$, y que deberían colocarse

por encima y por debajo del plano del anillo macrocíclico, denota una gran deslocalización de la densidad electrónica. Además, dado que la molécula de porfirina, *Por*, se puede representar como dos subunidades aisladas de pirrol, C_4H_4NH , y de dos aniones pirrol, $[C_4H_4N]^-$, la topología de la ELF para las moléculas de *Por*, C_4H_4NH y $[C_4H_4N]^-$ se comparó con la del anillo porfirínico presente en el Cpd I. Los resultados de esta comparación mostraron que, teniendo en cuenta una serie de atractores y su sinapticidad, la topología de la ELF para el anión pirrol es la más similar de los tres a la encontrada en el Cpd I. Este hecho apoya la observación de que la cuenca $V(Fe, N)_{i=1-4}$ se compone principalmente de electrones provenientes del par solitario del átomo de nitrógeno. Sin embargo, la distancia entre los atractores $V(Fe, N)_{i=1-4}$ y $C(N)_{i=1-4}$ es más corta en el Cpd I que entre los atractores $V(N)$ y $C(N)$ en el anión pirrol. Este hecho puede justificarse como resultado de la repulsión de Pauli entre electrones “contenidos” en las cuencas $C(Fe)$ y $V(Fe, N)_{i=1-4}$ y a la correspondiente “compresión” de la densidad electrónica en esta región. La presencia de estas cuatro cuencas $V(Fe, N)_{i=1-4}$ localizadas entre el átomo de hierro y los cuatro átomos de nitrógeno se puede interpretar como enlaces covalentes entre estos átomos. Sin embargo, el análisis topográfico (2D) de la función $\eta(\vec{r})$ realizado en el plano molecular del Cpd I, muestra que los dominios de valencia observados entre los dominios de los núcleos de hierro y nitrógeno están bien separadas, y que la función de los valores de $\eta(\vec{r})$ en las regiones $Fe \cdots N_{i=1-4}$ se aproximan a cero. Ésto indicaría que la región del par solitario de nitrógeno debería estar asociado con una cuenca monosináptica no enlazante $V(N)_{i=1-4}$ y se interpreta como un par solitario. De hecho, la falta de una cuenca disináptica enlazante implica que la naturaleza del enlace $Fe - N$ se rige por interacciones electrostáticas, $Fe^{\delta+} \cdots N^{\delta-}$, y no se trata de un enlace covalente. Dado que estas interpretaciones parecen ser contradictorias, se llevó a cabo un análisis más detallado. De este modo, el examen de la matriz de covarianza de la función ELF mostró un término cruzado grande para los índices de localización entre las cuencas $V(Fe, N)_{i=1-4}$ y $V(C, N)_{i=1-8}$. Ésto significaría que el anillo de porfirina estaría donando electrones en forma de enlaces dativos $N \rightarrow Fe$. Además, se observó que cada cuenca de valencia entre el nitrógeno y el hierro pertenece a una cuenca de tipo disináptico $V(Fe, N)_{i=1-4}$, dado que tiene una superficie común tanto con las cuencas de core del nitrógeno, $C(N)_{i=1-4}$, como con la del átomo de hierro, $C(Fe)$. Por lo tanto, el análisis del orden sináptico implica que el enlace $Fe - N$ pertenece al tipo de enlace covalente dativo ($N \rightarrow Fe$), dado que la distribución de los valores de ELF muestra que la densidad electrónica del enlace está formado principalmente por el par solitario del nitrógeno. El mismo patrón se observó para el enlace $Fe - S$ cuando se estudió la cuenca disináptica $V(Fe, S)$. En este caso, el análisis topográfico de la función $\eta(\vec{r})$ también mostró que los dominios de valencia que hay entre el dominio de los núcleos de hierro y azufre están bastante separados. Sin embargo, un análisis basado en los índices obtenidos a partir de la matriz de covarianza, así como la ob-

servación de las superficies comunes entre la cuenca disináptica $V(Fe, S)$ y las de core $C(Fe)$ y $C(S)$, confirma la presencia de un enlace covalente dativo ($S \rightarrow Fe$), el cual estaría formado principalmente por el par solitario del azufre. Por otro lado, el átomo de oxígeno que está formalmente unido al átomo de hierro por el doble enlace $Fe = O$, está descrito por un atractor de core $C(O)$ y dos atractores monosinápticos no enlazantes $V_1(O)$ y $V_2(O)$. Estos dos últimos atractores reflejan la densidad de electrones no enlazantes del oxígeno, el cual, de acuerdo con la simetría del complejo, está representado por dos máximos locales. Dicha topología de $\eta(\vec{r})$ se mantiene de acuerdo con la estructura de Lewis predicha para el doble enlace $Fe = O$, con dos pares libres en el oxígeno. Las posiciones de los atractores $V_1(O)$ y $V_2(O)$ sugieren que la nube de electrones de la capa de valencia del átomo de oxígeno está polarizado para minimizar la repulsión de Pauli con las nubes de electrones asociados con las cuencas de localización $V(N)_{i=1-4}$. Cabe resaltar que en la región $Fe \cdots O$ donde se esperaría el doble enlace $Fe = O$, no se encuentra ningún atractor de enlace. Así pues, se puede observar una clara diferencia entre el enlace formado entre los átomos de hierro y oxígeno y los enlaces $Fe - N$ y $Fe - S$. Por lo tanto, se descarta que haya un enlace covalente ($Fe - O$) o un enlace covalente dativo ($O \rightarrow Fe$). Desde el punto de vista del análisis topológico de la ELF, la naturaleza del enlace químico tendría que atribuirse principalmente a interacciones electrostáticas $Fe^{\delta+} \cdots O^{\delta-}$. El análisis del orden sináptico realizado para las cuencas de valencia del átomo de oxígeno muestra que ésta también tiene superficies comunes con el núcleo de hierro y por tanto, podría ser clasificada como cuenca disináptica $V(Fe, O)$. Sin embargo, este resultado no certifica el carácter covalente del enlace dado que las cuencas (y atractores) se asocian claramente con los pares solitarios. No obstante, el análisis de la matriz de covarianza de los cálculos de la ELF revela la presencia de 0,43 electrones localizados entre los pares solitarios de hierro y oxígeno. Este hecho podría ser un indicador de una gran fluctuación de la densidad electrónica entre estas cuencas, que a su vez podrían indicar la presencia de un enlace de desplazamiento de carga (charge-shift). Por último, el análisis de la densidad de espín calculado para las $\eta(\vec{r})$ -cuencas entre los estados electrónicos doblete y cuadruplete confirma la naturaleza radicalaria del anillo de la porfirina y del grupo sulfhidrilo ($S - H$), que da lugar a las diferentes multiplicidades presentes en el Cpd I.

- El segundo estudio que se presenta tiene un título que se traduce por: “**Modelado QM/MM de la Hidroxilación del Sustrato Androstenediona Catalizado por el Citocromo P450 Aromatasa (CYP19A1)**”.

En este trabajo, la hidroxilación del sustrato natural ASD para dar 19-hidroxi-ASD, que tiene lugar durante el primer subciclo catalítico de la enzima aromatasa, se estudió por medio de simulaciones de dinámica molecular y de cálculos híbridos QM/MM. El punto de partida del estudio fue la estructura cristalina de rayos

X de la aromataasa placentaria humana, en complejo con el sustrato natural ASD (código PDB 3EQM). Puesto que en el archivo PDB se encontró que el cofactor era un grupo *hemo B* pentacoordinado, éste se modeló a la especie hexacoordinada Compound I, que es la especie oxidante de la reacción. Una vez que se completó la configuración del sistema, se llevó a cabo una simulación de dinámica molecular de 20 ns con el fin de equilibrar el sistema enzimático. Para investigar este paso de hidroxilación se adoptó el mecanismo de rebote de oxígeno, propuesto originalmente por Groves y otros. De acuerdo con este mecanismo, el Cpd I abstrae un átomo hidrógeno del sustrato para dar lugar a un radical de carbono ($R\bullet$) como intermedio de la reacción. Seguidamente, este carbono radical se recombina con el equivalente formal de un radical hidroxilo, que está unido al átomo de hierro [$Fe(IV) - OH$] para dar lugar a un alcohol como producto final de la reacción. La hidroxilación del ASD se estudió en dos estados de espín diferentes: doblete y cuadruplete, que conducen a dos superficies de energía potencial (PES) diferentes. Para poder explorar dichas PESs, el subsistema QM (125 átomos) fue descrito mediante el funcional de densidad B3LYP, mientras que el subsistema de MM (57.127 átomos) se trató por medio de los campos de fuerza OPLS-AA y TIP3P. De acuerdo con los resultados obtenidos a partir de los análisis realizados en las PESs, la etapa limitante de la reacción corresponde al proceso de abstracción de hidrógeno. Las barreras de energía libre de Gibbs calculadas para esta etapa (obtenidas por medio de la aproximación RRHO) fueron ligeramente más bajas para el sistema enzimático (22,3 kcal/mol) que para el modelo QM en fase gas (25,7 kcal/mol), lo que demuestra un efecto catalítico moderado por parte de la enzima en esta reacción bioquímica. Un estudio adicional de esta etapa, basado en el análisis de los efectos cinéticos isotópicos, 2H -KIE, mostró valores de KIEs prácticamente iguales en ambos modelos estudiados (modelo QM en fase gas y el sistema enzimático), lo que indica que el valor de KIE es intrínseco a la reacción y no se ve muy afectado por el entorno enzimático. Por otra parte, los resultados muestran una ligera diferencia en los valores de los KIEs para los diferentes estados de espín, lo cual se explica por la gran similitud geométrica que existe entre los estados de transición (TS) obtenidos para cada multiplicidad. Una vez que se incluyó la corrección del efecto túnel, los valores de los KIEs excedieron el límite semiclásico, denotando así un efecto túnel sustancial (como ya se ha reportado en algunos estudios sobre procesos de abstracción de hidrógeno en enlaces $C-H$). Además, en este estudio incorporamos, por primera vez en esta isoforma P450, la diversidad conformacional enzimática en los cálculos por medio de la metodología de perturbación de la energía libre (FEP). Dado que la técnica FEP tiene en cuenta la flexibilidad de la enzima, ésta proporciona una visión más realista del papel desempeñado por la enzima en el proceso bioquímico. De acuerdo con los resultados obtenidos mediante la FEP, las barreras de energía libre para el paso de la hidroxilación son inferiores a los calculados por medio de la aproximación RRHO, estando los valores obte-

nidos dentro del rango esperado para este tipo de procesos biológicos. El paso de abstracción de hidrógeno tiene lugar a través de un proceso endergónico con una barrera de activación de 13,5 kcal/mol, lo que lleva a la formación de un radical alquilo en el sustrato y a un grupo hidroxilo unido al cofactor hemo. Cabe señalar que el complejo de hierro-hidroxo formado en la especie intermedia se encontró en dos configuraciones electroméricas diferentes para el estado de espín cuadruplete (${}^4Fe^{III}Por^{\cdot+}$ y ${}^4Fe^{IV}Por$) y sólo una para el estado doblete (${}^2Fe^{III}Por^{\cdot+}$). Los dos electrómeros encontrados en el estado cuadruplete comparten la misma geometría, pero difieren en la distribución de los electrones en sus átomos. A partir de estos productos intermedios, tiene lugar una reorientación del grupo hidroxilo que se produce previamente a la etapa de rebote de oxígeno. La energía de este proceso de reorientación depende del estado de espín del sustrato, siendo el estado doblete el más favorecido. De esta manera, la reconsolidación del oxígeno tiene lugar a través de la rotación del grupo hidroxilo y de la formación del nuevo enlace C–O. El proceso acoplado de reorientación-rebote de oxígeno se caracteriza por una barrera de energía libre de activación bastante baja, de aproximadamente 1,5 kcal/mol, que da lugar a la formación de los productos a través de una reacción altamente exergónica. Finalmente, el análisis de la descomposición de la barrera de energía libre (obtenida por cálculos de la FEP) en sus diferentes contribuciones, sugiere que el campo eléctrico producido por la enzima durante la reacción bioquímica no está directamente implicada en la estabilización de los TSs en comparación con las especies reaccionantes. De hecho, hay que destacar que tanto la polarización de la función de onda como el término de Lennard-Jones son los únicos responsables de tal estabilización. Estos hallazgos son consistentes con el tipo de proceso químico y/o sustrato. Por lo tanto, el papel principal de la enzima aromatasasa durante la hidroxilación de la ASD consiste en la estabilización del estado de transición a través de la participación de los efectos de dispersión y polarización.

- El tercer estudio que se presenta tiene un título que se traduce por: **“Uso Conjunto de la Teoría de la Evolución de Enlace y del Método Híbrido QM/MM para la comprensión del Mecanismo de Abstracción de Hidrógeno, a través del Citocromo p450 Aromatasasa”**.

Con el fin de entender mejor el proceso que ocurre en la hidroxilación de la ASD, concretamente en la etapa mecanística correspondiente a la abstracción de hidrógeno, se aplicó una nueva metodología consistente en el uso conjunto de la teoría de la evolución de enlace y del método QM/MM. La BET combina la función de localización electrónica con la teoría de las catástrofes (CT) de René Thom, o dicho de otra forma, el comportamiento topológico del gradiente del campo de la ELF se puede estudiar en el marco proporcionado por la CT. De esta manera, la BET permite encontrar una relación directa entre la topología de la ELF y la evolución de los procesos de ruptura/formación de enlaces químicos y de los reordenamientos

de los pares de electrones, a medida que avanza la reacción química. De acuerdo con esto, podemos ofrecer una representación alternativa de la redistribución de la densidad electrónica para el proceso de abstracción de hidrógeno del sustrato ASD por medio de la enzima aromatasas, que tiene lugar en cualquiera de las dos superficies competitivas correspondientes a los estados de espín: doblete y cuadruplete. De hecho, este es el primer estudio en el que la BET se ha utilizado en un sistema biológico grande, donde se emplean las funciones de onda electrónicas obtenidas a partir de cálculos QM/MM, permitiendo incorporar de este modo el efecto de la polarización producido por las cargas del entorno enzimático.

Por comodidad, únicamente se consideraron para la discusión las poblaciones de las cuencas directamente relacionadas con los reordenamientos de los enlaces $C_{19} - H_{19}$ y $O - H_{19}$. En cuanto a la descripción topológica de la ELF, se ha encontrado que la ruptura del enlace $C_{19} - H_{19}$ y la posterior formación del enlace $O - H_{19}$ tiene lugar a lo largo de cinco dominios de estabilidad estructural de la ELF (ELF-SSDs), que están conectados por sus respectivos puntos de inflexión. Esta respuesta topológica se encontró tanto en la configuración de bajo espín (doblete) como en la de alto espín (cuadruplete). La evolución de la ELF a lo largo del camino de reacción muestra una cierta sincronía en ambas multiplicidades: (i) Los SSD-I y SSD-II se localizaron antes de los respectivos estados de transición; (ii) El TS se localizó en el SSD-III; (iii) Los SSD-IV y SSD-V se encontraron una vez sobrepasado el TS.

El primer dominio de estabilidad, SSD-I, se caracteriza por la presencia de dos cuencas que involucran a los átomos C_{19} , H_{19} y O , que son la cuenca disínáptica $V(C_{19}, H_{19})$ y la cuenca monosínáptica $V(O)$. El primer cambio topológico (catástrofe) que conecta la SSD-I y SSD-II se lleva a cabo a través de una catástrofe de tipo *fold* y se asocia con la creación de la cuenca monosínáptica $V_x(O)$. Cabe destacar que el coste energético asociado con los dos primeros dominios de estabilidad, SSD-I y SSD-II, no comprende únicamente el coste energético asociado con el reordenamiento de la densidad electrónica, el cual anticipa el proceso de rotura del enlace $C_{19} - H_{19}$, sino que también implica la rotación del grupo metilo, la cual permite la orientación más próxima entre los átomos H_{19} y O . Al avanzar en la coordenada de reacción, tiene lugar un segundo cambio topológico que conecta los dominios SSD-II y SSD-III, el cual se representa por una catástrofe de tipo *cusp*. Esta catástrofe está asociada al proceso de rotura del enlace $C_{19} - H_{19}$ y conlleva la escisión de la cuenca disínáptica $V(C_{19}, H_{19})$ para dar lugar a dos cuencas monosínápticas $V(H_{19})$ y $V(C_{19})$. Vale la pena mencionar que el cambio topológico brusco del campo ELF producido por esta catástrofe, no parece afectar a las cuencas cercanas $V(C_{19}, C_{18})$, $V(C_{19}, H_{18})$ y $V(C_{19}, H_{17})$, pertenecientes a la ASD. Dentro del dominio de la SSD-III, el sistema alcanza el estado de transición, y, en este punto, aumenta el carácter radical de los átomos C_{19} y O . Curiosamente,

desde un punto de vista topológico de la ELF, cuando el sistema alcanza el TS no hay evidencia de la formación del enlace $O - H_{19}$. Por tanto, el papel del SSD-III consiste en preparar el sistema para la inminente formación de este enlace. Seguidamente, se localiza un nuevo punto de inflexión que conecta los SSD-III y SSD-IV, el cual corresponde a una catástrofe de tipo *cusp*. En este punto de inflexión, los atractores monosinápticos no enlazantes $V(H_{19})$ y $V_x(O)$ son sustituidos por un único atractor disináptico enlazante $V(O, H_{19})$. Por lo tanto, desde un punto de vista topológico de la ELF, ésta es la primera evidencia de la formación del enlace $O - H_{19}$, el cual tiene lugar una vez que se ha sobrepasado el TS. Posteriormente, durante el progreso del dominio SSD-IV, la población de la cuenca monosináptica $V(C_{19})$ se mantiene prácticamente constante, mientras que la población de la cuenca disináptica $V(C_{19}, H_{18})$ aumenta. Este hecho indica que a lo largo del dominio SSD-IV, una determinada densidad de carga parcial se concentra en la cuenca disináptica $V(C_{19}, H_{18})$ antes que en la cuenca $V(C_{19})$. Por lo tanto, durante el progreso del dominio SSD-IV, el flujo electrónico se dirige hacia el átomo C_{19} , aumentando el carácter radical de este átomo. Finalmente, cuando el sistema alcanza el último punto de inflexión, que conecta los dominios SSD-IV y SSD-V, se localiza una catástrofe de tipo *fold*. En este punto de inflexión se crea un nuevo atractor monosináptico $V_2(C_{19})$, debido a un exceso de la densidad de carga alrededor del átomo de C_{19} . Este exceso se produce por un flujo electrónico permanente hacia este átomo desde la cuenca monosináptica $V_1(C_{19})$, así como desde las cuencas disinápticas $V(C_{19}, H_{17})$ y $V(C_{19}, H_{18})$.

Cabe mencionar también, que en el estado cuádruplete tiene lugar una bifurcación en la PES posterior al estado de transición, la cual conduce hacia dos estados electroméricos diferentes en las especies intermedias. Estos electrómeros, ${}^4Fe^{III}Por^{\bullet+}$ y ${}^4Fe^{IV}Por$, difieren en el estado de oxidación del átomo de hierro y del ligando de porfirina, y se observan en el sistema enzimático, pero no en el modelo de fase de gas. La principal diferencia entre estos electrómeros desde el punto de vista de la BET, es que el dominio SSD-V localizado para el electrómero ${}^4Fe^{IV}Por$ se encuentra más retrasado a lo largo del camino de reacción, que el respectivo dominio localizado para el electrómero ${}^4Fe^{III}Por^{\bullet+}$. Este hecho es consistente con las diferentes estructuras electrónicas presentes en cada uno de los electrómeros. Por otro lado, los resultados obtenidos al realizar el análisis BET en fase gas son similares a los del sistema enzimático. De este modo, en el modelo en fase gas se encontraron el mismo número y tipo de dominios de estabilidad (SSD-I–SSD-V), que se caracterizan por el mismo tipo de catástrofes (dos de tipo *cusp* y dos de tipo *fold*). La única diferencia que existe entre ambos modelos radica en que estas catástrofes aparecen en diferentes valores de la coordenada de reacción, lo que demuestra que el reordenamiento electrónico se produce de forma ligeramente diferente en fase gas y en la enzima. En particular, se puede observar que el último punto de inflexión, que conecta los dominios SSD-IV y SSD-V a través de

una catástrofe de tipo *fold*, se encuentra en valores más bajos de la coordenada de reacción en el modelo de fase de gas que en el sistema enzimático. Estos resultados también muestran que el estado cuadruplete en el modelo de fase gas es más similar al electrómero ${}^4Fe^{III}Por^{\bullet+}$, que al electrómero ${}^4Fe^{IV}Por$. Sin embargo, los dos primeros puntos de inflexión que dan lugar a los dominios de estabilidad desde el SSD-I al SSD-III, están localizados en valores similares de la coordenada de reacción, tanto para el modelo de fase gas como para el sistema enzimático. Dado que estos tres SSD están involucrados en la activación del enlace $C_{19} - H_{19}$, esto sugiere que el proceso electrónico que tiene lugar durante la activación del enlace $C_{19} - H_{19}$, está mínimamente afectado por el campo electrostático generado por la enzima.

- El cuarto estudio que se presenta tiene un título que se traduce por: **“Estudio Teórico de la Hidroxilación del Exemestano Catalizada por la Enzima Aromatasa Humana”**.

La hidroxilación del sustrato EXE a 19-hidroxi-EXE catalizada por la enzima aromatasa se estudió por medio de simulaciones de dinámica molecular y de cálculos híbridos QM/MM. El objetivo de este estudio fue comparar el mecanismo de hidroxilación de este inhibidor suicida de la aromatasa, con la que se produce durante el primer subciclo catalítico del sustrato natural ASD. Con este fin, el mecanismo de abstracción de hidrógeno – rebote de oxígeno que tiene lugar durante la hidroxilación de la ASD, fue también adoptado para el mecanismo de la hidroxilación del EXE. Al igual que en el estudio relativo a la ASD, se utilizó el PDB con código 3EQM como geometría de partida en las simulaciones. El cofactor *hemo B* pentacoordinado existente en el archivo PDB original fue alterado para obtener la especie porfirina de oxo-hierro, Cpd I, al igual que el sustrato ASD alojado en el centro activo de la aromatasa, el cual se modeló para dar lugar al EXE. Cabe señalar que el sustrato EXE se mantuvo en la misma posición en la que se encontró la ASD cristalizada. El resto del proceso de configuración del sistema se llevó a cabo de una manera similar a la realizada en el estudio con el sustrato ASD. La PES se obtuvo aplicando métodos híbridos QM/MM, en los cuales el subsistema QM (123 átomos) fue descrito mediante el funcional de densidad B3LYP, mientras que el subsistema de MM (57.160 átomos) se trató por medio de los campos de fuerza OPLS-AA y TIP3P. De acuerdo con los resultados obtenidos a partir de los análisis realizados sobre la PES, la ΔG de activación para el paso de abstracción de hidrógeno (etapa limitante de la velocidad de la reacción) cae en el rango de 18 a 19 kcal/mol. Las especies intermedias generados en este paso se corresponden con el complejo hierro-hidroxo y la especie alquilo en el sustrato. Al igual que en el estudio con la ASD, el primero se encuentra en dos configuraciones electroméricas diferentes para el estado de espín cuadruplete (${}^4Fe^{III}Por^{\bullet+}$ y ${}^4Fe^{IV}Por$), y sólo una para el estado doblete (${}^2Fe^{III}Por^{\bullet+}$). La ΔG de activación

para el paso de rebote de oxígeno muestra valores de 1,7 kcal/mol para el doblete y de 5,5 kcal/mol para el cuadruplete, que conduce a la formación del producto 19-hidroxi-EXE a través de un proceso altamente exergónico (-42 y -47 kcal/mol para el doblete y el cuadruplete respectivamente). Las energías libres también se calcularon por medio de técnicas de FEP para el paso de abstracción de hidrógeno, proporcionando valores en el rango de 17 a 19 kcal/mol. Curiosamente, cuando se comparan los valores calculados a partir de las PESs (ΔG) con los calculados a través de técnicas de la FEP (ΔF), se puede observar que la inclusión del espacio conformacional de la enzima no tiene un impacto significativo en las barreras de activación. Sorprendentemente, esta tendencia contrasta claramente con la observada en la investigación que incluía al sustrato ASD, donde las barreras de activación calculadas para la abstracción del átomo H_{19} por medio de la FEP, fueron aproximadamente 9 kcal/mol más pequeñas en promedio que las calculados a partir de las PESs. Además, nuestros resultados indicaron que la etapa de hidroxilación es más lenta para el EXE que para el sustrato natural ASD, dado que la barrera para el paso limitante de la velocidad de la reacción es 3,4 kcal/mol más alta para el EXE (16,9 kcal/mol) que para la ASD (13,5 kcal/mol). De hecho, esta observación está respaldada por los valores obtenidos experimentalmente, concretamente cuando se compara con los datos proporcionados por estudios cinéticos de la hidroxilación de ambos sustratos. De acuerdo con estos datos experimentales, la energía libre de activación para el EXE sería de alrededor de 2,4 kcal/mol más alta que para la ASD. Aunque las barreras de activación que obtuvimos en nuestros estudios subestiman en alrededor de 5 kcal/mol los valores experimentales, la diferencia en la barrera de activación que hemos obtenido entre el EXE y la ASD cae en el rango observado experimentalmente. Además, la hidroxilación de el EXE es más exergónica que la hidroxilación de la ASD (4,8 kcal/mol para el doblete y 3,3 kcal/mol para el cuadruplete respectivamente). Esta observación sugiere que el EXE podría actuar como un inhibidor competitivo en el primer subciclo catalítico de la aromatasas humana.

Al comparar las interacciones que se establecen entre los sustratos EXE y ASD con el centro activo de la aromatasas, vemos que la principal diferencia reside en las interacciones entre carbono C_6 sustituido y los aminoácidos circundantes más cercanos a este carbono. En nuestras simulaciones, observamos que el grupo metilideno sustituido en el C_6 del EXE interactúa con los restos Phe-221, Thr-310, Ser-478, Val-369, Val-370, y con una molécula de agua. Curiosamente, también observamos un puente de la naturaleza polar formada por el grupo carbonilo peptídico del Asp-309, la molécula de agua, y el hidrógeno H_γ de la Ser-478, que se encuentra en la proximidad del canal de acceso al centro activo. Este puente de naturaleza polar está alineado con el grupo metilideno y proporcionaría una región de unión de naturaleza electrostática. Además, éste introduciría cierto carácter electrostático a la interacción entre el grupo metilideno y el centro activo. Por otra parte, pudimos

observar la presencia de un conformero del residuo Thr-310, en el cual el resto hidroxilo O_{γ_1} se encontraba entre el sustrato EXE y el grupo *hemo*. La disposición espacial de este residuo también reveló una interacción entre el átomo O_{γ_1} con el hidrógeno del C_{19} -metilo del sustrato EXE, la cual podría asistir la orientación del átomo C_{19} , facilitando de este modo el proceso de abstracción de hidrógeno. Cabe destacar que esta disposición difiere de la que se encontró en los cálculos que llevamos a cabo para el sustrato ASD. La estabilidad de este conformero se verificó mediante la realización de simulaciones de dinámica molecular de 30 ns.

Finalmente, el examen de los diferentes términos en los que se puede descomponer la energía de activación, sugiere que el término electrostático es el único implicado en la estabilización del estado de transición, en comparación con las especies reaccionantes. Este hallazgo obtenido para el sustrato EXE no está en concordancia con los resultados obtenidos previamente para la ASD, donde se observó el comportamiento contrario. En este último caso, tanto el término Lennard-Jones como la polarización de la función de onda son los responsables de dicha estabilización. Por lo tanto, el papel de la actividad catalítica de la aromatasasa durante la hidroxilación del sustrato EXE reside en la estabilización de los TSs por medio del campo eléctrico producido por la enzima, mientras que el papel de esta enzima en la hidroxilación de la ASD consiste en una estabilización de los TSs, a través de los efectos tanto de polarización como de dispersión. Un análisis más detallado reveló que la diferencia en la fuente de la actividad catalítica entre ambos sustratos radica en dos razones: (i) la presencia de un grupo metilideno sustituido en el C_6 del sustrato EXE, que interactúa principalmente con los aminoácidos circundantes de naturaleza hidrófoba, y en parte con un puente polar lábil que proporciona una región de unión de carácter electrostático; y (ii) la diferente orientación del residuo Thr-310 en ambos sustratos. En el caso de la EXE, el O_{γ_1} de este residuo forma un enlace de hidrógeno con un hidrógeno del grupo C_{19} -metilo, lo que podría ajustar la posición de este grupo a lo largo de la reacción, facilitando de ese modo la abstracción de hidrógeno por parte del Cpd I.

Conclusiones

Las principales conclusiones extraídas de las investigaciones realizadas en la presente tesis pueden resumirse de la siguiente forma:

- Se ha estudiado por primera vez la estructura electrónica y la naturaleza de los enlaces del Cpd I desde el punto de vista de la ELF. De acuerdo con este estudio, la única diferencia existente entre los estados de espín doblete y cuadruplete del Cpd I se basa en las diferentes posiciones espaciales de los atractores no enlazantes del átomo de oxígeno. El macrociclo de porfirina del Cpd I se compone de una red deslocalizada de enlaces carbono-carbono y carbono-nitrógeno, ya que no hay

evidencia de dobles enlaces localizados en el plano de la porfirina. Por otro lado, los enlaces $Fe - N$ y $Fe - S$ son enlaces covalentes dativos, aunque las respectivas cuencas de localización se forman por la densidad electrónica de pares solitarios. Finalmente, la naturaleza del enlace $Fe - O$ predicho por el análisis ELF se atribuye principalmente a la interacción electrostática, dado que no se observa ningún atractor enlazante entre los atractores de core del Fe y del S . Por el contrario, los índices de localización obtenidos a partir de la matriz de covarianza para las cuencas de los pares solitarios de hierro y oxígeno indican la presencia de un enlace de desplazamiento de carga (charge-shift).

- La hidroxilación del sustrato ASD a 19-hidroxi-ASD, que tiene lugar durante el primer subciclo catalítico de la enzima aromatasasa, es compatible con el mecanismo de rebote de oxígeno propuesto por Groves y otros en 1976. Por lo tanto, este mecanismo comienza con la abstracción de un átomo de hidrógeno de la ASD por medio de las especies oxidantes, el Cpd I, que tiene lugar a través de un proceso endergónico con una barrera de activación de 13,5 kcal/mol. Este paso químico, que corresponde al paso limitante de la reacción, conduce a la formación de un radical alquilo en el sustrato y un grupo hidroxilo unido al cofactor hemo. Un análisis de los 2H -KIEs calculados para esta etapa muestra valores prácticamente idénticos, no sólo para los dos modelos estudiados (QM fase gas y sistema enzimático), sino también en los dos estados de espín. Ésto indica que el valor del KIE es intrínseco a la reacción y no se ve apenas afectado por el entorno enzimático, ni tampoco por la estructura electrónica del Cpd I. Una vez que se incluye la corrección del efecto túnel, los valores exceden el límite semiclásico denotando así un efecto túnel sustancial. A partir de los productos intermedios de la abstracción de hidrógeno, tiene lugar una reorientación del grupo hidroxilo previa a la etapa de rebote de oxígeno. La energía de este proceso de reorientación depende del estado de espín del sustrato, siendo el estado doblete el más favorecido. De este modo, la reconstitución del oxígeno tiene lugar a través de la rotación de hidroxilo y la formación del nuevo enlace C-O. El proceso de reorientación-rebote acoplado se caracteriza por una barrera de activación libre muy baja, 1,5 kcal/mol, que da lugar a la formación de los productos, 19-hidroxi-ASD, a través de una reacción altamente exergónica. Finalmente, el análisis de la descomposición de la barrera de energía libre en sus diferentes contribuciones, sugiere que el campo eléctrico producido por la enzima durante la reacción bioquímica no está involucrado directamente en la estabilización de los TS en comparación con las especies reaccionantes. De hecho, hay que destacar que tanto la polarización de la función de onda, como el término Lennard-Jones, son los únicos responsables de tal estabilización. Por lo tanto, el papel principal de la enzima aromatasasa durante la hidroxilación de ASD consiste en una estabilización de los TSs a través de la participación de efectos tanto de dispersión como de polarización.

• Hemos introducido un nuevo enfoque para analizar el progreso de una reacción bioquímica basado en el uso conjunto de la BET y del método híbrido QM/MM, a la vez que hemos demostrado su aplicabilidad en las reacciones que implican grandes sistemas biológicos. De acuerdo con esta metodología, la abstracción de un átomo de hidrógeno del sustrato ASD catalizada por la enzima aromatasasa es un proceso no concertado, puesto que los procesos de ruptura del enlace $C-H$ y de formación de enlace $O-H$ no tienen lugar al mismo tiempo. Los eventos químicos proporcionados por la BET permiten caracterizar 5 dominios de estabilidad estructural diferentes para este proceso, que están conectados a través de cuatro catástrofes, dos de ellas de tipo *fold* y las otras dos de tipo *cusp*. Como resultado, el camino de reacción asociado al proceso de abstracción de hidrógeno implica primero una rotación del grupo metilo, seguido por el proceso de ruptura del enlace $C-H$, y, finalmente, la formación del enlace $O-H$. Este último se produce una vez que se ha sobrepasado el estado de transición. Además, se han encontrado dos configuraciones electroméricas diferentes para el estado de alto espín (cuadruplete) en sistema enzimático (no en el modelo de fase gas); sin embargo, estos electrómeros no afectan a la evolución del campo de la ELF a lo largo de la coordenada de reacción. Ambos electrómeros conducen al mismo intermedio radical; sin embargo, el SSD-V aparece ligeramente retrasado para el electrómero $^4Fe^{IV}Por$. Al comparar los resultados proporcionados por la BET en el sistema enzimático y en el modelo de fase de gas, se ha observado que los mismos dominios (SSD-I–SSD-V) están presentes en ambos modelos, los cuales se caracterizan por el mismo tipo de catástrofes. La única diferencia existente entre estos modelos reside en los diferentes valores de la coordenada de reacción, en los cuales aparecen estas catástrofes. Esta diferencia, que no es grande, muestra que el reordenamiento electrónico se produce de forma ligeramente diferente en la enzima y en fase gas, lo que sugiere que la reacción química es casi electrónicamente equivalente en ambos modelos.

• Se ha propuesto que la hidroxilación del inhibidor EXE a 19-hidroxi-EXE catalizada por la enzima aromatasasa, procede a través del mecanismo de rebote de oxígeno, al igual que la hidroxilación de la ASD. De acuerdo con los resultados derivados de los análisis realizados en la PES, la etapa limitante de la velocidad durante la reacción de hidroxilación de EXE también se corresponde con el proceso de abstracción de hidrógeno. Las barreras de activación para este paso revelan que éste transcurre a través de un proceso endergónico con barreras de activación en el intervalo de 16,9 kcal/mol (doblete) a 19,2 kcal/mol (cuadruplete), que son alrededor de entre 3 y 5 kcal/mol más altas que las que se encuentran en el mismo proceso para la ASD. De acuerdo con estos hallazgos, la reacción de hidroxilación para el inhibidor procede a través de un proceso más lento que para el sustrato natural. Estos resultados son consistentes con los estudios experimentales, de los cuales se puede estimar una barrera de activación para la hidroxilación del EXE de 2,4 kcal/mol más alta que para la hidroxilación de la ASD. Por lo tanto, la

hidroxilación del sustrato EXE es compatible con el mecanismo de Groves y otros y se muestran barreras de activación que están en el rango de valores esperados. De este modo, y dado que no se observa una unión irreversible del EXE al centro activo de la enzima en ninguna etapa de la reacción, desestimamos el primer subciclo catalítico de la aromatasa como la etapa catalítica donde se produce la inhibición. Además, el hecho de que los productos de la hidroxilación del EXE muestran energías libres menores que los obtenidos para la ASD, sugeriría una inhibición competitiva en la primer subciclo de la aromatasa humana. Finalmente, el examen de los diferentes términos en los que se puede descomponer la energía de activación, sugiere que el término electrostático es el único implicado en la estabilización del estado de transición, en comparación con las especies reaccionantes. Este hallazgo obtenido para el sustrato EXE contrasta con lo observado para la ASD, donde el comportamiento es el opuesto. De este modo, el papel de la actividad catalítica de la aromatasa durante la hidroxilación del sustrato EXE reside en la estabilización de los TSs por medio del campo eléctrico producido por la enzima. Un análisis más detallado revela que la diferencia en la fuente de la actividad catalítica entre ambos sustratos radica en dos razones: (i) la presencia del grupo metilideno del sustrato EXE, que en su mayoría interactúa con los aminoácidos circundantes de naturaleza hidrófoba, y en parte con un puente polar lábil que proporciona una región de unión de carácter electrostático; y (ii) la diferente orientación del residuo Thr-310 en ambos sustratos. En el caso del EXE, este residuo podría ajustar la posición del grupo C_{19} -metilo a lo largo de la reacción, facilitando de ese modo la abstracción de hidrógeno por parte del Cpd I.

Appendices

Appendices

A. SoftwareUsed	281
B. ScientificProduction	289

Appendix *A*

Software Used



In this appendix, the main programs and software that have been used for the present thesis are shown. Such programs have been classified into those that have been used for scientific computation, those that have been used for the representation and visualization of molecular geometries and, finally, other useful programs that have been utilized for imaging, for drawing chemical structures, and so on. A brief description of each program, as well as the official website where it can be found, is displayed.

Calculation Programs:

- **Gaussian03 (G03)**
- **Gaussian09 (G09)**

Both are versions of the Gaussian program initially released in 1970 by John Pople and his research group at Carnegie-Mellon University. Currently, it is a commercial computational chemistry software application developed, licensed, and supported by *Gaussian Inc.*. The Gaussian program is based on the basic laws of quantum chemistry and provides state-of-the-art capabilities for electronic structure modeling. This program is very useful for predicting energies, molecular structures, vibrational frequencies as well as many other molecular properties. It is also widely used to study molecules and reactions in a broad range of conditions in both gas-phase and condensed media.

<http://www.gaussian.com>

- **fDYNAMO (v. 2.2)**

This is an open source program library written in Fortran 90 and initially developed by Martin Field in the early 1990s. This library of modules has been designed for the simulation of molecular systems using *QM*, *MM* and hybrid *QM/MM* methods and has the following capabilities: Semi-empirical *QC* methods of the *MNDO* type, including *AM1* and *PM3* Hamiltonians, support for the *OPLS-AA* force field, energy calculations, optimization of geometries, searching of transition states, computation of reaction paths, normal mode analysis, and calculation of properties such as charges or dipoles. It is also used to perform Molecular Dynamics simulations or Monte Carlo simulations.

<https://sites.google.com/site/pdynamomodeling/>

- **NAMD Scalable Molecular Dynamics (v. 2.10)**

NAMD (NANoscale Molecular Dynamics program) is an open source molecular dynamics simulation package written using the *Charm++* parallel programming model. Its code has been designed for high-performance simulation of large biomolecular systems (millions of atoms) and is noted for its parallel efficiency. It has been developed by the joint collaboration of the Theoretical and Computational Biophysics Group (TCB) and the Parallel Programming Laboratory (PPL) at the University of Illinois at Urbana-Champaign.

<http://www.charmm.org>

- **PROPKA** (v. 3.0)

PROPKA is a freeware program developed by Hui Li, Andrew D. Robertson and Jan H. Jensen which is available as a python code or via a web interface. This program provides a very fast empirical method to predict the pK_a values of ionizable groups in proteins and protein-ligand complexes based on the 3-D structure.

<http://propka.org>

- **CAMVIB/CAMISO**

Both are open source programs used to compute Kinetic Isotope Effects (KIE) and were developed by Ian H. Williams at the University of Bath. They are based on the Rigid Rotor/Harmonic Oscillator ($RRHO$) approximation and also include the effect of tunneling calculated with either Wigner (KIE_{WIGNER}) or Bell (KIE_{BELL}) models.

<http://www.bath.ac.uk/~chsihw/grace/camiso.html>

<http://www.bath.ac.uk/~chsihw/grace/camvib.html>

- **DGrid** (v. 4.6)

DGrid is a program developed by Miroslav Kohout which is used to compute the density, density gradient, density Laplacian, one-electron potential, and electron localization function (ELF) on an equidistant grid, starting from wave functions generated by quantum chemical packages. Dgrid is open source code and available for Windows and Linux platforms.

<http://www2.cpfs.mpg.de/~kohout/dgrid.html>

- **TopMoD 09**

TopMoD is an open source package of programs written in FORTRAN 90 developed by Bernard Silvi. TopMoD allows ELF and AIM topological analyses to be carried out on a 3-dimensional grid for molecular wave functions expressed in terms of Gaussian primitives. It uses wave functions written in a wfn file, which is available as an output file generated by Gaussian92/94/98/03 and GAMESS *ab initio* software packages.

http://www.lct.jussieu.fr/pagesperso/silvi/topmod_english.html

Molecular Visualization Programs:

- **Molden** (v. 5.0)

MOLDEN is an open source program for the pre- and post-processing of molecular and electronic structures. This package makes it possible to display Molecular Orbitals, Electronic Densities, Molecular Atomic Densities and other properties obtained from the output file generated by *ab Initio* programs such as GAMESS-UK, GAMESS-US, GAUSSIAN or ORCA. It can import and show the coordinates from many different kinds of files such as *XYZ*, *mol* and *PDB* or from the geometries obtained from the *ab Initio* calculations, as well as to show reaction paths and molecular vibrations in movement. Molden also has a powerful Z-matrix editor that allows molecules to be constructed from scratch, including polypeptides.

<http://www.cmbi.ru.nl/molden/>

- **Jmol** (v. 13.0.12)

Jmol is an open-source molecule viewer for students, educators, and researchers in chemistry and biochemistry. It is written in Java and thus it is cross-platform, running on Windows, Mac OS X, and Linux/Unix systems; in addition, the Jmol Applet is a web browser applet that can be integrated into web pages. Jmol is a viewer for chemical structures in *3-D* with features for chemicals, crystals, materials and biomolecules that supports a wide range of molecular file formats, including *PDB*, Crystallographic Information File (*cif*), MDL MolFile (*mol*), and Chemical Markup Language (*CML*).

<http://www.jmol.org>

- **VMD** (v. 1.9.2)

VMD (Visual Molecular Dynamic) is designed for modeling, visualizing, and analyzing biological systems such as proteins, nucleic acids or lipid bilayer assemblies, although it may be used to view more general molecules. VMD can read many kinds of files and display the structure contained with them. In addition, it provides a wide variety of methods for rendering and coloring a molecule and can be used to display, animate and analyze the trajectory that comes from a molecular dynamics (*MD*) simulation.

<http://www.ks.uiuc.edu/Research/vmd/>

- **pyMOL** (v. 1.7.2)

PyMOL is an open-source, user-sponsored, molecular visualization system written in python language. It was initially developed by Warren Lyford although it is currently commercialized by Schrödinger, Inc. PyMOL can produce high-quality 3-D images of small molecules and biological macromolecules, such as proteins.

<https://www.pymol.org>

- **UCSF Chimera** (v. 1.10)

UCSF Chimera is a highly extensible program for the interactive visualization and analysis of molecular structures and related data, including density maps, supramolecular assemblies, sequence alignments, docking results, trajectories, and conformational ensembles. High-quality images and animations can be generated. Chimera is free of charge for academic, government, non-profit, and personal use. Chimera is developed by the Resource for Biocomputing, Visualization, and Informatics, funded by the National Institutes of Health.

<https://www.cgl.ucsf.edu/chimera/>

Other Tools:

- **Gnuplot** (v. 4.6.6)

Gnuplot is a portable command-line driven graphing utility that can generate two- and three-dimensional plots of functions, data, and data fits. The source code is copyrighted but freely distributed and it is frequently used for publication-quality graphics as well as education. The program is multi-platform and can run on several operating systems (GNU/Linux, Unix, Microsoft Windows, Mac OS X, and others). It is a program with a fairly long history, and has been supported and under active development since 1986.

<http://www.gnuplot.info>

- **POV-Ray** (v. 3.7)

The Persistence of Vision Raytracer, or POV-Ray, is a ray-tracing program which generates stunning three-dimensional graphics from a text-based scene description, and is available for a variety of computer platforms. It was originally based on DKBTrace, written by David Kirk Buck and Aaron A. Collins for the Amiga computers. POV-Ray is free and open-source software with the source code available under the AGPLv3.

<http://www.povray.org>

- **ChemDraw Ultra** (v. 12)

Chemdraw is a molecular editor that provides the tools to draw chemical structures and reactions as well as a full set of chemical structure essentials such as rings, bonds, chains, atoms and functional groups. It was first developed in 1985 by David A. Evans and Stewart Rubenstein and commercialized by the cheminformatics company CambridgeSoft; currently it is commercialized by the company PerkinElmer.

<https://www.cambridgesoft.com/software/overview.aspx>

- **Gimp** (v. 2.8)

GIMP (GNU Image Manipulation Program) is a freely distributed program for such tasks as photo retouching, image composition and image authoring. It can be used as a simple paint program, an expert quality photo retouching program, an online batch processing system, a mass production image renderer, an image format converter, etc. GIMP is written and developed under *X11* on UNIX platforms, but the same code also runs on MS Windows and Mac OS X.

<http://www.gimp.org>

- **Libreoffice Draw** (v. 1.10)

Draw is an application within the office suite LibreOffice, which is free and open source software. Draw is a powerful package that allows graphical objects to be manipulated, grouped, and cropped, as well as allowing the use of objects in *3-D* and much more, for producing technical drawings, general posters, etc.

<https://www.libreoffice.org/discover/draw/>

Appendix *B*

Scientific Production



In this appendix, the main scientific production during the research and writing of the thesis is shown. This has been divided into papers published in scientific journals, book chapters, oral presentations at international scientific conferences and poster communications at national and international scientific conferences.

Papers:

- Ignacio Viciano; Slawomir Berski; Sergio Martí; Juan Andrés. *New Insight into the Electronic Structure of Iron(IV)-oxo Porphyrin Compound I. A Quantum Chemical Topological Analysis*. J. Comput. Chem. **2013**, 34, 780–789. DOI: 10.1002/jcc.23201
- Ignacio Viciano; Raquel Castillo; Sergio Martí. *QM/MM Modeling of the Hydroxylation of the Androstenedione Substrate Catalyzed by Cytochrome P450 Aromatase (CYP19A1)*. J. Comput. Chem. **2015**, 36, 1736–1747. DOI: 10.1002/jcc.23967
- Ignacio Viciano; Patricio González-Navarrete; Juan Andrés; Sergio Martí. *Joint Use of Bonding Evolution Theory and QM/MM Hybrid Method for Understanding the Hydrogen Abstraction Mechanism via Cytochrome P450 Aromatase*. J. Chem. Theory Comput. **2015**, 11, 1470–1480. DOI: 10.1021/ct501030q
- Ignacio Viciano; Sergio Martí. *Theoretical Study of the Mechanism of Exemestane Hydroxylation Catalyzed by Human Aromatase Enzyme*. J. Phys. Chem. B **2016**, 120, 3331–3343. DOI: 10.1021/acs.jpcc.6b01014
- Ignacio Viciano; Raquel Castillo; Sergio Martí. *New Insights into the Mechanism of Human Aromatase (CYP19A1): A QM/MM Study of the hitherto Unknown Second Catalytic Subcycle*. **In preparation**.

Book Chapters:

- Ignacio Viciano ; Sergio Martí; Olivier Parisel. *DFT/MM Studies of the Substrates Androstendione and Exemestane in the First Hydroxylation Step of Cytochrome P450 Aromatase*. HPC-Europa2: Science and Supercomputing in Europe research highlights 2011, Publisher: CINECA Consorzio Interuniversitario, Editor: Monfardini, Silvia, pp.31. ISBN 978-88-86037-25-9

Oral Presentations:

- *Estudio Teórico QM/MM de la Hidroxilación de Sustratos Catalizada por el Citocromo P450 Aromatasa (CYP19A1)*. Ignacio Viciano, Raquel Castillo, Sergio Martí, XXXVIII Congress of Theoretical Chemists of Latin Expression (QUITEL 2012), 2-7/12/2012, Natal (Brasil)

- *Bonding Evolution Theory Analysis of the Hydrogen Abstraction Mechanism via Cytochrome P450 Aromatase using QM/MM methods.* Ignacio Viciano, Patricio González-Navarrete, Juan Andrés, Sergio Martí. 6th Theoretical Biophysics International Symposium (Theobio 2013), 24-27/06/2013, Gothenburg (Sweden)

Poster Communications:

- *DFT/MM Studies of the Exemestane as an Aromatase Inhibitor in the Treatment of Female Breast Cancer.* Ignacio Viciano, Raquel Castillo, Sergio Martí, Vicente Moliner. 4th Theoretical Biophysics International Symposium (Theobio' 09), 29/06/2009-3/07/2009, Roscoff (France)
- *DFT/MM and Gas Phase Studies of the Exemestane as an Aromatase Inhibitor.* Ignacio Viciano, Raquel Castillo, Sergio Martí, Vicente Moliner. VII Escola de química teórica de la UIB, 6-10/07/2009, Palma de Mallorca (Spain)
- *DFT/MM Studies of the Cytochrome P450 Aromatase Inhibition by Exemestane in the Treatment of Breast Cancer.* Ignacio Viciano, Raquel Castillo, Sergio Martí, Vicente Moliner. Expanding the Frontiers of Molecular Dynamics Simulations in Biology, 23-25/11/2009, Barcelona (Spain)
- *DFT Study of Compound I and Theoretical Study of Two-State Reactivity of Exemestane in the First Hydroxylation Step of Cytochrome P450 Aromatase.* Ignacio Viciano, Raquel Castillo, Sergio Martí. 7th Congress on Electronic Structure: Principles and Applications (ESPA 2010), 29/06/2010-2/07/2010, Oviedo (Spain)
- *DFT/MM Studies about the First Hydroxylation Step of the Cytochrome P450 Aromatase.* Ignacio Viciano, Raquel Castillo, Sergio Martí. Ninth Triennial Congress of the World Association of Theoretical and Computational Chemists (WATOC 2011), 17-22/07/2011, Santiago de Compostela (Spain)
- *Theoretical Study of Molecular Mechanism of Cytochrome P450 Aromatase. New Insights in the Breast Cancer Treatment.* Ignacio Viciano, Raquel Castillo, Sergio Martí. 8th Congress on Electronic Structure: Principles and Applications (ESPA 2012), 26-29/06/2012, Barcelona (Spain)

- *Biochemical and Biophysical Study of Molecular Mechanism of Cytochrome P450 Aromatase.* Ignacio Viciano, Fatima Lucas, Raquel Castillo, Sergio Martí, Victor Guallar, *Frontiers in Dynamics Simulations of Biological Molecules*, 4-6/11/2013, Barcelona (Spain)
- *Hydrogen Atom Abstraction via Cytochrome P450 Aromatase: Mechanistic Insights from the Analysis of Electron Localization Function and Catastrophe Theory.* Ignacio Viciano, Patricio González-Navarrete, Juan Andrés, Sergio Martí. 9th Congress on Electronic Structure: Principles and Applications (ESPA 2014), 2-4/07/2014, Badajoz (Spain)
- *New Theoretical Insights into the Mechanism of Human Aromatase (CYP19A1). Exploration of its Hitherto Unknown Second Catalytic Subcycle.* Ignacio Viciano, Raquel Castillo, Sergio Martí. **Accepted for POSTER presentation at the conference:** 10th Congress on Electronic Structure: Principles and Applications (ESPA 2016), 28/06/2016-1/07/2016, Castellón de la Plana (Spain)

$$\hat{H}_{\text{int}} \approx \sum_{\text{pairs}} K_{ij} (r-r_{ij})^2 + \sum_{\text{angles}} K_{ij} (\theta-\theta_0)^2 + \sum_{\text{dihedrals}} K_{ij} [1 + \cos(n\phi-\delta)] + \sum_{i,j} f_{ij} \left[\frac{q_i q_j}{4\pi\epsilon_0\epsilon r_{ij}} + 4\epsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^6 \right] \right]$$

$$\hat{H}_{\text{non-bond}} = \sum_{M \neq N} \frac{q_M q_N}{R_{MN}} + \sum_{\text{O-M}} \frac{Z_{\text{O}} q_{\text{O}}}{R_{\text{O-M}}} + \sum_{\text{A-M}} \frac{A_{\text{A-M}}}{R_{\text{A-M}}} + \sum_{\text{B-M}} \frac{B_{\text{B-M}}}{R_{\text{B-M}}}$$

$$\hat{H}_{\text{rot}} = \sum_{\alpha} \frac{1}{2} I_{\alpha} \omega_{\alpha}^2 + \sum_{\alpha} \sum_{\beta} \frac{1}{2} I_{\alpha\beta} \omega_{\alpha} \omega_{\beta} + \sum_{\alpha} \sum_{\beta} \sum_{\gamma} \frac{1}{6} I_{\alpha\beta\gamma} \omega_{\alpha} \omega_{\beta} \omega_{\gamma} + \sum_{\alpha} \sum_{\beta} \sum_{\gamma} \frac{1}{24} I_{\alpha\beta\gamma\delta} \omega_{\alpha} \omega_{\beta} \omega_{\gamma} \omega_{\delta} + \sum_{\alpha} \sum_{\beta} \sum_{\gamma} \sum_{\delta} \frac{1}{24} I_{\alpha\beta\gamma\delta} \omega_{\alpha} \omega_{\beta} \omega_{\gamma} \omega_{\delta}$$

$$\hat{H}_{\text{eff}} = \hat{H}_{\text{QM}} + \hat{H}_{\text{MM}} + \hat{H}_{\text{QM/MM}} + \hat{H}_{\text{BC}}$$

