

UNIVERSITÀ DEGLI STUDI DI BOLOGNA
DIPARTIMENTO DI CHIMICA ORGANICA “A. MANGINI”

COMPUTATIONAL MODELS IN ORGANIC AND BIO-ORGANIC CHEMISTRY

AUTHOR: MARCO STENTA
SUPERVISOR: PROF. D. SPINELLI
SUPERVISOR: PROF. A. BOTTONI

Dottorato di Ricerca in Scienze Farmaceutiche (XX Ciclo)
Settore Disciplinare: CHIM/06
Coordinatore: Chiar.mo Prof. Maurizio Recanatini
Relatore: Chiar.mo Prof. Domenico Spinelli
Correlatore: Chiar.mo Prof. Andrea Bottoni

Declaration

The work described in this dissertation was carried out at the Dipartimento di Chimica Organica “A. Mangini” and the Dipartimento di Chimica “G. Ciamician” of the Università degli Studi di Bologna between January 2005 and December 2007. The present thesis is the results of the original works carried out by the candidate under the supervision of Professor A. Bottoni, for the computational part, and Professor D. Spinelli, as a coordinator of the collaboration with some experimentalist groups. A profitable and tight teamwork with Dr. P. Altoè and with Dr. M. Garavelli led to the development of the new computational methods presented here. Dr. F. Benfatti must also be acknowledged as a needful *trait d'union* with Professor G. Cardillo experimentalist group. Some of work described in this dissertation is the outcome of a collaboration with Dr. M. Calvaresi and Dr. G. Miscione. Some of the results presented here have been published or submitted to journals for the referee procedure, as stated in the Publications Section.

Marco Stenta
Bologna, December 2006

Introduction to the Ph.D. dissertation

In this thesis the outcome of a Medicinal Chemistry Ph.D. course ranging from Physical to Bio-Organic Chemistry, is reported. A multidisciplinary approach allowed the development of several fruitful research projects and led to many innovative results in different fields. The present dissertation focuses on the description of a newly developed computational method. The most important feature of this method, founded on well tested models and theories and on accurately verified approximations, is its reliability and the capability of handling very different problems. This thesis mainly deals with the original and unpublished aspects of the work carried out during the Ph.D. period and only short notes and comments (essential to understand the present work) concerning the models developed by others researchers are given. An exhaustive list of references is provided where more detailed descriptions of these previous work can be easily found) Commercially available packages for molecular computations have been used to study a number of chemically relevant problems and to achieve a better understanding of the inborn approximations of models and theories.

The thesis is composed by three main sections. In the first section a brief overview on molecular modelling is given (Chapter 1). Also, some hints concerning the most common computational methods are provided (Chapter 2). Then, the new computational approach developed by the author is described in details (Chapter 3) and some test examples are presented (Chapter 4) to illustrate the capabilities of the new method and show its fields of applicability. In the second section are presented the results obtained by applying the developed method in the study of the reaction mechanism of two interesting enzymes, the proline racemase (Chapter 5) and the diaminopimelate epimerase (Chapter 6). The mechanistic details of the enzymatic catalysis of both these enzymes are presented and the role of some key-residues is stressed. The third section (Chapter 7-10) contains the results of some combined experimental-computational studies, where a satisfactory rationalisation of chemical processes is given by a skilful combination of experimental data and accurate computational models. In the same section is reported (Chapter 11) a full QM mechanistic study on the reaction mechanism of the enzyme Human Carbonic Anhydrase (HCAII).

Dr. Piero Altoè must be acknowledged for the time and the knowledge invested in the common research project aimed at the development and testing phase of this computational method. Dr. M. Calvaresi also participated to the study of enzymes with a precious contribute to the set-up phase and to the construction of physically reliable model systems of these amazingly overcomplicated macro-molecules. Dr G. Miscione must be acknowledged for his essential contribution the QM study of the HCA II enzyme.

Contents

DECLARATION	2
INTRODUCTION TO THE PH.D. DISSERTATION	3
CONTENTS	4
LIST OF ABBREVIATIONS	7
PART I: COMPUTATIONAL METHODS	9
1 AN OVERVIEW ON MOLECULAR MODELLING.....	11
1.1 INTRODUCTION	11
1.2 MODELLING A MOLECULAR SYSTEM	11
1.3 THE POTENTIAL ENERGY SURFACE (PES).....	12
1.3.1 PES description.....	13
1.3.2 PES exploration: geometry optimisation.....	16
1.3.3 PES exploration: the NEB approach.....	18
1.3.4 Molecular vibrations.....	18
1.4 MOLECULAR DYNAMICS	20
2 COMPUTING THE ENERGY OF A MOLECULAR SYSTEM.....	23
2.1 INTRODUCTION	23
2.2 MOLECULAR MECHANICS APPROACH.....	23
2.3 QUANTUM MECHANICS APPROACH	25
2.3.1 Hartree-Fock (HF).....	26
2.3.2 Møller-Plesset perturbation theory (MP).....	29
2.3.3 DFT.....	30
2.3.4 DFTB.....	32
2.3.5 The Resolution of Identity approximation	33
3 COBRAMM	34
3.1 INTRODUCTION	34
3.2 THE QM/MM PARTITIONING SCHEME.....	36
3.3 THE BOUNDARY REGION	39
3.4 THE ATOM-LINK POSITION.....	40
3.5 HANDLING QM/MM NON-BONDING CROSS-TERMS	41
3.6 QM/MM FORMALISM.	42
3.7 IMPLEMENTATION.....	43
3.8 FREQUENCY CALCULATION WITH A GENERAL QM/MM HYBRID POTENTIAL	45
3.9 FINGERPRINT ANALYSIS	45
3.10 THE “SET-UP” PROBLEM: A SHORT DISCUSSION.....	52
4 TEST EXAMPLES.....	54
4.1 INTRODUCTION	54
4.2 OPTIMISATION OF A COMPLEX MOLECULAR STRUCTURE: A FUMARAMIDE BASED ROTAXANE	55
4.3 THE ANOMERIC EQUILIBRIUM OF D-GLUCOPYRANOSE.....	56
4.3.1 The L level optimisation.	58
4.3.2 The HL level optimisation.	58
4.3.3 MM Molecular Dynamics: the ML level	59
4.3.4 QM/MM Molecular Dynamics: the HML level	60
4.4 RHODOPSIN AND GFP CHROMOPHORES: SOLVENT AND PROTEIN	61
4.4.1 Structure and optical properties for solvated HBDI and GFP: flexible QM/MM optimisations	62
4.4.2 Solvation by HL and HML QM/MM computations	62
4.4.3 GFP by HML QM/MM computations	64
4.4.4 11-cis retinal PSB in solution.....	65
4.4.5 Solvated ion pairs: structure and properties.....	67
4.5 COMPARISON BETWEEN QM AND QM/MM RESULTS FOR AN ENZYMATIC REACTION MECHANISM.....	68
4.5.1 QM reaction profile.....	68
4.5.2 QM/MM reaction profile.....	70
4.6 STRUCTURAL INVESTIGATION OF CHELATED OCTAHEDRAL COMPLEXES OF FE AND ZN.....	72

4.6.1 Computational methods	76
4.6.2 Results and discussion	76
PART II: USING COBRAMM TO STUDY BIO-ORGANIC SYSTEMS.....	83
5 THE ENZYME PROLINE RACEMASE	85
5.1 INTRODUCTION	85
5.2 SET-UP OF THE BIO-MOLECULAR SYSTEM	86
5.3 PARTITIONING THE SYSTEM INTO LAYERS.....	87
5.4 PRELIMINARY STUDIES TO TEST THE SET-UP PROCEDURE	90
5.5 EXPLORING OF THE PES WITH A QM/MM POTENTIAL.....	93
5.6 UNDERSTANDING THE ENZYMATIC CATALYSIS.....	96
5.6.1 DFP analysis applied to M1a and TSCa.....	96
5.6.2 DFP analysis applied to M1a and M2a.....	101
5.7 CONCLUSIONS	106
6 THE ENZYME DIAMINOPIMELATE EPIMERASE.....	108
6.1 INTRODUCTION	108
6.2 SETTING-UP THE SYSTEM	109
6.2.1 Cluster Analysis	110
6.2.2 Simulated Annealing	111
6.2.3 Docking.....	111
6.3 THE STRUCTURE OF THE COMPLEX	111
6.4 THE MODEL1 MODEL SYSTEM	112
6.5 THE MODEL2 MODEL SYSTEM	123
6.6 CONCLUSIONS	130
PART III: USING STANDARD COMPUTATIONAL APPROACHES TO STUDY	
ORGANIC AND BIO-ORGANIC SYSTEMS.....	131
7 INVESTIGATIONS ON THE ELECTROCHEMICAL BEHAVIOUR OF SOME ORTHO-	
SUBSTITUTED (ARYL)(2-NITROBENZO[B]THIOPHEN-3-YL)AMINES.....	133
7.1 INTRODUCTION	133
7.2 ELECTROCHEMICAL BEHAVIOUR OF 1AR ₂ O	133
7.3 COMPUTATIONAL INVESTIGATION OF THE BEHAVIOUR OF 1AR ₂ O(1)	136
7.4 CONCLUSIONS	145
8 UNPRECEDENTED 1,2-MIGRATION OF A NITRO GROUP WITHIN A	
α-ARYL-β-NITROETHENYL MOIETY	146
8.1 INTRODUCTION	146
8.2 EXPERIMENTAL SECTION	146
8.3 COMPUTATIONAL SECTION.....	149
9 RATIONALISATION OF COMPLEX MOLECULAR REARRANGEMENT	151
9.1 INTRODUCTION	151
9.2 DISCUSSION	152
10 THE CYCLOADDITION REACTION BETWEEN VINYLKETENES AND IMINES.....	156
10.1 INTRODUCTION	156
10.2 EXPERIMENTAL SECTION	157
10.2.1 Reaction of linear α -bromo vinylketenes with imines.....	157
10.2.2 Reaction of γ,γ -disubstituted- α -bromo vinylketenes with imines	160
10.2.3 Reaction of α -bromo vinylketenes with ketimines.....	161
10.3 COMPUTATIONAL SECTION.....	162
10.3.1 Choice of the model system and description of the computational method.....	162
10.4 RESULTS AND DISCUSSION	164
10.5 CONCLUSIONS	174
11 NEW COMPUTATIONAL EVIDENCE FOR THE CATALYTIC MECHANISM OF CARBONIC	
ANHYDRASE.....	176
11.1 INTRODUCTION	176
11.2 COMPUTATIONAL DETAILS AND CHOICE OF THE MODEL	180

11.3 RESULTS AND DISCUSSION	182
11.3.1 <i>The starting complex, the nucleophilic attack and the origin of the barrier</i>	182
11.3.2 <i>The rearrangement of the zinc-bicarbonate complex</i>	187
11.3.3 <i>The attack of the water molecule and the formation of a penta-coordinated zinc complex.</i>	190
11.3.4 <i>The effect of the protein environment.</i>	193
11.4 CONCLUSIONS.....	194
CONCLUSIONS.....	196
APPENDICES.....	197
(1) HANDLING THE CHARGES IN THE “REAL” SYSTEM.....	197
REFERENCES	199
LIST OF FIGURES.....	211
LIST OF SCHEMES	213
LIST OF TABLES.....	214
PUBLICATIONS.....	216

List of Abbreviations

AFF	Amber Force Field
AMBER	Assisted Model Building with Energy Refinement
AO	Atomic Orbital
BFGS	Broyden-Fletcher-Goldfarb-Shanno minimisation algorithm
bipy	2,2-bypiridine
BO	Born-Oppenheimer approximation
COBRAM	Main core of the COBRAMM suite of programs (COmputational BRidge from Ab-initio and Molecular mechanics) (Computations OBTained by Running Ab-initio and Molecular mechanics) (Complete OBServation of the Reality with Ab-initio and Molecular mechanics)
COBRAMM	Suite of programs for computational chemistry (COmputations at Bologna Relating Ab-initio and Molecular mechanics Methods)
DAP	Diaminopimelate
DFP	Direct fingerprint analysis method (for evaluation of electrostatic contributions)
DFT	Density Functional Theory
D-NEB	Double-ended Nudged Elastic Band
Freq	Frequency calculation
GAFF	Generalised Amber Force Field
<i>h</i>	High layer of a system treated with COBRAMM (QM potential and accurate OPT/MD algorithm)
H	High level COBRAMM calculation type (pure QM system)
HF	Hartree-Fock
HL	High-Medium level COBRAMM calculation type (hibrid QM/MM system)
HM	High-Medium level COBRAMM calculation type (hibrid QM/MM system)
HML	High-Medium-Low level COBRAMM calculation type (hibrid QM/MM system)
<i>l</i>	Low layer of a system treated with COBRAMM (MM potential and fast OPT algorithm)
L-BFGS	Limited memory BFGS minimisation algorithm
LCAO	Linear Combination of Atomic Orbitals
M	High-Medium level COBRAMM calculation type (pure MM system)
<i>m</i>	Medium layer of a system treated with COBRAMM (MM potential and accurate OPT/MD algorithm)
MD	Molecular Dynamics method
ML	High-Medium level COBRAMM calculation type (pure MM system)
MM	Molecular Mechanics based method

MO	Molecular Orbital
MP2	Møller-Plesset second order perturbation theory
NEB	Nudged Elastic Band
ONIOM	our Own N-layered Integrated molecular Orbital and molecular Mechanics
OPT	Optimisation (or minimisation)
PCM	Polarizable Continuum Model
PDB	Protein Data Bank
PES	Potential Energy Surface
phen	1,10-phenanthroline
QM	Quantum Mechanics based method
QM/MM	Quantum Mechanics / Molecular Mechanics based method (hybrid)
RFP	Reverse fingerprint analysis method (for evaluating electrostatic contributions)
RMSD	Root Mean Square Deviation
SCC-DFTB	Self-Consistent-Charge Density-Functional Tight-Binding
SCF	Self Consistent Field
TcPRAC	<i>Trypanosoma Cruzi</i> Proline Racemase
TS	Transition State
vdWFP	van der Waals fingerprint analysis method

Part I:

Computational Methods

1 An Overview on Molecular Modelling

1.1 Introduction

“Molecular Modelling used to be restricted to a small number of scientists who had access to the necessary computer hardware and software”. From its beginnings, as described by A. Leach in the preface of his book,¹ the discipline evolved with unpredictable velocity. The reliability of the obtained results strongly improved throughout the last decades. During this period Theoretical Chemists developed new strategies to describe the reality and Computational Chemists were able to implement and test their models. Nowadays many Experimental Chemists, working either in organic or physical chemistry, can easily take advantage of modern commercial software for both research and teaching purposes. Many books have been published “to provide an introduction to some of the techniques used in molecular modelling and computational chemistry, and to illustrate how these technique can be used to study physical, chemical and biological phenomena” (A. Leach).¹

1.2 Modelling a molecular system

The systems that can be considered in molecular modelling¹⁻³ range from small isolated molecules to biological macromolecules (like proteins and DNA) and solids. Most molecular modelling studies involves three stages. The first one requires the choice of the theoretical approach suitable to describe the system under examination. This choice reflects the nature of the system as well as its dimension because sometimes one needs to sacrifice some accuracy to study a larger system or to obtain the results faster. The two most common computational approaches are quantum mechanics (QM) and molecular mechanics (MM), the first being based on an approximate solution of the Schrödinger equations and the second on a classical description of the atoms (Newton’s laws). Both computational approaches provide the total energy for a given atomic configuration. Thus it is possible to determine the energy change as a function of atomic and molecular motions (Potential Energy Surface: PES). The QM and MM approaches have been implemented here using different levels of theory and different algorithms and in Chapter 2 a short discussion on the most common ones will be given. The

second stage of a molecular modelling study is the calculation itself, namely the computational procedure chosen to obtain the best geometrical arrangement (the one with the lowest energy), the reaction pathway, the behaviour of systems as a function of time, or the value of various observables useful to rationalize, or even predict, experimental data. A description of these computational procedures (i.e. geometry optimisation, scan of potential energy surfaces, molecular dynamics, spectra prediction, etc...) are described in the following sections of this chapter. The third stage in a molecular modelling investigation can be the accurate analysis of the results and the construction of a semi-quantitative or qualitative interpretative model. This stage is not obvious and not always accomplished.

1.3 The potential energy surface (PES)

A common assumption for both quantum mechanics (QM) and molecular mechanics (MM) methods is the Born-Oppenheimer (B.O.) approximation.^{3,4} This basic approximation enables electronic and nuclear motions to be separated. Being the ratio electron-mass/proton-mass about 1/1836, the electronic motion is much faster than the nuclear motion. Consequently, the electrons of a molecule can rapidly adjust to any change of the nuclear positions and the energy of a molecule in its ground state can be considered as function of the nuclear coordinates only. The form of this function (and its accuracy) is determined by the particular method (QM and MM) used (see Chapter 2). In particular in QM calculations¹ the energy of electrons is computed for a fixed nuclear configuration and the nuclear term is treated as an additive factor to the electronic energy. On the contrary, in MM calculations³ one considers explicitly the motions of the nuclei and the electronic contribution is indirectly taken into account by the potential describing the nuclear interactions (See Chapter 2). In both cases, it is possible to obtain an energy value for each nuclear configuration and one can build a diagram of the energy as a function of the nuclear coordinates. These diagrams (Figure 1.1-a), often referred to as Potential Energy Surfaces (PES), are of paramount importance in describing molecular structure and reactivity.

The first step in drawing a PES is the choice of an appropriate coordinate system to describe the configurations of the nuclei. Even if the choice is, in principle, arbitrary only in a few cases it can be useful to adopt a coordinate system different from the Cartesian or Internal ones (two other coordinate systems rarely used are Spherical and Cylindrical coordinates). In the Cartesian coordinate system each atom is described by 3 coordinates that specify its position with respect to an arbitrary point (origin of the coordinate system). Thus, a molecular system of N nuclei is described by a set of $3N$ coordinates. Since 3 coordinates must describe the translational motions and 3 the rotational motion of the whole system, the relative

positions of N nuclei can be determined by $3N-6$ internal coordinates (or $3N-5$ if the system, being linear lacks a rotational degree of freedom). The Internal coordinate system can be obtained by choosing $3N-6$ linearly independent coordinates that coincide with bond lengths, angles (plane angles) and dihedral angles (solid angles) between atoms.^{1,3} Except for diatomic molecules, possessing only one ($[3*2]-5=1$) internal coordinates, the PES of the most common systems are rather complicated and correspond to many-dimensional surfaces (hypersurfaces). Carbonic dioxide (CO_2), a triatomic linear molecule, can be described by four ($[3*3]-5=4$) internal coordinates, and the related PES can not be visualised as in an unique graphical representation. When the number of internal coordinates is higher than 2, the PES can not be represented in a single three-dimensional graphic. It is possible to have partial representations of the PES constituted by bi- or three-dimensional (the energy plus one or two coordinates) sections (cross-sections) of the whole surface (in Figure 1.1-a and Figure 1.1-b a three- and a bi-dimensional cross-section of a multi-dimensional surface are represented). Ozone (O_3), a triatomic non linear molecule, can be described by three ($[3*3]-6=3$) internal coordinates. To visualize the corresponding PES the energy can be considered as a function, for instance, of the OOO angle or, alternatively, of the O-O bond distance etc.

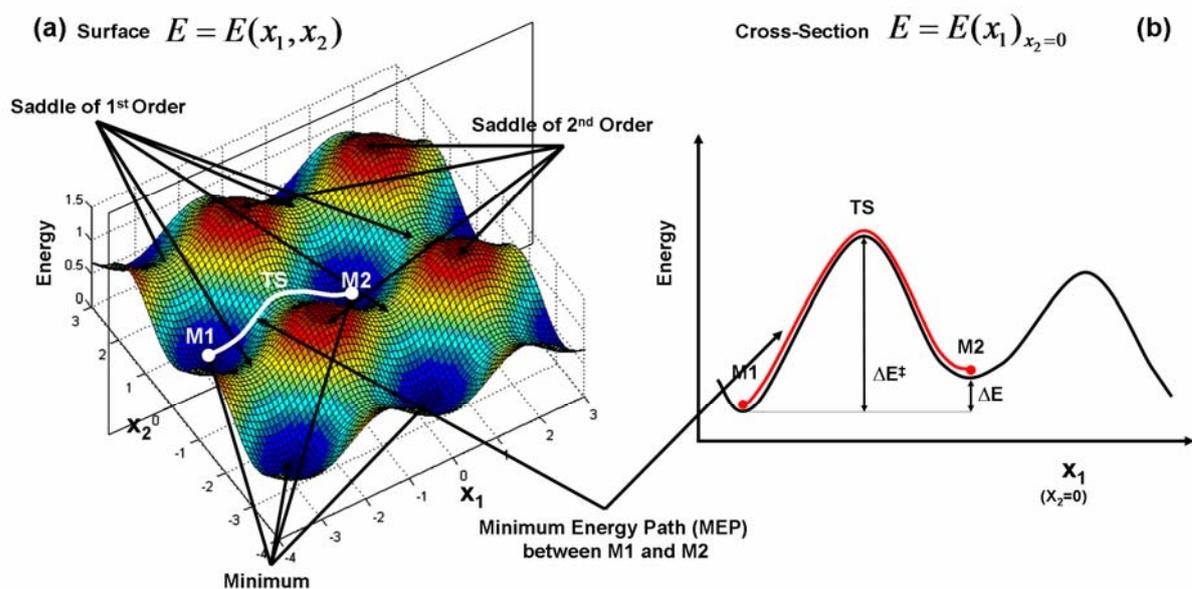


Figure 1.1 a) Graphical representation of a three-dimensional energy surface; b) bi-dimensional cross-section. The MEP (Minimum Energy Path) connecting the two critical points M1 and M2 is showed.

1.3.1 PES description

The PES (or a cross-section) can be represented as a diagram where each point corresponds to an unique arrangement of coordinates and, thus, to a particular geometrical structure of the system. Only a few between these structures are interesting and to determine the position of the corresponding points on the PES is usually a challenging task. The PES can be used to

study either the structural features of a molecular system or its reactivity. If we consider the energy of the system as a function (Equation 1.1) of n variables (for example the $3N-6$ internal coordinates), the interesting points (called critical points) are characterized by a null first derivative of the energy with respect to all the n variables. Only a few of these points have a chemical significance. In particular the points with null first derivatives and all positive second derivatives can be either one of the many local minima or the unique global minimum of the PES. It can correspond to one meta-stable nuclear configuration or to the most stable configuration, respectively. Points with null first derivatives and negative second derivatives with respect to k variables are denoted as saddle point of index k . They have a chemical sense only if $k=1$ and in this case they correspond to transition structures. An elementary reaction step is described as a transition from one equilibrium state (minimum) to a neighbouring one *via* a single transition state. The reaction mechanism is given by the sequence of steps involved in a chemical process and corresponds on the PES, to the Minimum Energy Path (MEP) connecting the two minima that represent reactants and products, respectively (see Figure 1.1).

Equation 1.1 $E = E(\mathbf{x})$

The localization of the critical points is the main target in exploring the PES in both structural and reactivity studies. Should the exact (analytical) shape of the PES be known, the stationary points would be (rather) easily obtained by applying the rules of mathematical analysis. However, the analytical expression of the PES is usually unknown and is rather different for each system. Thus, to locate a point on the unknown PES it is necessary to use an approximate representation of the surface itself. If we consider a bi-dimensional surface where the energy is a function of one variable x (Equation 1.2), we can represent this function as a Taylor series around a critical point x_0 according to Equation 1.3. This expression contains an infinite number of terms which can be reduced to the first two, by truncating the series after the quadratic term (quadratic approximation). Moreover the first term must be zero because the first derivative at the critical point x_0 is null; so, by putting $\Delta\mathbf{x}=\mathbf{x}-\mathbf{x}_0$ we obtain Equation 1.4. If we generalize Equation 1.3 to a system of n coordinates (Equation 1.5) we obtain the general Taylor expansion (Equation 1.6), which can be simplified to Equation 1.7. A most compact representation (Equation 1.8) can be obtained by means of matrix formalism where \mathbf{G} and $\Delta\mathbf{x}$ are used to denote the second derivative (or Hessian) matrix and the displacement vector, respectively (Equation 1.9).

Equation 1.2 $E = E(x)$

Equation 1.3 $E = E^0 + \left(\frac{\partial E}{\partial x}\right)_0 (x - x_0) + \frac{1}{2} \left(\frac{\partial^2 E}{\partial x^2}\right)_0 (x - x_0)^2 + \frac{1}{6} \left(\frac{\partial^3 E}{\partial x^3}\right)_0 (x - x_0)^3 + \dots$

Equation 1.4 $E = E^0 + \frac{1}{2} \left(\frac{\partial^2 E}{\partial x^2}\right)_0 \Delta x^2$

Equation 1.5 $E = E(x_1, \dots, x_n)$

Equation 1.6 $E = E^0 + \sum_i \left(\frac{\partial E}{\partial x_i}\right)_0 (x - x_0) + \frac{1}{2} \sum_{i,j} \left(\frac{\partial^2 E}{\partial x_i \partial x_j}\right)_0 \Delta x_i \Delta x_j +$

$+ \frac{1}{6} \sum_{i,j,k} \left(\frac{\partial^3 E}{\partial x_i \partial x_j \partial x_k}\right)_0 \Delta x_i \Delta x_j \Delta x_k + \dots$

Equation 1.7 $E = E^0 + \frac{1}{2} \sum_{i,j} \left(\frac{\partial^2 E}{\partial x_i \partial x_j}\right)_0 \Delta x_i \Delta x_j$

Equation 1.8
$$E = E^0 + \frac{1}{2} (\Delta x_1 \quad \dots \quad \Delta x_n) \begin{pmatrix} \left(\frac{\partial^2 E}{\partial x_1^2}\right)_0 & \dots & \left(\frac{\partial^2 E}{\partial x_1^2 \partial x_n^2}\right)_0 \\ \vdots & \ddots & \vdots \\ \left(\frac{\partial^2 E}{\partial x_n^2 \partial x_1^2}\right)_0 & \dots & \left(\frac{\partial^2 E}{\partial x_n^2}\right)_0 \end{pmatrix} \begin{pmatrix} \Delta x_1 \\ \vdots \\ \Delta x_n \end{pmatrix} = E^0 + \frac{1}{2} \Delta \mathbf{x}^T \mathbf{G} \Delta \mathbf{x}$$

Equation 1.9
$$\mathbf{G} = \begin{pmatrix} \left(\frac{\partial^2 E}{\partial x_1^2}\right)_0 & \dots & \left(\frac{\partial^2 E}{\partial x_1^2 \partial x_n^2}\right)_0 \\ \vdots & \ddots & \vdots \\ \left(\frac{\partial^2 E}{\partial x_n^2 \partial x_1^2}\right)_0 & \dots & \left(\frac{\partial^2 E}{\partial x_n^2}\right)_0 \end{pmatrix}; \Delta \mathbf{x} = \begin{pmatrix} \Delta x_1 \\ \vdots \\ \Delta x_n \end{pmatrix}$$

The Hessian matrix provides both chemical and topological information. It provides an estimate of the coupling between the coordinates adopted to describe the system. The Hessian matrix \mathbf{G} can also give information concerning the nature of the various critical points of a surface. However, to obtain this type of information we must carry out a diagonalization of the matrix \mathbf{G} , that is to transform it into a new different matrix where only diagonal elements are different from zero (diagonal matrix \mathbf{H}). A matrix \mathbf{U} , satisfying Equation 1.10, is needed to diagonalize \mathbf{G} , which is a real symmetric matrix. \mathbf{U} is the matrix of *eigenvectors* and \mathbf{U}^{-1} its inverse (which coincides in this case with its transpose \mathbf{U}^T). This transformation, which is

equivalent to a change of the basis set used to represent the matrix, does not determine any loss of information. The set of coordinates that makes diagonal the Hessian matrix is usually referred to as “normal coordinates” The matrix \mathbf{H} (Equation 1.11) is a diagonal $n \times n$ matrix formed by n *eigenvalues*. If all *eigenvalues* of the Hessian, computed in a given critical point on the PES, are positive, then the point is a minimum of the surface and the corresponding structure describes a chemical species (reactants, products or intermediate). If only one *eigenvalue* is negative, the point is a saddle point of index 1 and corresponds to a transition state. When k ($k > 1$) negative *eigenvalues* are detected the point is a saddle point of index k and, as stated above, the corresponding structure is not of chemical interest.

Equation 1.10 $\mathbf{U}^{-1}\mathbf{A}\mathbf{U} = \mathbf{H}$

Equation 1.11 $\mathbf{H} = \begin{pmatrix} h_1 & 0 & \dots & 0 & 0 \\ 0 & h_2 & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & h_{n-1} & 0 \\ 0 & 0 & \dots & 0 & h_n \end{pmatrix}$

1.3.2 PES exploration: geometry optimisation

Since the PES complexity rapidly increases with the increasing number of coordinates, a crucial point is the search and location of the various critical points. Efficient algorithms for locating equilibrium and transition structures are now available in modern molecular software. These algorithms are based on the calculation of the first (gradient) and second derivatives (Hessian matrix) and allow to perform a simultaneous optimisation of the whole set of coordinates. The problem consists in finding a critical points (and the corresponding geometry), given a structure represented by a point on the PES potentially far away from the goal. In general the search algorithm is iterative and the geometry is gradually modified till the wanted critical point is obtained. The fundamental equations to compute the coordinate variations at each step of the search procedure can be derived assuming a quadratic shape of the PES. This can be done after development of a Taylor expansion (see Equation 1.1) on a generic point of the surface x_0 up to second order. For the one-dimensional case $E = E(x)$ (Equation 1.12) can be used to locate a critical point on the surface.

Equation 1.12 $E = E^0 + \left(\frac{\partial E}{\partial x}\right)_0 + \frac{1}{2}\left(\frac{\partial^2 E}{\partial x^2}\right)_0 \Delta x^2$

$\Delta x = x - x_0$ is a displacement coordinate and E_0 is the energy value at the point x_0 . Since in a critical point the first energy derivative must be zero, so we can use Equation 1.13 (where the first and second derivative are indicated, respectively with g and G) to determine the critical point x . Thus we obtain Equation 1.14 and Equation 1.15 (Newton-Raphson equations) for one variable case. The extension to a n variable problem can be easily given using a matrix formulation (Equation 1.16). Should the surface a real quadratic surface, the wanted critical point would be obtained in a single step. However, since in most cases the surface is far from being quadratic, a sequence of motions (sometimes several) on the surface (optimization step) is usually required to locate the critical point. Thus, the most correct and general form of the Newton-Raphson equations is represented by Equation 1.17, where $\mathbf{x}^{(i+1)}$ is the new position on the surface as computed from the previous one at the i^{th} displacement.

$$\text{Equation 1.13} \quad \frac{\partial E}{\partial \Delta x} = g + G = 0$$

$$\text{Equation 1.14} \quad \Delta x = x - x_0 = -\frac{g}{G}$$

$$\text{Equation 1.15} \quad x = x_0 - \frac{g}{G}$$

$$\text{Equation 1.16} \quad \mathbf{x} = \mathbf{x}_0 - \mathbf{G}^{-1} \mathbf{g}$$

$$\text{Equation 1.17} \quad \mathbf{x}^{(i+1)} = \mathbf{x}^i - \mathbf{G}^{-1} \mathbf{g}$$

Being the calculation of the Hessian matrix computationally expensive, approximate forms of the Newton-Raphson equations (involving approximate Hessian matrices) are usually employed. These methods are often referred to as “*quasi-Newton*” methods. An example is given in Equation 1.18, where the Hessian is approximated by a unitary matrix and λ_i is an appropriate scale factor used to modulate the amount of the i^{th} displacements on the surface, which always follows the opposite direction of the gradient. In this particular case, where the Hessian is a unit matrix, the method is known as a “steepest descent” method)

$$\text{Equation 1.18} \quad \mathbf{x}^{(i+1)} = \mathbf{x}^i - \lambda_i \mathbf{g}$$

The “steepest descent” method is far from being accurate but is very fast. It can be efficiently used to decide the first moves on a non quadratic region of the PES, far away from the critical point. Then, in the vicinity of the critical point, the search algorithm can be switched to the Newton-Raphson method or to a more accurate quasi-Newton scheme. Examples of quasi-

Newton methods are the widely used BSGF⁵⁻⁸ scheme (Broyden-Fletcher-Goldfarb-Shanno), the L-BFGS^{9,10} scheme (Limited memory-BFGS) and the PSB¹¹ scheme (Powell-symmetric-Broyden).

1.3.3 PES exploration: the NEB approach

An alternative technique that can be used for transition structure geometry optimization is represented by the NEB¹²⁻¹⁴ (Nudged-Elastic-Band) approach. This method can be used with good results when the structures of two minima are known and we need to locate the transition state¹⁵ that connects them. The modern implementations of the NEB method allow to find the MEP (Minimum Energy Path) relating two stable structures, even if separated by more than one transition state.¹⁶ The method requires the optimization of a number of intermediate images (Figure 1.2) along the reaction path. The images are obtained by interpolation from the geometries of the two given minima. Each image finds the lowest energy possible while maintaining equal spacing to neighbouring images. This constrained optimization is done by adding spring forces along the band between images and by projecting out the component of the force due to the potential perpendicular to the band.

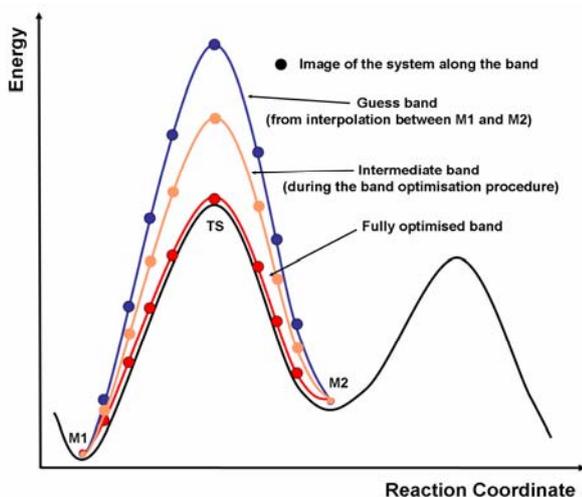


Figure 1.2 Schematic representation of the NEB approach

1.3.4 Molecular vibrations

The vibrational states¹⁷ of a molecule are experimentally observed by means of IR (Infra-Red) and Raman spectroscopy¹⁸ and give precious information about molecular structure and environment. However, to achieve this information is often an hard work because it can be difficult to assign each observed peak to a defined molecular motion. The calculation of vibrational frequencies can be of great help in the peak assignment and also in the computation of some important thermodynamical parameters (molecular enthalpy, entropy

and free energy), using some results of statistical mechanics.¹⁹ The simplest description of a vibration is an harmonic oscillator²⁰ defined by a quadratic (harmonic) potential energy function (E_{harm}) where the energy is a function of the square of the displacement with respect to an equilibrium position (x_0), as indicated in Equation 1.19.

Equation 1.19 $E_{\text{harm}} = k(x - x_0)^2 = k\Delta x^2$

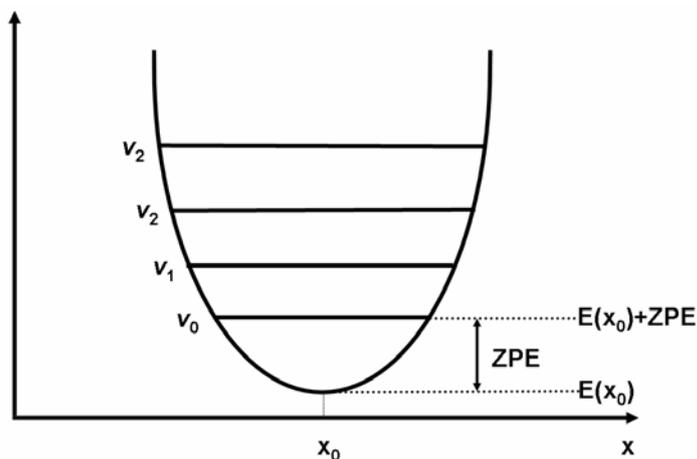


Figure 1.3 Schematic representation of a harmonic potential energy function and first vibrational levels (v_0 to v_2). The Zero Point Energy (ZPE) is indicated.

Better results can be achieved by using a more sophisticated potential like the one given using the Morse functional form.

Vibrational frequencies and vibrational energy levels for a molecular system can be directly obtained from the Hessian matrix, which can be computed either analytically (for the most part of the methods used in molecular investigation) or numerically by means of a finite difference procedure applied to analytical first derivatives (a practical task to compute Cartesian numerical frequencies on a given point consists in displacing each atom in the 6 directions of the Cartesian space; this requires a total of $6N+1$ energy and gradient evaluations, N being the number of atoms of the system).

A different technique to obtain the vibrational frequencies is bound to the molecular dynamics (MD) (see the next section). The vibrational motions observed in a MD simulation is a superposition of all the normal modes of vibration, so, to obtain the frequencies, it is necessary to apply the Fourier transform to all the motions of the MD motions; the motion corresponding to each peak of the so obtained spectrum is obtained by applying to it an inverse Fourier transform.

The energy value obtained by means of geometry optimization of a stable species coincides with the minimum of the PES. Actually, the system can never reach this energy value because it always maintains some vibrational motion. It is possible to evaluate the energy associated

with the lowest vibrational level (Zero-Point energy or ZPE, see Figure 1.3) and sum it to the potential energy to obtain a more accurate value. This is of particular importance when comparing the energy of different critical points of the surface (different conformers or in the investigation of a chemical reaction).

1.4 Molecular dynamics

Molecular dynamics (MD)^{1-3,21} is one of the main tools²² available for the theoretical study of both large biological molecules and smaller organic system. This computational method calculates the time dependent behaviour of a molecular system as a function of time. MD simulations can provide detailed information on the fluctuations and conformational changes of proteins²³ and nucleic acids,²⁴ and on the properties²⁵ (and their dependence on temperature and pressure) of condensed matter systems, like fluids^{26,27} or solids. MD methods are now routinely used to investigate the structure, dynamics and thermodynamics of many molecular and supra-molecular systems. They are also used in the structural determination from X-ray crystallography and from NMR experiments. During a MD simulation atoms and molecules are allowed to interact for a period of time under known laws of physics, giving a view of the motion of the atoms. Because molecular systems generally consist of a vast number of particles, it is impossible to find the properties of such complex systems analytically; MD simulation circumvents this problem by using numerical methods. Given an initial configuration of the system, the subsequent ones are generated by integrating Newton's laws of motion. The results is a trajectory that indicates how the positions and the velocities of the particles vary with time.¹ MD simulations can be based either on a MM (classical MD) or a QM potential²⁸ (AI-MD, *ab initio*-MD), or, alternatively, an hybrid method^{29,30} (QM/MM-MD). Given the second Newton law (Equation 1.20), it is possible to generate a differential equation (Equation 1.21), which describes the motion of a particle on mass m_i along the coordinate x_i , F_{x_i} being the force acting on the particle along that direction.

Equation 1.20 $F = ma$

Equation 1.21 $\frac{dx_i}{dt^2} = \frac{F_{x_i}}{m_i}$

A trivial case is that of a moving particle with no forces acting in the time between collisions: the position changes by $\mathbf{v}_i \delta t$ (\mathbf{v}_i is the constant velocity, and δt is the time between collisions); A different situation is that of a particle that experiences a constant force between collision:

for instance, a charged particle moving in an uniform electric field. A third case is that where the force acting on the particle depends on its position with respect to the position of other particles. The latter type of motion is very difficult to handle and often impossible to describe analytically, due to the coupled nature of the particle's motion. Under this circumstance, to overcome the many-body problem, the equations are integrated using the *finite difference method*.¹ The trajectories are generated under the assumption of a continuous potential and the approximation of breaking the integration into small stages, each separated by a small and fixed amount of time δt . In all algorithms based on the *finite difference method* both position (\mathbf{r}) and dynamical properties (velocities, accelerations, etc.) are approximated by a Taylor series expansion. See Equation 1.22 where \mathbf{v} is the velocity (first derivative of position with respect to the time), \mathbf{a} is the acceleration (second derivative), \mathbf{c} is the third derivative and so on.

$$\mathbf{r}(t + \delta t) = \mathbf{r}(t) + \delta t \mathbf{v}(t) + \frac{1}{2} \delta t^2 \mathbf{a}(t) + \frac{1}{6} \delta t^3 \mathbf{b}(t) + \frac{1}{24} \delta t^4 \mathbf{c}(t) + \dots$$

Equation 1.22 $\mathbf{v}(t + \delta t) = \mathbf{v}(t) + \delta t \mathbf{a}(t) + \frac{1}{2} \delta t^2 \mathbf{b}(t) + \frac{1}{6} \delta t^3 \mathbf{c}(t) + \dots$

$$\mathbf{a}(t + \delta t) = \mathbf{a}(t) + \delta t \mathbf{b}(t) + \frac{1}{2} \delta t^2 \mathbf{c}(t) + \dots$$

One of the most common implementations is given by the *Verlet algorithm*,³¹ which uses the positions and accelerations at time t , and the positions from the previous step, $\mathbf{r}(t - \delta t)$, to calculate the new positions $\mathbf{r}(t + \delta t)$ at the time $t + \delta t$. The key-equation is Equation 1.25, obtained by summing Equation 1.24 to Equation 1.23. To compute the velocities it is possible to use Equation 1.26 or Equation 1.27, thus obtaining an estimate of the velocity for a double step or for half a step, respectively.¹

Equation 1.23 $\mathbf{r}(t + \delta t) = \mathbf{r}(t) + \delta t \mathbf{v}(t) + \frac{1}{2} \delta t^2 \mathbf{a}(t)$

Equation 1.24 $\mathbf{r}(t - \delta t) = \mathbf{r}(t) - \delta t \mathbf{v}(t) + \frac{1}{2} \delta t^2 \mathbf{a}(t)$

Equation 1.25 $\mathbf{r}(t + \delta t) = 2\mathbf{r}(t) - \mathbf{r}(t - \delta t) + \delta t \mathbf{v}(t) + \delta t^2 \mathbf{a}(t)$

Equation 1.26 $\mathbf{v}(t) = \frac{\mathbf{r}(t + \delta t) - \mathbf{r}(t - \delta t)}{2\delta t}$

Equation 1.27 $\mathbf{v}(t + \frac{1}{2} \delta t) = \frac{\mathbf{r}(t + \delta t) - \mathbf{r}(t)}{\delta t}$

After its first appearance, the *Verlet* algorithm has been highly improved and new versions have been developed to increase its reliability and enhance its performances. An explicit inclusion of velocities, without making differences on large numbers, was accomplished by the *leap frog* algorithm,³² and further developments lead to the *velocity Verlet*³³ and *Beeman*³⁴ procedures.

The time step integration δt is a key-parameter for MD simulations. A too small time step would result in a short trajectory unsuitable for statistical analysis (it would provide a poor sampling of the phase space).¹ On the other hand, a too large time step can cause algorithm instabilities, due to unrealistic sampling of high energy regions of the phase space.¹ Even if a rigorous rule is not available, the experience suggests to choose, in the case of fluid systems, a δt value which is small compared to the mean time between collisions; in the case of flexible molecules (proteins, DNA, etc.) the time step should be approximately one-tenth of the shorter period of motion. The highest frequency motions are often associated with bond stretching. The C-H stretching motions, for instance, have the shortest period (about 10 fs); thus, in this case the optimum time step is about 1 fs and can be increased up to 2 or 3 fs by freezing the C-H bond stretching to the initial (equilibrium) value. Constrained dynamics¹ enables individual internal coordinates (or combinations of specific coordinates) to be fixed during the simulation. According to a different scheme, i.e. restrained dynamics¹, a further potential contribution is added for a given coordinate to reduce its variation without keeping it frozen. The SHAKE procedure^{35,36} is the most common tool for applying constraints in MD simulations. Some modern and complex algorithms are able to handle multiple step³⁷⁻⁴¹ with no loss of accuracy; they allow longer trajectories as well as more complete sampling of the phase space.

2 Computing the Energy of a Molecular System

2.1 Introduction

As stated in the previous section, the first step to accomplish in the study of a molecular system is the choice of the method (level of theory) to compute the energy associated with a given nuclear configuration (B.O. approximation).⁴ Many computational approaches are available for this purpose. They can be either quantum mechanical (QM) or molecular mechanics (MM) methods. The MM methods, also referred to as “classical methods”, are based on the Newton’s laws of physics, while the QM ones provide, in general, approximate solutions of the wave equation (Schrödinger equation). The MM methods are much faster than the QM ones and can be used to handle systems composed by thousands of atoms. The QM methods, because of their higher computational cost, are rarely used for systems composed by more than one hundred atoms. The performance and reliability of MM methods strongly depend on their parameterisation, i.e. the assignment of fixed values to some parameters that enter the functional form of the MM energy equations. These parameters vary from system to system and are strongly affected by the atom type and its valence state. Even if a number of valid parameter sets have been developed and are available in the literature, the MM methods must be used with caution, after a careful analysis of the class of molecules under investigation. On the other hand QM methods can, in principle, be applied to any type of molecular systems. In particular, *ab initio* QM methods are fully general and do not need any parameterisation.

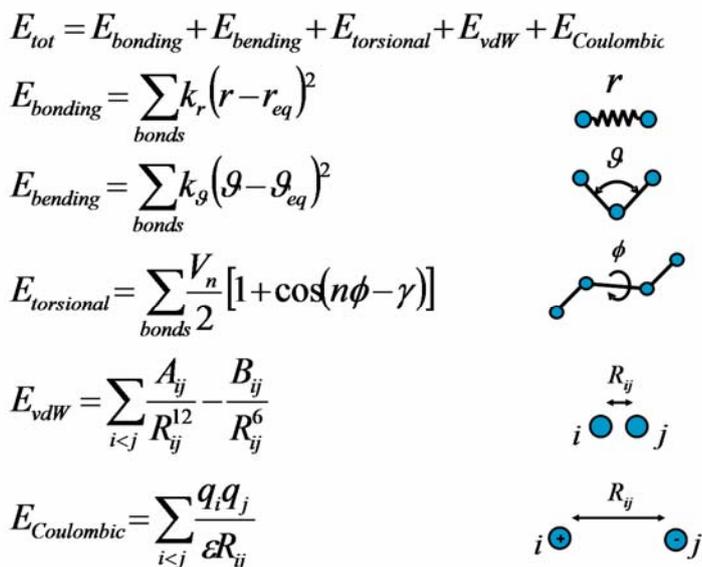
2.2 Molecular mechanics approach

The MM approaches allow to compute energy and properties of large molecular system using very simple models. The atoms are handled as charged sphere interacting under the Newton’s laws of motion. These spheres are connected by elastic springs along the direction of the chemical bond. The functional form¹⁻³ of the MM energy is composed by various terms, each ones taking into account the contribution arising from various bonding (stretching, bending

and torsional) and non-bonding interactions (van der Waals, Coulombic). In Equation 2.1 and Scheme 2.1 the functional form of the widely used Amber Force Field (AFF) is reported.⁴² Here both stretching and bending contributions (first and second terms, respectively) are expressed by simple harmonic potentials. The torsional contribution is described by a more complex form (third term) to account for multiple conformational minima. The classical 6-12 Lennard-Jones potential and the Coulombic function with a defined permittivity constant ϵ have been used for non-bonding interactions, i.e. van der Waals and Coulombic (fourth and fifth terms) terms, respectively. In the MM equations the equilibrium bond lengths and angles (r_{eq}, ϑ_{eq}) and the spring constants (k_r, k_ϑ) are parameterised to reproduce experimental frequencies of some sets of molecules. V_n is fitted to reproduce *ab initio* energies, A_{ij} and B_{ij} are obtained from *Monte Carlo*¹⁻³ simulations, the atomic point charges (q_i, q_j) are derived from *ab initio* calculation and subsequent RESP⁴³ fitting.⁴⁴

Equation 2.1

$$E_{tot} = \sum_{bonds} k_r (r - r_{eq})^2 + \sum_{bonds} k_\vartheta (\vartheta - \vartheta_{eq})^2 + \sum_{bonds} \frac{V_n}{2} [1 + \cos(n\phi - \gamma)] + \sum_{i < j} \frac{A_{ij}}{R_{ij}^{12}} - \frac{B_{ij}}{R_{ij}^6} + \sum_{i < j} \frac{q_i q_j}{\epsilon R_{ij}}$$



Scheme 2.1

A specific set of parameters is assigned to each couple (or triplet and quartet, for bending and torsions) of atoms. The definition of atoms within MM methods is more complicated than in QM approaches. More precisely to obtain reliable values for MM calculations a new atom

definition has been adopted: all atoms in a molecule are classified as different Atom Types not only on the base of the atomic number, but also according to their immediate environment. This lead to the development of different parameters, for instance, for aliphatic and aromatic Carbon atoms, for carbonylic or alcoholic Oxygen atoms, and so on. A particular MM Force Field is defined on the basis of the adopted functional form for the energy expression, of the specific values of the chosen parameters and of the available Atom Types. As stated above, a commonly used Force Field is the AFF, firstly implemented in the AMBER simulation package⁴⁵ but now available in many others. AFF gives extremely reliable results when used to study proteic^{46,47} or nucleic acid⁴⁸ systems, because of its accurate parameterisation focused on bio-molecules. An extension of the application field was obtained by increasing the size of the Atom Type library (with the related parameters. This lead to the development of the Generalised Amber Force Field (GAFF).⁴⁹ Other common Force field are the Gromacs⁵⁰ and the Charmm^{51,52} ones, implemented in the homonymous packages.

2.3 Quantum mechanics approach

Of the three, *ab initio* (first-principles) methods are the most accurate and consistent because they provide the best mathematical approximation to the actual system. The term *ab initio* implies that these methods, based on the laws of quantum mechanics, only require the knowledge of a few fundamental constants: the electron mass, the electron and nuclear charges and the values of fundamental physical constants, such as the speed of light and the Planck's constant. The time-independent Schrödinger equation is reported in Equation 2.2, where Ψ is the wavefunction describing the system, E is the energy of the system and \hat{H} is the Hamiltonian operator. For a system of charged particles (electrons and nuclei), \hat{H} can be written as in Equation 2.3, where \hat{T} is the kinetic energy operator, \hat{V} is the potential energy operator, m_a is the mass of the particle a , ∇_a^2 is the Laplacian operator for particle a , q_a and q_b are the charges of a and b , and \vec{r}_a and \vec{r}_b the positions of a and b . According to the B.O. approximation the total Hamiltonian can be split in the nuclear and electronic term to give Equation 2.4 and Equation 2.5, where the indices μ, ν and i, j indicate the coordinates of nuclei and electrons, respectively. Thus, the total Hamiltonian is written in term of one-, two- and zero-electrons operators.

Equation 2.2 $\hat{H}\Psi = E\Psi$

$$\text{Equation 2.3} \quad \hat{H} = \hat{T} + \hat{V} = \sum_a -\frac{1}{2m_a} \nabla_a^2 + \sum_{a < b} \frac{q_a q_b}{|\vec{r}_a - \vec{r}_b|}$$

$$\text{Equation 2.4} \quad \hat{H}_{nuc} = \sum_{\mu} -\frac{1}{2m_{\mu}} \nabla_{\mu}^2$$

$$\text{Equation 2.5} \quad \hat{H}_{el} = \sum_i -\frac{1}{2m_i} \nabla_i^2 - \sum_{\mu, i} \frac{Z_{\mu}}{|\vec{r}_{\mu} - \vec{r}_i|} + \sum_{i < j} \frac{1}{|\vec{r}_i - \vec{r}_j|} + \sum_{\mu, \nu} \frac{Z_{\mu} Z_{\nu}}{|\vec{r}_{\mu} - \vec{r}_{\nu}|} = \sum_i \hat{h}_i + \sum_{i < j} \hat{g}_{ij} + \hat{h}_0$$

According to the Molecular Orbital (MO) approach the wavefunction can be written in terms of one-electron functions (spin-orbitals) $\psi_i(j)$ defined as a product of a spatial part $\phi_i(j)$ and a spin function $\alpha_i(j)$ or $\beta_i(j)$. The i and j indices are running on the spatial functions and the electrons, respectively. The anti-symmetric electronic wave-function can be approximated as a product of spin-orbitals, in the form of a Slater Determinant (Equation 2.6).⁵³

$$\text{Equation 2.6} \quad \Psi = (n!)^{-\frac{1}{2}} \begin{vmatrix} \psi_1(1)\alpha(1) & \psi_1(1)\beta(1) & \psi_2(1)\beta(1) & \cdots & \psi_{n/2}(1)\beta(1) \\ \psi_1(2)\alpha(2) & \psi_1(2)\beta(2) & \psi_2(2)\beta(2) & \cdots & \psi_{n/2}(2)\beta(2) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \psi_1(n)\alpha(n) & \psi_1(n)\beta(n) & \psi_2(n)\beta(n) & \cdots & \psi_{n/2}(n)\beta(n) \end{vmatrix}$$

Each molecular orbital ϕ_i of a system can be expanded in terms of a set of N pre-defined one-electron functions $\chi_{\mu i}$ denoted as basis functions (basis set) according to the Equation 2.7, where the $c_{\mu i}$ are the molecular orbital expansion coefficients. When atomic orbitals (AO) are used as basis functions this approach is often referred to as Linear Combination of Atomic Orbital (LCAO) approximation.

$$\text{Equation 2.7} \quad \phi_i = \sum_{\mu=1}^N c_{\mu i} \chi_{\mu}$$

2.3.1 Hartree-Fock (HF)

The Hartree-Fock⁵⁴ method is one of the most widely used computational approach for the determination of the wave-function within the molecular orbital model. The wave function is written in the form of a Slater Determinant and the related energy is obtained by minimization of its expression with respect to the molecular orbitals ϕ_i . This procedure leads to the Hartree-Fock eigenvalue equations (Equation 2.8) where the Fock operator \hat{F} acts on the orbitals ϕ_i to give the same orbital function multiplied by a constant ε_i , which represents

the energy of the orbital. The LCAO approximation can be applied to obtain Equation 2.9 and, after rearrangement, Equation 2.10. If both terms of Equation 2.10 are multiplied by χ_ν , summed to ν and then integrated over all the configuration space we obtain the Roothaan-Hall formulation⁵⁵ of Hartree-Fock equations (Equation 2.11 and Equation 2.12). Roothaan-Hall formulation is given in the χ_μ basis. $F_{\nu\mu}$ and $S_{\nu\mu}$ are the elements of the Fock and overlap matrix in the atomic representation, respectively.

$$\text{Equation 2.8} \quad \hat{F}\phi_i = \varepsilon_i\phi_i \quad (i = 1, 2, \dots, N)$$

$$\text{Equation 2.9} \quad \hat{F}\sum_{\mu=1}^N c_{\mu i}\chi_\mu = \varepsilon_i\sum_{\mu=1}^N c_{\mu i}\chi_\mu$$

$$\text{Equation 2.10} \quad \sum_{\mu=1}^N c_{\mu i}\hat{F}\chi_\mu = \varepsilon_i\sum_{\mu=1}^N c_{\mu i}\chi_\mu$$

$$\text{Equation 2.11} \quad \sum_{\nu=1}^N \sum_{\mu=1}^N c_{\mu i}\langle \chi_\nu | \hat{F} | \chi_\mu \rangle = \varepsilon_i \sum_{\nu=1}^N \sum_{\mu=1}^N c_{\mu i}\langle \chi_\nu | \chi_\mu \rangle$$

$$\text{Equation 2.12} \quad \sum_{\nu=1}^N \sum_{\mu=1}^N c_{\mu i}F_{\nu\mu} = \varepsilon_i \sum_{\nu=1}^N \sum_{\mu=1}^N c_{\mu i}S_{\nu\mu}$$

To obtain a more compact form (Equation 2.14), the matrix notation (Equation 2.13) can be used, by considering $F_{\nu\mu}$ and $S_{\nu\mu}$ as the elements of two $N \times N$ matrices and $c_{\mu i}$ as the elements of a vector.

$$\text{Equation 2.13} \quad \begin{pmatrix} F_{11} & F_{12} & \cdots & F_{1N} \\ F_{21} & F_{22} & \cdots & F_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ F_{N1} & F_{N2} & \cdots & F_{NN} \end{pmatrix} \begin{pmatrix} c_{1i} \\ c_{2i} \\ \vdots \\ c_{Ni} \end{pmatrix} = \varepsilon_i \begin{pmatrix} S_{11} & S_{12} & \cdots & S_{1N} \\ S_{21} & S_{22} & \cdots & S_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ S_{N1} & S_{N2} & \cdots & S_{NN} \end{pmatrix} \begin{pmatrix} c_{1i} \\ c_{2i} \\ \vdots \\ c_{Ni} \end{pmatrix}$$

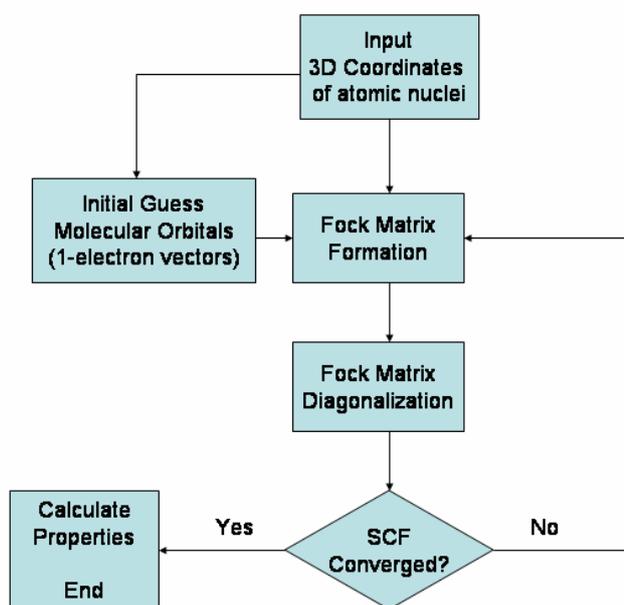
$$\text{Equation 2.14} \quad \mathbf{F}\mathbf{c}_i = \varepsilon_i\mathbf{S}\mathbf{c}_i$$

The latter equation refers to a specific i^{th} orbital. If we consider all N molecular orbitals that can be obtained as linear combinations of the basis functions, we have a more general form (Equation 2.15), where \mathbf{c} is a square matrix containing all vectors corresponding to the N molecular orbitals and $\boldsymbol{\varepsilon}$ is a diagonal matrix containing the orbital energies (Equation 2.16 and Equation 2.17).

Equation 2.15 $\mathbf{F}\mathbf{c} = \epsilon\mathbf{S}\mathbf{c}$

Equation 2.16 $\mathbf{c} = \begin{pmatrix} c_{11} & c_{12} & \cdots & c_{1N} \\ c_{21} & c_{22} & \cdots & c_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ c_{N1} & c_{N2} & \cdots & c_{NN} \end{pmatrix}$

Equation 2.17 $\epsilon = \begin{pmatrix} \epsilon_1 & & & \\ & \epsilon_2 & & \\ & & \ddots & \\ & & & \epsilon_N \end{pmatrix}$



Scheme 2.2 HF-SCF approach

The solution of the Hartree-Fock equations, even in the Roothaan-Hall formulation, is not trivial, being the Fock operator a function of the molecular orbitals themselves corresponding to the solutions of the eigenvalues equations. The algorithm used to solve the Roothaan equations is an iterative approach (Scheme 2.2) known as Self Consistent Field (SCF) procedure. According to this algorithm an initial guess for the Fock matrix is built using the molecular orbitals obtained at a lower level of theory (for instance Extended Huckel¹ or semiempirical methods,¹ not discussed here). Then, the Roothaan-Hall equations are solved to give atomic coefficients for each orbital and the relative energies and, consequently, the total energy. The procedure is repeated using the new coefficients to build a new Fock matrix and a new value of energy is obtained. The new total energy value is compared to the previous one. The procedure stops when this energy difference is smaller than a chosen value.

We can summarise the main approximations used in the HF-SCF approach:

- 1) The Born-Oppenheimer approximation is inherently assumed. The true wavefunction should also be a function of the coordinates of each of the nuclei.
- 2) Typically, relativistic effects are completely neglected. The momentum operator is assumed to be completely classical.
- 3) The representation of molecular orbitals is composed of a finite number of orthogonal one-electron wave-functions (molecular orbitals). The true wavefunction would need a complete (infinite) representation.
- 4) The energy eigenfunctions (wave-functions) are assumed to be products of one-electron functions (Slater determinants). The effects of electron correlation, beyond that of exchange energy resulting from the anti-symmetry of the wavefunction, are neglected.

The missing electronic correlation effects of the HF-SCF method, lead to large deviations with respect to experimental results. A number of approaches, usually denoted as post-Hartree-Fock methods, have been devised to include electron correlation in the multi-electron wave function. One of these approaches, Møller-Plesset perturbation theory,^{54,56} treats correlation as a perturbation of the Fock operator. Others methods expand the multi-electron wave-function in terms of a linear combination of Slater determinants, such as Multi-Configurational Self Consistent Field (MC-SCF),^{2,3} Configuration Interaction (CI),⁵⁴ Complete active space SCF (CAS-SCF).^{2,3} Other approaches (less popular) modify the Hartree-Fock wave-function by multiplying it by a correlation function ("Jastrow" factor), a term which is explicitly a function of inter-electron coordinates and cannot be decomposed into independent single-particle functions.

2.3.2 Møller-Plesset perturbation theory (MP)

Møller-Plesset perturbation theory (MP) is one of the most popular quantum chemistry post-Hartree-Fock *ab-initio* method (the main idea was published in 1934⁵⁶). Rayleigh-Schrödinger perturbation theory (RS-PT) is used to improve the Hartree-Fock energy. MP2, MP3, MP4 and so on acronyms are used to denote a truncation of the perturbation series up to second, third, fourth order, respectively. The MP2 method, even if more computationally expensive than HF, is often used to treat systems where electronic correlation effects play a key role. For instance, weak interactions like dispersion (London) forces or π -stacking interactions between aromatic compounds can be properly described at MP2 level.

2.3.3 DFT

Density Functional Method (DFT) is somewhere between the *ab initio* and the semiempirical approaches. According to its formalism the basic quantity is not a many-body wavefunction (which depends on $3N$ spatial variables), but the molecular electron density (dependent on three variables only). The most common implementation of density functional theory is the Kohn-Sham⁵⁷ approach (KS-DFT), where the intractable many-body problem of interacting electrons in a static external potential is reduced to a tractable problem of non-interacting electrons moving in an effective potential. This is composed by the external potential and the effects of the Coulomb interactions between the electrons, namely exchange and correlation interactions. In Equation 2.18 the first term represents the electromagnetic interaction of the electron density with the external potential (it often corresponds to the Coulombic interaction between electrons and nuclei), the second represents the repulsion between the electron density and itself (Coulomb energy), the third term approximates the electronic kinetic energy (by expanding the density in a set of orbitals and computing the kinetic energy for the hypothetical system of not interacting electrons) and the latter the exchange-correlation functional which corrects the former ones. The major problem within KS-DFT is the modelling of exchange and correlation interactions. The older and simplest approximation, called Local Density Approximation (LDA), was based on the Thomas-Fermi model of a uniform electron gas. In local exchange–correlation functionals, correlation and exchange interactions are functions which depend only on the scalar value of the electron density at a given point in space (Equation 2.19). A simple exchange functional (Slater)⁵⁸ is reported in Equation 2.20: it accounts for stabilisation of high electron density correcting the electron-electron energy, which is overestimated by the Coulomb term. Combination of this exchange functional with that proposed by Vosko, Wilk and Nusair leads to the SVWN functional (which is synonymous of LDA).

$$\text{Equation 2.18} \quad E[\rho] = \int v(r)\rho(r) \, dr + J(\rho) + T_s(\rho) + E_{EX}(\rho)$$

$$\text{Equation 2.19} \quad E_{EX}[\rho] = \int [\varepsilon_X\{\rho(r)\} + \varepsilon_C\{\rho(r)\}]\rho(r) \, dr$$

$$\text{Equation 2.20} \quad \varepsilon_X\{\rho(r)\} = -\frac{3}{2} \left(\frac{3}{\pi} \rho(r) \right)^{\frac{1}{3}}$$

$$\text{Equation 2.21} \quad \varepsilon_X\{\rho(r)\} = -\frac{3}{2} \left(\frac{3}{\pi} \rho(r) \right)^{\frac{1}{3}}$$

A second generation of exchange–correlation functionals include not only functions of the scalar density in the integrand, but also functions of the gradient of the density. They are also referred to as “non-local” as the gradient introduces a certain degree of non locality in the energy expression. They should be more properly referred as “gradient-corrected” (adopting the Generalised Gradient Approximation or GGA) because the exchange correlation term is still an integral relating a defined energy contribution with the electron density of an infinitesimal volume. Since Becke proposal⁵⁹ of a gradient-corrected exchange functional many attempts have been made to improve the reliability of t of GGA DFT by developing correlation functionals with parameters obtained by fitting experimental data or generated to reproduce well-known physical principles. A popular GGA functional combines the Becke exchange expression with the Lee, Yang Parr correlation formulation,⁶⁰ giving the commonly used BLYP functional. Other functional forms for both exchange and correlation expression have been proposed and all their possible combinations constitute he large family of GGA DFT functionals. The next GGA functional generation, called meta-GGA, like the Tao, Perdew, Staroverov and Scuseria (TPSS),⁶¹ include two more functions of the density, the Laplacian of the total density (or of the densities of spin-up and -down electrons) and the sum of the kinetic energy densities of the Kohn–Sham orbitals.

The observation that atomisation energy are underestimated by HF method and overestimated by both LDA and GGA methods suggested some combined treatment to improve the results. Many and sophisticated hybrid functional have been developed. An early formulation⁶² adopted a three-parameter functional where the exchange–correlation energy was expressed as a combination of the local exchange–correlation energy, the HF exchange energy, and the gradient corrections to the exchange and correlation energies as shown in Equation 2.22, where Becke exchange⁵⁹ and Perdew *et al.* correlation⁶³ are used. The substitution of the Perdew correlation with the LYP⁶⁰ one lead to the well known B3LYP functional.

Equation 2.22
$$E_{XC} = E_{XC}^{LSDA} + a_0(E_X^{HF} - E_X^{LSDA}) + a_x \Delta E_X^{B88} + a_c \Delta E_C^{PW}$$

Despite the continuous development of new functionals some problems still affect DFT accuracy. Due to deficiencies in the treatment of exchange–correlation the pure dispersion interactions between unbound chemical species are not well reproduced by common functionals. This problem can be treated by adding an empirical correction to the functional, as proposed by Grimme,^{64,65} enhancing the results with no added computational cost. The second problem relates to the poor cancellation between the electron self-interaction present

in the Coulomb term of Equation 2.18 and the exchange energy. The third problem is that even the best current exchange–correlation functionals still lead to unacceptably large energy errors for a significant number of “outliers” species,⁶⁶ even when dispersion and self-interaction do not appear to be involved and normal bonding is expected to occur.

A very new family of functionals has been developed in 2006 by Grimme⁶⁷ to deal somehow with these problems and, despite the lack of a very large benchmark, they are believed to improve the performance of DFT particularly for weak interactions. They consist of a semi-empirical mixture of DFT components and the MP2 correlation energy calculated with the DFT orbitals. Grimme referred to his functional as B2PLYP⁶⁸⁻⁷¹ (B88 exchange, 2 fitted parameters and perturbative mixture of MP2 and LYP); a version with improved performance (in particular for weak interactions) is mPW2PLYP.⁷² From the extensive calibration work, the new functionals appear to give better energetics and a narrower error distribution (in relation to the third error mentioned above) than B3LYP. Thus, the additional cost of the calculation of the MP2 energy may be well invested. The use of the DFT method has become very popular in the last two decades, also between experimental chemists, due to its good cost/efficiency ratio. A large number of publication review the main application fields of the various functional.^{66,73-75}

2.3.4 DFTB

The density functional tight binding (DFTB) method is an alternative approach to the quantum chemical semi-empirical¹ methods. It corresponds to an approximate DFT scheme and is characterized by a computational speed similar to that of traditional semi-empirical quantum chemical methods (like MNDO, AM1, and PM3) but without having a large number of empirical parameters. The approximate DFTB method is derived from DFT by a second-order expansion of the DFT Kohn-Sham total energy with respect to charge density fluctuations. The zero-order approach is equivalent to a common standard non-self-consistent (Tight-Binding: TB) scheme, while at the second order a transparent, parameter-free and readily calculable expression for generalized Hamiltonian matrix elements may be derived. DFTB can be seen as a tight binding method, parametrised from DFT. DFTB was augmented by a self-consistency treatment based on atomic charges in the so-called self-consistent charge density-functional tight-binding (SCC-DFTB) method.⁷⁶⁻⁷⁹ The DFTB method has been applied for a large variety of problems in chemistry,⁸⁰⁻⁸² physics,^{83,84} materials science,^{76,85} and biochemistry.⁸⁶⁻⁹² During the past decade it has been continually further developed to improve its accuracy and the applicability.

2.3.5 The Resolution of Identity approximation

The Resolution of Identity (RI) approximation⁹³⁻⁹⁵ was developed to speed-up Hartree-Fock,⁹⁶ MP2⁹⁷⁻⁹⁹ and DFT¹⁰⁰⁻¹⁰² calculations. According to the LCAO approach the molecular orbitals are approximated as linear combinations of a finite number of atomic orbital (basis set). Products of basis function $\nu(r)\mu(r)$ are approximated by linear expansion of so-called auxiliary basis function $P_i(r)$ (Equation 2.23), leading to an approximation of two-electron integrals.

Equation 2.23 $\nu(r)\mu(r) \approx \sum_i c_{\nu\mu}^i P_i(r)$

A further level of approximation is the Multipole Accelerated Resolution of Identity (MARI),¹⁰³ developed to increase the speed of RI-DFT computations. The use of RI and RI+MARI in conjunction with MP2 and DFT methods allows the treatment of large systems with an high level of accuracy. A practical example is the possibility to study with the perturbative approach (MP2) systems or processes where the electronic correlation is not correctly accounted for by DFT methods: the RI-MP2 provides results of MP2 comparable accuracy at the computational cost of a non-RI GGA DFT method.

3 COBRAMM

3.1 Introduction

In the last decade combined quantum mechanics/molecular mechanics (QM/MM) methods have been applied to a large variety of chemical problems. This section describes a new QM/MM implementation that operates as a flexible computational environment. In particular, geometry optimizations, frequency calculations and molecular dynamics can be performed on a given molecular system, that can be divided in two or three layers corresponding to different levels of computational accuracy. Here we report a detailed description of the method and its implementation. In the following sections examples of applications of our QM/MM method are presented, to demonstrate its capability in treating very different chemical problems, (from ground to excited states topics). Flexibility, general applicability and accuracy of the new hybrid approach are also pointed out. Finally, it will be shown that the current implementation (called COBRAMM: Computations at Bologna Relating Ab-initio and Molecular Mechanics Methods)¹⁰⁴ is more than a simple QM/MM method, but a more general hybrid approach with a modular structure that is able to integrate some specialized programs, which may increase the flexibility/efficiency of QM, MM and QM/MM calculations.

The COBRAMM package is developed at the Università degli Studi di Bologna (University of Bologna) by Marco Stenta (the Ph.D. candidate) and by Dr. Piero Altoè under the supervision of Professor Andrea Bottoni and Dr. Marco Garavelli. This suite of programs contains a main interface (to perform QM, MM and QM/MM calculation) as well as some tools used to prepare inputs, to check their consistency, and to analyse the results. The Ph.D course of the candidate focused on the development and testing phase of the COBRAMM package.

The study of large molecular systems takes advantage of the use of computational methods based on Molecular Mechanics (MM).^{1,3} These methods simulate atoms as rigid charged spheres. The interactions between atoms are modelled on the basis of chemical connectivity, using simple harmonic potentials (or sometimes more complex functions) to describe bonds,

bending and torsions. Non-bonded electrostatic and Van der Waals interactions are taken into account on the basis of the charge (or dipole) assigned to atoms using a simple Coulombic potential and a Lennard-Jones potential (or similar), respectively. The analytic functional form of the equations used to compute energies and forces make the MM calculation fast even for large molecules. However, the drawback of these methods is their inability in describing processes involving a change in the “nature” of atoms. Thus, chemical reactions (where bonds are broken and formed, and where connectivity and atom-types are not preserved) cannot be described by MM methods. On the other hand, computational methods based on Quantum Mechanics (QM)³ have proved to be successful in dealing with such problems because they explicitly treat the electrons (and their couplings) by computing the associated wave function. Unfortunately, the QM methods are much more expensive (in terms of computational time) than the MM ones and cannot be used to study very large systems. Thus, a problem arises when studying the chemical reactivity involving large molecules. To overcome this problem several theoretical studies for the development of new computational methods have been carried out during the last three decades. A promising technique is the partitioning of the whole system (called *real* in the following discussion) in two regions (see Figure 3.1): a small part, containing the atoms involved in the chemical process, which is described at the QM level and a large region containing all the remaining atoms that are treated at the MM level. This approach allows to speed-up the calculation and simulate (although at a lower level) the influence of the environment on the reactive core. This hybrid approach is usually denoted as “QM/MM” approach¹⁰⁵⁻¹⁰⁹. However, when using this method a difficult technical problem, which is often a source of significant errors and approximations, must be overcome i.e. the correct description of the boundary region. Consequently, great care is addressed to this problem when developing QM/MM methods,¹⁰⁸⁻¹¹⁰ because it strongly influences the capability of reproducing the effects of the surrounding regions on the investigated process.

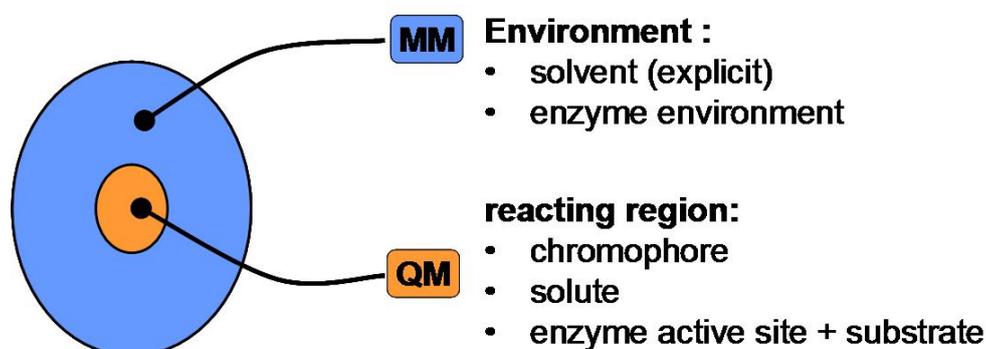


Figure 3.1 Schematic representation of the hybrid QM/MM approach.

The wide literature¹⁰⁵⁻¹¹³ on hybrid methods gave us a complete landscape on past attempts to develop a solid and general QM/MM approach. We took advantage of this experience for developing a new versatile QM/MM algorithm and in this work we describe the details of our implementation, focusing into its advantages as well as its limitations. More generally, our algorithm (called COBRAMM hereafter) is a hybrid approach that acts as an interface between different programs. These programs execute the single phases of the whole calculation and can communicate through the interface, whilst the partial data are gathered and manipulated to give the final result. Modular structure is the main feature of this approach, allowing the user to tailor the computational level by selecting (and combining together) specific programs according to the specific requirements of the investigation. The final purpose is to perform the computational task with the more appropriate (and efficient) computational tools available. A similar approach was used by P. Sherwood and co-workers for ChemShell.¹¹⁴

3.2 The QM/MM partitioning scheme

In order to construct a versatile QM/MM scheme COBRAMM adopts a modular approach, combining different available codes to exploit their recognized ability in dealing with a specific aspect of the whole problem. Such an approach allows a great flexibility in the partitioning of the system, because the computation of E (energy) and \mathbf{F} (forces) (see section 2.1) on one hand, and the projection of the new geometry on the other, are performed separately at different points of the whole calculation (see Figure 3.2). More specifically, E and \mathbf{F} are calculated according to a two-region partitioning of the system, where a QM and an MM calculation are independently performed on the two aforementioned sub-regions. This information is then processed (see Scheme 3.1 for the details) to obtain the total energy and the complete set of forces. These data are then used to generate the new geometry; during this step the system is also partitioned into two independent regions, which do not necessarily coincide with the first ones: the external (and often larger) one (**Opt1**) is handled by means of a fast and rough algorithm, like the Steepest Descents, while the smaller region (**Opt2/MD**) (including the **QM** sub-region) is treated with a more sophisticated optimisation algorithm, like BFGS¹¹⁵, or with a molecular dynamics code based on the Beeman or the Velocity-Verlet algorithms¹.

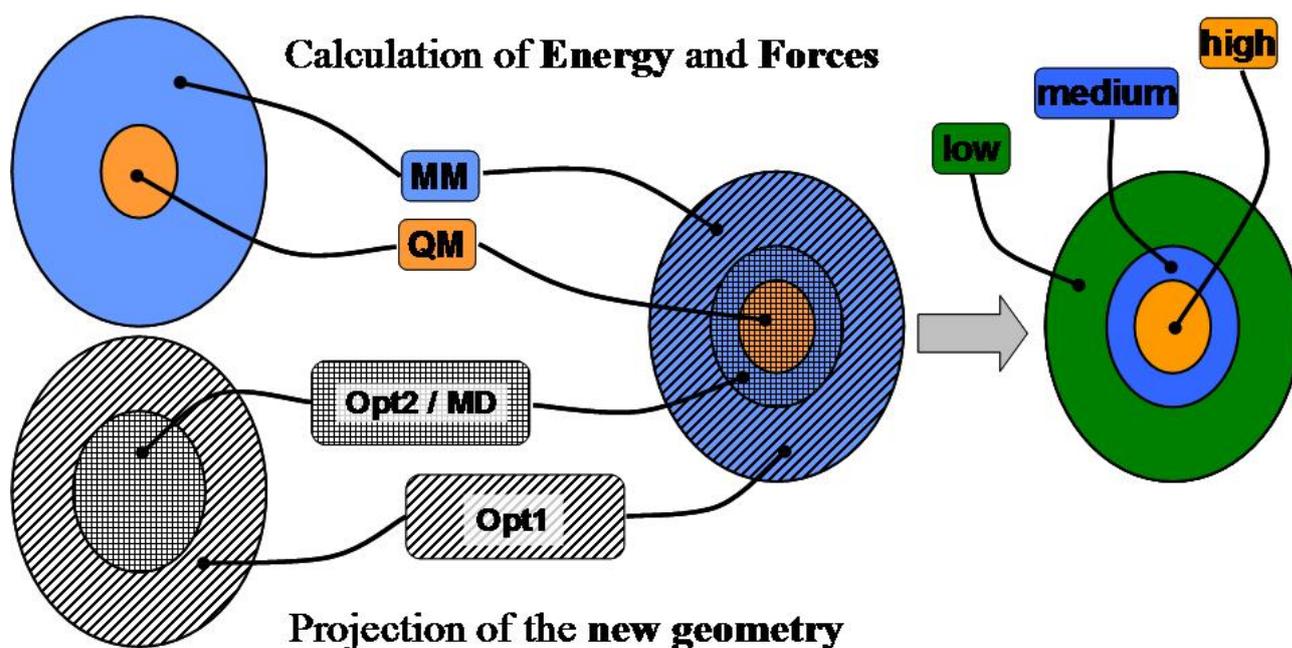


Figure 3.2 Graphical representation of the tree-layer scheme adopted in our code.

In our approach the **QM** region is a subset of the **Opt2/MD** region. When **QM** and **Opt2/MD** do coincide, then the system is divided in two layers called respectively *high*, treated at QM level and optimised with the accurate algorithm, and *low*, treated at the MM level and optimised independently using the fast algorithm. When **QM** and **Opt2/MD** do not coincide (i.e. **Opt2/MD** is larger than **QM**), then some MM atoms are involved, together with the QM atoms, in the accurate geometry optimisation or MD (i.e. these MM and QM atoms are coupled together). This buffer (intermediate) region of MM atoms, called *medium*, represents a significant improvement and makes the difference in respect to the three layers ONIOM(MO1:MO2:MM) approach from Morokuma,^{112,113} where the system is divided in three layers: the inner one (MO1, which is treated at high QM level), the outer (MM, which is treated at MM level) and the intermediate layer (MO2, which is treated at a low QM level, often semi-empirical). This scheme is intended to improve the simulation of the electrostatic interaction between QM and the MM regions by interposing a buffer layer treated at a low QM level. In our experience we found that a simpler scheme, like our one, is efficient if an appropriate MM force field is used and if an Electrostatic Embedding^{106,109} scheme is adopted. In our approach the intermediate region has a different function in respect to Morokuma's approach and it is used to improve the reliability and efficiency in the optimisation (or molecular dynamic) phase. The capability of handling coupled together the *high-medium* region (**Opt2/MD**) allows the description and the study of large molecular motions around the reactive region without increase of the computational cost; indeed **E** and **F**

of the non-reacting (MM) atoms included in the **Opt2/MD** region are computed at the MM level.

As mentioned above, our scheme uses the Electrostatic Embedding^{106,109} approach to account for the electrostatic influence of the surrounding MM region on the QM layer. Analogously, the effect of the charge variations occurring during the chemical process in the *high* (QM) layer is accounted for by using in the MM calculations the QM atomic point charges coming from QM computations. Furthermore, our approach also allows a full independent optimisation of the *low* layer (using the fast algorithm) per each optimisation step of the *high-medium* layer; this feature nearly resembles the so called “micro-iteration”¹¹⁶⁻¹¹⁸ technique (also used in the ONIOM¹¹⁶ approach) that, for example, can be useful in the study of a QM solute in a MM solvent. Some authors¹¹⁹ pointed out the need of taking into account, when studying complex systems like enzymes, the contribution from the protein reorganization energy. We developed the aforementioned optimisation scheme to deal with “the linear response of the protein to the movement along the reaction coordinate”, in the aim of improving the description of this phenomenon.

The algorithm we developed is intended to deal with all the possible combinations of layers in order to customize calculations according to the specific problem under investigation. The full set of possible combinations actually supported is reported in Table 3.1; the calculation level is indicated using up to three figures, which are the first figures of the name of the layers used. This notation will be used all through the paper.

Table 3.1 Calculation types.

Calculation type	Layer used		
	<i>high</i>	<i>medium</i>	<i>low</i>
HML	X	X	X
HM	X	X	
HL	X		X
ML		X	X
H	X		
M		X	

For example, HML refers to a three layers system: this can be useful, for instance, to study the reaction mechanism of an enzymatic process when including a small part of the active site and the substrate in the *high* layer, assigning the remaining part of the active site to the *medium* layer and leaving all the other residues of the enzyme in the *low* layer. An HM calculation can be employed for the accurate optimisation, or for the molecular dynamics, of the whole molecular system and allows a great saving in computational resources by treating at the QM level (*high* layer) only the reactive sub-region, while the remaining part of the system is

treated at the MM level (*medium* layer). As mentioned above, the HL scheme can give good results for the study of a solute (*high* layer) in a solvent (*low* layer) composed by a large number of explicit molecules. The ML calculation is fully performed at the MM level; it can be used, for example, to obtain a preliminary guess for a critical point on the Potential Energy Surface (PES) associated to the reaction under study. In the H (full QM) calculation the energy and forces evaluation from a code is combined with the optimisation (or MD) algorithm of a different code to study, for example, photoinduced ultra-fast processes with QM-MD simulations. The M (fully MM) calculation can be used to get a guess geometry for an H calculation. ML and M calculations can be used only for minima or conformational transition states (with no change in the chemical connectivity); they also give the opportunity to use the force field from an MM code and the optimisation or MD algorithm from a different one.

3.3 The boundary region

As stated in the introduction, handling the boundary between the QM and MM regions need extreme caution, because wrong assumptions can easily lead to unphysical results. In some cases the boundary does not go through a covalent bond: this is the case of a solute (QM level) immersed in a large number of explicit solvent molecules (MM level); this case is very easy to handle and do not need special assumptions. However, in many cases, one cannot avoid passing the QM-MM boundary crosswise one (or more) covalent bond(s), like it is for enzymatic systems. Two strategies have been developed to overcome this problem: a) the *atom-link* approach^{109,120-122} and b) the modified orbital methods.^{107,123-125} We will focus on the former approach because we adopted it to handle boundary regions in our QM/MM method.

When the QM-MM boundary cuts a covalent bond, then the valence of the frontier QM atom (Q1 in Figure 3.3) remains unfilled; we call *model* this dangling QM structure and we can observe that it coincides with the *high* layer, as mentioned above. To perform the QM calculation we need to saturate the dangling bond with a “frontier atom”. This link atom is usually a hydrogen atom, in which case from *model* we obtain *model-H* (Figure 3.4). The *atom-link* approach has, like each boundary treatment, strong advantages as well as weaknesses,¹⁰⁹ it is widely used and proved to give good results, it is straightforward to understand and it is easy to implement. However, the introduction of artificial atoms that are not present in the original system, can give problems in obtaining the total QM/MM energy value and in geometry optimisations.

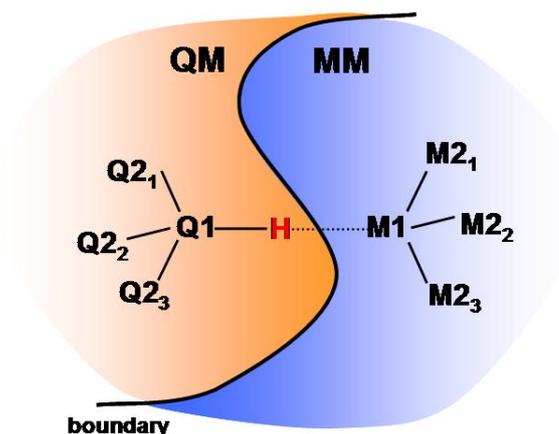


Figure 3.3 Boundary region crossing a covalent bond.

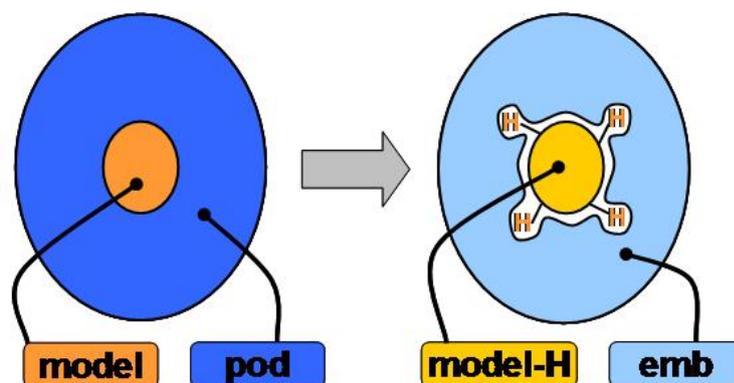


Figure 3.4 The atom-link approach implies a change in the original charges (*pod*) of the MM region to be introduced in the QM calculation, according to the Electrostatic Embedding scheme. The procedure to obtain the new set of charges (*emb*) is described in the text.

3.4 The atom-link position

Obtaining the correct QM/MM energy can be difficult because *model-H*, used for QM calculation, do not coincide with the *high* layer, due to the presence of additional hydrogen link atoms and because the MM atomic point charge at the boundary have to be modified.¹⁰⁹ We adopted a subtractive approach, similar to the ONIOM scheme,^{112,113} but not identical to it, because of a different partitioning scheme.

To improve the geometry optimisation procedure and avoid the artifacts arising from the presence of foreign H atoms, we tried to eliminate their extra degrees of freedom by making their position depending by the position of the Q1 (QM) and M1 (MM) atoms (see Figure 3.3). Thus, during a geometry optimisation or a molecular dynamics the position of the hydrogen link atom is reassigned at each step and is chosen to be on the line connecting Q1 and M1 atoms, removing two degrees of freedom. The last degree of freedom can be eliminated in several ways, but the simplest way is to keep fixed the Q1-H distance. Herein, we impose a fixed distance of 1.09 Å, which is a standard value for a C(sp³)-H bond. To reduce the error associated with this technique, this choice implies the boundary bond type to be only C(sp³)-

C(sp³). It is worth noting that the bond itself between Q1 and M1 is calculated molecular mechanically.

3.5 Handling QM/MM non-bonding cross-terms

In the presented algorithm, all the energy components of the *high* layer (QM region) are given by a QM calculation on *model-H*; the bonded (stretching, bending, torsions) and non bonded (electrostatic and Van der Waals) interactions within the *medium* and *low* layers (MM region) are handled at the MM level. Managing the cross terms is more difficult and can be done with different approaches. In our scheme, all the bonding and the Van der Waals terms are treated at the MM level, while electrostatic interactions between the QM and the MM region are computed at the high QM level. For this purpose, an Electrostatic Embedding¹⁰⁹ scheme was adopted: QM computation are performed on *model-H* surrounded by the atomic point charges of the MM layers (i.e. the *emb* charges). In our notation *pod* and *emb* are two subset of atomic point charges of the MM regions. The *pod* set contains the unmodified atomic point charges from the MM force field, while the *emb* set of charges do differ from the *pod* one (see Figure 3.4) only at the boundary region. Indeed, it has been shown that *pod* charges, when coupled to the *atom-link* method, lead to an unrealistic hyperpolarisation of the wave function at the Q1-H bond. This unphysical effect is mainly due to the presence of the atomic point charge on the M1 atom near the hydrogen atom link (see Figure 3.3). To avoid this problem, the charge on M1 is redistributed (i.e. it is spread) on the neighbouring MM atoms (M2) while it is set to zero on M1, so that the total charge of the system is preserved. It is worth to specify that the charge is not equally redistributed on the M2 atoms, but the added fractional charge is proportional to the module of the original MM charge hosted on that atom (see Appendix A for a detailed description of the adopted charge redistribution scheme). We realized that this charge redistribution scheme gives better results if compared to other strategies, like the zeroing of all the charges on the M2 atoms. Anyway, basing on what suggested recently in publications,^{106,109} some attempts have been made to improve this redistribution scheme in order to preserve the polarity of the original Q1-M1 bond as well as the one of the M1-M2 bonds. This will be particularly important for the study of systems in which the frontier does cut bonds other than the standard (substantially non-polar) C(sp³)-C(sp³) ones.

3.6 QM/MM formalism.

The QM/MM total energy follows a subtractive scheme¹⁰⁶ and can be written as:

$$\text{Equation 3.1} \quad E^{tot} = E_{QM}^{model-H} + E_{MM}^{real} + E_{QM}^{el,model-H/emb} - E_{MM}^{el,model/pod} - E_{MM}^{model-H}$$

Where the first and third terms are calculated at the QM level. $E_{QM}^{model-H}$ is the QM energy of *model-H in vacuo*, while $E_{QM}^{el,model-H/emb}$ is the electrostatic interaction between *model-H* and the charge embedding (*emb*). In this way the wave function is perturbed by the surrounding charges, ensuring the polarization of the QM part by the MM environment. This method is usually referred to as Electrostatic Embedding.^{109,126} In our scheme the two $E_{QM}^{model-H}$ and $E_{QM}^{el,model-H/emb}$ terms are computed together via a single QM computation on *model-H* in the bath of *emb* charges.

The second, fourth and fifth terms of Equation 3.1 (E_{MM}^{real} , $E_{MM}^{el,model/pod}$, $E_{MM}^{model-H}$) are MM energies. In particular, they can be divided into the following single contributions:

$$\begin{aligned} \text{Equation 3.2} \quad E_{MM} &= E_{MM}^{real} - E_{MM}^{el,model/pod} - E_{MM}^{model-H} = \\ &= E_{MM}^{pod} + E_{MM}^{VdW,model/pod} + E_{MM}^{bond,bend.,tors,model/pod} \end{aligned}$$

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

Two approximations are implied in this description: a) the *atom-link* scheme (i.e. the atom added to saturate the QM part) is a hydrogen link atom, b) the charges on M1 atoms (i.e. the frontier MM atoms covalently bonded to the QM region) are redistributed to their nearest-neighbours (M2).

Analytical derivatives of E^{tot} lead to the forces used for the QM part, while the MM region is optimised on the basis of a pure MM calculation. During this later (classical MM) step, the electrostatic interaction between the QM and MM part is computed using point charges derived from the QM wave function, according to the CHELPG¹²⁷ or ESP^{128,129} schemes. The charge of the H *atom-link* is redistributed over all the QM atoms in order to preserve the total charge (see Appendix A for technical details about the adopted charge redistribution schemes).

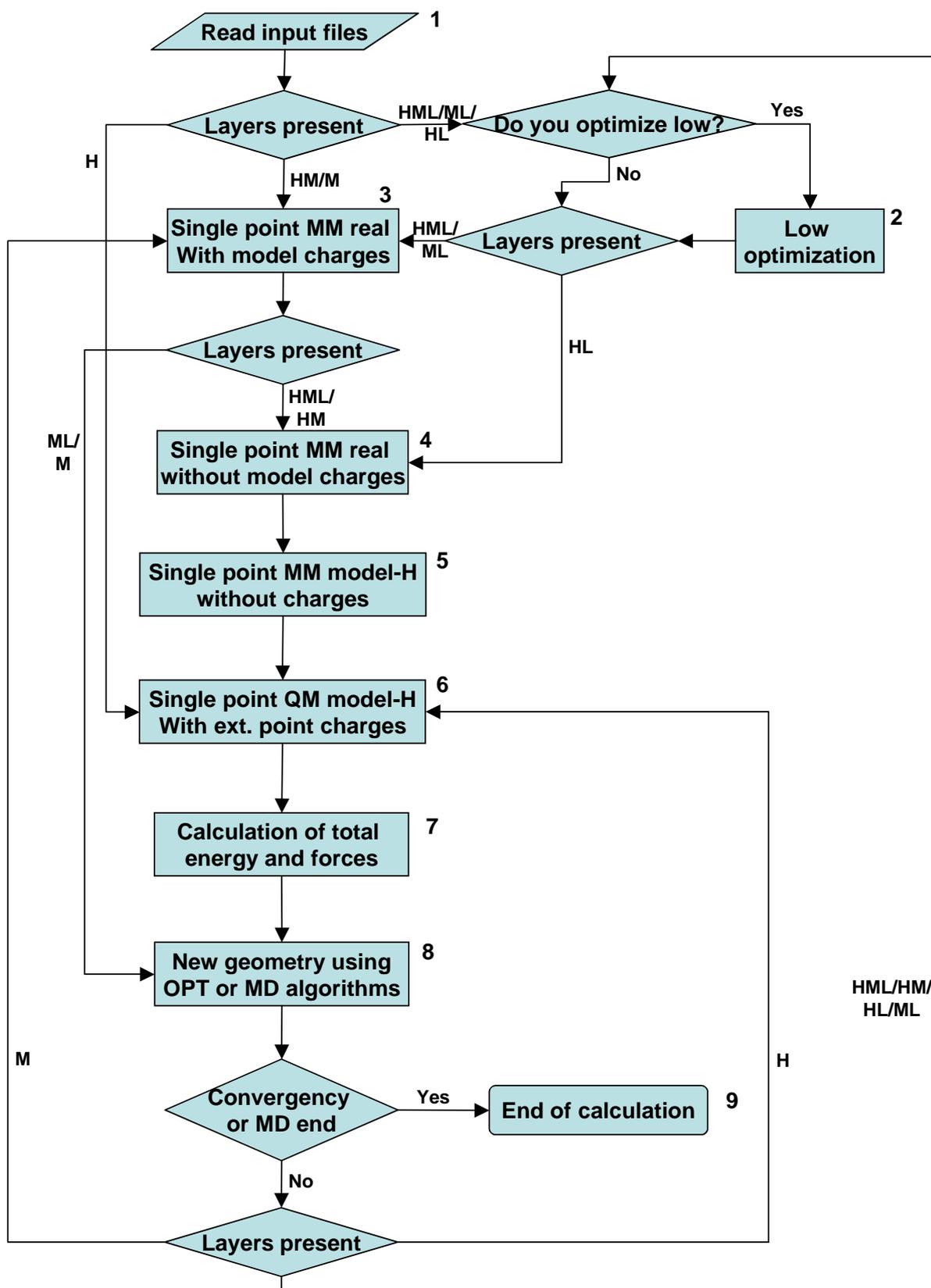
3.7 Implementation

A series of standard commercial packages such as MOLCAS,¹³⁰ Gaussian 03,¹³¹ Turbomole,¹³² ORCA^{133,134} (for QM calculations), Amber⁴⁵ and Tinker⁴⁷ (for the MM part) may be plugged to our software. In particular, an efficient strategy resulted when using Gaussian 03 and Tinker algorithms to perform, respectively, geometry optimisations or molecular dynamics on the *high-medium* region, while Amber for MM calculations (i.e. optimisation of the *low* layer and evaluation of MM energy and forces), since it allows the treatment of systems whose parameters are not included in many force fields: indeed, the standard Amber force field (*ff03*)⁴² is fully compatible with the Generalized Amber Force Field (GAFF)⁴⁹ developed to describe almost every organic molecule. More generally we found the use of the optimisation drive contained in the Gaussian 03 package to be fairly robust and performing to be used in both QM/MM and pure QM or MM optimisation, when coupled (*via* COBRAMM interface) to different packages. This allows the user to exploit some features, in the wave function calculation, for example, which are not included in the Gaussian 03 package.

To explore the Potential Energy Surface associated to a chemical problem some approaches different from geometry optimisation of the critical points are available. To perform NEB and D-NEB we interfaced to COBRAMM two of the most performing programs available, namely GMAN¹² and OPTIM.¹³⁵

The COBRAMM suite of programs have been coded using the Python scripting language.¹³⁶ The flowchart of COBRAMM is presented in Scheme 3.1: in particular, points 2, 3, 4, 5 do need external MM programs to calculate energies and forces, while point 6 uses a QM software to get the electrostatically embedded wave function and its energy and first derivatives. Point 7 merges all these results to give the total energy and gradient. Finally, in the 9th section, an external program for geometry optimisation or molecular dynamics is used to generate the new geometry to restart the cycle.

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



Scheme 3.1 QM/MM implementation flowchart.

3.8 Frequency calculation with a general QM/MM hybrid potential

To ascertain the nature of a critical point, which can be a minimum (reactants, products, intermediate) or a saddle points of index 1 (transition state) or higher, it is mandatory to compute vibrational frequencies.

Numerical frequency calculations can be carried out at the QM/MM level on the whole system (for instance the whole enzyme). In our approach only the *high-medium* region is moved in the presence of the MM potential determined by the frozen *low* region. A complete numerical frequency run (we call it “**fullfreq**” calculation) would require a total of $1+6N^{hm}$ QM/MM energy evaluations, N^{hm} being the number of atoms of the *high-medium* region. A further level of approximation, denoted here as “**freq**” calculation, has been tested in the computation of frequencies. Within this approximation we hypothesize that the small movement of an MM atom can have only a tiny effect on the wave-function. Thus, it is possible to save $6N^m$ QM computation (N^m being the number of MM atoms) by a simple sum of the current MM energy value to the reference (initial) QM value when an MM atom is moved. The QM computation is then made only when an atom belonging to the *high* layer is moved. In this way only $1+6N^h$ wave-function evaluations are required.

3.9 The fingerprint analysis

According to the adopted QM/MM approach, the bonding and non-bonding terms within the **H** region (QM region, *high* layer) can be computed by means of a single point calculation on a system (called *model-H*) composed by the *model* region with the hydrogen atoms properly placed to saturate the dangling bonds at the QM-MM boundary (according to the H-atom link approach described above). Bonding and non bonding terms of **M** and **L** layers are computed at the MM level. Some care is needed to take into account the QM-MM cross terms and many solutions are available to overcome the arising problems. We followed a general and widely applied strategy consisting in treating all cross terms (van der Waals, bonding, bending, torsions) at the MM level, except for the electrostatic ones. We adopted the Electrostatic Embedding Scheme (EES), consisting in computing the QM wave-function in the presence of the atomic point charges of **M** and **L** layers. The polarization of the wave function is believed to account for the electrostatic cross-term interaction. Under these approximations it is easy to derive a procedure able to split the cross-terms of the energy into single-residue contributions. We briefly describe here two general procedures, called Direct and Reverse Finger Print (**DFP** and **RFP**), to rank the electrostatic effects of single residues, and one (**vdWFP**) to evaluate the van der Waals contributions.

All analysis performed here give semi-quantitative information about the ability of a certain residue (or group of residues) to stabilize/destabilize one critical point with respect to another one. If, for instance, a TS is compared to the nearest minimum we can rank the effects of each residue on the barrier height for the process.

Let us consider two critical points A and B located on the QM/MM Potential Energy Surface. The overall electrostatic contribution can be easily computed as follows. QM calculations *in vacuo* on the QM region (i.e. *model-H*) alone gives the energy values E_0^A and E_0^B . From QM calculations in the presence of all atomic point charges we obtain the two energy values E_t^A and E_t^B (see Equation 3.3 and Equation 3.4). From these values, after subtraction of the corresponding “self energy of the charges” values (e_t^A and e_t^B), we obtain the two terms E_{QM}^A and E_{QM}^B . These represent, for A and B, respectively, the sum of pure QM and electrostatic cross terms, as included in our QM/MM potential.

Equation 3.3
$$E_{QM}^A = E_t^A - e_t^A$$

Equation 3.4
$$E_{QM}^B = E_t^B - e_t^B$$

The electrostatic effect of the MM regions on the QM wave-function can be estimated by E_{pol}^A and E_{pol}^B (Equation 3.5 and Equation 3.6). It is worth to notice that the lower they are, the greater is the stabilization effect played by the charges. If these quantities are lesser than zero, the polarization contribution from charges is stabilizing, while, when they are greater than zero, they represent a destabilizing contribution arising from charge interaction.

Equation 3.5
$$E_{pol}^A = E_{QM}^A - E_0^A$$

Equation 3.6
$$E_{pol}^B = E_{QM}^B - E_0^B$$

From the obtained energy values we can operate a first comparison between A and B. $\Delta E_0(A, B)$ and $\Delta E_{QM}(A, B)$ (Equation 3.7 and Equation 3.8) are the (QM) energy difference between A and B, in the absence and in the presence of the MM atomic point charges, respectively. Equation 3.9 stresses the connection between the differential stabilization effect of charges on A and B with the energetic contribution that characterizes the A to B transition. The stability factor $S_{tot}(A, B)$ (Equation 3.9) represents the magnitude of the point charge’s effect in promoting (Equation 3.10 and Equation 3.12) or discouraging (Equation 3.11 and Equation 3.13) the passage from A to B.

Equation 3.7
$$\Delta E_0(A, B) = E_0^B - E_0^A$$

Equation 3.8
$$\Delta E_{QM}(A, B) = E_{QM}^B - E_{QM}^A$$

Equation 3.9
$$S_{tot}(A, B) = \Delta E_{QM}(A, B) - \Delta E_0(A, B) = E_{QM}^B - E_{QM}^A - E_0^B + E_0^A = E_{pol}^B - E_{pol}^A$$

Equation 3.10 $E_{pol}^B > E_{pol}^A \Rightarrow S_{tot}(A, B) > 0$

Equation 3.11 $E_{pol}^B < E_{pol}^A \Rightarrow S_{tot}(A, B) < 0$

Equation 3.12 $\Delta E_{QM}(A, B) > \Delta E_0(A, B) \Rightarrow S_{tot}(A, B) > 0$

Equation 3.13 $\Delta E_{QM}(A, B) < \Delta E_0(A, B) \Rightarrow S_{tot}(A, B) < 0$

This gives a precious information about the influence of the surrounding enzyme, treated at the MM level, on the reaction process occurring within the small QM region. This data can be useful if one is able to estimate, even if qualitatively, the contribution coming from the single residues. In this way, if A and B are a minimum and the near transition state, it is possible to emphasize the residues playing the main catalytic effect,.

To handle this problem we adopted two different decomposition strategies, both able to rank the influence of the enzyme residues on a reaction step (for instance, the transformation from A to B).

The first method, called Direct Finger Print (**DFP**) analysis, is performed by means of a series of single point QM calculations (SPc) on the QM region (i.e. *model-H*) for A and B (using the optimized QM/MM geometry).

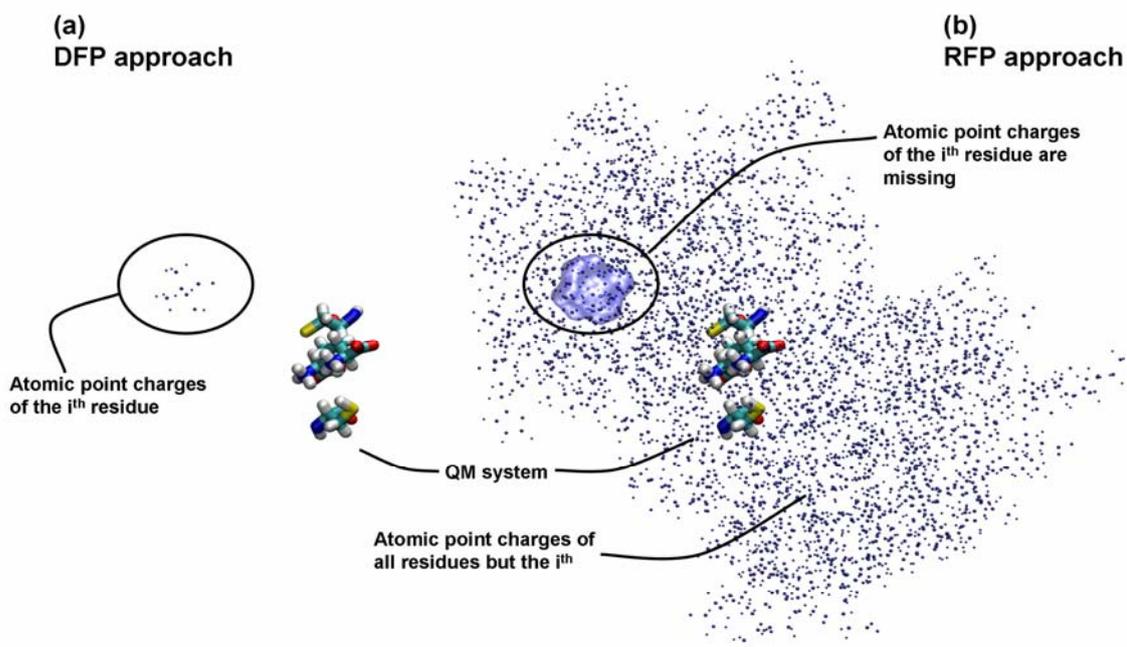


Figure 3.5 Schematic representation of the N (total number of residues) SP calculations performed in a) DFP and b) RFP analysis

After evaluation of the terms needed to compute $S_{tot}(A,B)$ (Equation 3.3 to Equation 3.13), we perform N SP calculations (where N is the total number of residues to be analyzed). In this calculations *model-H* is surrounded by the atomic point charges of the i^{th} residue only (see Figure 3.5-a). The charges are placed according to the coordinates of the atoms of the i^{th} residues. This gives N E_i^A and N E_i^B energy values, as well as the corresponding “self energy of the charges” values e_i^A and e_i^B (which represents the pure electrostatic contribution among the point charges of the i^{th} residue only). As showed above (Equation 3.5 and Equation 3.6) for the total electrostatic effects, we compute the electrostatic (i.e. Coulombic) influence ($E_{pol,i}^A$ and $E_{pol,i}^B$) of the i^{th} residue on the QM region (Equation 3.14 and Equation 3.15), which is stabilizing or destabilizing if the value is lesser or greater than zero, respectively.

$$\text{Equation 3.14} \quad E_{pol,i}^A = (E_i^A - e_i^A) - E_0^A$$

$$\text{Equation 3.15} \quad E_{pol,i}^B = (E_i^B - e_i^B) - E_0^B$$

If we compare for each i^{th} residue the values of $E_{pol,i}^A$ and $E_{pol,i}^B$, we obtain the stability parameter S_i (Equation 3.16) for the i^{th} residue. If $S_i < 0$ (Equation 3.18) then the i^{th} residue favours the transition from A to B, *vice versa*, if $S_i > 0$ (Equation 3.17), it disfavours the transition.

$$\text{Equation 3.16} \quad S_i = E_{pol,i}^B - E_{pol,i}^A$$

$$\text{Equation 3.17} \quad E_{pol,i}^B > E_{pol,i}^A \Rightarrow S_i > 0$$

$$\text{Equation 3.18} \quad E_{pol,i}^B < E_{pol,i}^A \Rightarrow S_i < 0$$

To stress the accuracy of the method in the decomposition of the total electrostatic effects into smaller components, a further passage has mandatory importance. For an ideal decomposition scheme we should have the perfect equivalence between the total electrostatic effect (E_{pol}^A and E_{pol}^B) and the summation values (E_{polsum}^A , E_{polsum}^B in Equation 3.19 and Equation 3.20) over the single contributions, namely $E_{pol,i}^A$ and $E_{pol,i}^B$.

$$\text{Equation 3.19} \quad E_{polsum}^A = \sum_{i=1}^N E_{pol,i}^A$$

$$\text{Equation 3.20} \quad E_{polsum}^B = \sum_{i=1}^N E_{pol,i}^B$$

It is easy to see that, under the adopted approximations, E_{polsum}^A , E_{polsum}^B are not equal to E_{pol}^A and E_{pol}^B , being the summation of all single-residue contributions not perfectly additive.

To estimate the entity of the error we introduce by means of Equation 3.21, Equation 3.22 and Equation 3.23 three error parameters.

$$\text{Equation 3.21} \quad X_{add,S}^A = E_{pol}^A - \sum_{i=1}^N E_{pol,i}^A$$

$$\text{Equation 3.22} \quad X_{add,S}^B = E_{pol}^B - \sum_{i=1}^N E_{pol,i}^B$$

$$\text{Equation 3.23} \quad X_{add,S}(A,B) = X_{add,S}^B - X_{add,S}^A$$

The $X_{add,S}(A,B)$ term plays an important role in estimating the reliability of our decomposition procedure. Since the errors made on A and B ($X_{add,S}^A$ and $X_{add,S}^B$) do differ, an evaluation of the error associated with the comparison between A and B ($X_{add,S}(A,B)$) is important to check the whole calculation.

The plotting of the S_i values against the corresponding residue number gives an immediate outlook of the whole enzyme influence on the A - B transformation. A further analysis based on the magnitude of the so called stabilization parameters can help in ranking the residues according to their importance in the process under examination.

The results obtained by the DFP approach can be affected by the adopted approximation. We must remember that the S_i factor is estimated by comparing the unperturbed *in vacuo* system (to obtain E_{pol}^A and E_{pol}^B) and the single-residue perturbed system (to obtain $E_{pol,i}^A$ and $E_{pol,i}^B$). To improve this description we can choose a different approach, based on an opposite scheme. This alternative approach will be referred as Reverse Finger Print (**RFP**) analysis.

As showed for **DFP** we compute the E_{QM}^A , E_{QM}^B , E_0^A , and E_0^B values, by means of single point calculation on the *model-H* system in the presence and in the absence of the whole set of atomic point charges, respectively. Here we take as reference energy values E_{dest}^A and E_{dest}^B (Equation 3.24 and Equation 3.25).

$$\text{Equation 3.24} \quad E_{dest}^A = -E_{pol}^A = E_0^A - E_{QM}^A$$

$$\text{Equation 3.25} \quad E_{dest}^B = -E_{pol}^B = E_0^B - E_{QM}^B$$

Then, we perform N SP calculations (where N is the total number of residues to be analyzed). In these calculations *model-H* is surrounded by all the atomic point charges of the system except for the ones of the i^{th} residue (see Figure 3.5-b). The charges are placed according to the QM/MM optimized geometries. This gives N $E_{h,i}^A$ and N $E_{h,i}^B$ energy values, as well as the corresponding “self-energy of the charges” values $e_{h,i}^A$ and $e_{h,i}^B$ (which represents the pure

electrostatic contribution among the point charges except for the “hole” coming from the missed i^{th} residue). As showed above (Equation 3.24 and Equation 3.25) for the total electrostatic effects, we compute the electrostatic (i.e. coulombian) influence ($E_{\text{pol},i}^A$ and $E_{\text{pol},i}^B$) due to the absence of i^{th} residue on the QM region (Equation 3.26 and Equation 3.27), which is destabilizing or stabilizing if the value is lesser or greater than zero, respectively.

$$\text{Equation 3.26} \quad E_{\text{dest},i}^A = (E_{h,i}^A - e_i^A) - E_{QM}^A$$

$$\text{Equation 3.27} \quad E_{\text{dest},i}^B = (E_{h,i}^B - e_i^B) - E_{QM}^B$$

If we compare the values of $E_{\text{dest},i}^A$ and $E_{\text{dest},i}^B$, we obtain the destabilization parameter D_i (Equation 3.28) for the i^{th} residue. If $D_i < 0$ (Equation 3.30), then the i^{th} residue disfavours the transition from A to B. On the contrary, if $S_i < 0$ (Equation 3.29), it favours the conversion itself.

$$\text{Equation 3.28} \quad D_i = E_{\text{dest},i}^B - E_{\text{dest},i}^A$$

$$\text{Equation 3.29} \quad E_{\text{dest},i}^B > E_{\text{dest},i}^A \Rightarrow D_i > 0$$

$$\text{Equation 3.30} \quad E_{\text{dest},i}^B < E_{\text{dest},i}^A \Rightarrow D_i < 0$$

For **RFP** also, in an ideal decomposition procedure we should have the perfect equivalence between the total electrostatic effect (E_{dest}^A and E_{dest}^B) and the sum over the single contributions (the $E_{\text{dest},i}^A$ and $E_{\text{dest},i}^B$ values), as stated in Equation 3.31 and Equation 3.32.

$$\text{Equation 3.31} \quad E_{\text{destsum}}^A = \sum_{i=1}^N E_{\text{dest},i}^A$$

$$\text{Equation 3.32} \quad E_{\text{destsum}}^B = \sum_{i=1}^N E_{\text{dest},i}^B$$

For the same reasons above discussed for **DFFP**, in our approach the E_{destsum}^A and E_{destsum}^B are not equivalent to E_{dest}^A and E_{dest}^B , because the contributions of the single-residues are not perfectly additive. To estimate the error we introduce three error parameters, as stated by Equation 3.33, Equation 3.34 and Equation 3.35.

$$\text{Equation 3.33} \quad X_{\text{add},D}^A = E_{\text{dest}}^A - \sum_{i=1}^N E_{\text{dest},i}^A$$

$$\text{Equation 3.34} \quad X_{\text{add},D}^B = E_{\text{dest}}^B - \sum_{i=1}^N E_{\text{dest},i}^B$$

Equation 3.35 $X_{add,D}(A,B) = X_{add,D}^B - X_{add,D}^A$

As showed before errors made on A and B ($X_{add,D}^A$ and $X_{add,D}^B$) do differ. Thus, a more meaningful evaluation of the error associated with the comparison between A and B is provided by $X_{add,D}(A,B)$.

The plotting of D_i values against the residue number shows the same information as the plot of S_i , i.e. a ranked influence of each residue in favouring/discouraging the transition from A to B. The two graphs look different, because, as previously explained, the favouring residues have a negative S -factor but a positive D -factor (the reverse for the disfavouring residues). Another significant difference is the different magnitude of the S_i and D_i single residue contribution because of the different method adopted in their computation.

Another useful tool to analyze the influence of the enzyme over the transition from A to B is the estimation of the van der Waals contribution. This analysis (called **vdWFP**) is quite straightforward, being this contribution included in the QM/MM potential at the MM level. Moreover, the single-residue contributions are additive for the MM force-field definition. To gather for two QM/MM optimized geometries (A and B) the information about the role played by the van der Waals interactions of each residue, we use a procedure similar to **DFP** and **RFP**. We computed separately the van der Waals interaction energy between each residue and the *high* layer (note that the system here is not *model*-H as in previous cases). We obtain N (the total number of residues) energy values ($E_{vdW,i}^A$ and $E_{vdW,i}^B$) for both A and B. The lower each contribution is, the greater is its stabilization. It is easy to compare the values obtained for the two structures in order to derive a stabilization factor W_i (Equation 3.36). In this case the contributions of each component are perfectly additive (this is due to the definition of the adopted MM force field). Thus, there is no need to stress on the accuracy.

Equation 3.36 $W_i = E_{vdW,i}^B - E_{vdW,i}^A$

Equation 3.37 $E_{vdW,i}^B > E_{vdW,i}^A \Rightarrow W_i > 0$

Equation 3.38 $E_{vdW,i}^B < E_{vdW,i}^A \Rightarrow W_i < 0$

As previously demonstrated, also for **DFP** and **RFP** the results of **vdWFP** can be easily represented using a plot of W_i versus the number of residues.

All tree method here described are performed by means of automatic tools integrated in the **COBRAMM** software package. Suggestion from previous works of Karplus¹³⁷ and others¹³⁸⁻¹⁴¹ have been taken to develop **DFP** and **RFP** procedures. The **vdWFP** analysis was

performed by means of multiple calculation using the software “anal” from the Amber8 package. Our computational tools were used in the present paper to derive a “per-residue” analysis, but they be can easily exploited to explore the effects arising (for a generic transformation $A \Rightarrow B$) from different groups of atoms or even to have a “per-atom analysis”. This can be useful to understand the role played in the catalysis by secondary structures (α -Helices or β -sheets or combinations of them) present in the enzyme under examination.

3.10 The “set-up” problem: a short discussion

As seen above, a QM/MM code is generally based on many assumptions and approximations and not all the molecular system can be successfully studied under these limitations. For example, extreme caution is needed in choosing the QM-MM boundary because most programs (like our own) may give unphysical results if the broken bond is not a $C(sp^3)$ - $C(sp^3)$ one. Thus, cutting a bond more polar than an aliphatic one gives problems in the charge redistribution and choosing a Q1 atom that is different from a carbon atom may affect the validity of the hydrogen atom link approach. Some technical features can be adopted to partially solve these problems, but a good idea is to find the best compromise between the right boundary choice and the size of the QM sub-region. This decision can be taken on the basis of simple considerations: a) we can extend the QM frontier until the boundary satisfies the conditions above, but one has to remember the proportionality between the size of the QM region and the cost of its calculation; b) the better system should have a small localized reactive region and some aliphatic QM-MM boundary bonds near (but not in contact with) the reactive core.

Another important issue concerns the availability of MM parameters for the system under investigation, including the QM part. In fact, given the subtractive approach used for energy computation (see Equation 3.1), all QM-MM bonded and non-bonded cross interaction terms (except for the electrostatic effects for which the electrostatic embedding procedure is adopted) are computed at the MM level. Thus, in order to properly describe these QM-MM cross interaction terms, it is important to assign good MM parameters also to the QM atoms, even if the energy and force contributions of the *high* layer are computed by mean of a QM calculation. For example, it is of fundamental importance to have the correct Lennard-Jones parameters for all the (QM) atoms so that the Van der Waals terms are correctly accounted for. Anyway, bonded cross terms (i.e. stretching, bending and torsional contributions) for a bond placed at the frontier involve only up to one, two and three atoms of the *high* (QM) layer, respectively. Thus, only the frontier atoms, and adjacent neighbours (one or two bonds distant) are important in the calculation of bonded cross interactions and MM parameters of

bonded terms are not necessary for the other QM atoms. This is important since, while the intimate nature of the atoms of the reactive core may change during the reactive process, frontier atoms are not expected (hopefully) to change, which allows the use of the same MM parameters for these atoms all throughout the reactive process.

On the basis of all these considerations, in order to get meaningful results, it would be preferable to put the frontier at least four bonds from the reacting region, if possible. Unfortunately, this is often hard to do, due to the overwhelming size of the corresponding QM region and one has to make a compromise.

Charges represent another delicate point in system set-up. Since the electrostatic QM-MM cross interactions are calculated at the QM level (in the Electrostatic Embedding scheme), the results may be very sensitive to the charge distribution, even away from the reactive region. Many residues in an enzyme have an acidic or basic character and, according to their protonation state (which is strongly influenced by the close environment), they can carry a net negative or positive charge. This may lead to a system with a net (non zero) charge. Two main questions arise: it is possible to systematically assign the protonation state of the residues? And how is the best way to deal with a net charge in the system? The answers are not unique and strongly depend on the system under study. In general, the easiest procedure is to manually assign the protonation state of the residues that are part of the active site (by means of some available experimental data, for example), which are the most important ones, while some more general procedures may be used to assign the protonation state of all other residues (e.g. based on the pH of the surrounding media or of the local environment, such as using the H++ algorithm¹⁴²): they are easy to use but care is needed for the active site region and a supervision of the final result is highly encouraged to correct for some incompetence of the software. Often, after this protonation phase, a net (non zero) charge exists in the system. To avoid problems in QM/MM calculations some authors have proposed to scale the charges on titratable groups in order to avoid the presence of a net charge on the system. We think this can be a good procedure if these residues do not play an interesting electrostatic effect on the reactive process under study; but it is not a general method. Another procedure, widely used in MM Molecular Dynamics, is to assign a right number of counter-ions (usually Na⁺ or Cl⁻) at the surface of the protein, in order to correct for the charge imbalance and have an electro-neutral system.

4 Test Examples

4.1 Introduction

To confirm the effectiveness of this implementation, comparative tests are mandatory. The first is a comparison of computed and experimental geometrical parameters for a complex molecular architecture such as that of a rotaxane, whose structure and functionality depends on weak (hydrogen bonds) interactions. Then, it will be shown that the optical properties of two well known and widely studied biological chromophores (i.e. the rhodopsin and green fluorescent protein chromophores) are nicely reproduced both in solution and in the protein. The flexibility and potentiality of our code (namely its capability to perform optimizations as well as molecular dynamics at both QM, MM and QM/MM level) will be exploited for the investigation of the delicate equilibrium in water of the two (α and β) anomers of D-glucopyranose. Then, the QM energy profile of the reaction mechanism of HIV1-Protease will be compared to the QM/MM one, analyzing and discussing the differences obtained and the reliability of the results; we will show how inclusion of the enzymatic environment is essential for the correct description of the enzymatic process. Finally the capability of different potentials in describing medium-sized (c.a. 70 atoms) organo-metallic complexes will be tested.

4.2 Optimisation of a complex molecular structure: a fumaramide based rotaxane

The supra/super molecules are often characterized by a very high flexibility; in fact, the weak interaction between subunits leave the system free to move in a quite flat potential energy surface. The investigation of this class of molecules can be very hard in particular using a QM/MM approach where the forces (which drive the research of the minima) are calculated using different methods in different regions of the molecule.

This section has no other purpose than showing the accuracy of the current QM/MM implementation in the description of a structural problem for a delicate supermolecular system, i.e. a rotaxane based on a fumaramidic group¹⁴³. Here, the thread and the ring interact by hydrogen bonds (which is where the QM-MM boundary is passing through) and π stacking (see Figure 4.1).

Only the central dimethyl-fumaride has been treated at the QM level (this is the photo-reactive part whose activity must be accounted for by QM methods) using HF/6-31G*, while the rest of the thread and the ring are computed using a classical (MM) force field (GAFF⁴⁹). The main part of the interactions between the thread and the ring comes from QM and MM cross terms.

Fully unconstrained optimisations have been done according to a two layers *high* and *medium* (HM) partitioning scheme.

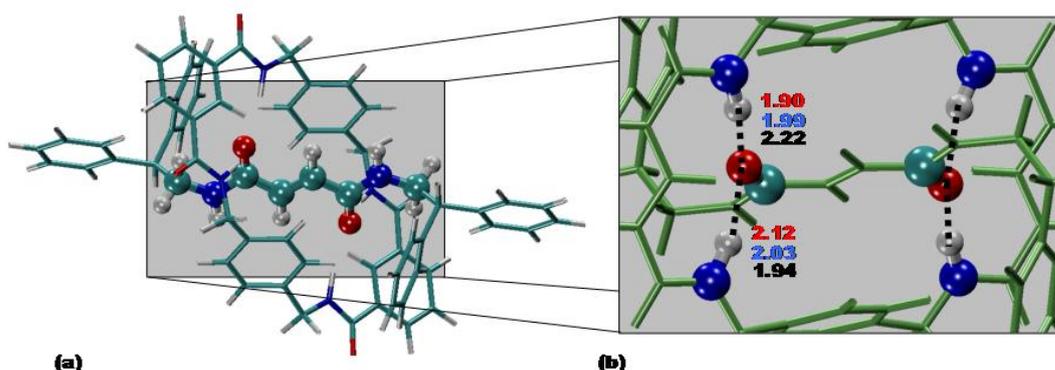


Figure 4.1 Rotaxane ground state structure. a) The dimethyl-fumaramide (ball&stick) is treated at the QM level while the rest (tube) is calculated at the MM level. b) Hydrogen bonds (dashed lines) linking the ring to the thread: QM/MM (red, up), mm3 (blue, centre) and experimental X-ray (black, low) values.

The results are in good agreement with the experimental X-ray¹⁴³ structure and previously published MM3¹⁴⁴ values.¹⁴³ Specifically, we have selected hydrogen bond distances as a measure of the accuracy; the errors in the hydrogen bond distances between the QM and MM regions are $\leq 0.3 \text{ \AA}$.

It can be concluded that this QM/MM approach may be a reliable method for structure determination also in cases where the molecular structure does depend by a subtle balance of factors, and small errors in the forces (such as the ones at the frontier region) may lead to large geometrical variations, like it is the case for this rotaxane.

4.3 The anomeric equilibrium of D-Glucopyranose.

The simulation of solvation is a crucial problem in general, and in biological systems in particular, since here polar protic solvents (i.e. water) are usually involved, which strongly interact with the solute. Two different computational ways can be used for simulating solvation: (i) the implicit solvent and (ii) the explicit solvent model. Polarizable continuum solvent (PCM)¹⁴⁵ for QM calculations and Generalized Born (GB) method¹⁴⁶ for MM calculations are examples of the former, while the latter implies that solvent molecules are explicitly taken into account to obtain the energy of the system. Obviously, the computation of explicit solvent is more time consuming with respect to the implicit approach and can be performed only by considering the solvent molecules at a low (such as MM) level of theory. Thus, a simple way to do this is to use QM/MM computations where the solvent bulk and the solute are treated at the MM and QM levels, respectively.

In this section, the code will be applied to a well known and delicate equilibrium problem: the relative stability of the two anomers of the pyranoid form of D-Glucose (i.e. α - and β -D-Glucopyranose, see Figure 4.2) in water. Rather than focusing on the result itself (which will be shown to go in the right direction, anyway), here we will stress on the flexibility of the implemented algorithm, showing how geometry optimisations and molecular dynamics computations by pure MM or QM/MM techniques may be successfully applied to the aforementioned study.

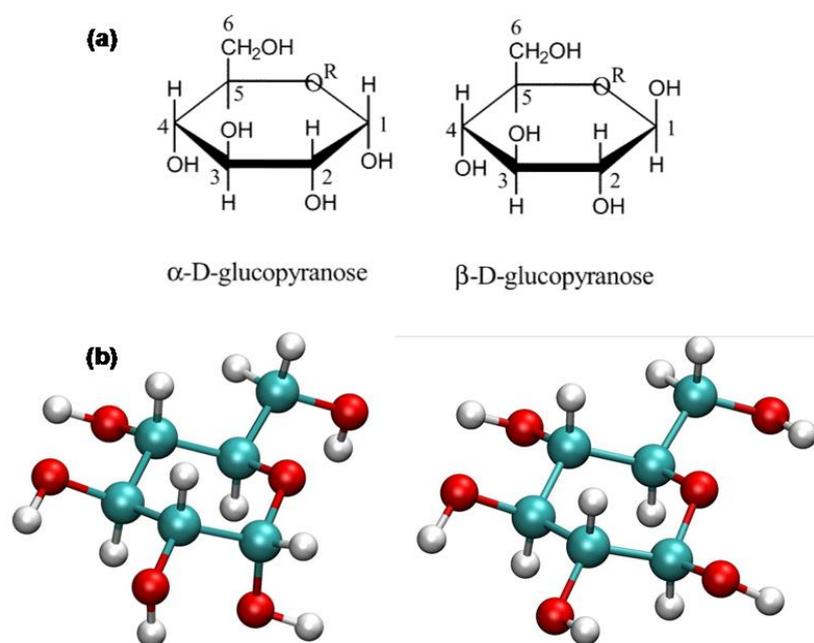


Figure 4.2 (a) Haworth and (b) ball & stick representations of the α - and β -D-glucopyranose.

Many studies¹⁴⁷⁻¹⁵⁰ in the past have focused on this topic revealing all the difficulties in simulating carbohydrates because of the high conformational freedom of their hydroxyl and hydroxymethyl groups; also, the interaction with polar protic solvents, like water, play a key role in determining the relative stability of the different isomers, but this is often a subtle and difficult term to estimate. Because of the high computational cost of a complete exploration of the conformational space of this system, we used only a small number of conformers chosen among the most stable ones. We found two minima, on the basis of geometry optimisations at the QM (DFT/B3LYP/6-31G*)^{75,151} level, simulating solvent by means of the SCRF-CPCM¹⁵² method. We used these geometries as a guess to create the model system used in the further studies. In all the model systems investigated here the solute molecule (α - or β -D-Glucopyranose) is surrounded by a 20 Å radius drop of water (1490 water molecules): a shell (**shell1**) of 10 Å around the solute is left free to move during the optimisation or MD runs, while the water molecules in the outer 10 Å shell (**shell2**) are fixed to their initial geometries (see Figure 4.3). Starting solvent geometries are obtained adding the water molecules to the solute by mean of the “leap” module of the AMBER8 program,⁴⁵ which uses a standard equilibrated TIP3P^{1,3} water library. It is worth noting here that **shell2** prevents molecules of **shell1** to spread out (which would be unphysical), thus acting as a physical barrier during all the calculations.

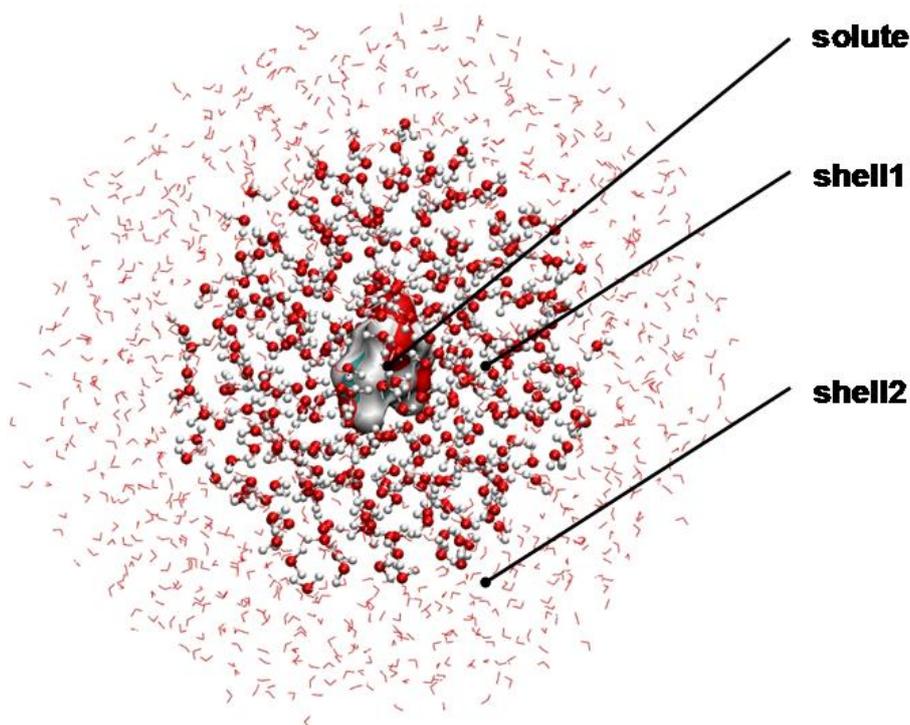


Figure 4.3 Schematic representation of the model system used to study the (α - and β -) D-glucopyranose (solute) in solution. The solvent molecules (water) are partitioned between shell1 (free to move) and shell2 (kept frozen to the initial geometry).

4.3.1 The L level optimisation.

We performed full MM minimization using the standard *ff03*⁴² force field and a flexible TIP3P model for the water molecules and the Glycam¹⁵³ parameters for the solute obtaining a $\Delta E=(E_{\beta}-E_{\alpha})$ value of about 21 kcal/mol in terms of Internal Energy (with α more stable than β), which is in complete disagreement with the experimental value of about -0.4 kcal/mol in terms of Gibbs Free Energy (the experimental ratio between α and β is 36:64). This result shows that a simple MM optimisation cannot reproduce the relative stability of the two anomers and that the force field is not accurate enough to account for solute/solvent interactions.

4.3.2 The HL level optimisation.

We performed QM/MM calculations with a two layer (*high* and *low*) approach (HL): *high* is constituted by one α - or β -D-Glucopyranose molecule treated at the QM (DFT/B3LYP/6-31G*)^{75,151} level; *low* contains all the MM solvent molecules (*ff03* force field and flexible TIP3P water model)⁴². After geometry optimisation of the solute (*high* layer) and **shell1** (which, together with **shell2**, does represent the *low* layer) we computed the energy difference between the two anomers. The calculations were repeated with different initial configurations

of the mobile water shell and we found the two anomers to be almost isoenergetic, with oscillations of the $\Delta E=(E_{\beta}-E_{\alpha})$, in terms of Internal Energy, between 1.5 and -1.5 kcal/mol. Considering the approximations involved in our approach, this is a quite good result for such a straightforward calculation because it is near to the experimental evidence. Anyway, the relative energy of the two conformers is strongly affected by the conformation assumed by their free -OH groups, which is in turn affected very much by the surrounding solvent molecules. From these simple considerations, it results that optimisations of local minima can only deliver snapshots of the real behaviour and that a dynamical treatment could be necessary.

4.3.3 MM Molecular Dynamics: the ML level

To explore the interaction between solute and solvent molecules we performed short MD simulations on a full MM system. The *ff03*⁴² force field and a flexible TIP3P model was used for the water molecules and the Glycam¹⁵³ parameters for the solute. The outer 10 Å-shell (**shell2**) water molecules were assigned to the *low* layer (which was kept frozen throughout all the computations), while the inner ones (**shell1**) plus the solute constituted the *medium* layer. The NVT Molecular Dynamics run of the *medium* layer was performed using the “Beeman” algorithm of the Tinker program with a time step of 1 fs for a total time of 3 ps at the temperature of 298 K. It is worth noting that the implemented code does exploit the Amber8 package for energy and first derivatives MM computations, while the Tinker program is used for MD: this grants a great flexibility to the procedure since many common MD algorithms do fail in handling a largely frozen system (e.g. our attempt to run a few ps MD run with the standard “sander” program - the molecular dynamics module of Amber8 - with the **shell2** kept frozen resulted in the blowing-up of the system). An ML calculation with our code gives the possibility to study large systems keeping frozen all (or part of) the *low* layer. This is possible because the MD run (or the optimisation) is performed on the *medium* layer in the presence of the geometry of *low*, which acts as a barrier (with its associated force field) around the smaller *medium* layer. The same considerations are valid if also the *high* layer is present.

The equilibration phase covers the first 1500 steps (1.5 ps) for both anomers (see Figure 4.4). From the obtained trajectories it is possible to observe the high mobility of all hydroxyl groups, while the methoxylic group conserves their initial conformation. Averaging the total energy (potential MM energy plus kinetic energy) over the last 2.5 ps we obtain a value for ΔE of 16.34 kcal/mol (see Figure 4.4-a). This value is very similar to the result obtained from

geometry optimisation at the MM level and is still in contrast with experimental evidence calling (unrealistically) for a more stable α anomer.

4.3.4 QM/MM Molecular Dynamics: the HML level

A three layers approach was used for QM/MM molecular dynamics computations: the solute (D-Glucopyranose) was assigned the *high* layer and the QM (DFT/B3LYP/DZVP) level; the **shell1** and **shell2** water molecules were assigned the *medium* (free to move) and the *low* (frozen) layer, respectively, and the MM level (*ff03*)⁴² was used. Starting from the pure MM MD results (geometries, velocities, accelerations) described in the previous section, we performed a MD (NVT) run of 3 ps at a temperature of 300 K with a time step of 1 fs using the same modular (AMBER//TINKER) approach as described above. The system equilibrates in 1.5 ps and the production takes 1.5 ps. By averaging the total energy (potential QM/MM energy plus kinetic energy) over the last 1.5 ps, a value of 6.47 kcal/mol for ΔE was obtained (see Figure 4.4-b). These results, although not yet in agreement with the experiments, are much better than the MM ones. In the present case we believe that a single and short (3 ps) QM/MM – MD run is inadequate to simulate the overall physical behaviour of solvated D-Glucopyranose, because the conformational space is not fully explored. The results can be improved by running multiple trajectories, each one starting from a different conformational minimum. However, a better reproduction of the experimental data has already been accomplished in the past¹⁴⁷ and is beyond the purpose of our short discussion. Herein, we wanted only to show a possible application of our code to a well known solvation problem, pointing its flexibility in treating a given system with different approaches. In the other sections some quantitative examples of the reliability of our code, as applied to both optimisation and molecular dynamics, are given in details.

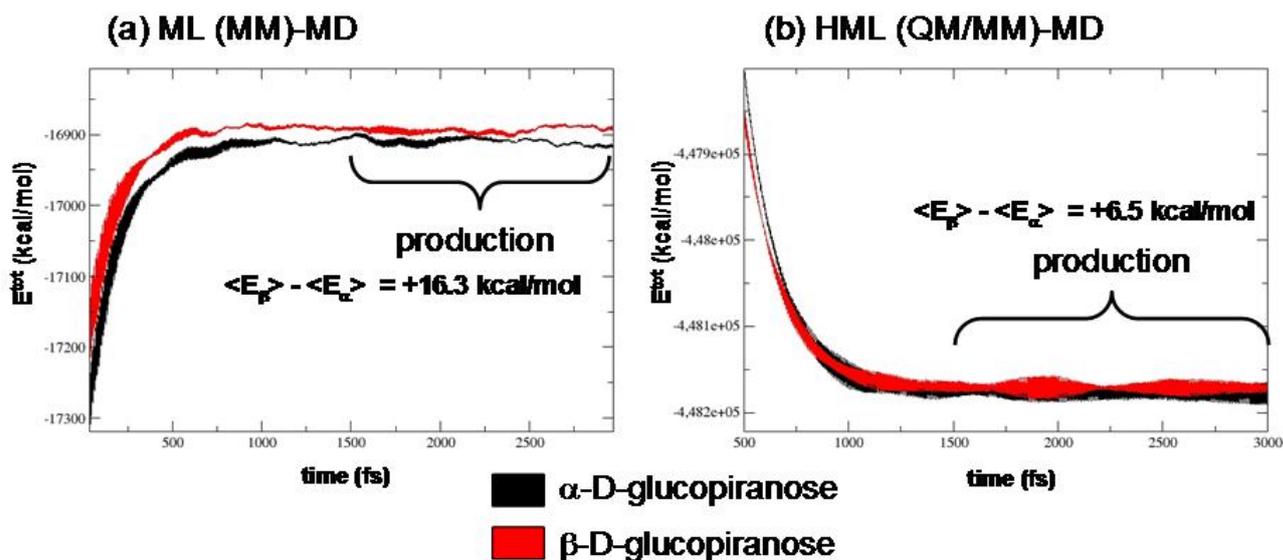


Figure 4.4 Molecular Dynamics results from (a) the MM system (ML calculation type) and (b) the QM-MM system (HML calculation type). The behavior of the α - (red line) and β - (black line) anomers is reported, highlighting the production part of each run. The difference between the averaged energies of the two anomers is also reported.

4.4 Rhodopsin and GFP chromophores: solvent and protein

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

Two biologically important chromophores belong to this class of compounds: the anionic 4'-hydroxybenzylidene-2,3-dimethyl-imidazolinone (HBDI, see Figure 4.5-a) and the protonated Schiff base of retinal (RPSB see Figure 4.5-b), which are the chromophores of the green fluorescent protein (GFP) and rhodopsin proteins, respectively. This section focuses on the description of the optical properties of these systems in different environments by means of QM/MM computations. An advantage is that both experimental and computed spectra are available in *vacuo*, solvents, as well as protein¹⁵⁴⁻¹⁵⁸. Thus, it is possible to compare (and validate) our QM/MM results against the ones already appeared in the literature.

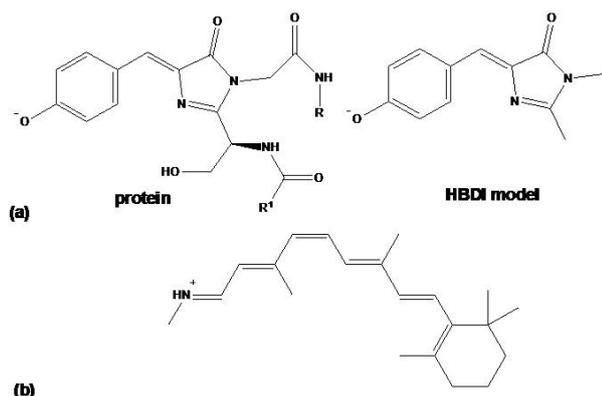


Figure 4.5 Chemical structure of the a) GFP chromophore and its isolated model (HBDI) and b) the protonated Schiff base of 11-*cis* retinal.

4.4.1 Structure and optical properties for solvated HBDI and GFP: flexible QM/MM optimisations

Thanks to its highly fluorescent chromophore and the fact that this is directly formed from a tri-peptide unit (Ser65-Thr66-Gly67)¹⁵⁶ during the protein folding without cofactors, GFP is widely used in biochemistry as a fluorescent marker¹⁵⁶. The anion is considered to be the active form of the chromophore and, while in protein it strongly fluoresces, in solution the fluorescence is quenched. Therefore, the study of this chromophore in different environments is crucial to understand its photochemical properties.

4.4.2 Solvation by HL and HML QM/MM computations

HBDI (Figure 4.4-a) has been chosen as a model for the GFP chromophore because it preserves all the conjugate system as well as the structure of the original protein-embedded chromophore and it has been widely investigated in water solution. QM/MM computations have been used to simulate solvation with the chromophore (anionic HBDI) embedded in a 16 Å TIP3P water box, using an MM Na⁺ as the counter ion (as it is experimentally).¹⁵⁴ The CASSCF/6-31G*¹ level has been extensively used for the chromophore during QM/MM optimisations and then the energy has been refined to account for correlation effects using single point CASPT2 computations. This CASPT2//CASSCF approach, which defines the *high* (QM) layer, has been widely used for the photochemistry of organic chromophores (proving to give reliable results of experimental accuracy)¹⁵⁹ and represents the state of the art to calculate excited state properties of organic systems. In order to obtain a good compromise between speed and accuracy, the full active π -space (16 electrons in 14 orbitals) has been reduced (according to a previous paper¹⁶⁰) to a 12 electrons in 11 orbitals, where the orbital localized on the amidic group has been removed together with the highest and lowest energy π -orbitals of the benzene ring.

To generate a good starting point we have performed a classical molecular dynamics with periodic boundary conditions. While the chromophore was kept frozen at its original QM geometry *in vacuo* with point charges obtained using the AM1-BCC procedure,^{161,162} the solvent and counter ion were kept free to move, so that during the 1 ns MD the counter ion is able to diffuse in the water bulk. The lowest potential energy geometry was selected along the trajectory as a representative point of the sampling and as starting guess for QM/MM computations.

Three different QM/MM optimisation types were performed on the system according to different and increasing levels of mobility (i.e. the parts comprising the atoms that are free to move and are optimised can be changed according to the partitioning scheme used): (i) HL optimisations, where only the *high* (QM) layer (i.e. the chromophore) is optimised, while all the other (MM) waters (i.e. the *low* layer) are kept frozen; (ii) HML optimisations, where the *high* (QM) layer (i.e. the chromophore) and a *medium* (MM) layer (i.e. the water molecules surrounding the chromophore up to a 3.5 Å radius) are optimised together, with all the other (MM) waters (i.e. the *low* layer) kept frozen; (iii) HML2 optimisations, where the *high* (QM) layer (i.e. the chromophore) and a *medium* (MM) layer (i.e. the water molecules surrounding the chromophore up to a 3.5 Å radius) are optimised together, while all the other (MM) waters (i.e. the *low* layer) are free to move (independently with respect to the first two ones) up to a 10 Å radius around the chromophore (and the others are kept frozen to their initial positions). These three different computational levels (which are identified in Table 4.1 as HL, HML1 and HML2, respectively) highlight the high flexibility of the optimisation procedure implemented in the current QM/MM code and, very remarkably, are shown to produce results of increasing accuracy (see Table 4.1).

Table 4.1 Absorption maxima for the anionic GFP chromophore.

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

a) ref. ¹⁶³; b) ref. ¹⁶⁰; c) ref. ¹⁵⁹; d) ref. ¹⁵⁵; e) ref. ¹⁶⁴; f) ref. ¹⁵⁴ g) ref. ¹⁶⁵

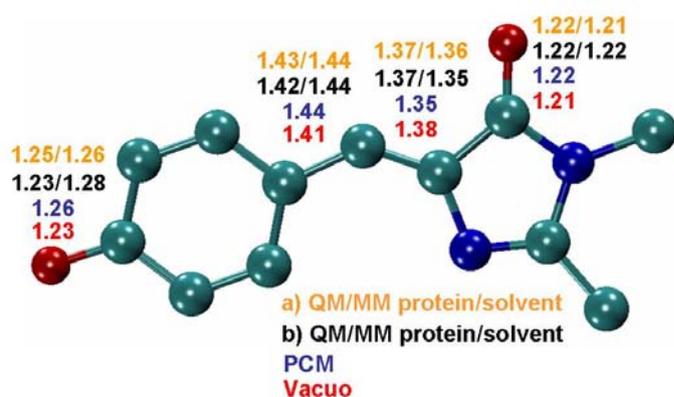


Figure 4.6 Bond distances values calculated for the GFP chromophore in *vacuo*, solvent (PCM and QM/MM) and protein: a) QM/MM values¹⁵⁹ and b) our QM/MM results.

It is worth noting that the counter ion (Na^+) is 7 Å away from the chromophore in all the QM/MM optimised structures, revealing a weakly bound ionic pair. Although this is only one of the many possible stable solute/solvent configurations, it is remarkable that the computed absorption energy (499 nm, 460 and 441 nm for (i), (ii) and (iii), respectively) is progressively getting very close to the experimental value recorded for the chromophore in water (425 nm) and does perfectly account for the blue shift observed on going from *vacuo* (where, again, the computed 465 nm absorption maximum^{159,160} does nicely reproduce the experimental value of 479 nm¹⁵⁴) to water (see Table 4.1). In Figure 4.6 are reported the relevant bond distances for HBDI *in vacuo*, explicit solvent and PCM.^{159,160,163} It is apparent that the solvent has also a key role in tuning the chromophore molecular structure, since it shifts the geometry to the individual resonance structure carrying the negative charge on the phenolic oxygen, while a resonance hybrid is more likely to represent the chromophore *in vacuo*. Furthermore, it results that implicit (PCM) as well as explicit (QM/MM) techniques do yield similar results.^{159,160}

4.4.3 GFP by HML QM/MM computations

Finally, QM/MM computations have been used to investigate the geometry and absorption maximum of the chromophore in the protein (GFP). The crystallographic structure (code 1GLF)¹⁶⁶ available in the protein data bank archive has been selected. After protonation, using the $\mathbf{H}++$ ¹⁴² procedure, hydrogen atoms have been relaxed in order to get a reliable starting structure. A three-layer (HML) partitioning scheme has been adopted: the *high* (QM) layer (i.e. the electrostatic embedded chromophore, see Figure 4.7) and the *medium* layer (see tube representation in Figure 4.7) constitute the mobile part (following the procedure of Sinicropi et al.¹⁵⁹), while all the remaining (MM) protein is kept frozen and represents the *low* layer. The same QM level as for HBDI in solution has been adopted for the chromophore, while

protein and water molecules are computed with the Amber force field. Very remarkably, the computed 504 nm absorption maximum for the I anionic state of wild type GFP does nicely reproduce the experimental value of 495 nm¹⁶⁵ (see Table 4.1). The same applies also for the chromophore structure (see Figure 4.6).

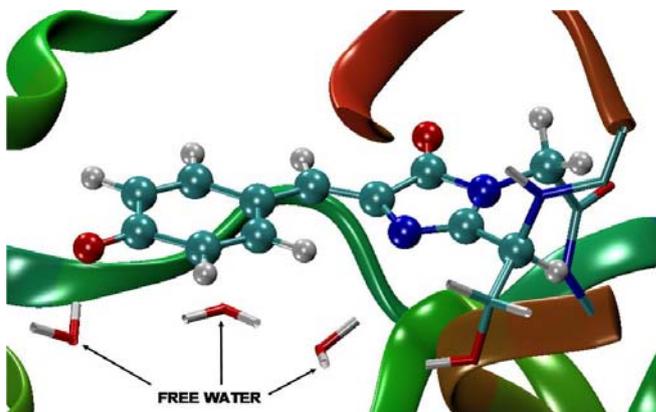


Figure 4.7 Model system used for the QM/MM study of the anionic GFP. Ball&Stick is the QM part, the tube is the movable *medium* layer, while the remaining atoms are kept frozen.

4.4.4 11-*cis* retinal PSB in solution

A similar QM/MM investigation as before has been performed for solvated RPSB. Specifically, in the present study our effort is focused to simulate the effect of methanol in the spectral tuning of the 11-*cis*-retinal chromophore. Thus, for this purpose, QM/MM computations have been used with the chromophore (RPSB) embedded in a cubic 45 Å methanol box, using an MM Cl⁻ as the counter ion (as it is experimentally).¹⁵⁸ Similarly to the examples above and previous computational works,^{167,168} a CASPT2//CASSCF(6-31G*)¹ approach (with electrostatic embedding) has been adopted for the chromophore (using the full π -space of 12 π -electrons in 12 π -orbitals), which defines the *high* (QM) layer, during QM/MM optimisations.

To generate a good starting point for QM/MM optimisations we have performed a classical molecular dynamics with periodic boundary conditions. While the chromophore was kept frozen at its original QM geometry *in vacuo*¹⁶⁹ with point charges obtained using AM1-BCC procedure,^{161,162} the solvent and counter ion were kept free to move, so that the counter ion is able to diffuse in the methanol solvent. After a MD run of 1 ns, we see (at least for this specific trajectory) that most of the time the chloride counter ion resides near the retinal nitrogen (i.e. close to the cationic head, in a tight ionic pair arrangement). Thus, a low energy snapshot with the counter ion close to nitrogen has been selected as a representative point of the sampling and as starting guess for QM/MM computations.

A two layers (HL) QM/MM optimisation has been performed: the *high* (QM) layer is given by the chromophore (as stated above), while all solvent molecules and the counter ion

constitute the *low* (MM) layer (thus excluding a *medium* buffer), which has been left free to move independently up to a distance of 10 Å from the chromophore (while the rest is kept frozen at its initial geometry).

Remarkably, the resulting absorption maximum (453 nm) is in good agreement with the experimental value of 440 nm¹⁵⁸ and with previous QM/MM computations¹⁷⁰ (see Table 4.2), but in this case the optimised geometry does not undergo significant variations on going from *vacuo* to solution (see Figure 4.8): while in HBDI geometrical modifications come from a different weight of the two resonance structures, in the retinal chromophore the structure does not change a lot, because the geometry is described only by one resonance structure (i.e. the one with the positive charge on nitrogen). This means that the change of the absorption spectrum on going from the *in vacuo* to the solvent (see Table 4.2) is addressed only to the perturbation of the wave function by the electrostatic effects of the external point charges and that electrostatic embedding is enough to get a correct description of the chemico-physical properties of the chromophore in solution.

Table 4.2 Absorption maxima for RPSB.

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

a) ref.¹⁷¹; b) ref.¹⁷⁰; c) ref.¹⁵⁷; d) ref.¹⁵⁸

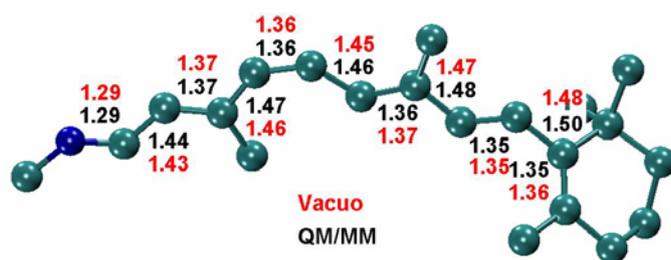
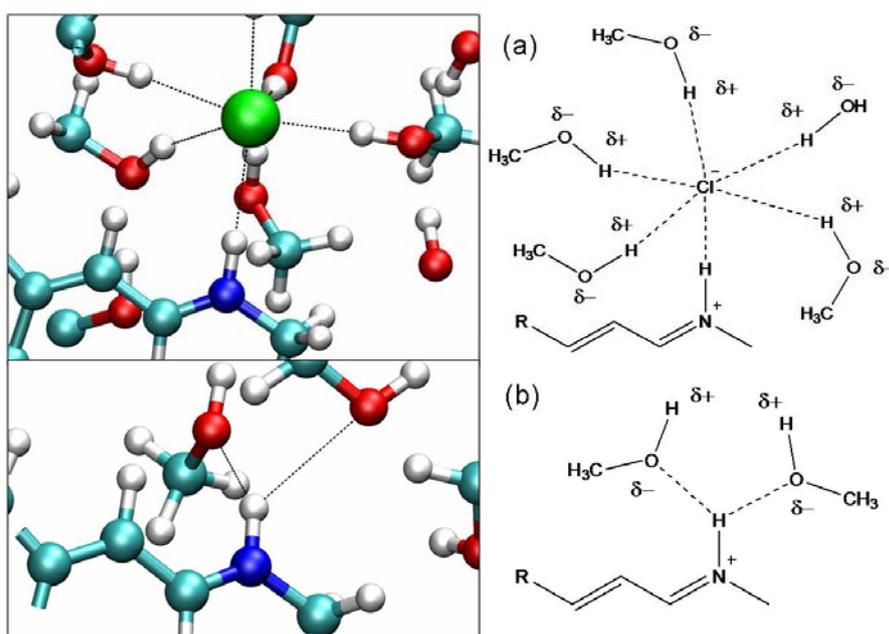


Figure 4.8 Bond distances computed for RPSB *in vacuo* (red) and in protein (black).

To explore the effects of the counter ion (Cl⁻) position on the absorption energy of the chromophore, we have also investigated the optical properties of a loose ionic pair: thus, a low energy snapshot with this arrangement has been selected along the previous MD run as a starting point for new QM/MM optimisations (which have been performed at the same HL level as before). Very interestingly, the absorption maximum computed for this QM/MM optimised loose ionic pair is substantially unchanged (442 nm) (see Table 4.2) and in agreement with the experimental value.

4.4.5 Solvated ion pairs: structure and properties

Interestingly, our results for the optimised geometry and the absorption maximum of solvated HBDI match perfectly (see Figure 4.6 and Table 4.1) with the ones reported in previous QM/MM computations¹⁵⁹ where an analogous CASPT2//CASSCF approach was employed (see Table 4.1). Anyway, it is worth noting that there the counter ion was forced to stay at the vicinity of HBDI (i.e. a tight bound ionic pair arrangement was selected), instead of being far away as it is in our case. The same results appear for solvated RPSB, where the close and the loose ion pairs do lead to very similar optical and structural properties. This leads to the remarkable conclusion that the solvent in the loose ionic pair interacts with the solute as if it were the real counter ion, i.e. solvation shells form a *virtual counter ion* that has the same (electrostatic) effect as the real one in a tight bound ionic pair. This effect is due to the reorientation (i.e. polarization) of the polar solvent close to the chromophore (see Scheme 4.1): the polarized permanent dipoles of the solvent act similarly to the counter ion, stabilizing the ground state respect to the charge transfer excited state.



Scheme 4.1 Schematic representation of the interaction between solvent, PSB and counter ion: a) tight bound ionic pair, b) dissociated couple.

This discussion implies that tight bound ionic pairs may be seen as qualitatively good models for solvated charged chromophores in general, as anticipated in our previous works.^{167,172} This study represents a further validation of the aforementioned statement and is in line with previous suggestions by Sakuray and coworkers.^{173,174}

4.5 Comparison between QM and QM/MM results for an enzymatic reaction mechanism

In this section QM and QM/MM results will be compared. The system under study is the homodimeric protease of the HIV1 virus (HIV1-Pr). This enzyme catalyzes the proteolysis of a long peptide¹⁷⁵⁻¹⁷⁷ synthesized from viral reverse-transcribed DNA and it is essential for viral replication. In the last two decades many studies^{138,178-185} have been addressed to understand the catalytic process of this enzyme in order to speedup the rational design of new inhibitors. The enzymatic reaction mechanism involves the Asp25 and Asp25' residues (each one belonging to one of the two identical sub-units of the dimeric enzyme); their side chains are symmetrically disposed in the active site, sharing a proton and a negative charge; they strictly bound a water molecule, forming a low barrier hydrogen bond system (see Figure 4.10). Both experimental¹⁸⁶⁻¹⁸⁸ and theoretical^{175,189} results suggest that the hydrolysis of the amide bond takes place after the formal addition of the water molecule, leading to a metastable gem-diolic intermediate. The aspartic dyad is believed to play the key role in the catalysis by activating both the amidic substrate and the water molecule via a double proton transfer (i.e. carbonylic group protonation and water molecule deprotonation, respectively).

4.5.1 QM reaction profile

To explore the details of the reaction mechanism we performed a pure QM investigation of the potential energy surface of a small model system by means of geometry optimisations. We included in this model the catalytic dyad, some nearby fragments, the nucleophilic water molecule, the Phe-Pro dipeptide (as a model of the substrate) and a structural water molecule bounded to the substrate (Figure 4.9-a). This model system (MOD_QM) was constructed on the basis of the crystallographic pdb data file (1G6L)¹⁹⁰ taking the Cartesian coordinates of the interesting atoms and discarding the rest of the enzyme. Missing hydrogen atoms were added to saturate the valence of dangling bonds. Our previous experience^{191,192} on QM studies of enzymatic systems encouraged us to partially consider the (steric) effect of the neglected part of the system by keeping frozen the border atoms of MOD_QM to their crystallographic position (Figure 4.9). This has proven to avoid unrealistic distortions of the model system due to additional degrees of freedom. The MOD_QM model system was investigated at the DFT/B3LYP level and, to reduce the computational cost, we assigned three different basis-sets to the atoms (see Figure 4.9-b), on the basis of their relevance in the reactive process. We assigned the most accurate basis-set (DZVP) to the atoms of the Asp dyad, to the nucleophilic water molecule and to the peptidic bond of the substrate, because they are involved in the

bond breaking/forming processes. The outer atoms were treated with a low level basis set (STO-3G), while for the other atoms we used an intermediate level (3-21G*).

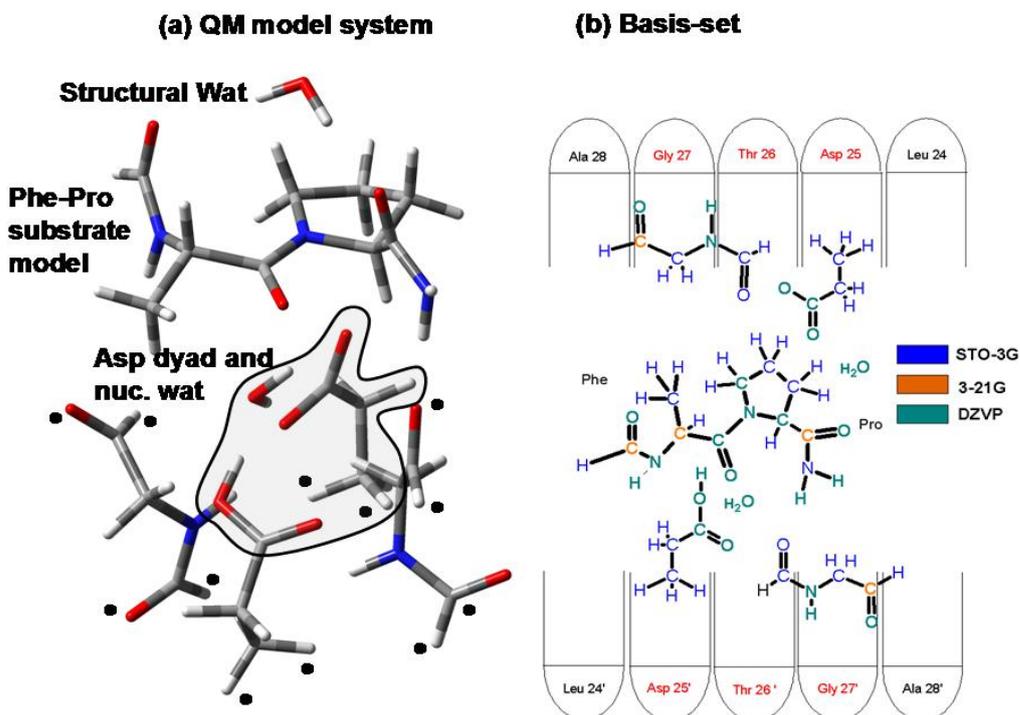


Figure 4.9 a) Model system used for the QM study; the • symbol is used to highlight the atoms kept frozen to their initial positions. b) The basis set adopted for the atoms of the QM system is shown using different colors. The *model-H* region used in QM/MM computations do coincide with the QM model system and the basis set adopted is the same.

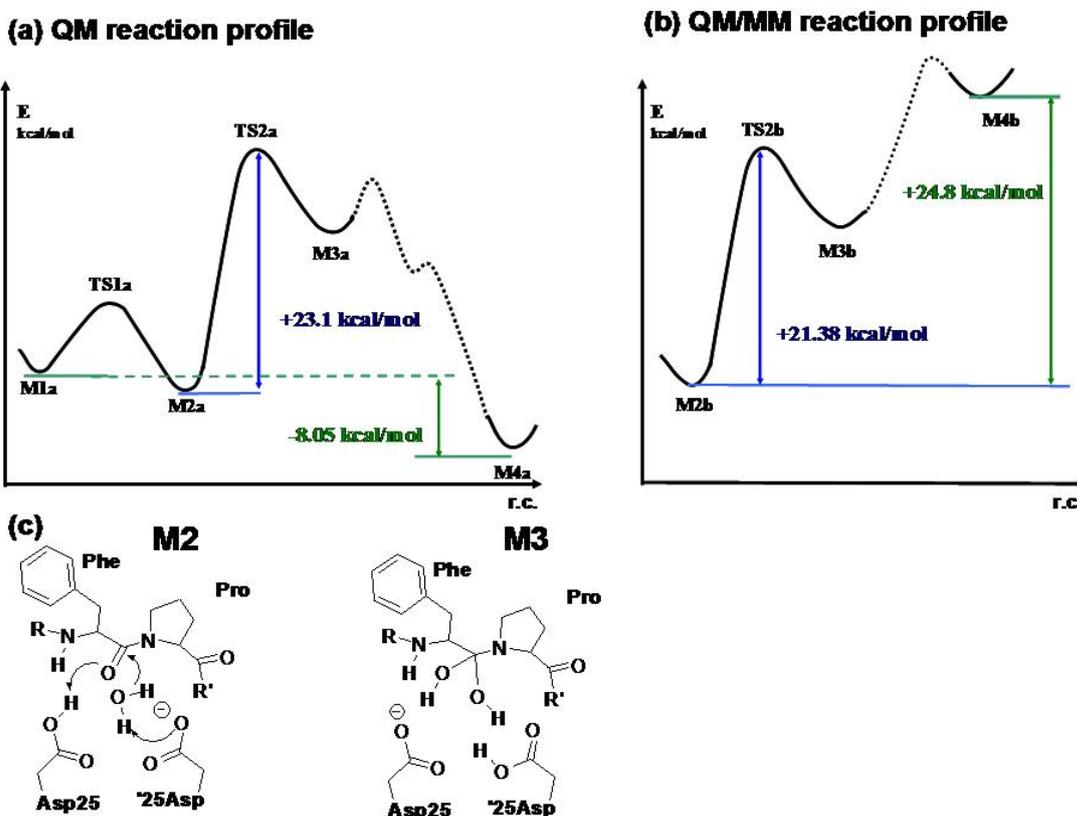


Figure 4.10 a) Pure QM and b) QM/MM reaction profiles. c) Schematic representation of the M2 and M3 species; the two Asp residues of the aspartic dyad (Asp25 and Asp25') belong to two identical chains.

The Figure 4.10-a shows a schematic representation of the reaction profile computed for MOD_QM and is briefly discussed here. The reaction path starts with the so called Michaelis complex (M1a) and after a small rearrangement of the nucleophilic water (TS1a) the rate limiting step takes place between the two minima M2a and M3a (Figure 4.10-a). The transition state TS2a corresponds to a concerted double proton transfer and the nucleophilic attack of the water molecule to the carbonyl of the amidic group (see Figure 4.10-c). Thus, water is activated (enhancing its nucleophilicity) by Asp25' deprotonation, while a concurrent protonation of the amidic carbonyl oxygen by Asp25, increases electrophilicity of the amidic carbon atom. The M3a minimum corresponds to a diolic intermediate with both hydroxylic groups that are H-bonded to the same unprotonated carboxylic group of the Asp dyad. The other steps from M3a to M4a correspond to structural rearrangements and proton displacements needed to complete the amide bond hydrolysis. The overall reaction path is apparently in good agreement with the observed behaviour of the enzyme (the reaction is spontaneous at room temperature), being the simulated process exo-ergonic (about 8 kcal/mol). Moreover, the taller energy barrier corresponds to the experimental value^{187,188} of the rate limiting step (i.e. formation of the diolic intermediate).

4.5.2 QM/MM reaction profile

A QM/MM model system (MOD_QMMM) is then used both to test our QM/MM code and to see the effects of the environment on the previous results. The system is constructed on the basis of the crystallographic coordinates used for the QM model (pdb code:1G6L); the protonation state is assigned by means of the H++ program,¹⁴² except for the Asp residues of the dyad.^{184,185} The substrate is simulated by an epta-peptide, in which the hydrolyzed bond is the Phe-Pro amidic bond. The system was fully minimized at the MM level using the Amber8 program and the *ff03*⁴² force field. Solvation was also simulated by means of the Born generalized method.¹⁴⁶ This represents the starting structure for QM/MM computations, which were performed according to a three layer (HML) partitioning (see Figure 4.11): the *high* layer was chosen to coincide with MOD_QM (also the basis-set is the same); few atoms (backbone only) of the residues near the *high* layer were assigned to the *medium* one; all the other atoms of the system were assigned to the *low* level. The residues of the *low* layer were kept frozen to their initial position because they are largely exposed to the surface and in our simulation solvent molecules are not included.

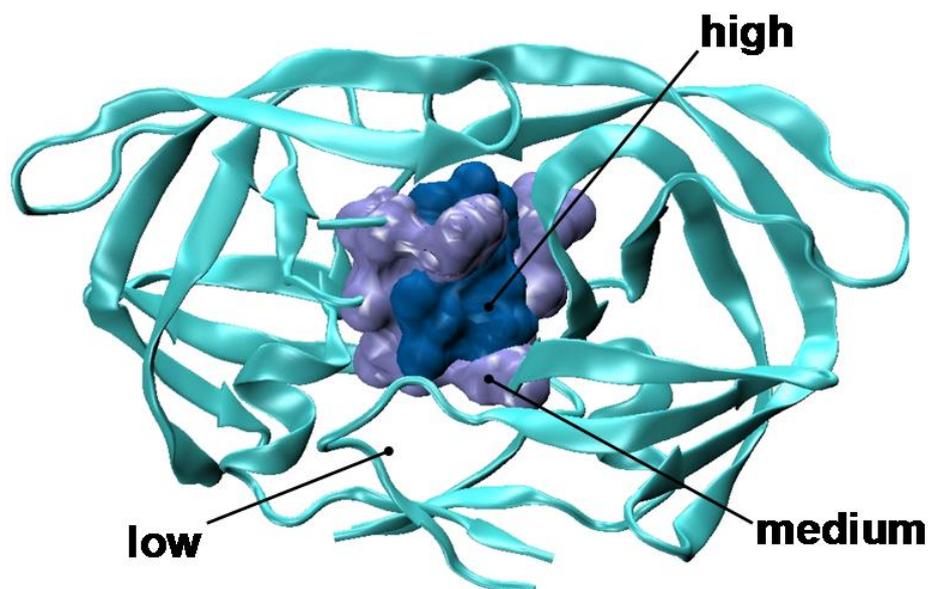


Figure 4.11 Model system used to study the QM/MM reaction profile of the enzyme HIV1-Pr.

The QM/MM PES showed interesting deviations from the behaviour observed with the pure QM model. The minimization of the Michaelis complex gave a structure similar to the reacting complex M2 of MOD_QM. The first reaction step (21.4 kcal/mol of activation energy) leads directly to the diolic intermediate, which is less stable than the reactant. This value is very similar to the one obtained for the QM model and also the transition state is very similar. Before mapping the whole reaction path we searched for the product, finding the process to be endo-ergonic by 25.8 kcal/mol (Figure 4.10-b).

These results may be accounted for if we consider the structure of the system and the difference between reactant and products. It is well known that the structure of the free enzyme with respect to that of the complex may be quite different. As a matter of fact, when the substrate binds to the open active site it causes a large conformational change on the mobile regions of the enzyme (called “flaps”),¹⁹³ leading to a closed form. Here, the peptide chain is fixed to the active site by an hydrogen bond network, but when the peptide bond is broken, the lid of the box opens again and the two proteic segments are released to the aqueous media: the lower affinity of the closed form of the enzyme for the hydrolysed product causes an opposite conformational change on the flaps to give the free enzyme in its open form and the product.

In the QM model the destabilizing interactions between the enzyme and the products are not taken into account properly, since most of the enzyme (like the “flaps”) are neglected along the whole reaction profile. Thus, the computed energies only reflect the enthalpic components of the hydrolytic process itself, because all the successive phases (e.g. enzyme rearrangement,

etc) cannot be explored with such a small system. On the other hand, in the QM/MM model most of the enzyme (like the “flaps”) is kept frozen (along the whole reaction profile) at the conformation of the starting complex (i.e. the closed form) and the computed energy for the M4 minimum (i.e. the products) has no physical sense, because such a structure is not reliable for the real biological system which would naturally relax by opening the active site via rotation of the “flaps”. This dynamical process cannot be accounted for in our QM/MM model and computations correctly predict a high energy value for such a stressed geometry. However the first part of the path is very similar for both the QM and QM/MM model systems (e.g. the barrier to the formation of the diolic intermediate is almost the same). Thus, we hypothesized that in the early region of the reaction coordinate the destabilizing factors are too low to play a key role, because the substrate backbone is substantially unchanged with respect to the products and the hydrogen bond network (responsible for substrate recognition) is still effective. As the system moves along the reaction coordinate, the breakable peptide bond becomes weaker until the substrate is broken in two fragments, which immediately go apart. We believe that the new hydrogen bonds configuration can destabilize the complex between enzyme and products, thus preparing the opening of the flaps and the release of the hydrolyzed product.^{193,194} The endo-ergonic nature of the computed QM/MM profile may be seen as a clue of that. Under this point of view, this result highlights the importance of the cooperative effect of the surrounding protein environment in triggering the final steps (i.e. product release) of the reaction.

4.6 Structural investigation of chelated octahedral complexes of Fe and Zn

Both 2,2'-bipyridine (bipy) and 1,10-phenanthroline (phen) (Figure 4.12) are bi-dentate chelating ligands; they form with many transition metals homoleptic complexes of general formula, $[M(\text{bipy})_3]^{n+}$ and $[M(\text{phen})_3]^{n+}$ (M = metal ion; Cr, Fe, Co, Ru, Rh, Ni). These complexes have six-coordinated, octahedral structures and they exist in two enantiomeric forms (Figure 4.13). In particular Fe, Zn and Ni complexes have been extensively studied in the last century and a large amount of experimental data have been collected.

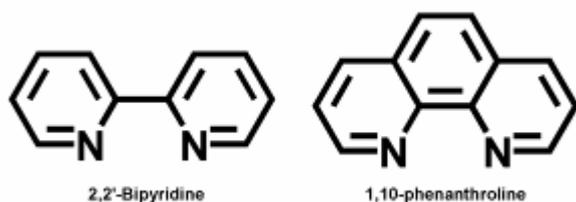


Figure 4.12 The chemical structure of the 2,2'-bipyridine (bipy) and 1,10-phenanthroline (phen) ligands.

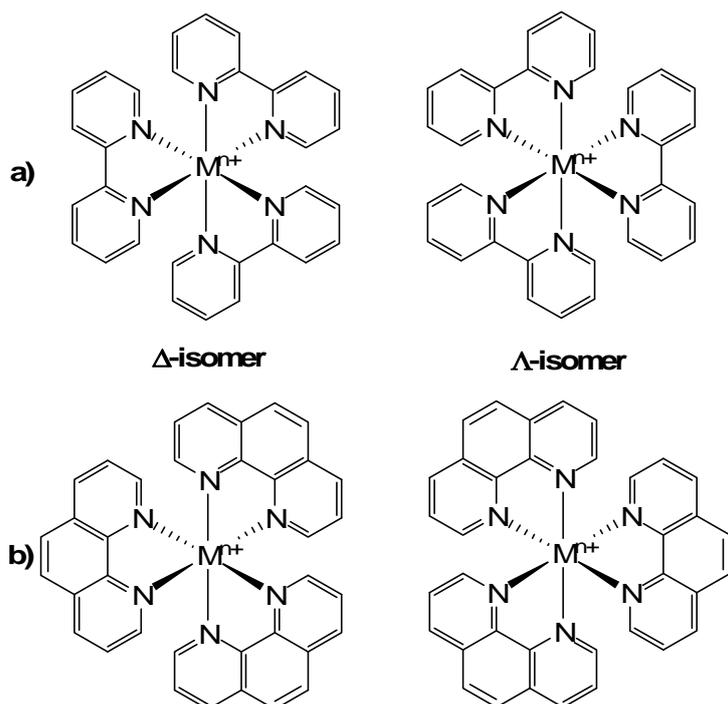


Figure 4.13 Structural representation of the two optical isomers of the octahedral complexes of a generic metal (M^{n+}) with a) 2,2-bipyridine and b) 1,10-phenanthroline.

Due to its high absorbance in the visible (the maximum absorbance is at 505 nm), the $[\text{Fe}(\text{phen})_3]^{2+}$ complex (ferroin) has been extensively used for more than half a century as a versatile redox indicator and iron detector. Also, it has been also found that the complex can act as co-reactant or catalyst in many organic reactions. The rigid and stable structure makes ferroin and its derivatives interesting building blocks for complex supra-molecular structures. Despite this large amount of data, some aspects of the physico-chemical properties of these compounds still remain unclear; ad example, the mechanistic details of the enantiomerisation process are not known. The $[\text{Fe}(\text{phen})_3]^{2+}$ has been extensively studied in the past by many authors¹⁹⁵⁻²⁰⁰ in order to understand the kinetic details of the racemisation process. At the best of our knowledge all the experimental evidences support an intra-molecular isomerisation mechanism instead of the inter-molecular one. The first mechanism accounts for the inversion of the chirotopic centre suggesting an achiral activated complex, which results from some intra-molecular rearrangements. The inter-molecular mechanism hypothesizes a dissociating phase, where one or more ligands are expelled into the solvent, and an associative one where the complex forms again in either Λ or Δ configurations. It has been showed²⁰¹ that, in a wide range of temperature and solvents, the rate of the dissociation process is about one order of magnitude lower than the rate of racemisation, making the intra-molecular mechanism dominant over the inter-molecular one. The effect of the solvent on the rate of bimolecular reactions, involving partial charge separation, has been extensively investigated²⁰², while some aspects of the solvation influence on unimolecular processes still

remain unclear. Racemisation reactions of metal-complexes involving an intra-molecular rearrangement of the ligands provides some interesting examples of unimolecular reaction. The behaviour of $[\text{Fe}(\text{phen})_3]^{2+}$ has been investigated under different solvent conditions^{200,201} and it has been demonstrated that there is no dependence of the racemisation rate on the dielectric constant. More intriguingly, an increasing of the reaction rate has been directly correlated to the methanol fraction in a methanol/water mixture. Some authors^{201,203} attributed a stabilization of the activated complex to van der Waals interactions between the ligands and the organic fraction of the solvent. In their opinion, given the greater solubility of the ligand in methanol, a structure exposing to the solvent a large portion of the same ligand would be better stabilized by a methanol-rich mixture. This belief was based upon the common mechanistic hypothesis about the intra-molecular rearrangement of octahedral complexes, which assumed the formation, during the isomerisation process, of an expanded structure with loosened Fe-N(ligand) bonds.

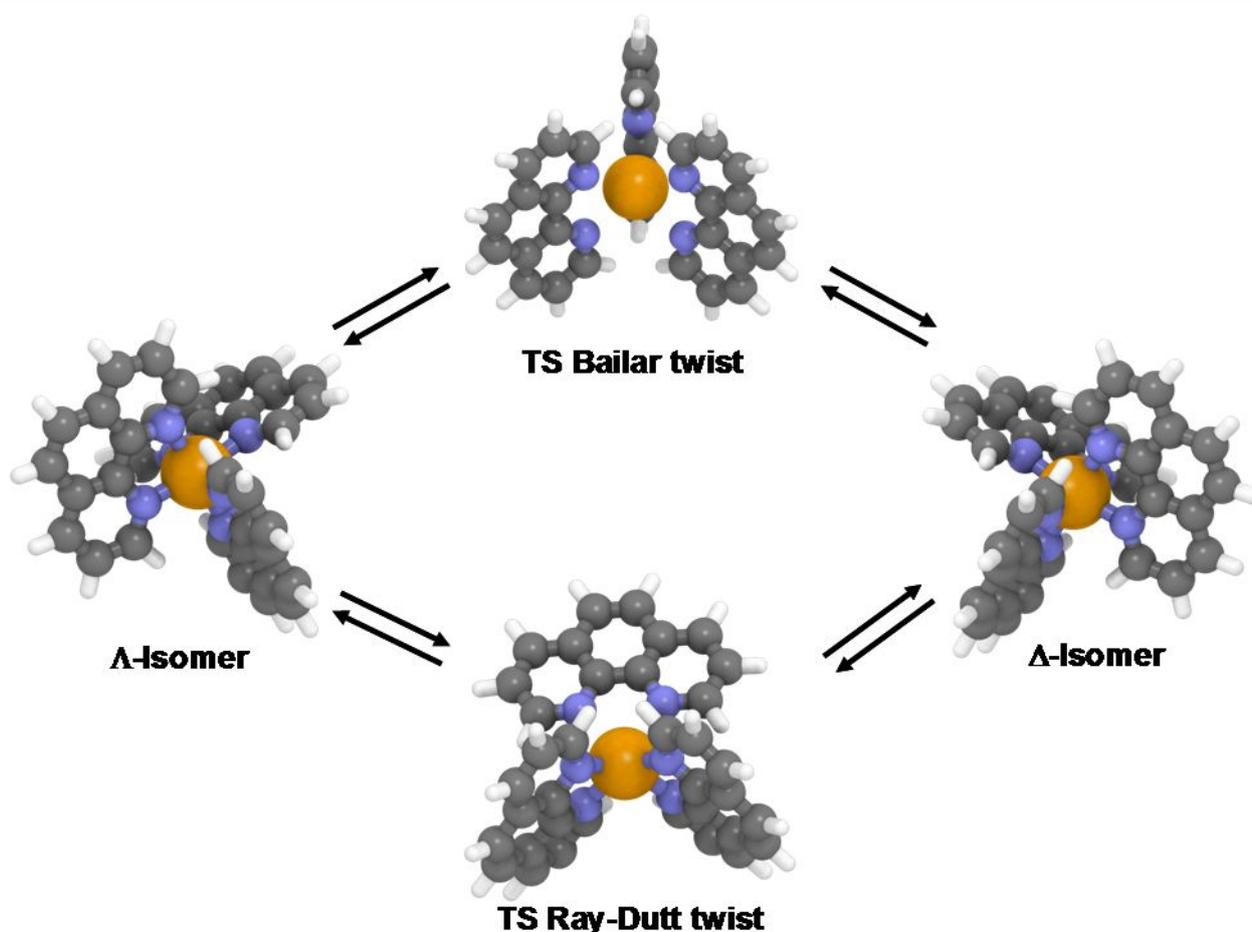


Figure 4.14 The two enantiomers of the $[\text{Fe}(\text{phen})_3]^{2+}$ and the two most plausible enantiomerisation TS.

A number of possible pathways have been successively proposed on the basis of a mathematically derived symmetry selection rule^{204,205}. The application of this theory to the isomerisation of metal complexes²⁰⁶ in the case of tris-bidentate complex (with general

formula $M(LL)_3$ of D_3 symmetry, gives four different rearrangement patterns of chiral homoconversion.²⁰⁶ The *push-through* mechanism, although symmetry allowed, is extremely unlikely on energy grounds, being all the three ligands stamped out on the plane, a unbearable geometry even for ligands smaller than phenanthroline. The *crossover twist* is another mechanism involving a highly hindered transition structure and for this reason it is considered hardly possible. In the remaining two mechanisms, the *Ray-Dutt twist* and the *Bailar twist* (Figure 4.14),²⁰⁷ the conversion involves transition states which are less hindered than the previous ones. Similar studies on $[Ni(phen)_3]^{2+}$ lead to different results, suggesting the prevalence, in the enantiomerisation process, of the dissociative-associative mechanism over the intra-molecular rearrangement. An accurate computational study could give a deeper insight into the mechanistic details of the enantiomerisation process of $[M(phen)_3]^{n+}$ and $[M(bipy)_3]^{n+}$ and could help to rationalise the effects of metal substitution and the role of the solvent. The choice of an appropriate computational method is the first step of a reliable theoretical investigation. In this section are reported the results of a preliminary study on phen and bipy complexes of Fe(II) and Zn(II). We tested and compared different computational methods to understand their effectiveness in the description of the structural parameters of these complexes.

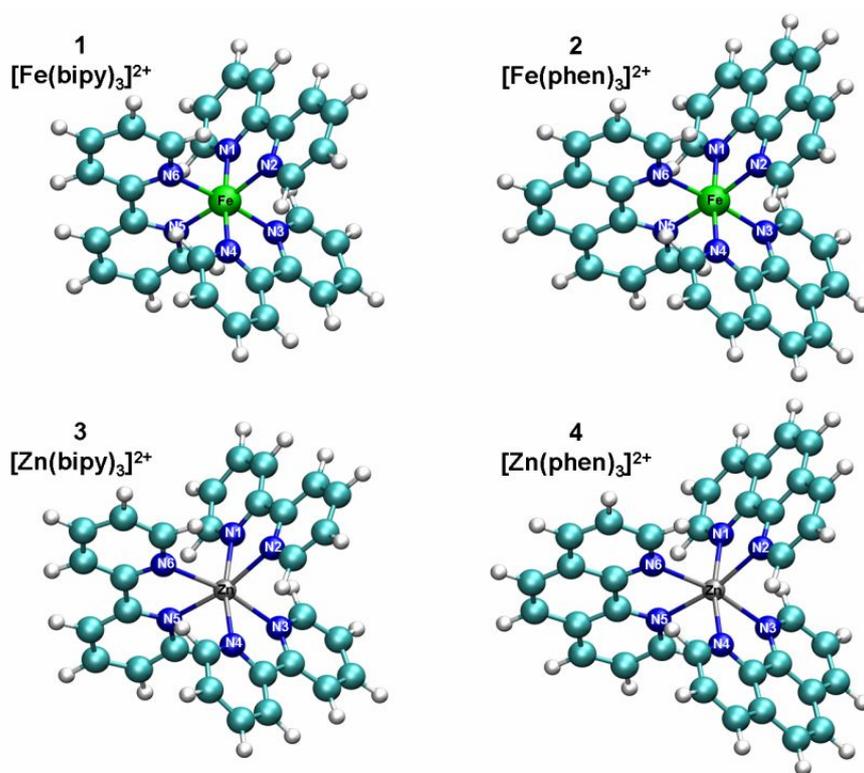


Figure 4.15 Ball & stick representation of the coordination complexes of Fe and Zn discussed here.

4.6.1 Computational methods

All the calculations have been carried out using the COBRAMM¹⁰⁴ suite of computational programs, as an interface between the Turbomole,¹³² the DFTB⁷⁶ and the Gaussian 03¹³¹ packages. Geometry optimisation of [Fe(bipy)₃]²⁺ (**1**), [Fe(phen)₃]²⁺ (**2**), [Zn(bipy)₃]²⁺ (**3**) and [Zn(phen)₃]²⁺ (**4**) (Figure 4.15) have been carried out with the Gaussian 03¹³¹ optimisation driver, using a redundant internal coordinates description²⁰⁸ within the powerful BFGS algorithm.⁵⁻⁸ Three different potentials have been tested: a) the DFT⁷⁵ method with the hybrid B3LYP¹⁵¹ functional (**b3lyp**); the DFT method with the pure GGA BP86²⁰⁹ functional (**bp86**); c) the SCC-DFTB⁷⁶ method (**dftb**). For DFT/BP86 and DFT/B3LYP (called in Turbomole, respectively, “b-p” and “b3-lyp_Gaussian”²¹⁰) calculations the Turbomole¹³² software was used, adopting the “m3” grid size for the density fitting and a SCF convergence criterion of 1x10⁻⁷ E_h. A balanced double ζ split valence (SP) basis set with polarisation functions (P) [referred in Turbomole as to def2-SV(P)]²¹¹ was adopted to describe H, C and N atoms, while the Fe and Zn atom were described using a triple ζ split valence (SP) basis set with polarisation functions (P) [referred in Turbomole as to def2-TZVP].²¹¹ The “resolution of identity” (RI)^{93,94} and the “multipole accelerated resolution of identity” (MARI)¹⁰³ approximations have been used to speed-up the DFT calculations, and a proper auxiliary basis set²¹² was chosen for the approximation of approximate the Coulomb term.^{93,94} The SCC-DFTB⁷⁶ method was used by assigning to H, C and N atoms the standard parameters and using the recently developed first-row metals parameter set⁸¹ for Fe and Zn.

4.6.2 Results and discussion

All geometric optimisations gave stable structures with all positive Hessian elements, as showed by numerical frequency calculations on the optimised geometries. The effectiveness of various potentials have been evaluated by cross-comparing the bond lengths in the optimised geometries of the four complexes. The bond length differences have been plotted for all the bonds of the molecules. The optimised geometries of all the complexes **1-4** are highly symmetric (D_3), even if no symmetry restrictions have been used; this is evident by examining the standard deviation (St. Dev.) of the M-N bonds lengths [see the **St. Dev. (Fe-N)** and **St. Dev. (Zn-N)** entries of Table 4.3, Table 4.4, Table 4.5 and Table 4.6], that is about 10⁻⁴ Å in all cases except for the **bp86** optimised geometry of **3**, where is about 10⁻⁴. The comparison of the two DFT functional (for **1** and **2**) showed a tendency of **bp86** to underestimate the Fe-N bond distances with respect of **b3lyp**, and to overestimate both C-H and C-N distances even if in a lower amount (see Table 4.3, Table 4.4, Figure 4.16-a and

Figure 4.17-a). In both Fe complexes **dftb** showed a better agreement with **b3lyp** than to **bp86** (see Table 4.3, Table 4.4, Figure 4.16-b,c and Figure 4.17-b,c), and in all cases the differences with both DFT methods are below the acceptable error of 0.1 Å, according to the parameter set developers.⁸¹

In the case of **3** and **4** the agreement of **b3lyp** and **bp86** is better than for **1** and **2** (see Table 4.5, Table 4.6, Figure 4.18-a and Figure 4.19-a): the C-H and C-N bond lengths are still overestimated by **bp86**, but the deviation in the M-N bond lengths with respect to **b3lyp** is lower if M = Zn (complexes **3** and **4**) than if M = Fe (complexes **1** and **2**). Even in this case the **dftb** structures is closer to the **b3lyp** ones (see Table 4.5, Table 4.6, Figure 4.18-b,c and Figure 4.19-b,c).

In conclusion, the reported data strongly suggest the reliability of the **dftb** approach to the study of ground state properties of Fe and Zn chelated complexes of bipy, phen. Moreover its cost/efficiency ratio is improved by its low computational cost, that is one or two order of magnitude smaller than, respectively, approximated (RI) and non approximated DFT.

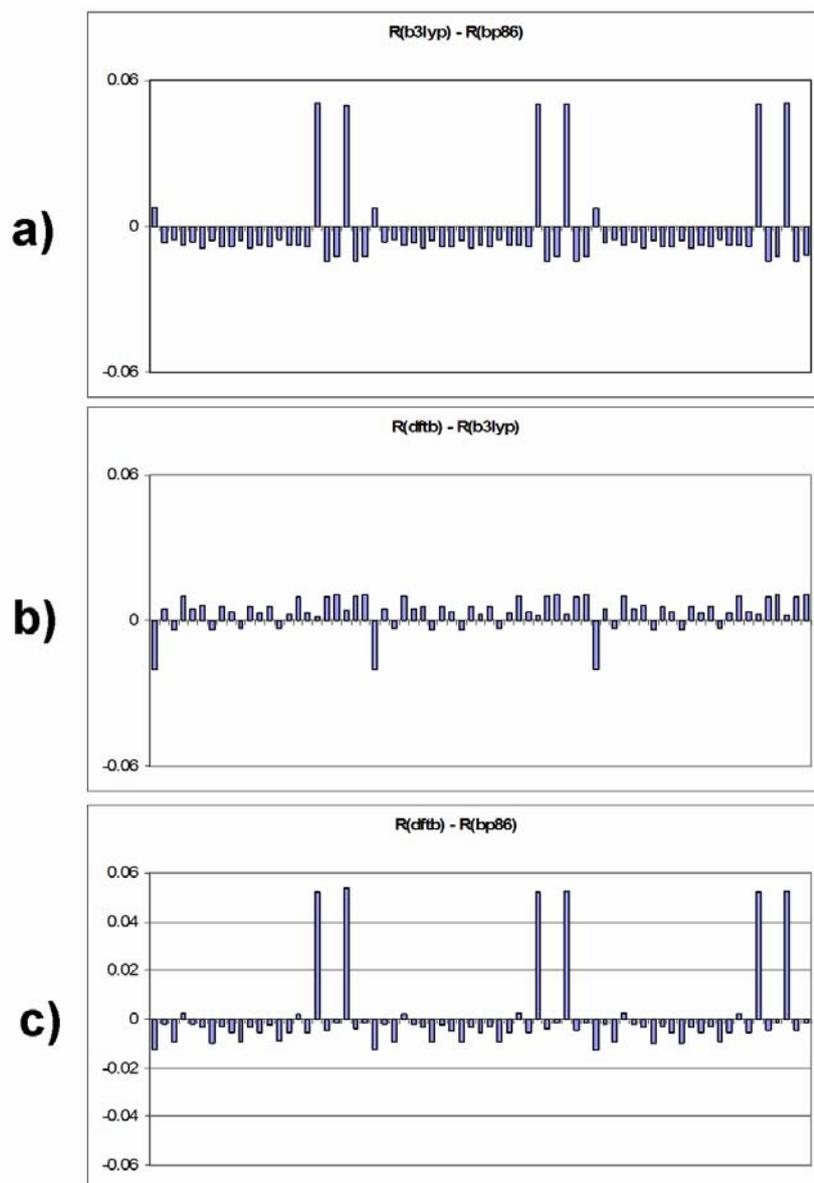


Figure 4.16 Bond distance differences for the $[\text{Fe}(\text{bipy})_3]^{2+}$ between a) **b3lyp** and **bp86**, b) **dftb** and **b3lyp**, **dftb** and **bp86** optimised geometries. The difference values are reported in Å for all the bonds.

Table 4.3 Relevant geometrical parameters and statistical analysis (in Å) for the $[\text{Fe}(\text{bipy})_3]^{2+}$ complex.

	Bond distances			Bond distance differences		
	R(b3lyp)	R(bp86)	R(dftb)	R(b3lyp) - R(bp86)	R(dftb) - R(b3lyp)	R(dftb) - R(bp86)
Fe-N1	2.0233	2.0219	1.9713	0.0506	0.0014	0.0520
Fe-N2	2.0233	2.0193	1.9696	0.0497	0.0040	0.0537
Fe-N3	2.0233	2.0212	1.9712	0.0500	0.0021	0.0521
Fe-N4	2.0233	2.0207	1.9705	0.0502	0.0026	0.0528
Fe-N5	2.0233	2.0210	1.9710	0.0500	0.0023	0.0523
Fe-N6	2.0232	2.0210	1.9705	0.0505	0.0022	0.0527
Average (Fe-N)^a	2.0233	2.0209	1.9707	0.0502	0.0024	0.0526
St. Dev. (Fe-N)^b	4.1E-05	8.6E-04	6.3E-04	3.4E-04	8.6E-04	6.3E-04
Average (all)^c				-0.0028	0.0033	0.0005
St. Dev. (all)^d				1.7E-02	6.9E-03	1.7E-02

a) Average on M-N distances (columns 1-3); Average on differences of M-N distances(columns 4-6).

b) Standard deviation on M-N distances (columns 1-3); Standard deviation on differences of M-N distances(columns 4-6).

c) Average on all differences of distances between bonded atoms.

d) Standard deviation on all between bonded atoms.

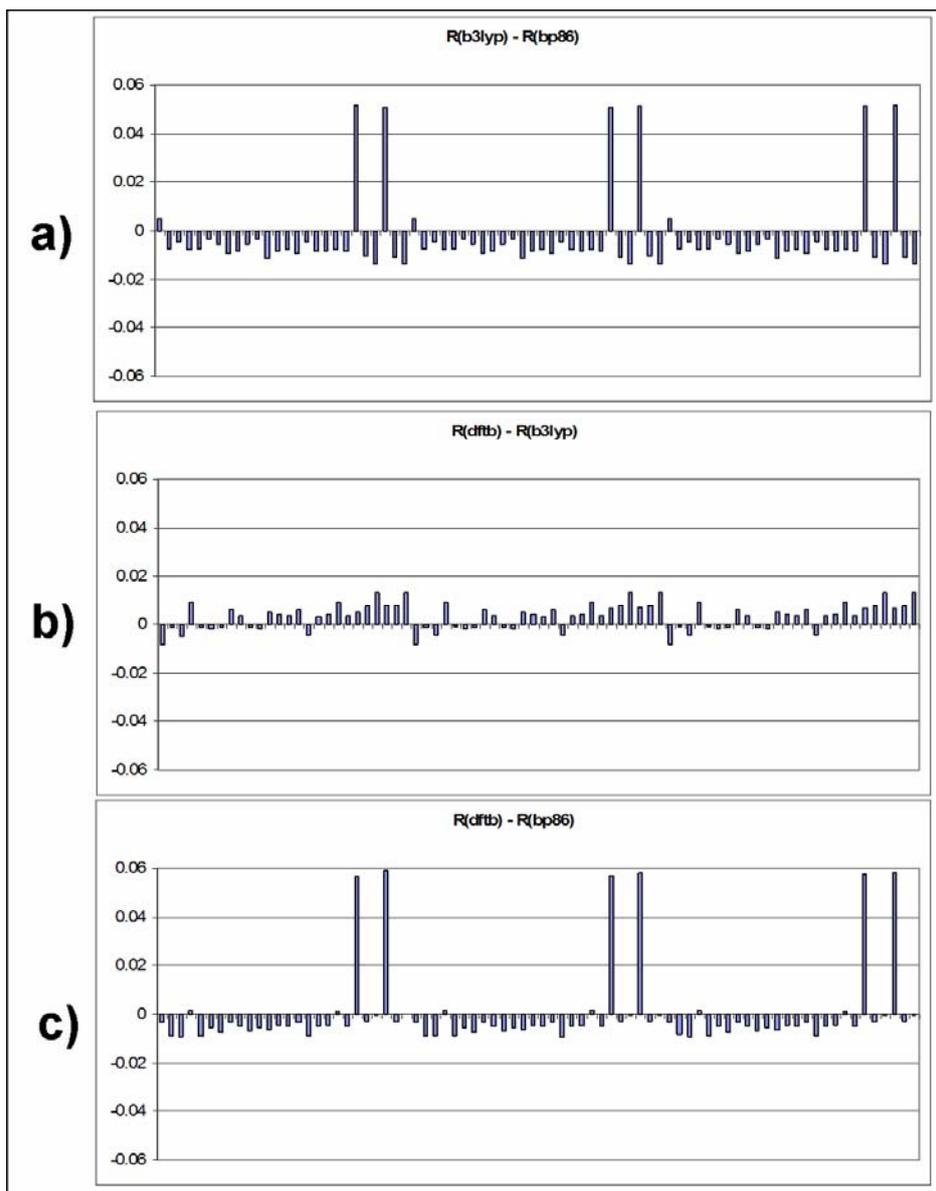


Figure 4.17 Bond distance differences for the $[\text{Fe}(\text{phen})_3]^{2+}$ between a) **b3lyp** and **bp86**, b) **dftb** and **b3lyp**, **dftb** and **bp86** optimised geometries. The difference values are reported in Å for all the bonds.

Table 4.4 Relevant geometrical parameters and statistical analysis (in Å) for the $[\text{Fe}(\text{phen})_3]^{2+}$ complex.

	Bond distances			Bond distance differences		
	R(b3lyp)	R(bp86)	R(dftb)	R(b3lyp) - R(bp86)	R(dftb) - R(b3lyp)	R(dftb) - R(bp86)
Fe-N1	2.0354	2.0305	1.9790	0.0515	0.0049	0.0564
Fe-N2	2.0355	2.0276	1.9767	0.0509	0.0079	0.0588
Fe-N3	2.0358	2.0294	1.9786	0.0508	0.0064	0.0572
Fe-N4	2.0356	2.0287	1.9774	0.0513	0.0069	0.0582
Fe-N5	2.0356	2.0292	1.9781	0.0511	0.0064	0.0575
Fe-N6	2.0358	2.0293	1.9779	0.0514	0.0065	0.0579
Average (Fe-N)^a	2.0356	2.0291	1.9780	0.0512	0.0065	0.0577
St. Dev. (Fe-N)^b	1.6E-04	9.5E-04	8.3E-04	2.8E-04	9.7E-04	8.3E-04
Average (all)^c				-0.0027	0.0037	0.0010
St. Dev. (all)^d				1.7E-02	5.3E-03	1.8E-02

a) Average on M-N distances (columns 1-3); Average on differences of M-N distances (columns 4-6).

b) Standard deviation on M-N distances (columns 1-3); Standard deviation on differences of M-N distances (columns 4-6).

c) Average on all differences of distances between bonded atoms.

d) Standard deviation on all between bonded atoms.

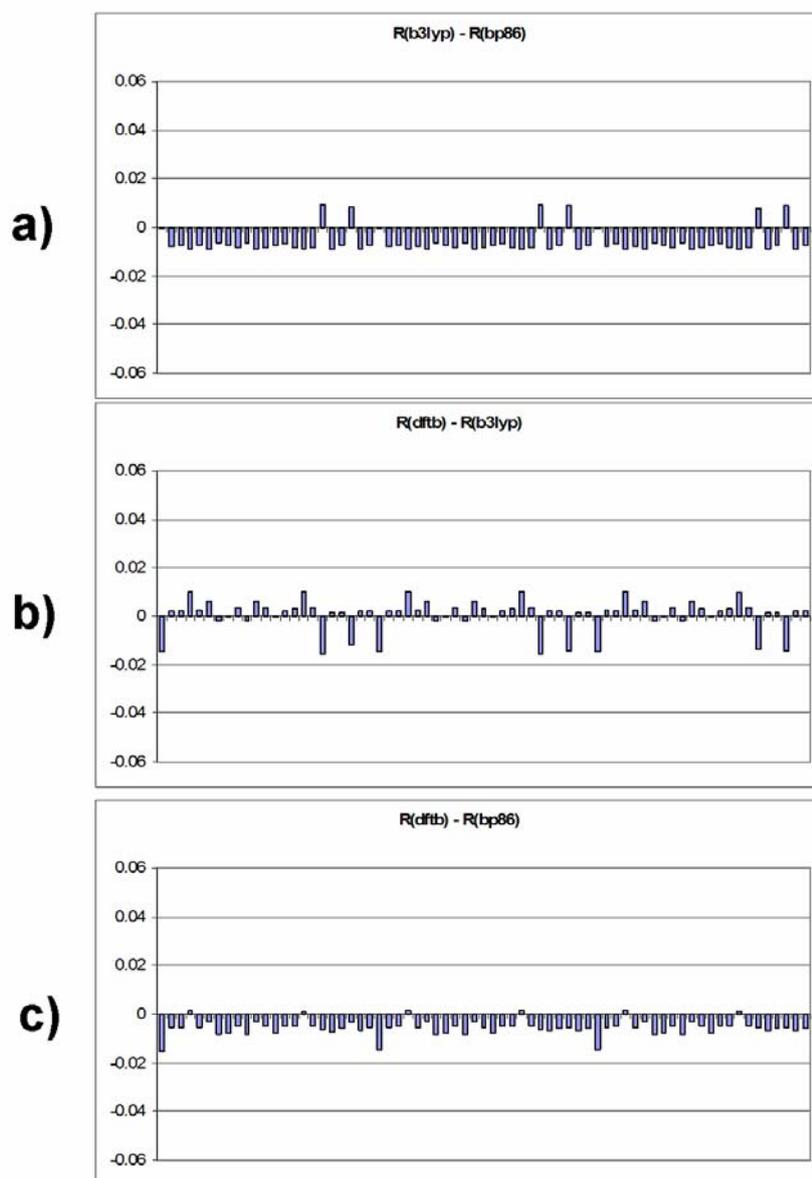


Figure 4.18 Bond distance differences for the $[\text{Zn}(\text{bipy})_3]^{2+}$ between a) **b3lyp** and **bp86**, b) **dftb** and **b3lyp**, **dftb** and **bp86** optimised geometries. The difference values are reported in Å for all the bonds.

Table 4.5 Relevant geometrical parameters and statistical analysis (in Å) for the $[\text{Zn}(\text{bipy})_3]^{2+}$ complex.

	Bond distances			Bond distance differences		
	R(b3lyp)	R(bp86)	R(dftb)	R(b3lyp) - R(bp86)	R(dftb) - R(b3lyp)	R(dftb) - R(bp86)
Zn-N1	2.1968	2.2127	2.2034	0.0093	-0.0159	-0.0066
Zn-N2	2.1968	2.2084	2.2000	0.0084	-0.0116	-0.0032
Zn-N3	2.1970	2.2125	2.2033	0.0092	-0.0155	-0.0063
Zn-N4	2.1969	2.2111	2.2023	0.0088	-0.0142	-0.0054
Zn-N5	2.1968	2.2104	2.2023	0.0081	-0.0136	-0.0055
Zn-N6	2.1975	2.2117	2.2028	0.0089	-0.0142	-0.0053
Average (Zn-N)^a	2.1970	2.2111	2.2024	0.0088	-0.0142	-0.0054
St. Dev. (Zn-N)^b	2.7E-04	1.6E-03	1.2E-03	4.6E-04	1.5E-03	1.2E-03
Average (all)^c				-0.0061	0.0005	-0.0056
St. Dev. (all)^d				4.9E-03	6.5E-03	3.2E-03

a) Average on M-N distances (columns 1-3); Average on differences of M-N distances(columns 4-6).

b) Standard deviation on M-N distances (columns 1-3); Standard deviation on differences of M-N distances(columns 4-6).

c) Average on all differences of distances between bonded atoms.

d) Standard deviation on all between bonded atoms.

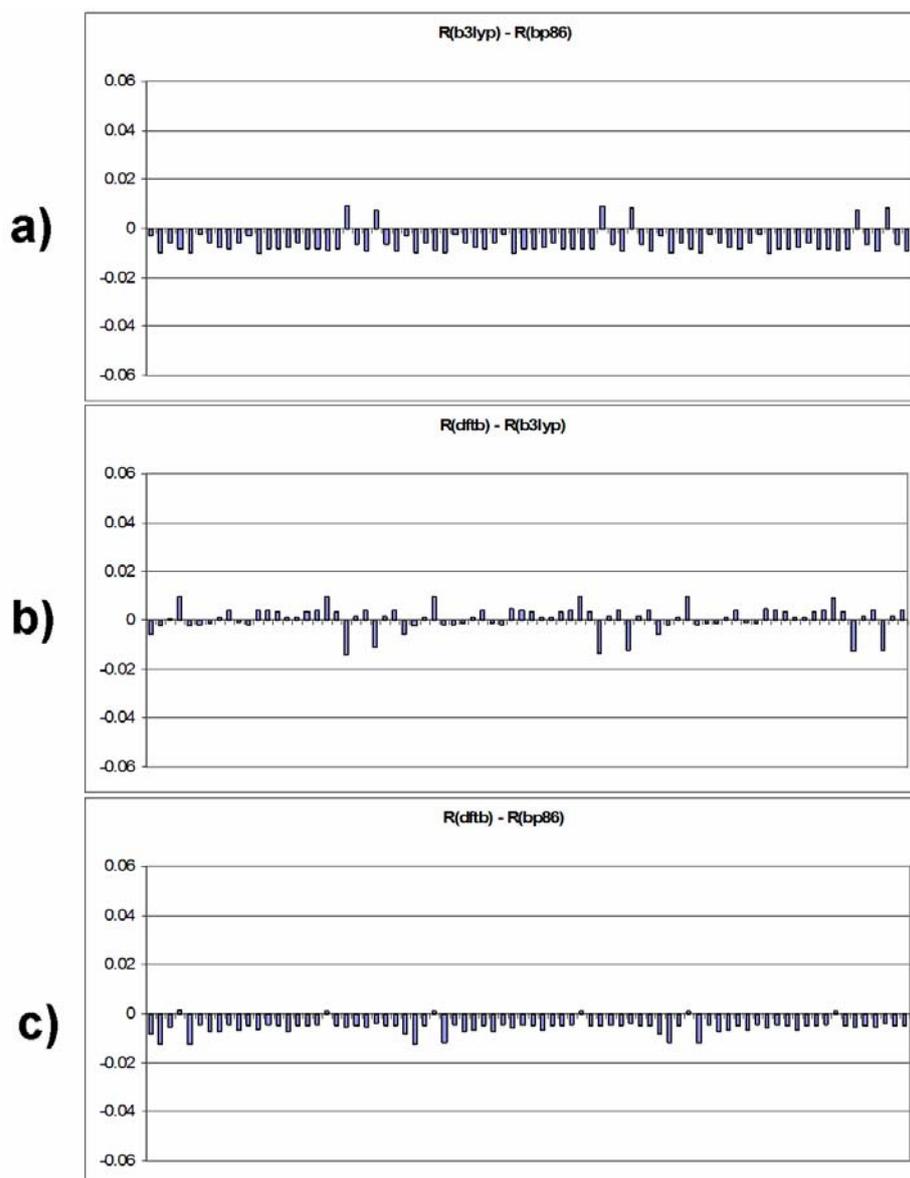


Figure 4.19 Bond distance differences for the $[\text{Zn}(\text{phen})_3]^{2+}$ between a) **b3lyp** and **bp86**, b) **dftb** and **b3lyp**, **dftb** and **bp86** optimised geometries. The difference values are reported in Å for all the bonds.

Table 4.6 Relevant geometrical parameters and statistical analysis (in Å) for the $[\text{Zn}(\text{phen})_3]^{2+}$ complex.

	Bond distances			Bond distance differences		
	R(b3lyp)	R(bp86)	R(dftb)	R(b3lyp) - R(bp86)	R(dftb) - R(b3lyp)	R(dftb) - R(bp86)
Zn-N1	2.2044	2.2188	2.2097	0.0091	-0.0144	-0.0053
Zn-N2	2.2042	2.2155	2.2080	0.0075	-0.0113	-0.0038
Zn-N3	2.2043	2.2180	2.2093	0.0087	-0.0137	-0.0050
Zn-N4	2.2046	2.2171	2.2086	0.0085	-0.0125	-0.0040
Zn-N5	2.2040	2.2170	2.2094	0.0076	-0.0130	-0.0054
Zn-N6	2.2052	2.2175	2.2091	0.0084	-0.0123	-0.0039
Average (Zn-N)^a	2.2045	2.2173	2.2090	0.0083	-0.0129	-0.0046
St. Dev. (Zn-N)^b	4.2E-04	1.1E-03	6.2E-04	6.3E-04	1.1E-03	7.4E-04
Average (all)^c				-0.0060	0.0008	-0.0053
St. Dev. (all)^d				4.9E-03	5.3E-03	2.6E-03

a) Average on M-N distances (columns 1-3); Average on differences of M-N distances(columns 4-6).

b) Standard deviation on M-N distances (columns 1-3); Standard deviation on differences of M-N distances(columns 4-6).

c) Average on all differences of distances between bonded atoms.

d) Standard deviation on all between bonded atoms.

Part II:

Using COBRAMM to Study Bio-

Organic Systems

5 The Enzyme Proline Racemase

5.1 Introduction

The enzyme proline racemase catalyzes the reversible stereoinversion of chiral C^α in proline. Within the large amino acid racemase²¹³ family, it belongs to an interesting subgroup of enzymes acting with a pyridoxal-phosphate-independent mechanism.²¹⁴ It is well-known that a large number of bacteria synthesize amino acid racemases to produce D-enantiomers directly from their L-counterparts. Thus, inactivation of amino acid racemase genes abolishes the pathogenicity of some bacteria, since it prevents the construction of the cell wall that contains D-amino acids. The common prokaryotic form, originally found in the bacterium *Clostridium sticklandi*,²¹⁵ has been studied for decades. Following the commonly accepted mechanistic hypothesis, on the basis of experimental evidences,²¹⁶⁻²¹⁸ two Cys residues in the active site are responsible for the inversion of the chiral center by deprotonating the C^α and reprotonating it on the opposite side. The recent discovery²¹⁹ of a proline racemase in the eukaryotic parasite *Trypanosoma cruzi* has renewed the interest for this class of enzymes. *T. cruzi*, which is responsible for the endemic Chagas' disease, produces two almost identical (96%) proline racemases (TcPRACA, TcPRACB), involved in immune evasion and in the regulation of intracellular metabolism.^{219,220} Topical researches indicate TcPRACA as a potential target for drug design. Recent crystallographic data²²¹ [pdb code: 1W61, resolution 2.1 Å] of secreted²¹⁹ TcPRACA in complex with an inhibitor molecule (pyrrole-2-carboxylic acid, PYC) revealed that a couple of Cys residues in the active site can exchange protons with the substrate. The enzyme acts as a homodimer formed by two chains (A and B), each characterized by an independent reaction center. The binding of the substrate promotes a large conformational rearrangement within each monomer, which results in a total exclusion of the active site from the solvent.

5.2 Set-up of the bio-molecular system

To build the model system we used the crystallographic coordinates of the pdb file with code 1W61 that describes both sub-units of the dimer, each one containing an inhibitor molecule (PYC). Since the two active sites work independently, we selected for our model the first sub-unit (chain A). In the following discussion the term “*real*” will indicate the whole A sub-unit where the PYC molecule is replaced by a zwitterionic L-proline molecule. The model system was obtained by protonating the selected crystallographic structure using the H++ software.¹⁴² This code uses an automated algorithm that computes pK_a values of ionizable groups in macromolecules and adds missing hydrogen atoms according to the specified pH of the environment. The position of the added H atoms are also optimized by this algorithm. The system was fully minimized at the MM level (Amber *ff99* force field) using a full conjugate gradient minimization approach and the GB method¹⁴⁶ to simulate the aqueous environment as implemented in the SANDER module of the AMBER 8.0 package.⁴⁵ The optimization procedure is carried out until the root mean square deviation (RMSD) of the Cartesian elements of the gradient is less than $0,0001 \text{ kcal mol}^{-1}$. Since the binding mode of L-proline was unknown (the protein has been crystallized with an inhibitor molecule), we carried out a conformational study of the substrate (L-proline) and product (D-proline) within the protein environment. We carried out high temperature Molecular Dynamics starting from the minimized structure of the complex. The region of 5 \AA around the substrate was free to move. The system was heated from 0 K to 800 K in 100 ps and then a trajectory of 2 ns was computed at constant temperature (800 K) and constant pressure (1 atm). The integration step of 2 fs was used in conjunction with the SHAKE algorithm³⁵ to constrain the stretching of bonds involving hydrogen atoms. The coordinates of the system were saved on a trajectory file every 2 ps, giving a total of 1000 structures. Solvation effects were taken into account using the GB model⁵⁸ with a dielectric constant of 78.5. To determine the most populated conformations of proline within the protein binding pocket, we performed a Cluster Analysis on the proline conformers during the simulation. To this purpose we used the MMTSB²²² toolset and we clustered the proline conformations by structural similarity (using *kclust*²²² and a fixed radius of 0.6 \AA on the Cartesian coordinate RMSD computed for all proline atoms). We determined the centroid of each cluster. For each cluster we chose the structure closest to the corresponding centroid as representative of the cluster itself (each structure was characterized by the smallest RMSD value with respect to the centroid). In this way we obtained some candidate structures to be used in the following QM/MM study. The main

differences between these structures are determined by the conformation of the proline five-membered ring.

5.3 Partitioning the system into layers

The “*real*” system was partitioned into three layers (Figure 3.2). to carry out the QM/MM computations. To study the stereo-inversion of L-proline played by the two residues Cys130 and Cys300, it was necessary to include in the QM region (*high* region) the whole substrate (we named PRL395) and the side-chains of Cys130 and Cys300. The free valences of boundary atoms were saturated with hydrogen atoms (H atom link approach). This provided the *model-H* sub-system, which is described at the QM computational level (see Figure 5.1 and Figure 5.2). The *medium* region includes atoms in a shell of about 5 Å from each proline atom. In Table 5.1 we list the residues included in this sub-system, and Figure 5.3 gives a graphical representation of its shape. All atoms not included in the *high* and *medium* layers belong to the *low* region. During the QM/MM optimization this region was kept frozen, except for the residues directly connected to the *medium* region. (see Table 5.2 and Figure 5.4) but not located on the enzyme surface (see Figure 5.5). The QM region was treated at the DFT/B3LYP/DZVP^{75,223} level using the Gaussian03¹³¹ series of programs, while the AMBER8⁴⁵ software and the *ff99* force field was used for the MM region.

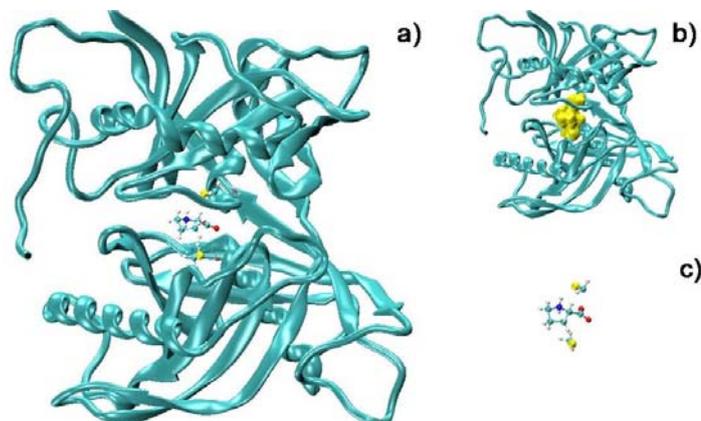


Figure 5.1 Different representation of the QM sub-region (*high* layer): a) ball & stick model within the whole system; b) yellow surface within the whole system; c) ball & stick *model-H* stand-alone.

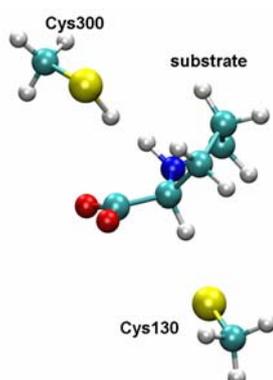


Figure 5.2 The *model-H* sub-system

Table 5.1 Residues in the *medium* region

Residue name	Residue number	Part included in the <i>medium</i> region
PHE	102	side-chain
LEU	127	side-chain
MET	129	C=O
CQD	130	backbone
GLY	131	all
HIE	132	side-chain
ASN	133	backbone
VAL	288	side-chain
PHE	290	side-chain
ASP	296	side-chain
ARG	297	C=O
SER	298	all
PRO	299	C=O
CQP	300	backbone
GLY	301	all
THR	302	all
GLY	303	all
THR	304	N-H

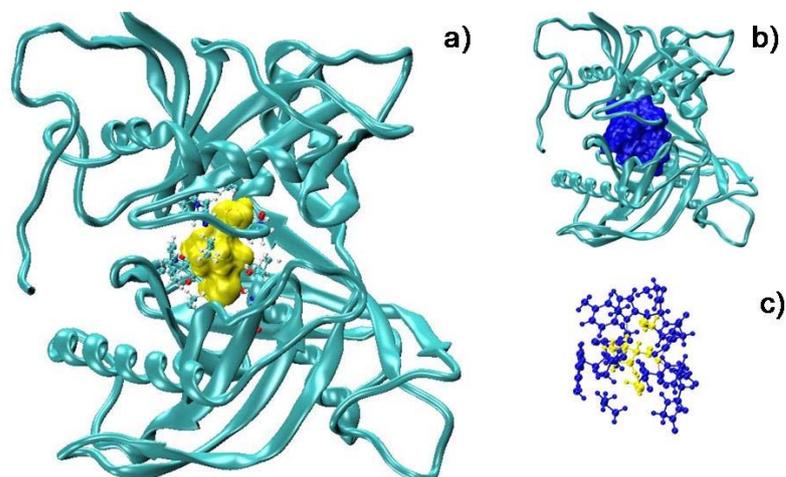


Figure 5.3 Different representations of the *medium* layer: a) ball & stick model (around the yellow surface representing the *high* layer) within the whole system; b) blue surface within the whole system; c) blue ball & stick model (around the yellow ball & stick model representing the *high* layer).

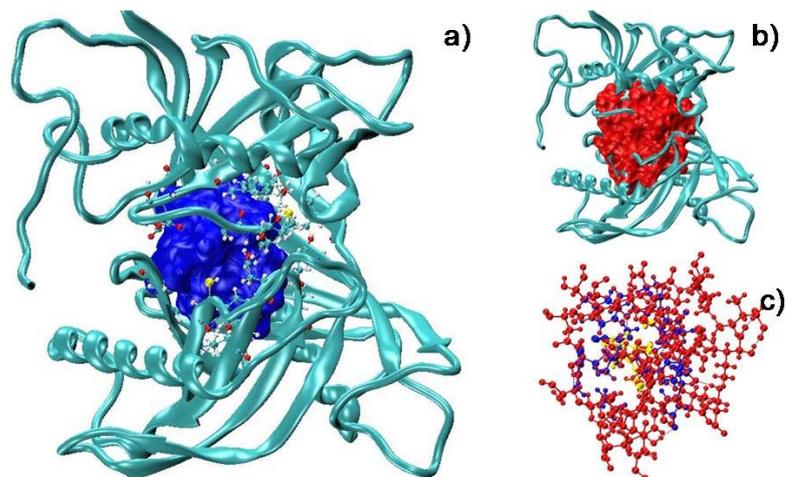


Figure 5.4 The *low* layer is all but the blue surface. The frozen *low* region is represented in ‘cartoon’, while the free *low* region is represented using: a) ball & stick model around the blue surface corresponding to the *medium* layer within the whole system; b) a red surface within the whole system; c) red ball & stick model around the blue and yellow ball & stick models corresponding to the *medium* and *high* layers, respectively.

Table 5.2 Residues in the moving *low* region

Residue number	Part included in the moving <i>low</i> region
56	all
102	backbone
120	all
122	all
127	backbone
128	all
129	all
132	backbone
133	side-chain
134	all
135	all
186	all
218	all
220	all
270	all
272	all
286	all
288	backbone
290	all
296	backbone
299	all
305	all

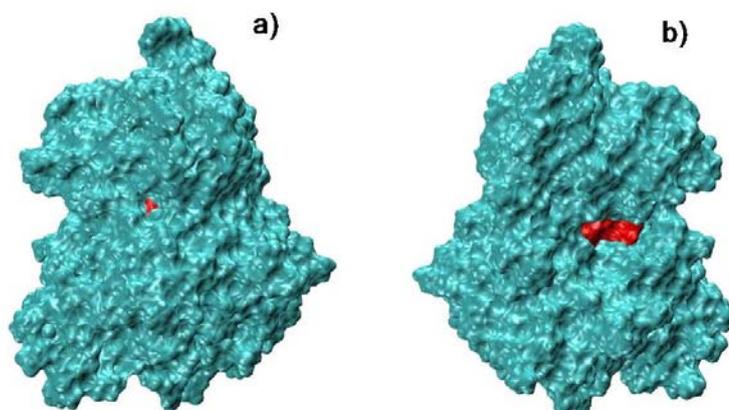


Figure 5.5 The residues belonging to the *low* layer and exposed to solvent are kept frozen during the calculations. The free residues of the *low* layer are represented by the red surface, while the frozen ones are represented by the cyan surface. Two views (a and b) of the enzyme show that the free *low* region is almost completely buried in the system.

5.4 Preliminary studies to test the set-up procedure

Some mutagenesis experiments²²⁰ showed that two residues of cysteine (Cys130 and Cys300) are involved in the catalytic activity and, as proved by iodometric titrations, one of them is unprotonated, bearing a net negative charge. The unprotonated Cys would take the H atom from the C^α of the L-Pro substrate meanwhile the other Cys residue would give its H atom on the opposite face of the substrate to give the product. According to this hypothesized reaction mechanism, the stereoinversion of the chiral center of the substrate would be accompanied by an inversion of the protonation state of the Cys couple. Before starting the exploration of the PES, which is fully described in the next section (Section 5.5), we explored¹⁰⁴ the relative stability of two structures, corresponding to the enzyme bound, respectively, to the L- and D-Pro, and having a different protonation state of the Cys couple. After placing the natural L-Pro into the natural active site we removed the thiolic H from Cys130 to obtain the ENZ-[L-Pro] complex (M1), while ENZ-[D-Pro] (M2) was obtained by placing D-Pro into the same active site where the thiolic H from Cys300 have been removed. Both structure were handled according to the procedure described in Section 5.2. The total charge of the system resulted to be -10, accounting for one negative charge in the active site plus the summation over the charges of ionisable residues which counts for 9 negative charges. In this section we describe the results obtained from different ways of handling this unbalanced charge, in order to test the set-up procedure used to build the system used to study the PES (Section 5.5). By using a tool (“leap”) of the AMBER8 package,⁴⁵ an appropriate number of positive ions was placed to neutralize the overall charge. This program places the counter-ions in a shell around the enzyme using a grid of Coulombic potential. Three different ways of balancing the 10 negative charges were tried, using: i) 10 Na⁺ ions (Mod1), ii) a set of 500 (Mod2) and iii) a set of 1000 (Mod3) pseudo-ions (with Na⁺ Lennard-Jones parameters) whose partial charges were of +0.02 and +0.01, respectively. Finally, also the unbalanced (i.e. not neutralized) -10 charged system (Mod0) was considered for reference (see Figure 5.6 and Table 5.3). For each set-up both structures were optimised at MM level and then at QM/MM method according to the partitioning scheme described in Section 5.3. The systems were optimised by keeping frozen (at their MM optimised position) all the residues on the surface, including the added ions or pseudo-ions. Two minima were identified for each of four systems, corresponding to the reactant (ENZ-[L-Pro]) and the product (ENZ-[D-Pro]) complex, called respectively M1a and M2a (see Figure 5.7). We evaluated the equilibrium between the two minima by computing the relative energy of M1a and M2a for each set-up: the three electro-neutral models plus the charged system.

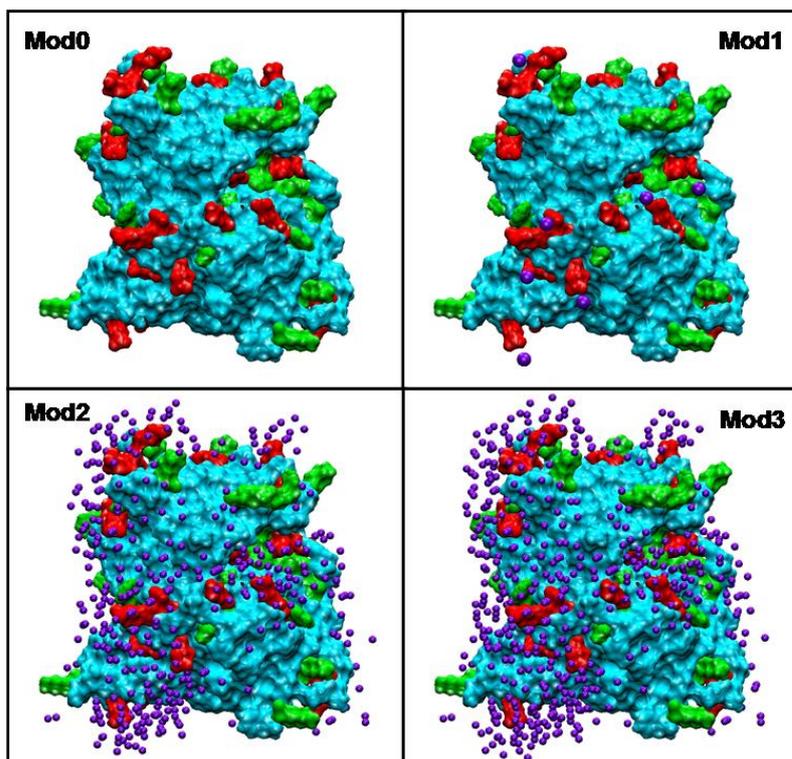


Figure 5.6 Distribution of the added point charges (purple spheres) around the TcPRAC enzyme. Negatively and positively charged residues are represented respectively with red and green surfaces. In Mod0 the total charge of the enzyme is not counterbalanced by any added charge.

Table 5.3 Number and value of the point charges added to the model system used for the QM/MM study of TcPRAC.

Model name	Overall charge	Added charges	Charge value
Mod1	0	10	+1
Mod2	0	500	+0.02
Mod3	0	1000	+0.01
Mod0	-10	-	-

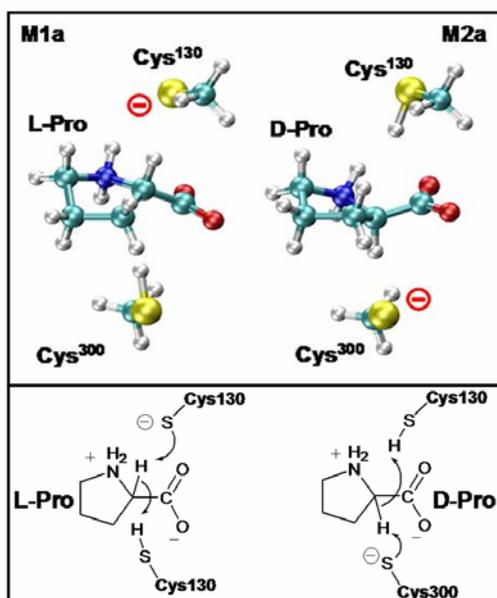


Figure 5.7 QM portion (i.e. *model-H*) of the QM/MM model system used for the study of TcPRAC.

In Table 5.4 and in Figure 5.8 we report the computed $\Delta E_{M2a-M1a}$ values. It is noticeable that in all cases M2a is less stable than M1a. We think this is a clue for the lower affinity of the enzyme for D-Pro and that this can be the basis for the protein conformational changes that allow the final release of the product to the aqueous medium. Under this respect, $\Delta E_{M2a-M1a}$ can be correlated to the binding constant between the enzyme and the two enantiomers of proline. It is worth noting that $\Delta E_{M2a-M1a}$ decreases (see Table 5.4) when increasing the number of the added charges to reach a value that is almost identical to the one obtained for the unbalanced system (Mod0). We noticed that the contribution to $\Delta E_{M2a-M1a}$ due to the enzyme residues is almost constant in all considered cases, and that the difference is mainly due to the added charges. In all the examined cases the contribution of the counter-ions was estimated by means of the direct fingerprint analysis (see Section 3.9), grouping all the ions in a unique group. As we can see in Figure 5.8, the effect of point charges decreases when increasing their number (keeping constant the total charge value). This behaviour can be explained if we consider the nature of the chemical reaction under study and the method we used to place the charges. As the reaction takes place, passing from M1a to M2a, a net charge moves from Cys130 to Cys300 (Figure 5.7 and Figure 5.11) thus changing the dipole moment of the molecule; since the ions (which are placed on the basis of the electrostatic potential at the surface) are not allowed to move, they stabilize the initial minimum (M1a) better than the other (M2a). When using only few point charges (Mod1), the effect of stabilizing the M1 minimum characterized by a net charge near the surface is very strong, but when the number of charges increases (Mod2 and Mod3), and the single value of each charge diminishes, then this stabilizing effect decreases, because the charge cloud is isotropically spread all around the enzyme. For this reason, we preferred to select the unbalanced system (Mod0) for the study of the enzymatic process (Section 5.5); in indeed we found that the standard method of balancing the net charge of the system using external point charges can lead to an error in this case. Of course, an even better model would be to consider the enzyme in a solvent box of explicit molecules with the right number of Na^+ ions fully solvated and equilibrated (i.e. after an appropriate MD run).

Table 5.4 Energy difference between the two minima: $\Delta E = E_{M2a} - E_{M1a}$; E_c represents the electrostatic contribution to ΔE due to the added external charges (ions or pseudo-ions); E_{Enz} represents the contribution from the enzyme itself.

Model name	ΔE (kcal mol ⁻¹)	E_c (kcal/mol ⁻¹)	E_{Enz} (kcal/mol ⁻¹)
Mod1	10.14	7.87	2.27
Mod2	3.37	1.20	2.17
Mod3	2.44	0.46	1.98
Mod0	2.68	0	2.68

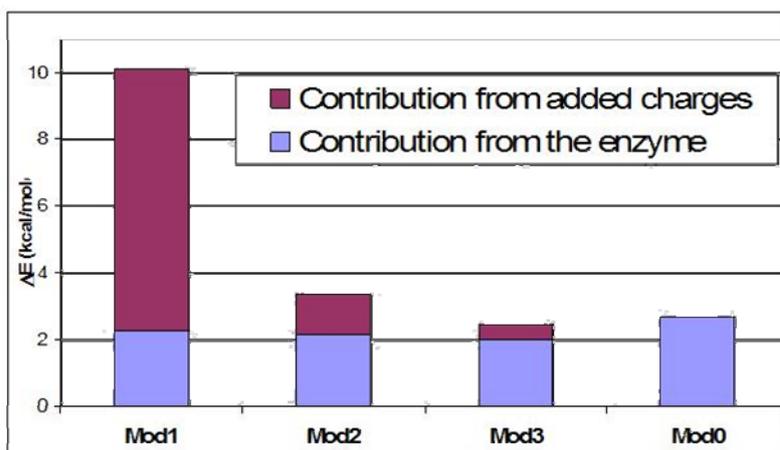


Figure 5.8 Bar diagram representing the energy difference between the M2a and the M1a minima for all the models studied. The contribution of added point charges is reported in each case as a red portion of the whole bar.

The treatment of electrostatic interactions within proteins has been extensively reviewed and discussed in some recent papers^{141,224,225} and herein we do not want to present a further comprehensive analysis of this problem neither a method of general validity. It is worth to point out that here we do not want to present our approach as a generally valid procedure. Anyhow, our data strongly suggest the need of an improved description of medium and long-range electrostatic interactions in order to improve the physical reliability of the simulations in this specific case. This clearly reveals the delicateness of this issue, and an alternative approach to the problem has been shown.

5.5 Exploring of the PES with a QM/MM potential

According to the discussion reported in Section 4.5 we used the unbalanced system Mod0, bearing a net charge of -10. The partitioning scheme described in Section 5.4 and used for the preliminary study on the relative stability of the minima M1a and M2a is used also to explore the whole Potential Energy Surface associated to the reaction process.

Three critical points have been located on the reaction surface: two minima (M1a and M2a), which correspond to the enzyme bound L- and D-proline, respectively, and a transition state (TSCa), for the inversion of chiral C^α. The Cys130 residue is not protonated in M1a and ready

to accept a proton from the substrate. Interestingly, the computed path reveals a highly asynchronous concerted process. In the corresponding transition state (TSCa) both Cys residues are nearly protonated and the unprotonated proline is almost planar (i.e., sp^2 -like C^α atom). Thus, our data strongly support the hypothesis that the PYC inhibitor binds to the active site by mimicking a carbanionic transition state rather than a carbanionic intermediate. In Figure 5.10 is reported the structure of the PYC inhibitor as given in crystallographic structure, superposed to the geometry of the found transition state. As a matter of fact, all our attempts to locate the unprotonated intermediate predicted for a stepwise reaction mechanism, failed. The two minima are closely isoenergetic, with the internal energy (E) and Gibbs free energy (G) difference being 2.7 and 1.5 kcal mol⁻¹, respectively. The computed energy barrier for the L \rightarrow D proline inversion is about 10.6 kcal mol⁻¹ (see Figure 5.9 and Table 5.5). G values were estimated by computing numerical frequencies on a region of the enzyme that comprises all QM atoms and nearby MM residues (the *high-medium* region) adopting the approximations described in Section 3.8; all other MM residues have been kept frozen so that they are able to exert their (electrostatic and steric) influence on the former ones.

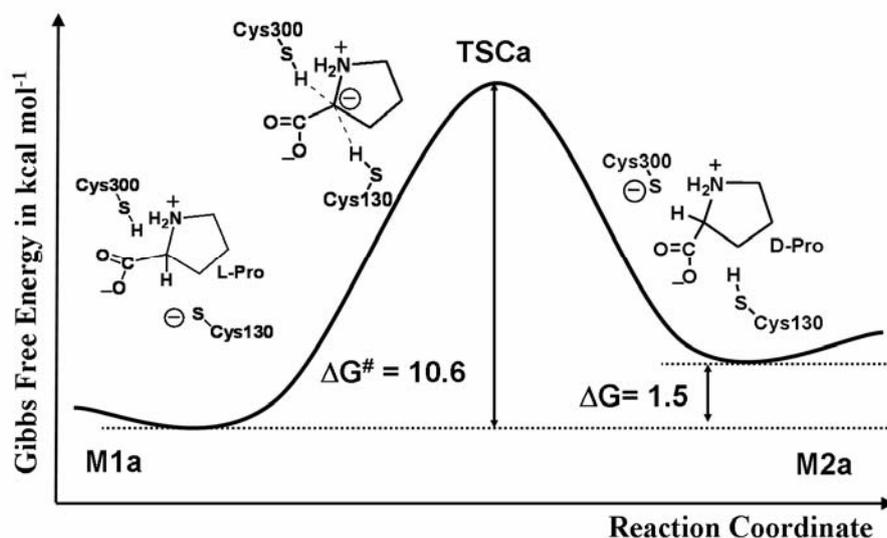


Figure 5.9 Gibbs Free Energy profile for the enzymatic reaction.

Table 5.5 QM/MM energy (E) and Gibbs Free Energy ($G = E + \text{Thermal Free Energies}$). The reported values (kcal mol⁻¹) are relative to M1a

	E	G	$E(\text{in vacuo})$	$E(\text{charges})$
M1a	0.0	0.0	0.0	0.0
	12.8	10.6	28.3	11.7
	2.7	1.5	-0.3	3.1

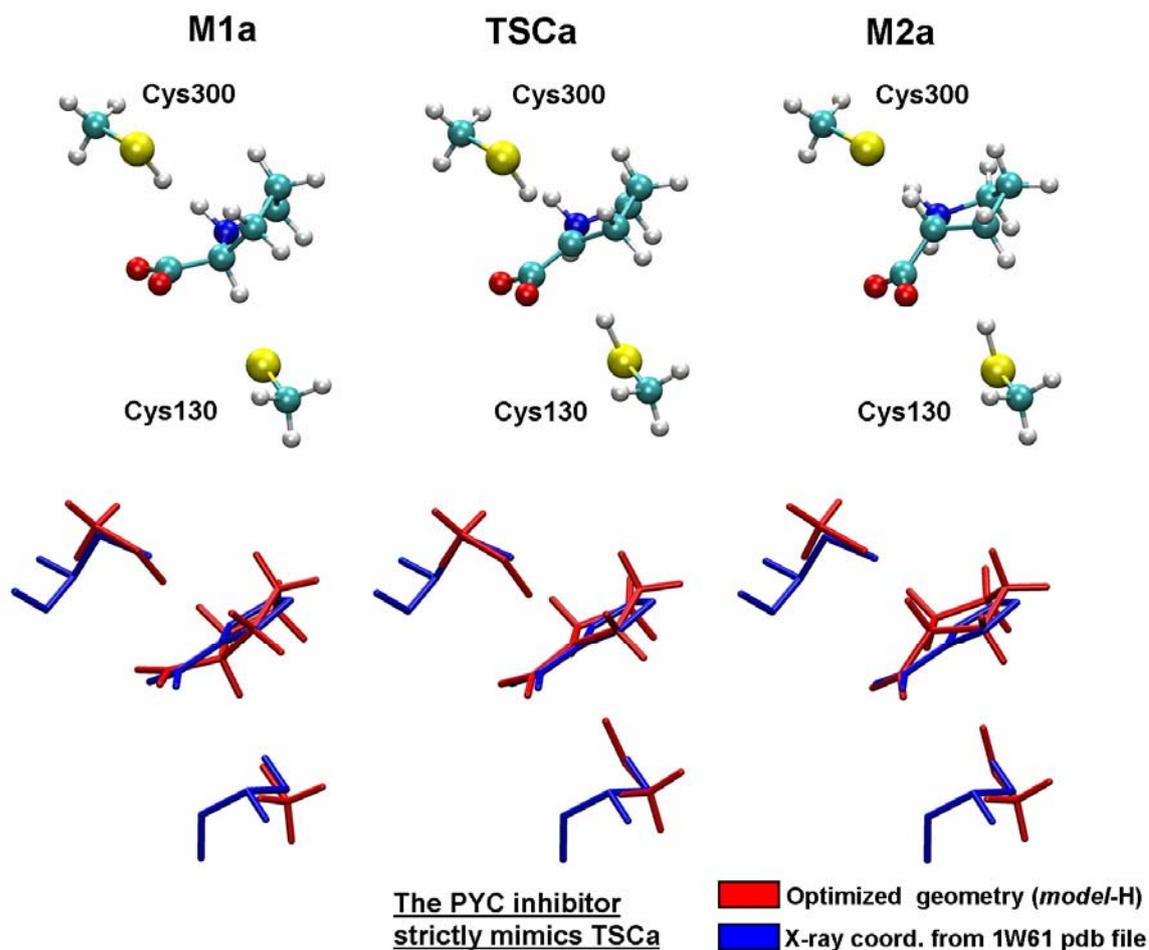


Figure 5.10 Representation of the geometry of *model-H* subsystem from QM/MM optimisations. In the lower part is reported a superposition with the corresponding atoms as taken from X-ray structure.

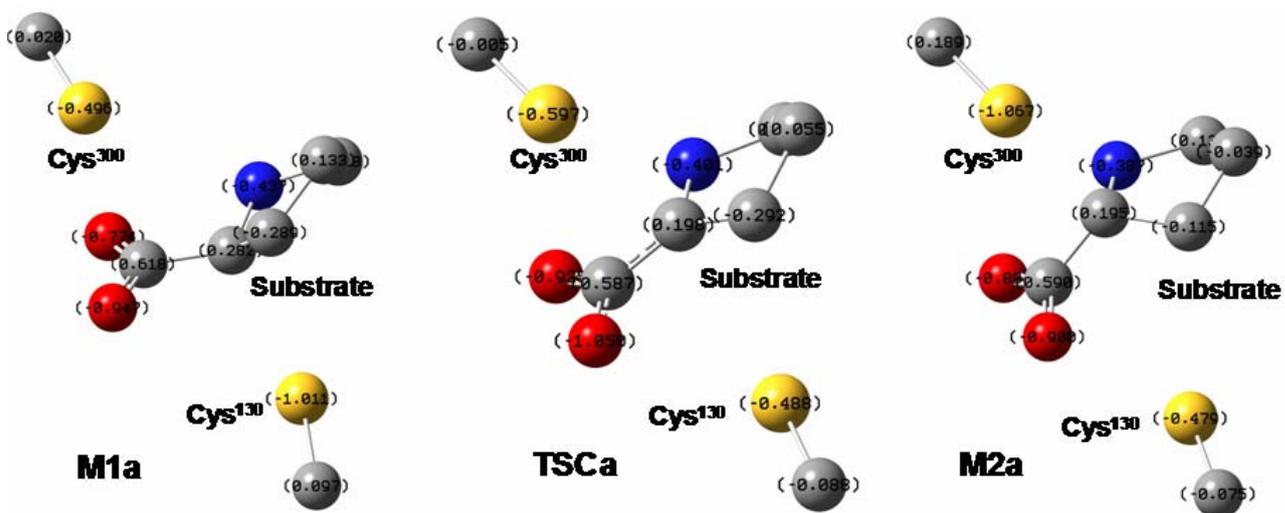


Figure 5.11 Atomic point charges of the QM region. H atoms are omitted for clarity.

5.6 Understanding the enzymatic catalysis

To evaluate the effect of the surrounding residues (Figure 5.12) on the energetic of the process, we performed *in vacuo* single-point calculations ($E(\text{in vacuo})$) on the stand-alone QM region (*model-H* sub-system) using its optimized QM/MM structure. Even if the two equilibrium structures remain almost isoenergetic, the computed energy barrier significantly increases ($28.3 \text{ kcal mol}^{-1}$) with respect to the QM/MM value, while it drops to $11.7 \text{ kcal mol}^{-1}$ when the enzyme atomic point charges are included in the calculations (**E(charges)**). This result unambiguously demonstrates that the enzymatic environment exerts a strong electrostatic stabilization on the transition state. To individuate the residues playing the most important role in stabilizing/destabilizing a point of the PES with respect to another we used the DFP method described in Section 3.9.

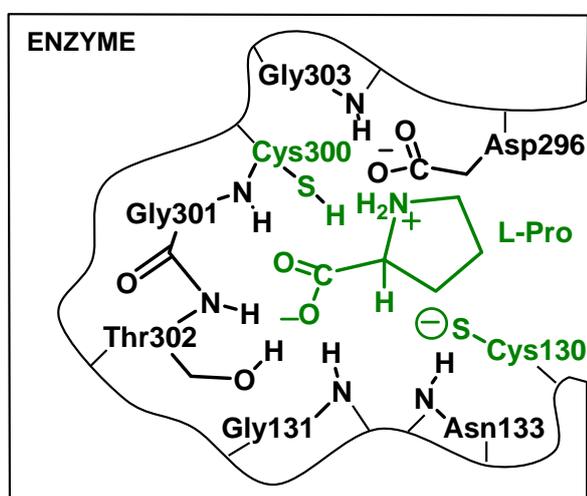


Figure 5.12 Schematic representation of the residues interacting with the QM sub-region (shown in green).

5.6.1 DFP analysis applied to M1a and TSCa

When applying the DFP analysis to the structures of M1a and TSCa we obtained an information about the residues able to promote (or oppose) the transition from the minimum (M1a) to the transition state, and thus favouring (or disfavouring) the enzymatic process by electrostatic lowering (or enhancing) of the associated barrier. In Figure 5.13 is reported the stability parameter S (expressed in kcal mol^{-1}) as a function of the residue number. It is easy to notice that only few residues exert a significant electrostatic effect; some of them are titratable residues bearing a net positive or negative charge, exerting their interaction far away from the active site (Figure 5.14-a and Table 5.6). More interesting are some residues which lie near the active site (Figure 5.14-b and Table 5.7), directly interacting with either the substrate or the Cys couple, or with both at once.

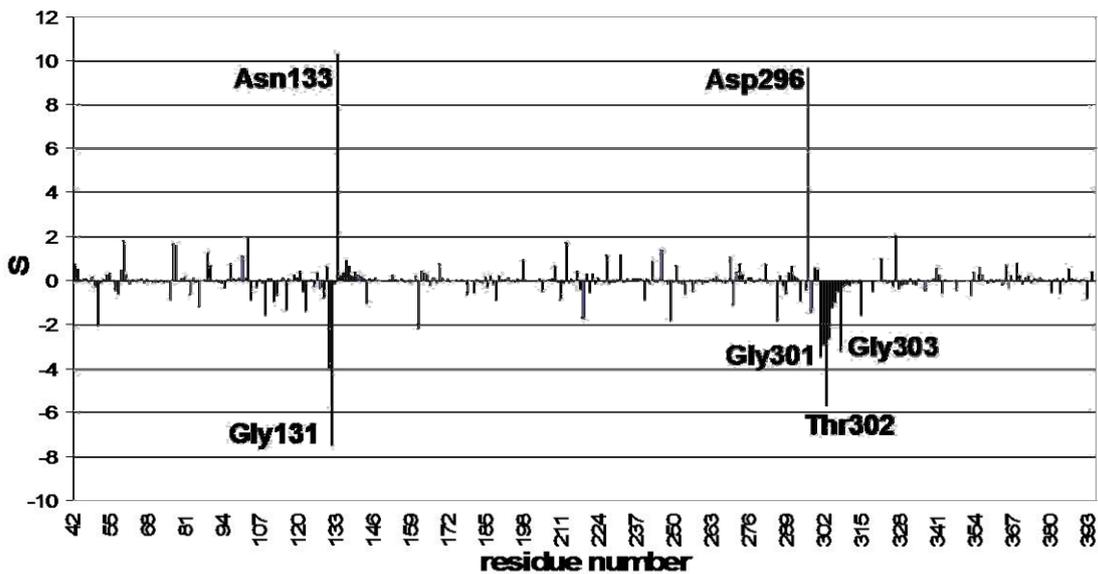


Figure 5.13 The stability parameter S (kcal mol⁻¹) as a function of the residue number.

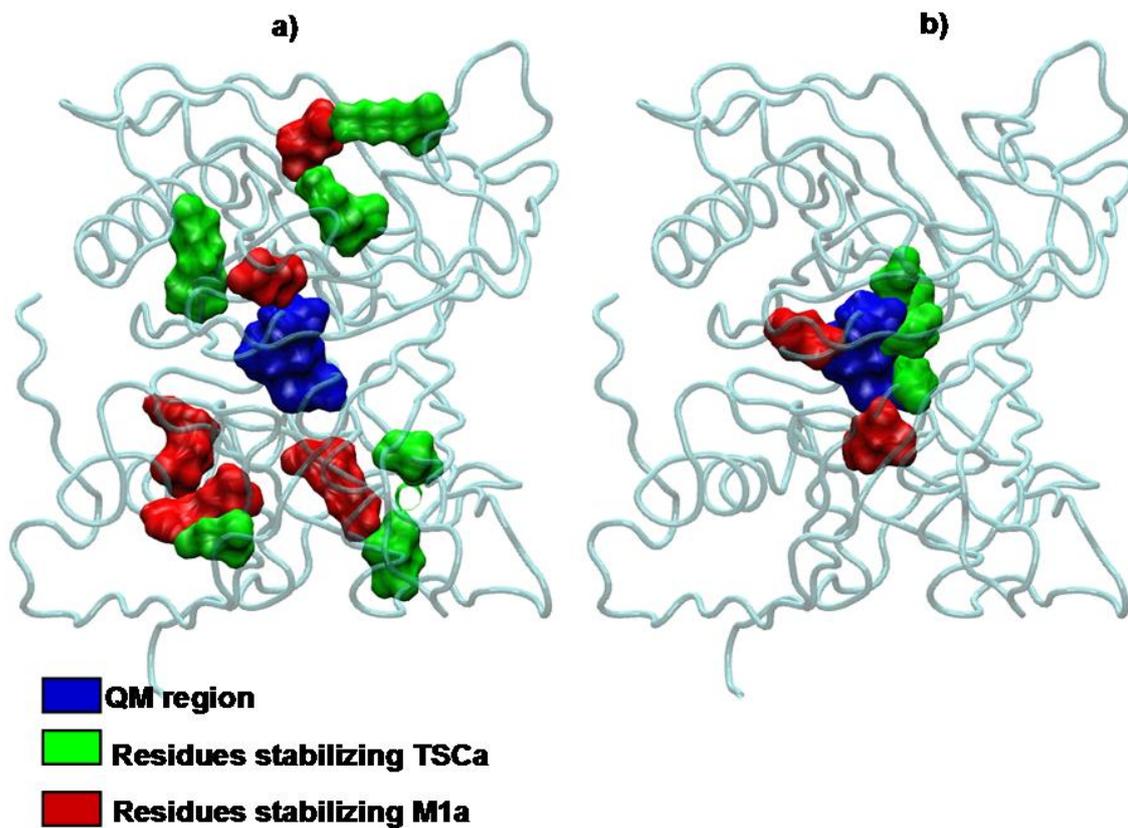


Figure 5.14 Residues having a significant effect on electrostatic stabilization/destabilization of TSCa with respect to M1a: a) charged long-range interacting residues (smaller effect, see Table 5.6); b) uncharged residues in close contact with the QM sub-region (higher effect, see Table 5.7).

Table 5.6 Long range interacting residues. If $S < 0$ the residue stabilizes TSCa.

Residue	S	charge
Arg248	-1.81	+
Lys307	-3.20	+
Lys314	-1.57	+
Asp50	-2.06	-
Asp161	-2.19	-
Asp108	-1.56	-
Arg59	1.82	-
Lys76	1.70	+
Arg88	1.29	+
Asp212	1.70	-
Glu326	2.05	-

Table 5.7 Short range interacting residues. If $S < 0$ the residue stabilizes TSCa.

Residue	S	Interacting with
Gly131	-7.46	Substrate-COO-
Cys301	-2.89	Substrate-COO-
Thr302	-5.65	Substrate-COO-
Cys303	-2.67	Cys300
Asn133	10.32	Cys130
Asp296	9.64	Cys300

A couple of residues (Asn133 and Asp296) have been found to stabilize M1a with respect TSCa. This can be explained if we consider that in TSCa the charge on the S atom of Cys130 is less negative than in M1a (Figure 5.11), thus the Asn133-N-H - - - S-Cys130 H-bond weakens in passing from M1a to TSCa (Figure 5.15). On the contrary the charge on S atom of Cys300 becomes more negative in TSCa and thus the destabilisation interaction with the negatively charged side-chain of Asp296 increases (Figure 5.16).

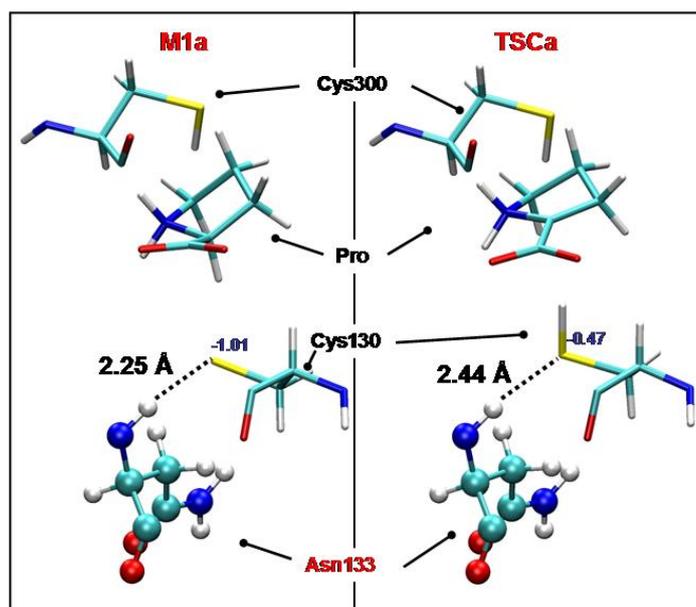


Figure 5.15 Asn133 stabilizes M1a with respect to TSCa. In TSCa the charge on the S atom of Cys130 is less negative than in M1a. The Asn133-N-H - - - S-Cys130 H-bond is weaker in TSCa with respect to M1a.

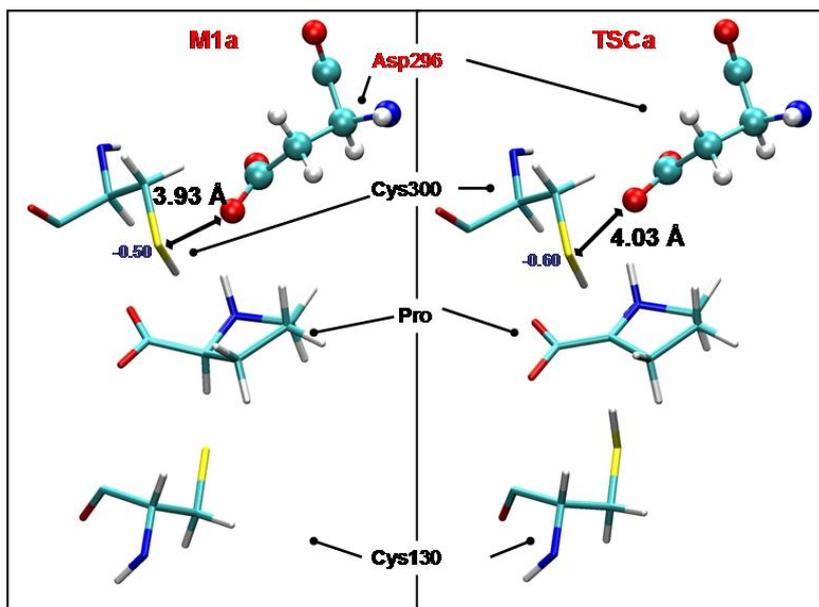


Figure 5.16 Asp296 stabilizes M1a with respect to TSCa. In TSCa the negative charge of the S atom in Cys300 increases. The Asp296-CO1-O2 - - - S-Cys300 distance increases because of the unfavorable interaction.

On the basis of the S diagram (Figure 5.13) it is possible to identify some residue exerting a strong stabilization of the transition structure. Their effect can be easily understood by comparing the geometries and the atomic point charge distributing in M1 and TSCa (Figure 5.11). Being the substrate almost completely unprotonated in TSCa, the partial negative charge left on it is not localised on the C $^{\alpha}$, but shared with the adjacent carboxylic group via an extended π -system (Figure 5.29) whose existence can be inferred by the loosening of C-O bonds concomitant to the augment of the negative charge on O atoms, when passing from M1a and TSCa. A residue capable of stabilizing this upcoming negative charge would play a stabilizing effect on the transition structure itself. The DFP analysis automatically identified these residues (Gly131, Gly303, Thr302) as the ones interacting with the carboxylic group of the substrate as H-bond donor. In particular the Gly131-N-H - - - O2-CO1-Pro and the Gly301-N-H - - - O1-CO2-Pro H-bonds are stronger in TSCa with respect to M1a (see, respectively, Figure 5.17 and Figure 5.18); Thr302 is a donor of two H-bond (Thr302-O-H - - - O2-CO1-Pro and Thr302-H-H - - - O2-CO1-Pro), which contribute stabilizing TSCa (Figure 5.19). Another important residue is Gly303, *via* the stabilisation of the upcoming negative charge on the S atom of Cys300 by formation of the Gly303-N-H - - - S-Cys300 H-bond (Figure 5.20).

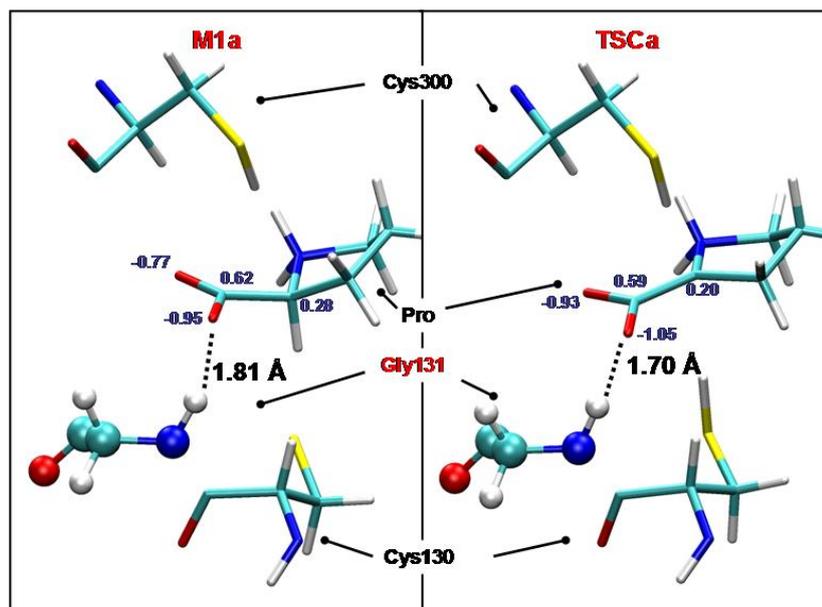


Figure 5.17 Gly131 stabilizes TSCa with respect to M1a. In TSCa the total negative charge of COO⁻ of the Pro residue increases. The Gly131-N-H ... O2-CO1-Pro H-bond is stronger in TSCa with respect to M1a.

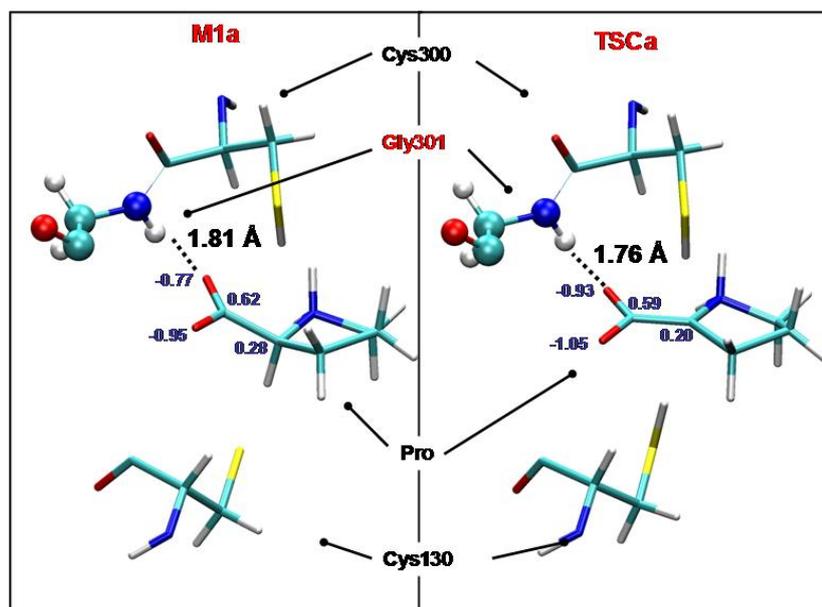


Figure 5.18 Gly301 stabilizes TSCa with respect to M1a. In TSCa the total negative charge of COO⁻ of the Pro residue increases. The Gly301-N-H ... O1-CO2-Pro H-bond is stronger in TSCa with respect to M1a.

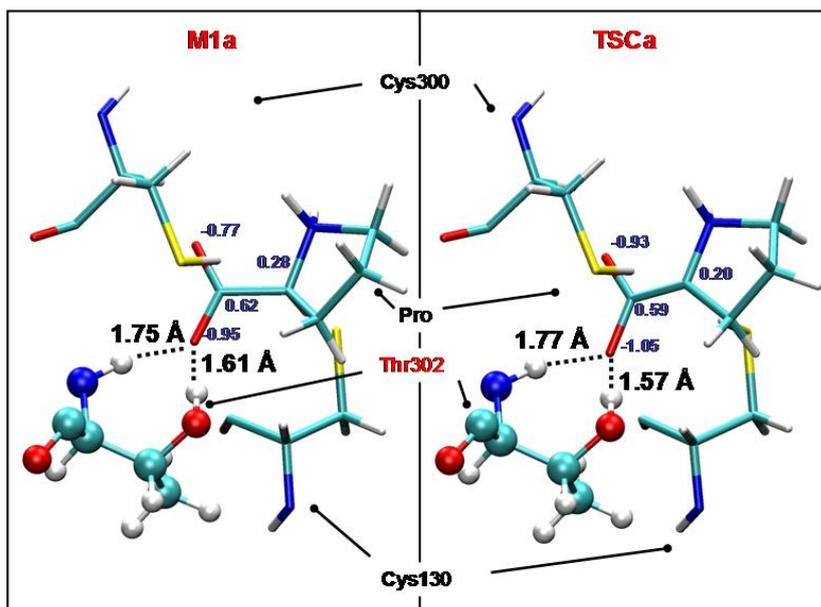


Figure 5.19 Thr302 stabilizes TSCa with respect to M1a. In TSCa the total negative charge of COO⁻ of the Pro residue increases; the Thr302-O-H - - O2-CO1-Pro H-bond is stronger in TSCa with respect to M1a.

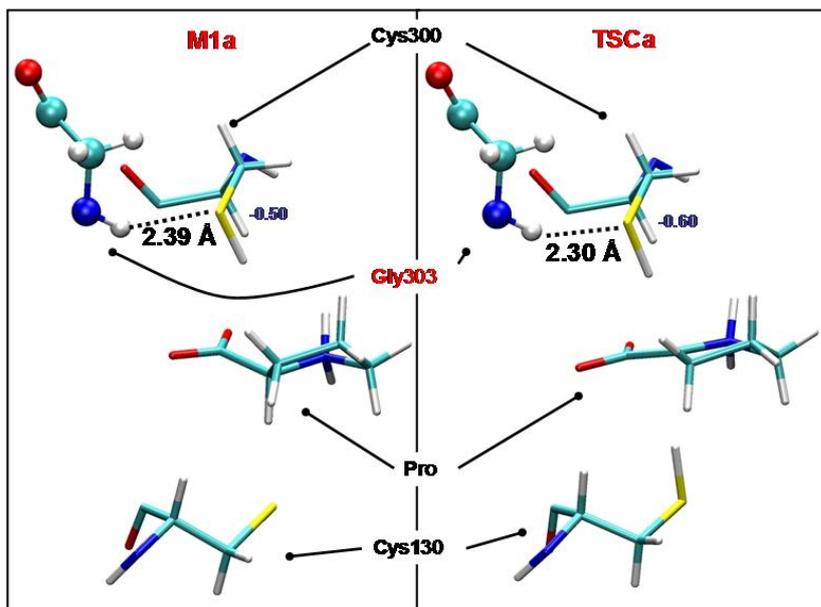


Figure 5.20 Gly303 stabilizes TSCa with respect to M1a. In TSCa a more negative charge accumulates on the S atom; the Gly303-N-H - - S-Cys300 H-bond is stronger in TSCa with respect to M1a.

5.6.2 DFP analysis applied to M1a and M2a

If the DFP analysis is applied to the structures of M1a and M2a it is possible to identify the residues playing a differential electrostatic stabilization of the two equilibrium stationary points of the reaction under examination. In Figure 5.21 is reported the stability parameter S (expressed in kcal mol⁻¹) as a function of the residue number. As seen for the M1a/TSCa transition, even in this case only few residues exert a significant electrostatic effect; some of them are titratable residues bearing a net positive or negative charge, exerting their interaction far away from the active site (Figure 5.22-a and Table 5.8). More interesting are some residues

which lie near the active site (Figure 5.22-b and Table 5.9), directly interacting with either the substrate or the Cys couple, or with both at once.

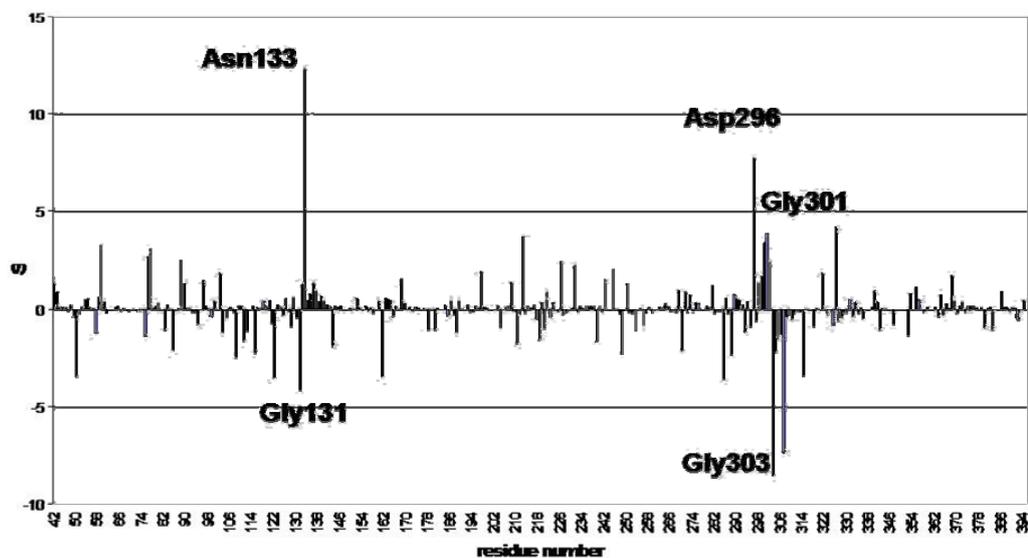


Figure 5.21 The stability parameter S (in kcal mol^{-1}) is plotted as a function of the various residues.

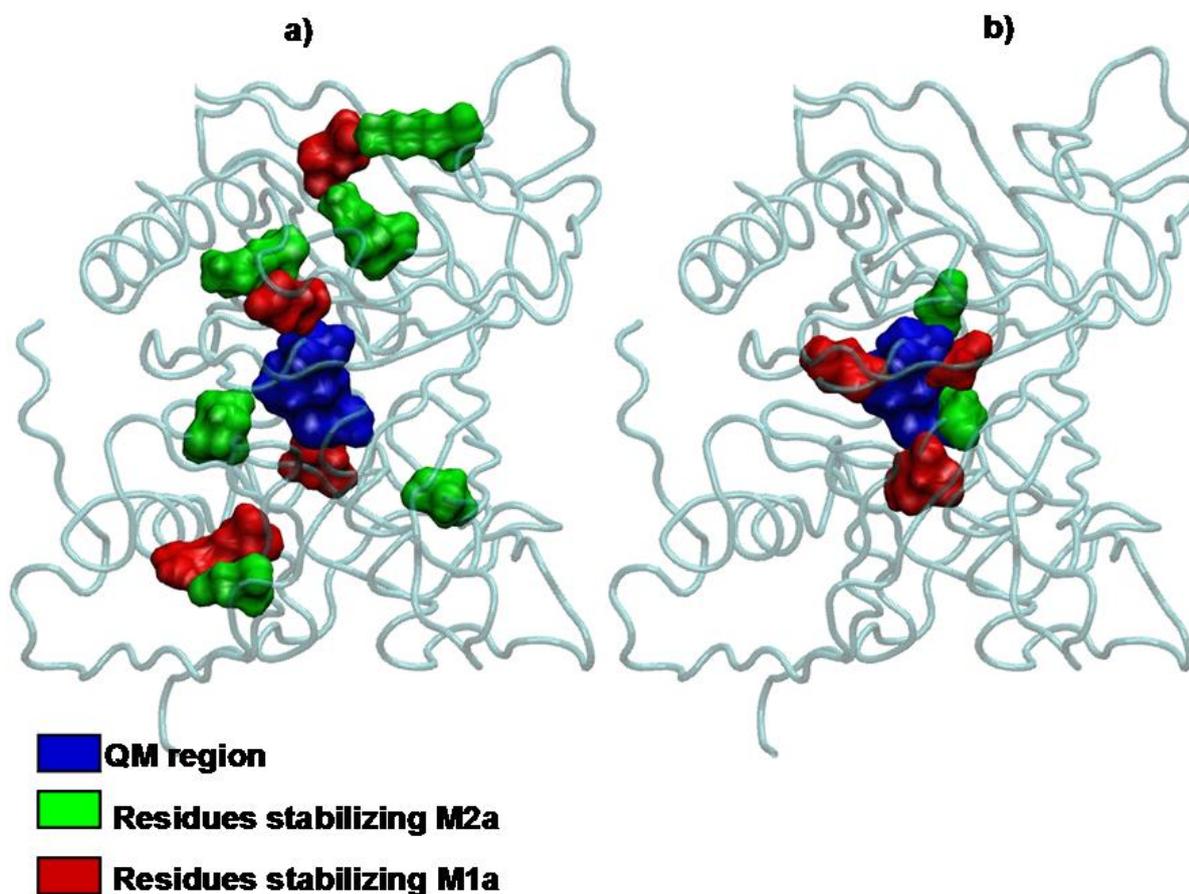


Figure 5.22 Residues having a noticeable effect on electrostatic stabilization/destabilization of M2a in respect to M1a: a) charged long-range interacting residues (Table 5.8); b) uncharged residues in close contact with QM sub-region (Table 5.9).

Table 5.8 Long range interacting residues. If $S < 0$ the residue stabilizes M2a.

Residue	S	charge
Lys307	-7.39	+
Lys314	-3.42	+
Lys285	-3.60	+
Asp161	-3.49	-
Asp122	-3.54	-
Asp50	-3.48	-
Glu326	4.18	-
Asp212	3.69	-
Arg59	3.19	+
Lys77	3.00	+

Table 5.9 Short range interacting residues. If $S < 0$ the residue stabilizes M2a.

Residue	S	Interacting with
Gly131	-4.15	Substrate-COO-
Cys303	-8.60	Cys300
Asn133	12.29	Cys130
Asp296	7.69	Cys300
Gly301	3.85	Substrate-COO-

As seen before the Asn133 stabilizes M1a by means of the Asn133-N-H - - - S-Cys130 H-bond, that is loosened in TSCa and mostly in M2a, because of the decrease in the negative character of the charge on the S atom of Cys130 side-chain (Figure 5.23). The M1a structure is also stabilized by Asp296, whose unfavourable interaction with Cys300 side-chain increases when passing from M1a to M2a (Figure 5.24), due to the more the negative charge on the S atom. Gly301 stabilizes M1a because, in this structure, the COO⁻ group of the Pro residue is better oriented for an H-bond interaction (Figure 5.25).

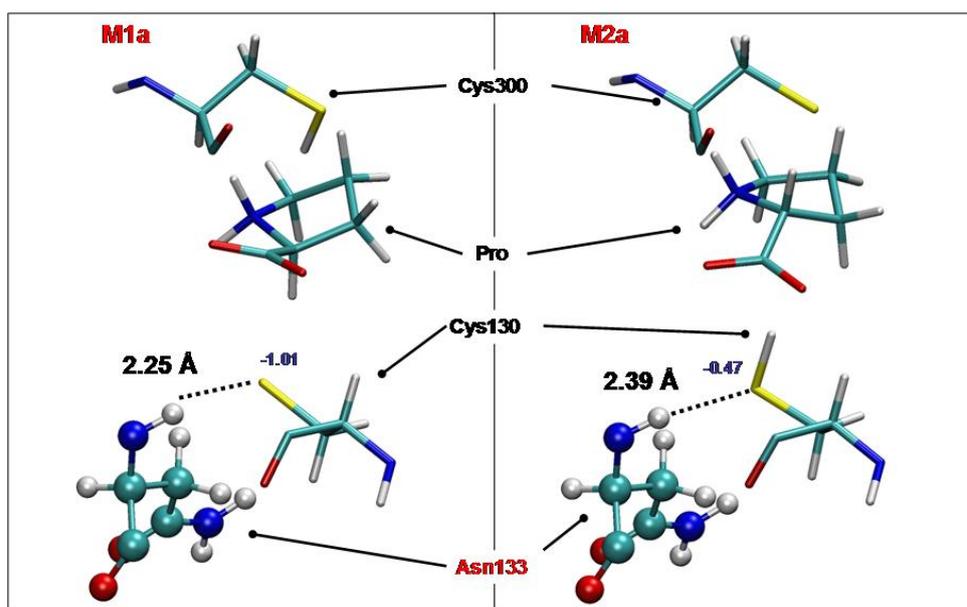


Figure 5.23 Asn133 stabilizes M1a with respect to M2a. In M2a the charge on the S atom of Cys130 is less negative than in M1a. The Asn133-N-H - - - S-Cys130 H-bond is weaker in M1a with respect to M2a.

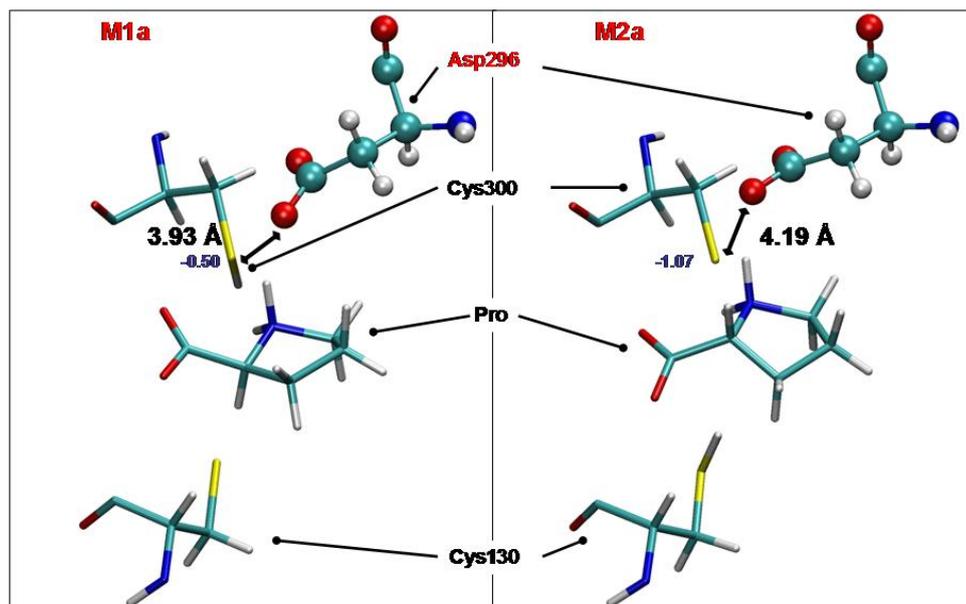


Figure 5.24 Asp296 stabilizes M1a with respect to M2a. In M2a the negative charge of the S atom in Cys300 increases. The Asp296-CO1-O2 - - - S-Cys300 distance increases for the unfavourable interaction.

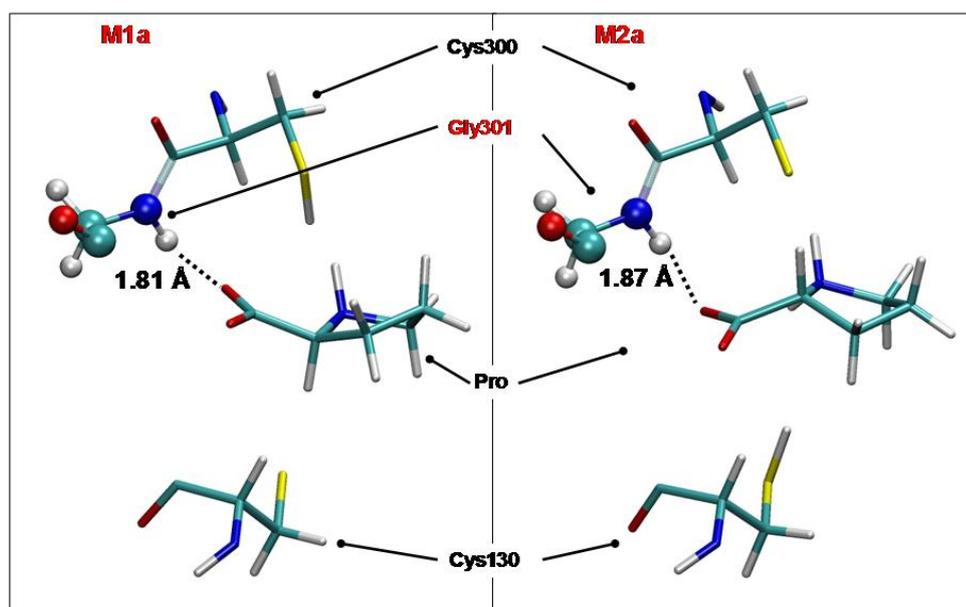


Figure 5.25 Gly301 stabilizes M1a. In M1a the COO⁻ group of the Pro residue is better oriented for the H-bond interaction with Gly301.

The M2s structure is favoured over M1a by Gly131 because of the better orientation of the Gly131-N-H - - - -OOC-Pro H-bond, resulting in a stronger interaction. In M2a the charge on the S atom of Cys300 is far more negative with respect to M1a, so the Gly133-N-H - - - S-Cys300 H-bond is stronger in M2a.

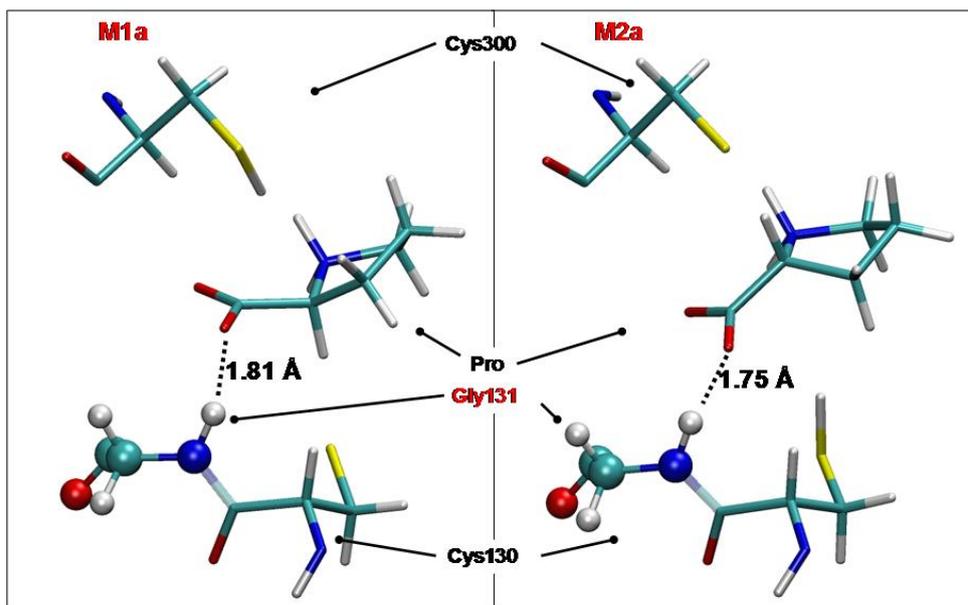


Figure 5.26 Gly131 stabilizes M2a. In M2a the geometry is more favorable for an effective Gly131-N-H - - - OOC-Pro H-bond interaction.

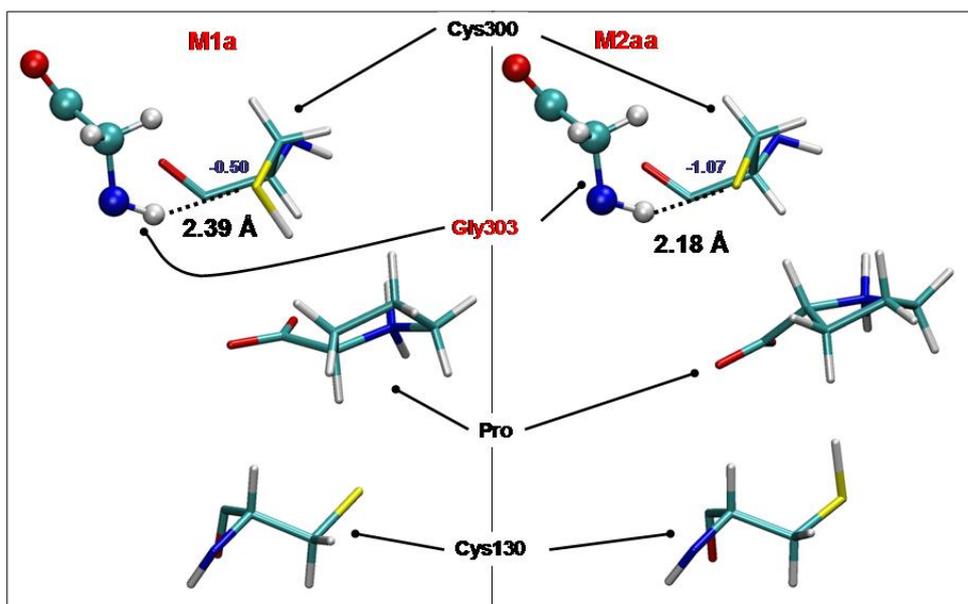


Figure 5.27 Gly303 stabilizes M2a. In M2a a more negative charge accumulates on the S atom. The Gly133-N-H - - - S-Cys300 H-bond is stronger in M2a with respect to M1a.

5.7 Conclusions

In this section an accurate computational study of the *Trypanosoma cruzi* proline racemase has been described. The set-up procedure has been described in details, pointing out the need of an accurate modelling of the system. The enzyme reaction mechanism has been described using an hybrid QM/MM potential, used to explore the PES associated to the stereoinversion of the C^α of the substrate (a zwitterionic proline). The calculations account for a concerted (but highly asynchronous) process of double proton transfer between the substrate and a couple of catalytically relevant cysteine residues (Cys couple). The data were used to show the importance of taking into account, for obtaining reliable results, both the reactive core and a larger region of the active site. Some residues have been individuated, by means of a general method (DFP analysis), as the most active in the stabilisation of the energy of the Transition Structure. By comparing the geometries and charge distribution of both the reactant (M1a) and the transition state (TCSa) the electrostatic catalysis of the enzyme have been fully interpreted. Being the process highly asynchronous, in the transition structure, the substrate is almost unprotonated and fully planar, and the partial charge left on the C^α is delocalised within a π-orbital system (Figure 5.29) containing the carboxylic group. The charge of this group is thus more negative in TSCa with respect to both M1a and M2a. The residues individuated by DFP analysis as stabilizing for TSCa (Figure 5.28) are H-bond donor toward the carboxylic group of the substrate, and thus able to stabilize the upcoming negative charge.

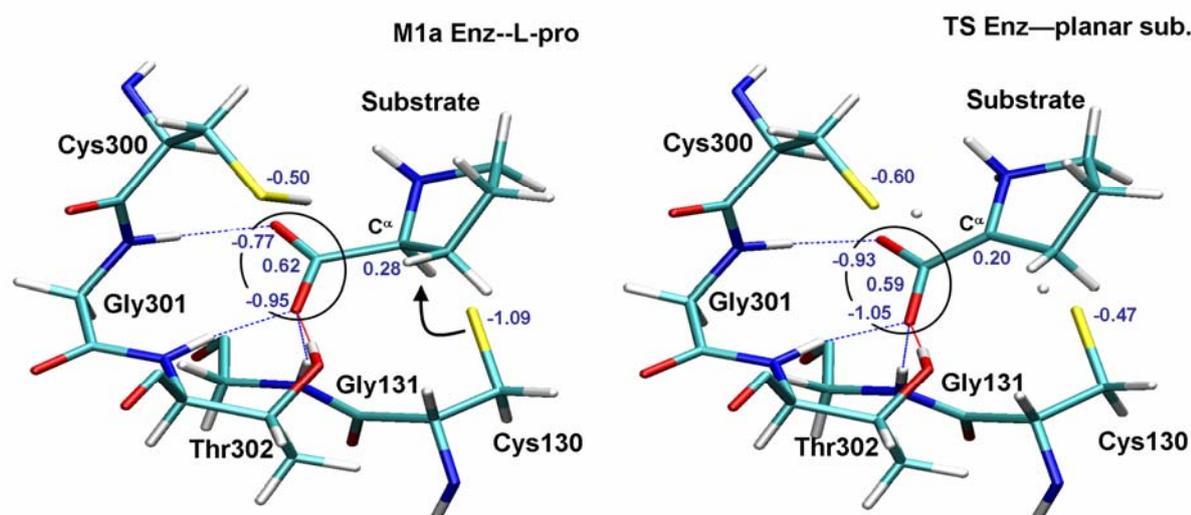


Figure 5.28 Representation of the residues stabilizing the transition state.

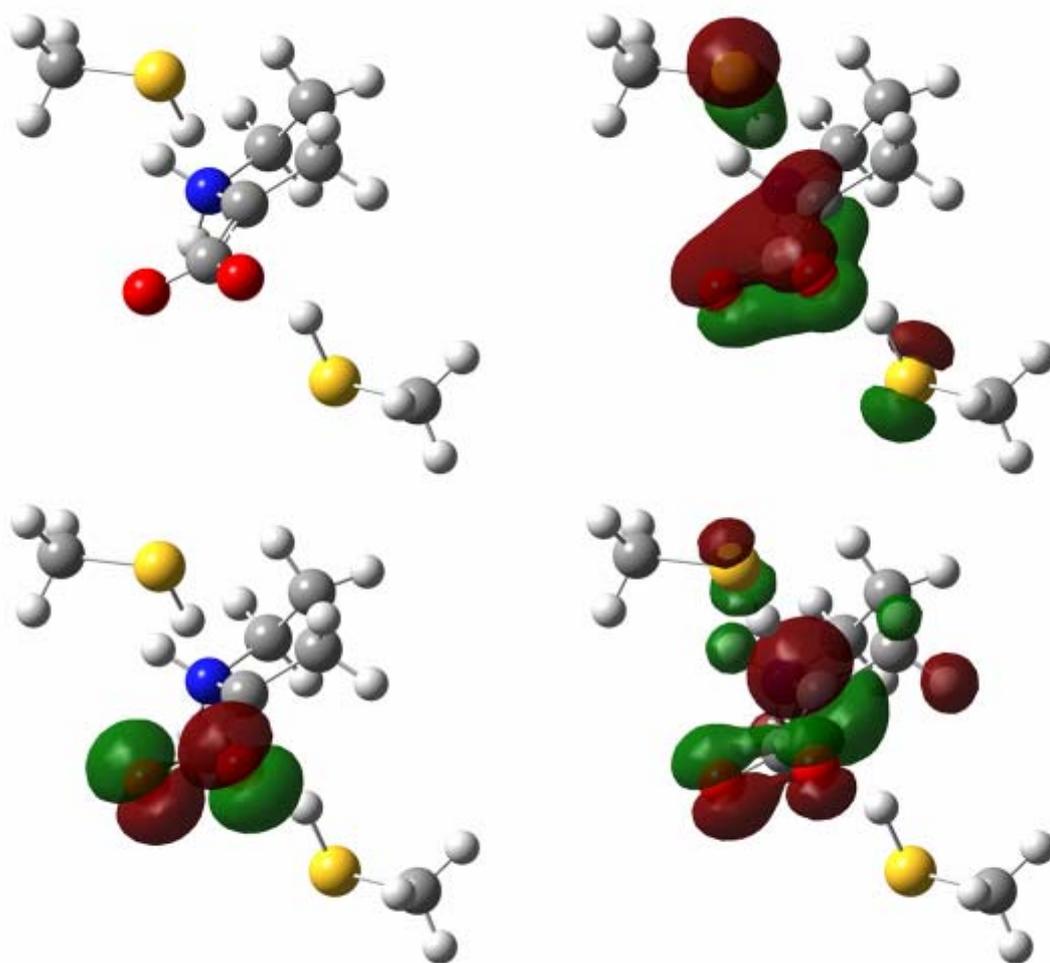


Figure 5.29 Representation of some π -symmetry molecular orbitals within the model-H system.

6 The Enzyme Diaminopimelate Epimerase

6.1 Introduction

The enzyme whose study is reported in this Chapter is the *Haemophilus influenzae* diaminopimelate epimerase.^{226,227} This enzyme is responsible for the epimerization of L,L- to D,L-*meso*-diaminopimelate (DAP). The racemization of the more common L-amino acids to the corresponding D-amino acids is of primary importance in the case of bacteria to provide substrates for the construction of cell walls which are resistant to the action of host protease. In particular D,L-*meso*-diaminopimelate is one of the most important component of the peptidoglycan layer. Diaminoepimelate epimerase belongs, like proline racemase (described in Chapter 5) to the class of pyridoxal phosphate-independent²¹⁴ racemases.²¹³ It operates via a “two base” mechanism that involve one active-site cysteine thiolate (Cys-73) acting as a base to deprotonate the α -carbon and a second cysteine thiol (Cys-217) that, through a general-acid catalysis, reprotonates the opposite site of the substrate. The knowledge of the mechanism of this enzyme is of primary importance especially for the possible applications in drug design. In particular specific inhibitors of the diaminopimelate epimerase may lead to a new class of antibiotics, that could be a very powerful weapon against the microbial resistance, more effective than the conventional ones. Moreover, it could become possible to determine the electrostatic interactions that control the substrate selectivity of DAP epimerase and determine the residues nearby the active site that have a significant effect on the catalysis.

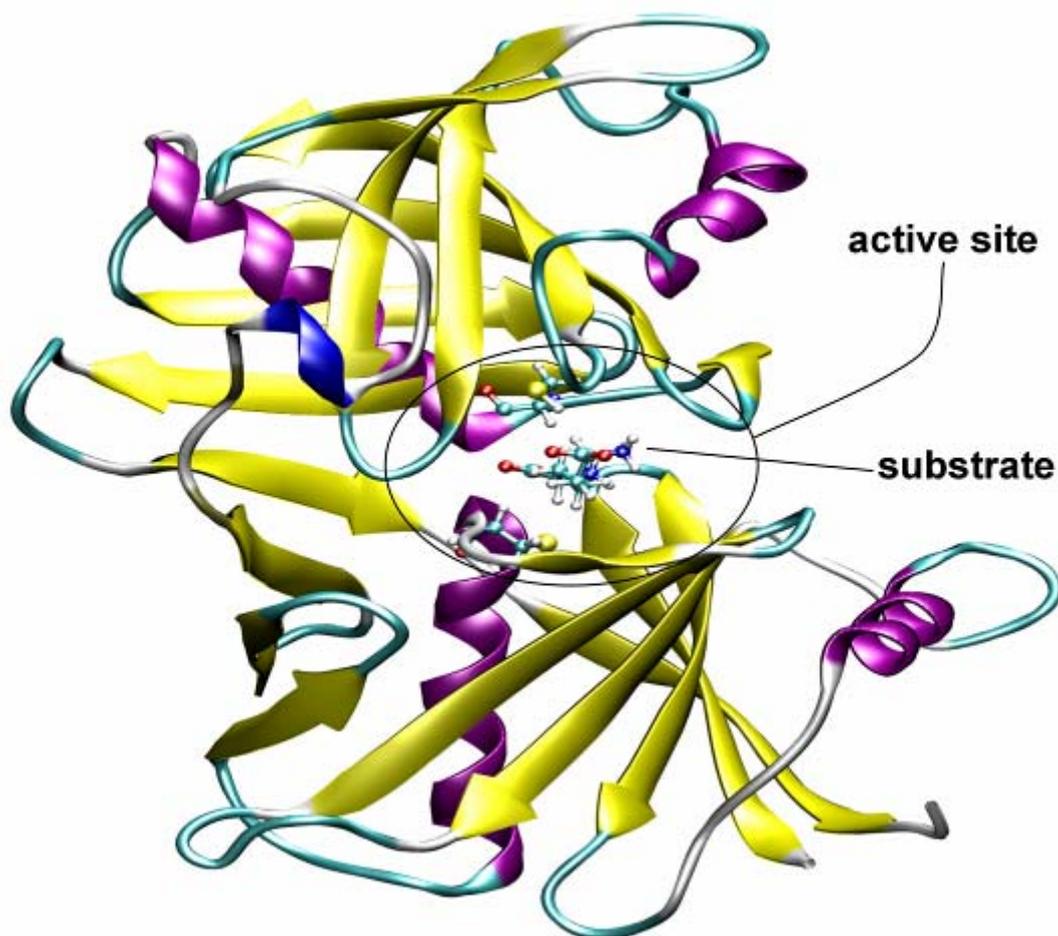


Figure 6.1 Sub-unit A of the diaminopimelate epimerase enzyme.

6.2 Setting-up the system

To model the system under examination, i.e. the diaminoepimelate (DAP) epimerase, we used two very recent crystal structures²²⁸ of DAP epimerase from *Haemophilus influenzae* bounded with different isomers of the irreversible inhibitor and substrate mimic aziridino analogues of DAP (AziDAP)²²⁹ [pdb codes: 2GKE (L,L-AziDAP), resolution 1.35 Å and 2GKJ (D,L-AziDAP), resolution 1.70 Å]. The two diastereomeric inhibitors present a covalent bond between the methylene carbon of the aziridine ring and the sulphur atom of Cys-73 or Cys-217. The irreversible inhibition of the enzyme takes place after the nucleophilic attack of the sulphur on the aziridine ring. There is little difference in the DAP epimerase backbone between the two crystal structures. The complexes under examination are excellent mimics of the reagent and of the product bounded to the active site of the enzyme. To build the model we use the coordinates of 2GKE pdb (for its better resolution with respect 2GKJ) changing the covalently bounded L,L-AziDAP, with the natural substrate L,L-DAP, and retaining only the A conformer of aminoacids present in the pdb with multiple conformations. The model-system was protonated using the H++¹⁴² software using the default parameters. This code uses

an automated algorithm that computes pKa values of ionizable groups in macromolecules and adds the missing hydrogen atoms according to the specified pH of the environment. Positions of added H atoms are also optimized by this algorithm.

The L,L-DAP and D,L-*meso*-DAP molecules were parametrised using the GAFF⁴⁹ force field. Partial atomic charges were assigned to atoms using the AM1-BCC^{161,162} method as implemented in the *antechamber*⁴⁵ module of the AMBER 8.0⁴⁵ package.

The model system was then minimized at the MM level using the AMBER 8.0 package. The minimization was carried out with the Amber(*ff99*)⁴⁵ force field. The system was fully minimized using a full conjugate gradient minimization algorithm and the GB¹⁴⁶ model to simulate the aqueous environment as implemented in the *sander* module of AMBER 8.0.⁴⁵ The optimization procedure is carried out until the root mean square deviation (RMSD) of the Cartesian elements of the gradient is less than 0,0001 kcal mol⁻¹.

Finally we checked if the crystal structures of 2GKE and 2GKJ strictly mimics the natural binding mode of the substrates or if the covalent bond between the enzyme and L,L-AziDAP and D,L-AziDAP can perturb the natural binding mode of L,L-DAP and D,L-*meso*-DAP.

For this reason we carried out a conformational study of the binding mode of L,L-DAP and D,L-*meso*-DAP within the protein environment using different methods.

6.2.1 Cluster Analysis

We carried out high temperature Molecular Dynamics starting from the minimized structure of the complex. The region of 5 Å around the substrate was free to move. The system was heated from 0 K to 800 K in 100 ps and then a trajectory of 2 ns was computed at constant temperature (800 K). The integration step of 2 fs was used in conjunction with the SHAKE^{35,36} algorithm to constrain the stretching of bonds involving hydrogen atoms. The coordinates of the system were saved on a trajectory file every 2 ps, giving a total of 1000 structures. Solvation effects were taken into account using the GB model with a dielectric constant of 78.5. To determine the most populated conformations of L,L-DAP and D,L-*meso*-DAP within the protein binding pocket, we performed a Cluster Analysis on the different conformations visited by the two molecules during the simulation. To this purpose we used the MMTSB²²² toolset and we clustered different conformations of the substrates by structural similarity (using the program *kclust* and adopting a fixed radius of 1.0 Å on the Cartesian coordinates RMSD computed for heavy atoms).

We determined the centroid of each cluster. For each cluster we chose the structure closest to the corresponding centroid as representative of the cluster itself (this structure is characterized by the smallest RMSD value with respect to the centroid).

6.2.2 Simulated Annealing

Using the same parameters of the previous simulation we performed 10 cycle of simulated annealing for each L,L-DAP and D,L-*meso*-DAP substrates, heating the system from 0 K to 1000 K in 30 ps, holding at 1000 K for equilibration for 10 ps, and cooling from 1000 K to 0 K in 60 ps. The heat bath coupling for the system is tight for heating and equilibration (0.1 ps), while the cooling phase is divided in three period, the first 48 ps are very slow (coupling of 5.0 ps) followed by 6 ps with coupling of 1.0 ps and the last 6 ps from 0.1 to 0.05 ps. At the end of the simulated annealing a complete minimization is carried out and the final coordinates of the complexes are retained.

6.2.3 Docking

The model derived by the pdb structure 2GKE is used for a docking calculation with the after deletion of the ligand from the cavity of the active site. The orientation sampling is based on spheres calculated by the *sphgen*²³⁰ program within 10.0 Å root mean square deviation (RMSD) from every atom of the minimized structure of the ligand. Partial atomic charges for the substrates are calculated with the AM1-BCC^{161,162} method. We carried out flexible ligand docking using the Anchor-and-Grow algorithm implemented in DOCK6²³¹ with the Grid-Based Score function as primary scoring function. The results are then rescored with the new AMBER score as secondary scoring function allowing a minimization of the ligand and residues within 5 Å (the same mobile residues of the previous calculations) for 100 steps. The parameters used in the grid construction and in the docking calculations are provided in the best 10 poses of the ranking of L,L-DAP and D,L-*meso*-DAP are used for subsequent QM/MM calculations.

6.3 The structure of the complex

The conformational study of the binding mode of LL-DAP and DL-DAP using the three different methods previously described above do converge all on complex models very similar to the crystallographic structures of 2GKE and 2GKJ. By observing the structure of the obtained complexes it is possible to understand the preference of the enzyme in binding only DAP isomers having the L configuration at the distal ϵ -carbon. The binding pocket is characterized by an asymmetric disposition of residues able to donate or accept H-bond, only to an L-aminoacidic group. The carboxyl group forms a salt bridge with the side positively charged chain of Arg-209 and three H-bond with side chains of Asn-64, Asn-154 and Asn-190. At the same time the positively charged amino group is hydrogen bonded with side chains of Asn-64 and Glu-208 and with the carbonyl oxygen of Arg-209. When the L,L-DAP

and D,L-DAP are bound to the enzyme the α -carboxyl group is bonded with the amidic hydrogen of Gly-74, Asn-75, Gly-218 and Ser-219 and with the side chain of the same Ser-219. The charged amino group forms instead hydrogen bond with the side chains of residues Asn-11, Gln-44 and Glu-208.

6.4 The Model1 model system

The construction of the model system is a crucial point in QM/MM simulation in order to obtain the best cost/efficiency compromise from the adopted hybrid potential. The choice of atoms to be assigned to the various layers has thus a fundamental importance. In the present case we reduced the **H** (*high*) layer,¹⁰⁴ treated at QM level (DFT/B3LYP/DZVP)^{75,223} to an as-smallest-as-possible region (see Figure 6.2-A). By knowing the enzyme plays the stereo-inversion by means of two residues of Cysteine without co-factor help, we introduced in the **H** layer the side-chains of Cys-73 and Cys-217, saturating the dangling bonds with hydrogen atoms (atom link approach).¹⁰⁴ The entire substrate Diaminopimelate (Zdp-275) is included in **H** and then treated at QM level. Some residues (Table 6.1 and Figure 6.2-B,C) surround the reacting core and are hypothesized to play a direct influence on the catalytic process. To improve the description of the system without increasing the simulation cost,¹⁰⁴ these residues have been included in the **M** (*medium*) layer. All the other residues have been included in the **L** layer. In the current simulation we used for the MM calculations the Amber(*ff99*)⁴⁵ force field for the proteic region and the Generalized Amber Force Field (GAFF)⁴⁹ for the DAP. In our simulation the solvent is taken into account neither explicitly nor implicitly; to avoid unrealistic deformations of the structure only a few residue of the **L** (*low*) layer around the **M** layer are free to move during the “microiterations”¹¹⁶ steps¹⁰⁴ of the optimization procedure, while all the other residues of **L** are kept frozen to their initial position. The free residues have been chosen among the ones not (or minimally) exposed to the surface of the enzyme (see Figure 6.2-D). In the following discussions we will refer to this model system as **Model1**.

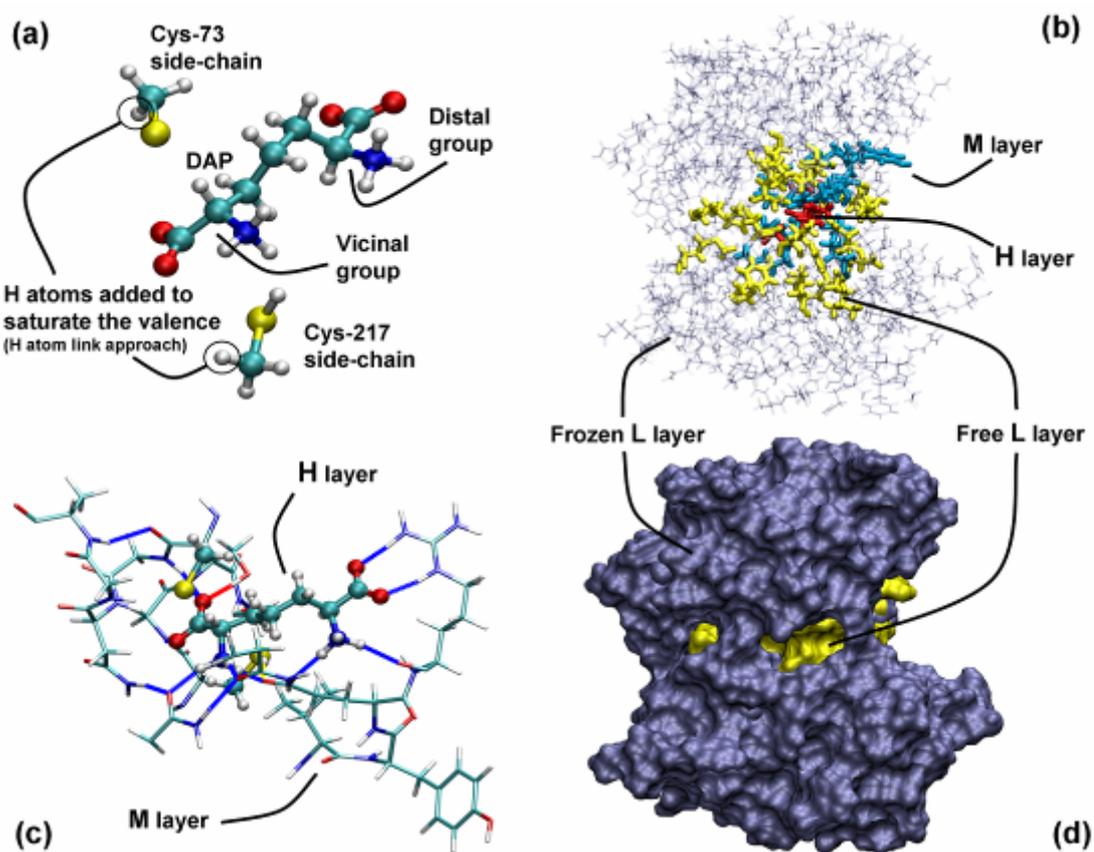


Figure 6.2 Partitioning scheme adopted to build the Model1 model system.

Table 6.1 Partitioning of the whole system into layers.

High layer (H)	Medium layer (M)	Free Low layer (H)	
Cys-73 ^a	Asn-11 ^a	Asn-11 ^b	His-159
Cys-217 ^a	Gln-44 ^a	Asp-12	Asn-190
Zdp-275 ^c	Cys-73 ^b	Phe-13	Gly-210
	Gly-74	Gln-44 ^b	Ala-211
	Asn-75	Leu-45	Thr-214
	Gly-76	Leu-46	Ala-216
	Ala-77	Tyr-60	Ala-221
	Val-206	Ile-62	Leu-242
	Phe-207	Asn-64	
	Glu-208	Val-70	
	Arg-209	Gln-72	
	Cys-217 ^b	Met-104	
	Gly-218	Met-118	
	Ser-219	Asn-157	
	Gly-220	Pro-158	

- a) Only the side-chain is included.
 b) Only the backbone is included.
 c) The substrate.

We performed on **Model1** geometry optimizations to characterize the Potential Energy Surface (PES) associated to the stereo-inversion of the outer stereocenter of the substrate. We

located two critical points on the surface M1 and M2 (Figure 6.3), corresponding to the enzyme bound respectively to the reactant (*L,L*-DAP) and to the product (*D,L-meso*-DAP), with an opposite protonation state of the catalytic Cys couple; in M1 the Cys-217 is protonated, while Cys-73 is un-protonated and optimally oriented to grab a proton from the substrate. In the hypothesized reaction mechanism the side chain of the formally negatively charged Cys-73 would take the proton from the substrate while the Cys-217 thiolic proton moves to protonate the same carbon atom on the opposite face; it is not known from experimental evidences if the stereo-inversion of the carbonic center follows a concerted path or a step-wise one and, consequently, if the PES under examination is characterized by a unique Transition State or more, with intermediates. Any attempt to locate the intermediate of the hypothesized step-wise reaction path failed, obtaining the reactant or the product starting from some guess geometries characterized by both Cys-73 and Cys-217 protonated and the planar deprotonated substrate. By searching for a Transition State we found a suitable geometry (TSC) where the two Cys residues are almost completely protonated (Figure 6.3), being the S γ -H distance of 1.43 Å and 1.36 Å respectively for Cys-73 and Cys-217, accounting thus for a concerted but highly asynchronous process. In TSC the substrate is deprotonated and planar in order to delocalize the partial negative charge in the extended π -orbital system. The charge delocalization can be quantified by means of the comparison (Table 6.2) of some relevant atomic distances and atomic point charges values (CHELPG¹²⁷ charges are reported). In TSC the negative charge, left on the substrate by its deprotonation, accumulates on the two oxygen atoms of the carboxyl group; in passing from the minima to the TS the carboxylic and the $-\text{NH}_3^+$ groups become, respectively, more negative and less positive, while the deprotonated C α atom does not undergo a variation in charge. The delocalization of the charge is witnessed by a decrease in the bond order of both C-O bond of carboxyl and by an increase of the bond order of the C α -C bond. The correct simulation of the reaction process is the first step to plan the construction of an efficient inhibitor of the enzyme under examination. The structure based drug design can take enormous advantage from the accurate knowledge of the TS geometry; also the charge distribution observed in the TS can be used to improve the drug development procedure. Many modern software do use geometries and atomic point charges to derive accurate pharmacophore definition in order to save time and money in the screening of candidate drugs.

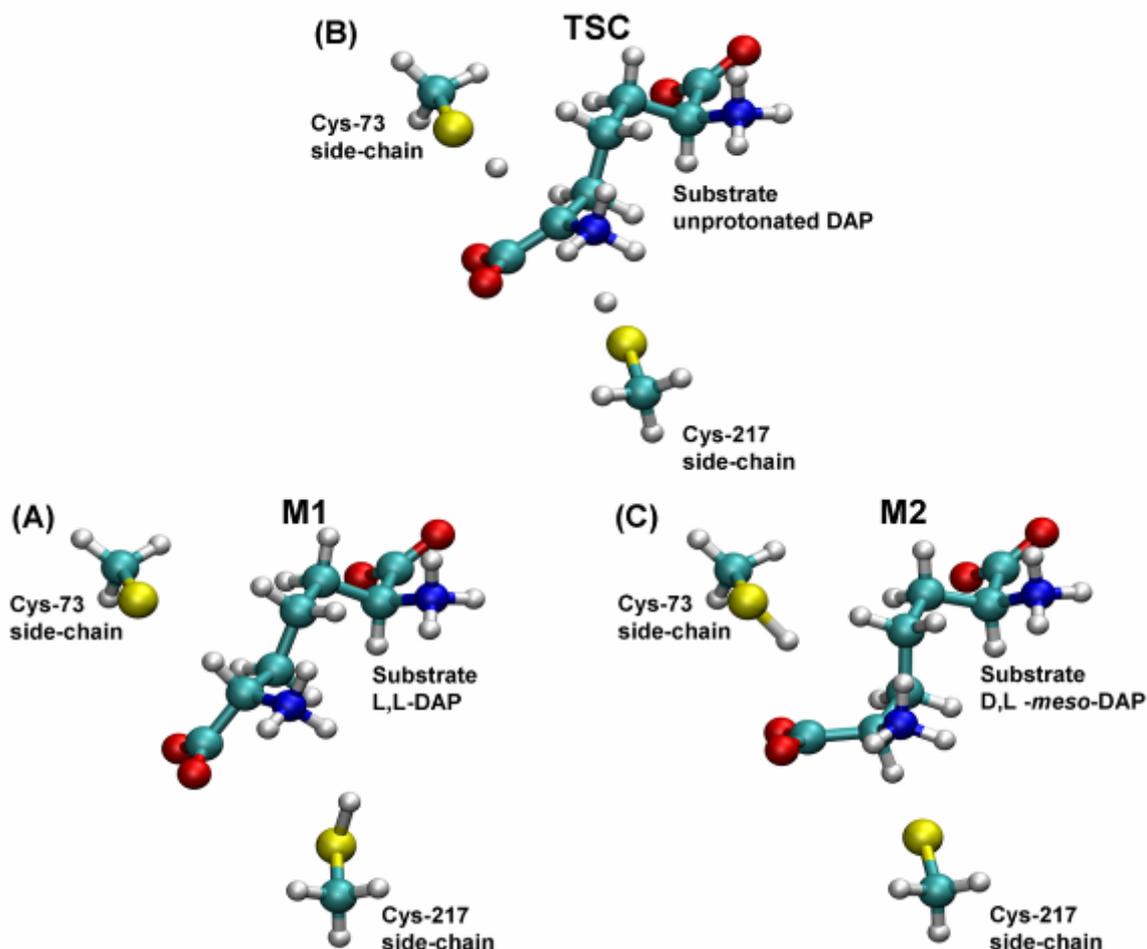


Figure 6.3 Representation of the model-H sub-system for the critical points found on the Model1 PES.

Table 6.2 Some relevant distances and atomic point charges^a for the three critical points found in the PES of Model1.

	M1	TSC	M2
Cys-73 $S\gamma$ -H	2.29 Å	1.43 Å	1.34 Å
Cys-217 $S\gamma$ -H	1.36 Å	1.36 Å	2.27 Å
Substrate $C\alpha$ -C	1.53 Å	1.42 Å	1.52 Å
Substrate $C\alpha$ -N	1.50 Å	1.49 Å	1.51 Å
Substrate $C\alpha$ - $C\beta$	1.53 Å	1.51 Å	1.54 Å
Substrate C-O1	1.26 Å	1.29 Å	1.27 Å
Substrate C-O2	1.28 Å	1.31 Å	1.28 Å
Charge on Cys-73 $S\gamma$ -	-1.008	-0.570	-0.425
Charge on Cys-217 $S\gamma$ -	-0.506	-0.449	-1.021
Charge on substrate $C\alpha$	0.353	0.204	0.188
Total charge on substrate $-NH_3$ group	0.611	0.529	0.599
Charge on substrate C	0.639	0.670	0.655
Charge on substrate O1	-0.769	-0.918	-0.818
Charge on substrate O2	-0.986	-1.122	-0.972
Total charge on substrate $-CO_1O_2$ group	-1.116	-1.370	-1.135

a) The reported charge values are computed with the CHELPG method.

The reaction profile we found for the epimerization of DAP is represented in Figure 6.4 and the relative data are reported in Table 6.3 (taking the energy value M1 as reference). We report also the energy values in terms of Gibbs Free Energy. The thermal corrections to energy (E) to get the Gibbs energy (G) are obtained by means of numerical frequencies calculations on the **HM** (*high-medium*) layer in the presence of the fixed **L** layer; results from both approximated (**freq**) and complete (**fullfreq**) frequencies calculations (see Section 3.8) are reported.

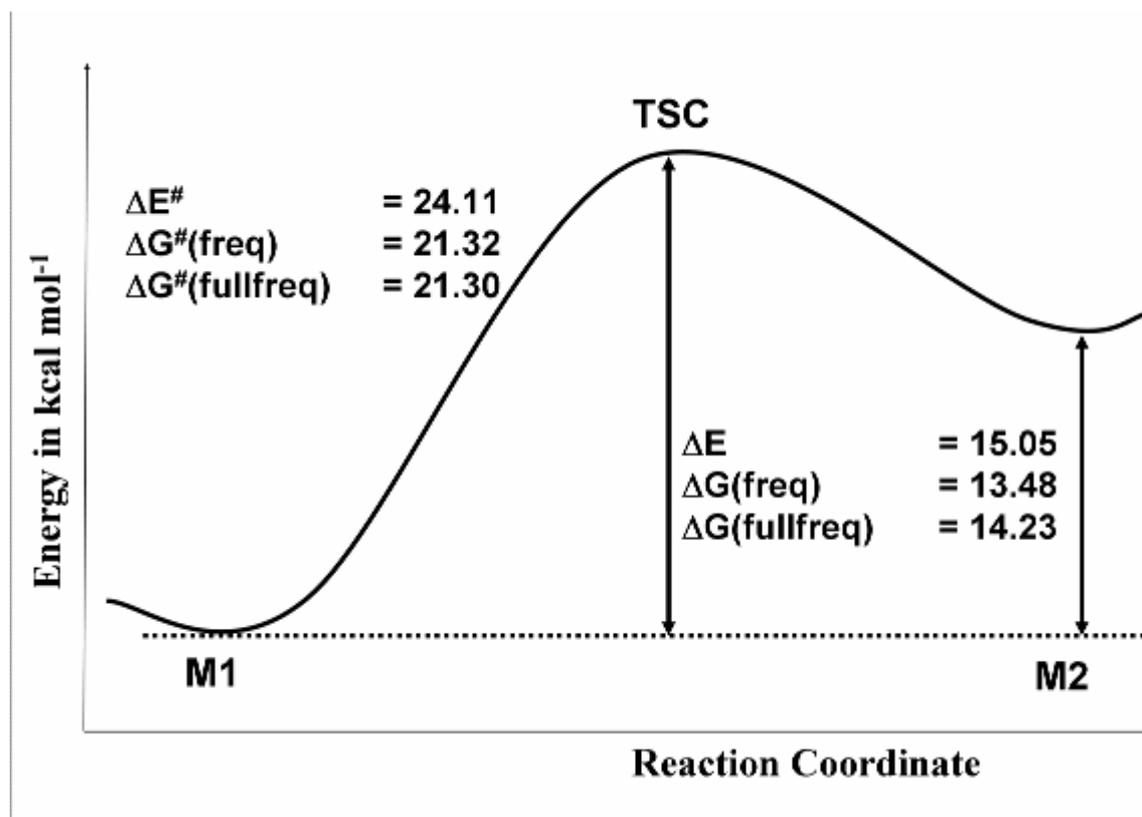


Figure 6.4 Reaction profile obtained for Model1.

Table 6.3 Energy difference between minima and reaction barrier for Model1.

ΔE	15.05 kcal mol ⁻¹
ΔE^\ddagger	24.11 kcal mol ⁻¹
ΔG (freq corrections)	13.48 kcal mol ⁻¹
ΔG^\ddagger (freq corrections)	21.32 kcal mol ⁻¹
ΔG (fullfreq corrections)	14.23 kcal mol ⁻¹
ΔG^\ddagger (fullfreq corrections)	21.30 kcal mol ⁻¹

Using the model system called **Model1** we obtained a reaction profile composed by three critical points (Figure 6.3 and Table 6.3); the product (Enz—D,L-*meso*-DAP) is less stable than the reactant (Enz—L,L-DAP) by 15.05 kcal mol⁻¹ (corrected to 13.48 and 14.23 kcal mol⁻¹ respectively by means of **freq** and **fullfreq** calculations). The computed barrier for the stereo-inversion is about 24.11 kcal mol⁻¹ (corrected to 21.32 and 21.30 kcal mol⁻¹

respectively by means of **freq** and **fullfreq** calculations). In TSC the normal mode associate to the negative frequency is very similar for both **freq** and **fullfreq** and accounts for the protonation/deprotonation of the planar substrate. The obtained results show that the computation of numerical frequencies by mean of the approximated procedure (**freq**) gives results which are comparable to the ones obtained from the **fullfreq** with a substantial saving of computation time. In the present case the **H** and **M** layer are composed respectively by 36 and 170 atoms, so the **fullfreq** calculation requires $1+(36+170)*6=1+(206)*6=1237$ wavefunction evaluation; on the opposite the **freq** calculation requires only $1+(36)*6=217$ saving about the 80% of computational time. It is worthwhile to note that in our simulation the use of frequency calculation was mainly intended to obtain a qualitative picture of the principal normal modes of the found Transition States structures, in order to recognize them as belonging to the MEP between the two stable states, or to discard them. The approximations of numerical frequency calculations (harmonic approximation, gas phase thermodynamic treatment, etc.) are too severe to obtain very accurate *G* values; some different methods have been developed to get more reliable Gibbs Free Energy values, but they were not applied in this work. Nevertheless their usefulness requires them to be implemented in the future versions of the **COBRAMM**¹⁰⁴ package, in order to improve the description of the system's energetic.

We performed the fingerprint analysis (see Section 3.9) over the found critical points in order to ascertain the reliability of the adopted system model and to get a complete pattern of the influences arising from the residues surrounding the reacting core. In the Table 6.4 we report the values of the M1/M2 and M1/TSC energy differences as obtained from QM calculations (single point) on the *model-H* system (from QM/MM optimized geometries) *in vacuo* and in the presence of the atomic point charges of the whole enzyme.

The two minima are almost isoenergetic for the *in vacuo* system ($\Delta E_0(\mathbf{M1},\mathbf{M2})$), while in the presence of the atomic point charges the energy difference ($\Delta E_{qm}(\mathbf{M1},\mathbf{M2})$) rise up to 16.50 kcal mol⁻¹ which is closer to the QM/MM value of 15.05 kcal mol⁻¹ confirming for a strong electrostatic effect of the surrounding enzyme on the M1/M2 equilibrium. The same effect, but lesser in magnitude, is still present for the M1/TSC couple; the computed barrier changes from 20.21 ($\Delta E_0(\mathbf{M1},\mathbf{TSC})$) to 26.42 kcal mol⁻¹ ($\Delta E_{qm}(\mathbf{M1},\mathbf{TSC})$) when atomic point charges are added to the naked QM core.

To individuate the residues which mainly influence the equilibrium between the critical point we performed both **DFP** and **RFP** (Section 3.9.) on M1, TSC and M2. In this section we

report the most important stabilization (S)/destabilization (D) values of **DFP/RFP** analysis for the couples M1/M2 (Table 6.5 and Figure 6.5)M1/TSC (Table 6.6 and Figure 6.7).

Table 6.4 Electrostatic interaction energy data for Model1.

$\Delta E_0(\text{M1}, \text{M2})^a$	-1.22 kcal mol ⁻¹
$\Delta E_0(\text{M1}, \text{TSC})^a$	20.21 kcal mol ⁻¹
$\Delta E_{qm}(\text{M1}, \text{M2})^b$	16.50 kcal mol ⁻¹
$\Delta E_{qm}(\text{M1}, \text{TSC})^b$	26.42 kcal mol ⁻¹

a) See Equation 3.7.

b) See Equation 3.8.

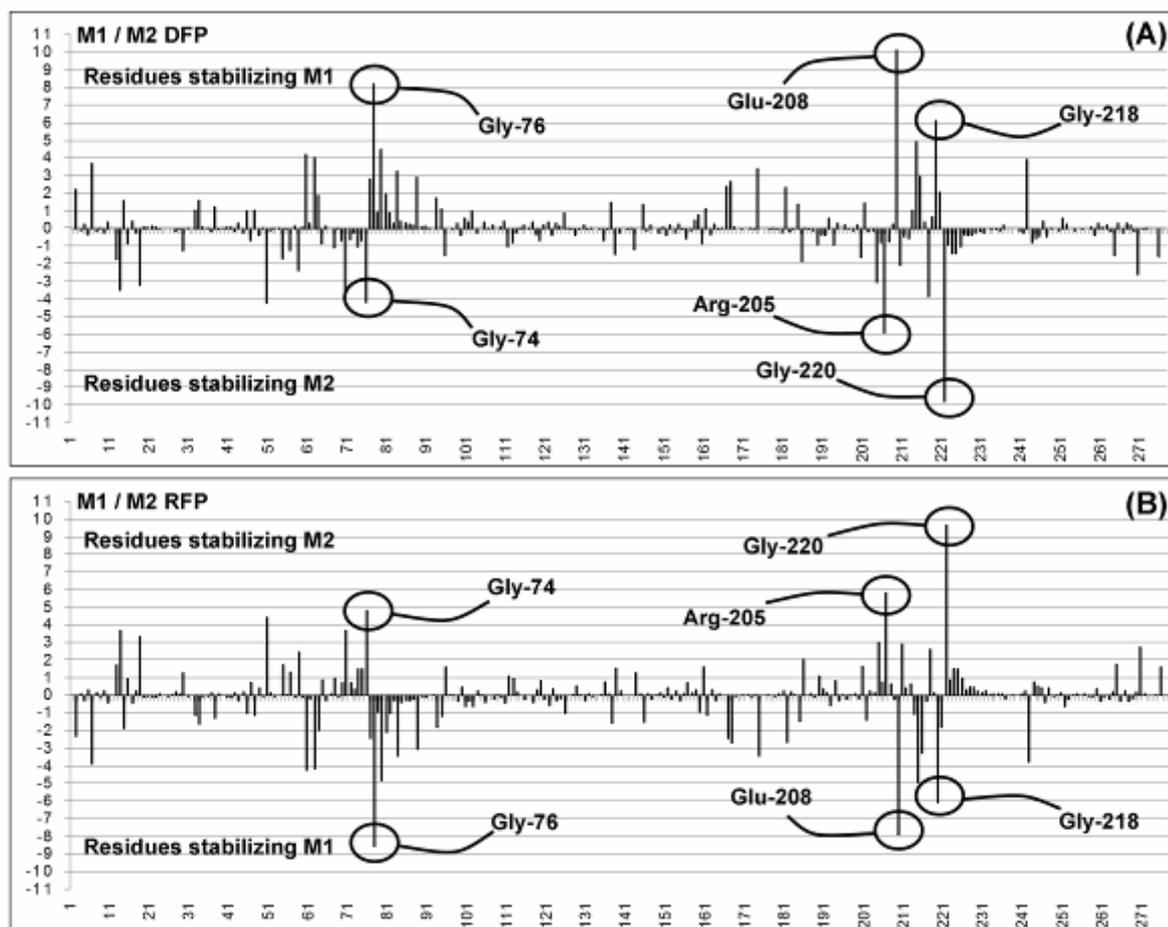


Figure 6.5 DFP and RFP diagrams for the M1/M2 transition.

Table 6.5 Stabilizing/destabilizing factor for the couple M1/M2 in Model1; values from selected residues are reported.

	Residue	S from DFP	D from RFP
Residues stabilizing M1	Lys-5	3.68	-3.92
	Hip-59	4.19	-4.28
	Arg-61	4.04	-4.17
	Gly-76	8.19	-8.58
	Arg-78	4.49	-4.87
	Arg-82	3.23	-3.48
	Glu-173	3.39	-3.44
	Glu-208	10.12	-7.89
	Glu-213	4.92	-4.97
	Gly-218	6.13	-6.12
	Asp-241	3.93	-3.81
Residues stabilizing M2	Asn-11	-3.53	3.68
	Asp-17	-3.23	3.35
	Glu-49	-4.24	4.43
	Glu-69	-3.69	3.69
	Gly-74	-4.17	4.79
	Lys-203	-3.09	2.98
	Arg-205	-5.96	5.82
	Ala-216	-3.84	2.62
	Gly-220	-9.80	9.66

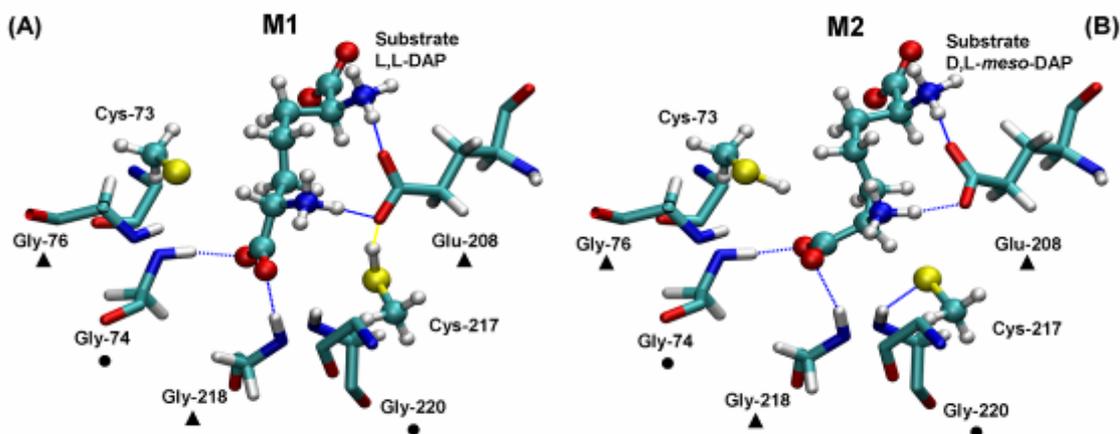


Figure 6.6 Schematic representation of the most important residues in the stabilisation/destabilisation of the M1/M2 couple.

The M1 is stabilized in respect of M2 (Figure 6.6) by Gly-76 which interacts as an H-bond donor with the S_{γ} atom of Cys-73; this interaction is stronger in M1 than in M2, because of the protonation of the S which grabs the proton from the substrate and loses part of its negative charge (see Table 6.2). The Gly-76-NH--- S_{γ} -Cys-73 distance increases from 2.34 Å to 2.29 Å as passing from M1 to M2. The residue Glu-208 plays a major role in destabilizing M2 with respect M1, because in M1 it interacts as a triple H-bond acceptor, by means of the

unprotonated carboxylic side chain, with both the -NH_3 groups of the substrate and with the thiolic H atom of the Cys-217; in M2 this atom is transferred to the substrate and so only two H-bond interactions are available for Glu208 to interact. The loss of an H-bond interaction and the negative charge on Cys-217-S γ account for the high destabilization effect of Glu-208 on M2. The Gly-218 is an H-bond donor to the carboxylic group of the substrate bonded to the C α which undergoes the stereo-inversion; the best interaction geometry is found in M1, where the Gly-218-NH--- ^^{\ominus} OCO-DAP distance is 1.81 Å; this value increases to 2.09 Å in M2 because of the inversion of the stereocenter, causing the weakening of the H-bond with a destabilization effect of M2 in respect to M1. Some residues do play an interesting role in stabilizing M2 in respect to M1 and thus they partially equilibrate the ratio between M1 and M2. As seen for Gly-218, Gly-74 interacts as a H-bond donor with the same carboxylic group of the substrate but, oppositely to it, interacts better with the D,L-*meso*-DAP (M2) than with the L,L-DAP (M1), as confirmed by the Gly-74-NH--- ^^{\ominus} OCO-DAP distance which shortens from 1.77 Å to 1.70 Å in passing from M1 to M2. For a similar reason Ala-216 stabilizes M2 because of an improved interaction geometry with the -NH_3^+ group of D,L-*meso*-DAP. The main effect is due to Gly-220 which is an H-bond donor to the S γ of Cys-217: when this atom is unprotonated (M2) the interaction is maximized and thus the structure stabilized.

In Table 6.6 we reported the main residues which, according to DFP and RFP results (Figure 6.7), play a positive or a negative role in the electrostatic catalysis, by stabilizing or destabilizing TSC in respect to M1 (Figure 6.8). As seen for the M1/M2 equilibrium, and for the same reasons, the residues Gly-76 and Gly-74 respectively stabilize and destabilize M1 in respect to TSC, but the magnitude of effects are smaller. The Glu-208 even in this case plays an important role in the stabilization of M1, because in the TSC structure one of the three H-bond (with the -NH_3 group of the substrate) is almost completely loosen. The carboxylic group of substrate interacts as a double H-bond acceptor with Ser-219, the better geometry interaction being found in TSC; as a matter of fact the H-bond between the Ser-219-OH and the carboxylic group passes from 1.60 Å (M1) to 1.53 Å (TSC) and 1.61 Å (M2). The H-bond between the Ser-219-OH and the carboxylic group does not play a very important role in stabilizing the TSC, passing from 1.76 Å (M1) to 1.77 Å (TSC) and 1.87 Å (M2).

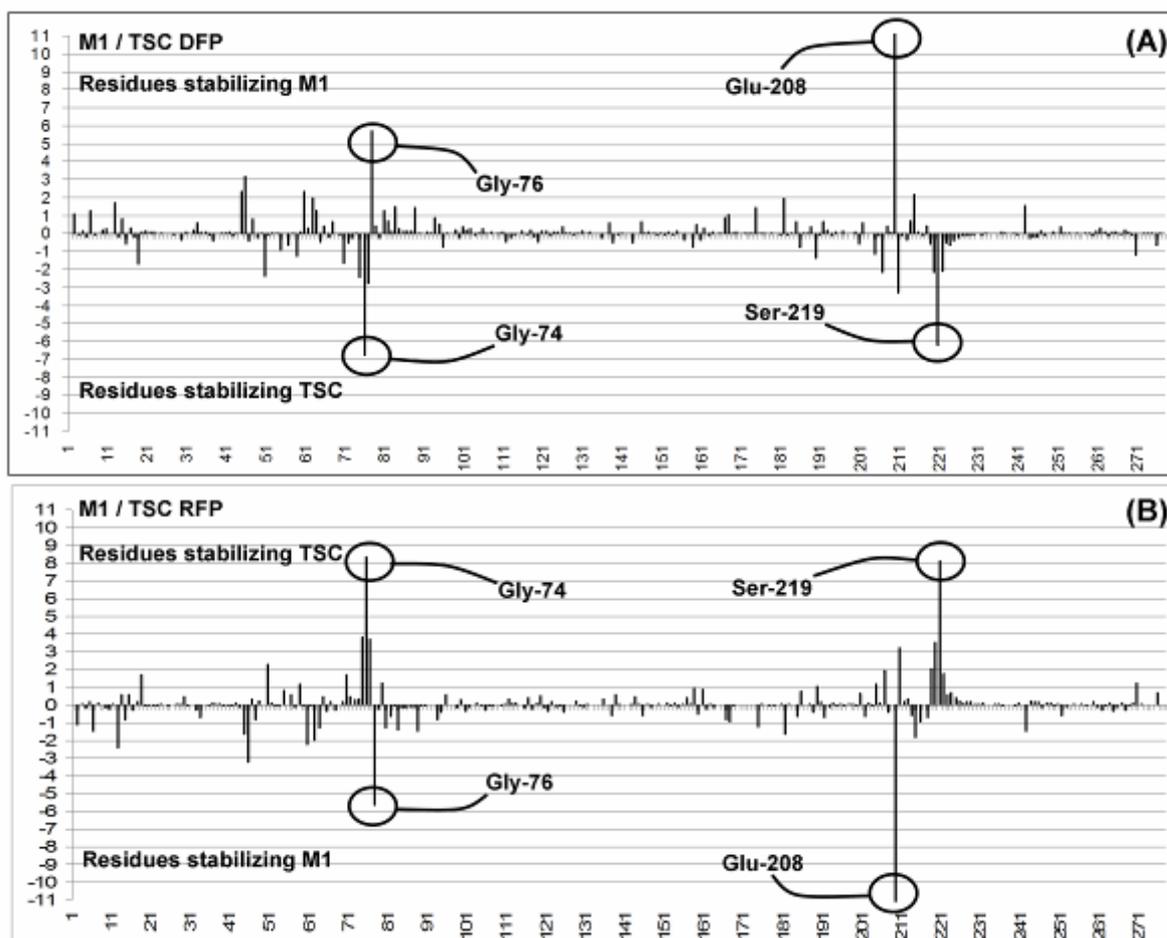


Figure 6.7 DFP and RFP diagrams for the M1/TSC transition.

Table 6.6 Stabilizing/destabilizing factor for the couple M1/TSC in Model1; values from selected residues are reported.

	Residue	S from DFP	D from RFP
Residues stabilizing M1	Gln-44	3.19	-3.21
	Gly-76	5.41	-5.65
	Glu-208	19.14	-15.6
Residues stabilizing TSC	Glu-49	-2.24	2.32
	Gly-74	-6.78	8.36
	Asn-75	-2.81	3.71
	Arg-209	-3.27	3.21
	Gly-218	-2.17	3.51
	Ser-219	-6.23	8.14

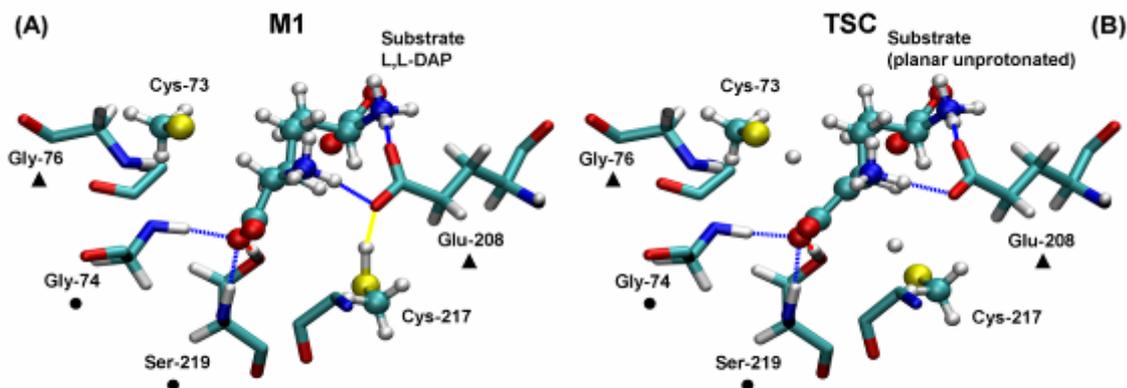


Figure 6.8 Schematic representation of the most important residues in the stabilisation/destabilisation of the M1/TSC couple.

From an analysis of the single residue contributions it is possible to observe the general convergence of both **DFP** and **RFP** in indicating the residues which play a role in the equilibrium between minima and/or in the stabilization of the TS.

Table 6.7 Stabilizing/destabilizing van der Waals factor for the couple M1/M2 in Model1; values from selected residues are reported.

	Residue	W from vdWFP
Residues stabilizing M1	Gly-74	1.11
	Asn-75	0.39
	Glu-208	0.47
	Gly-220	1.29
Residues stabilizing M2	Ans-11	-0.53
	Gly-76	-0.61
	Gly-218	-1.74
	Ser-219	-1.29

Table 6.8 Stabilizing/destabilizing van der Waals factor for the couple M1/TSC in Model1; values from selected residues are reported.

	Residue	W from vdWFP
Residues stabilizing M1	Gly-74	1.19
	Ans-75	0.64
	Ser-219	1.17
Residues stabilizing TSC	Asn-11	-0.48
	Gly-76	-0.95
	Glu-208	-1.54

In the Table 6.7 and Table 6.8 are reported some van der Waals single residue contributions to the equilibrium between the two minima and to the stabilization of TS. The magnitude of these contributions are smaller in respect to the electrostatic contributions estimated by means

of DFP and RFP. As a matter of fact we can consider, for the current problem, only the electrostatic contribution as worth of discussion.

6.5 The Model2 model system

On the basis of the analysis performed on the **Model1** model system we can derive some conclusion about the role of the single residue in determining the shape of the computed PES. As discussed above both **DFP** and **RFP** are able to automatically individuate the residues acting the strongest influence. Besides our analysis can be used to improve the model system in order to obtain better results. In this particular case we observed the residue Glu-208 acting the strongest effect in stabilizing M1 with respect to both TSC and M2. As described above Glu-208 is fully described by an MM potential, while its interaction on the QM core is accounted for by means of electrostatic and van der Waals QM-MM cross terms. We focus here on the electrostatic effects, being the van der Waals terms smaller. To check the correctness of the model and to ascertain if the interactions between the residue Glu-208 and the substrate are correctly accounted for it is of mandatory importance do construct a different model system; we will refer to this new model as **Model2**. This new model system differs from the previous one in the treatment of the Glu-208, whose side-chain is now included in the **H** layer, and so treated at QM (DFT/B3LYP/DZVP)^{75,223} level as well as the side-chains of Cys-73 and Cys-217 and the substrate (Figure 6.9). This allows a full QM treatment of the negatively charged Glu-208 side-chain and for its interactions with both substrate and Cys-217. We performed on **Model2** geometry optimizations to find the three critical points M1, M2 and TSC starting from the geometries obtained with **Model1**. Then on the found critical points we performed both **freq** and **fullfreq** computations, as well as **DFP** and **RFP** analysis. Some relevant parameters are reported in Table 6.9 and graphically represented in Figure 6.10, in order to drive a first comparison between the three critical points.

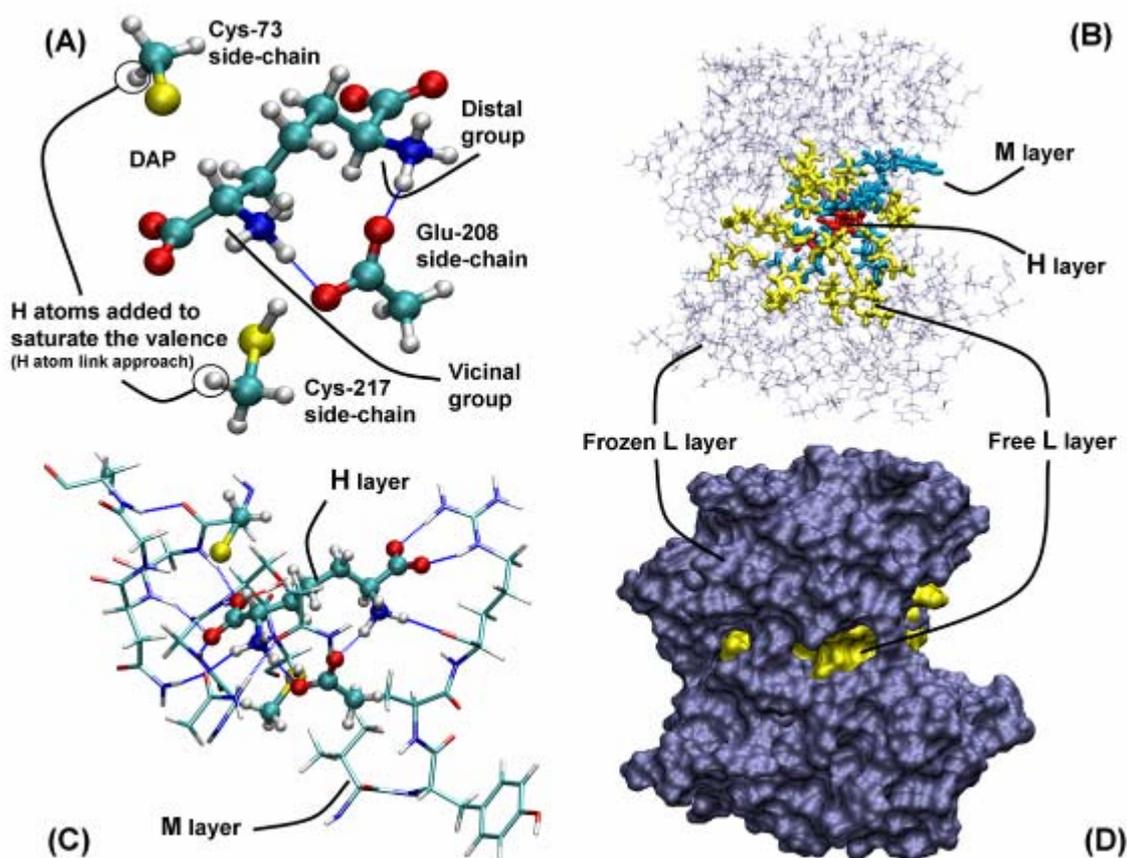


Figure 6.9 Partitioning scheme adopted to build the Model2 model system.

In the new reaction profile (Figure 6.11) the product (Enz—D,L-*meso*-DAP) is less stable than the reactant (Enz—L,L-DAP) by 11.74 kcal mol⁻¹ (corrected to 11.56 and 10.80 kcal mol⁻¹ respectively by means of **freq** and **fullfreq** calculations). The computed barrier for the stereo-inversion is about 25.31 kcal mol⁻¹ (corrected to 23.00 and 22.48 kcal mol⁻¹ respectively by means of **freq** and **fullfreq** calculations). In TSC the normal mode associated to the negative frequency is very similar for both **freq** and **fullfreq** and accounts for the protonation/deprotonation of the planar substrate.

We can easily compare the results obtained with the two model systems (see Table 6.3 and Table 6.10). The barrier associated to the stereo-inversion is unchanged passing from **Model1** (24.11 kcal mol⁻¹) to **Model2** (25.31 kcal mol⁻¹), meaning that the Glu-208 effect in destabilizing the TSC in respect to M1 is purely electrostatic and that it is correctly accounted for, even if the residue is treated at MM level, as in **Model1**. The inclusion of Glu-208 in the QM layer has a strong influence on the equilibrium between the two minima: the M2 is still higher in energy with respect to M1, but the difference decreased to 11.74 kcal mol⁻¹, indicating the insufficient description provided by **Model1**. From the data reported in Table

6.11, in comparison with Table 6.4, we can realize more about the effect played by Glu-208. In **Model1** the inclusion of the enzyme atomic point charges (Table 6.4) causes the energy difference between M1 and M2 to increase from -1.22 to 16.21 kcal mol⁻¹, while in **Model2** the increase is smaller (Table 6.11), from 7.17 to 10.20 kcal mol⁻¹. The magnitude of this effect in **Model1** is mainly due to the presence of the charges of Glu-208 (treated at MM level) which strongly stabilize M1. In **Model2**, where Glu-208 side-chain is a part of the QM region, the effect of charges on the equilibrium between minima is strongly reduced. This description is better than the former one, because the strong ionic interactions between the carboxylic group of Glu-208 and the two protonated aminic groups of the substrate are now treated at full QM level, as well as the H-bond interaction between the same carboxylic group and the side-chain of Cys-217. From **DFP** and **RFP** we obtained results perfectly comparable with the ones obtained from **Model1**, being the ranking order almost identical, except for Glu-208 which is now included in the QM region. In Table 6.12, Figure 6.12 and Figure 6.13 the comparison between M1 and M2 are reported, while in Table 6.13, Figure 6.14 and Figure 6.15 are reported the data relative to the M1/TSC couple.

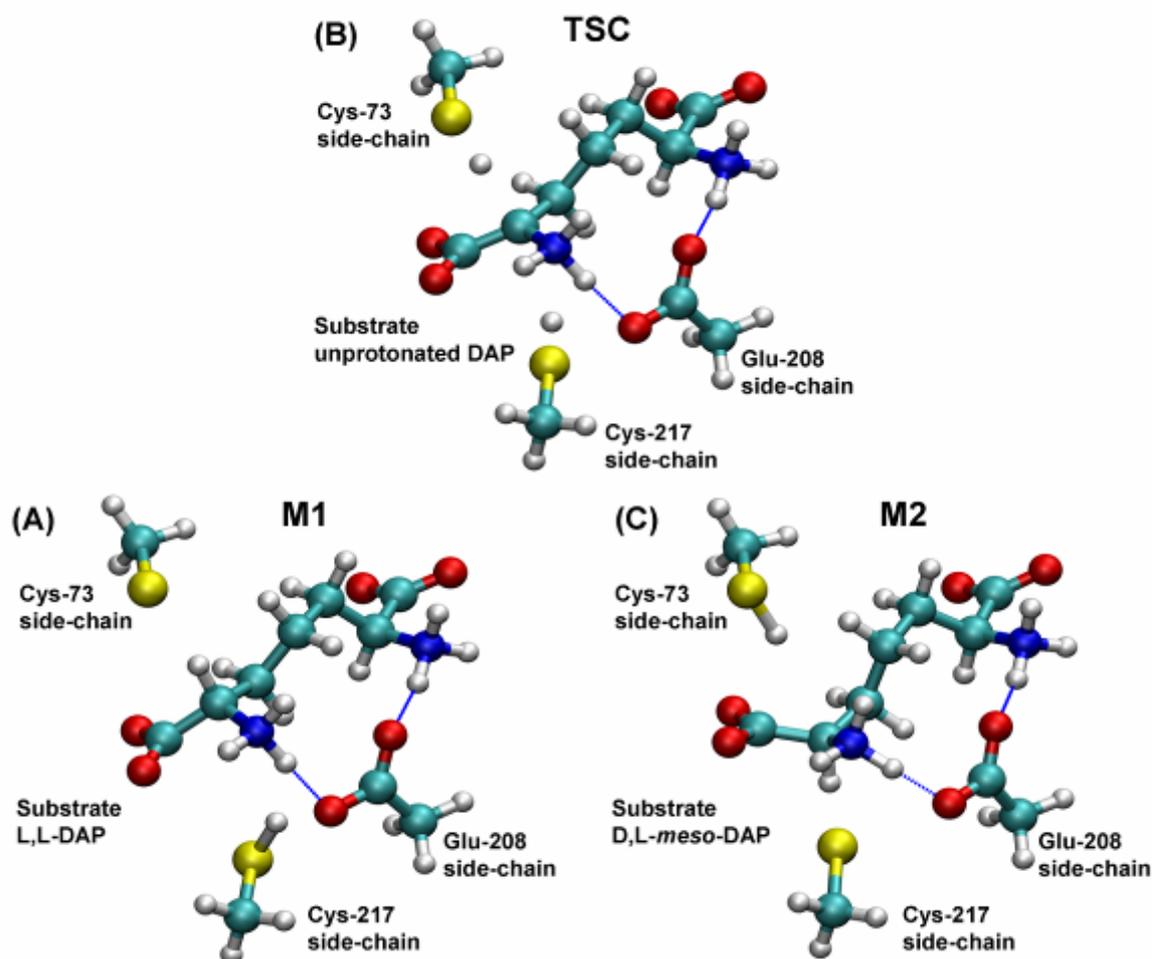


Figure 6.10 Representation of the model-H sub-system for the critical points found on the Model2 PES.

Table 6.9 Some relevant distances and atomic point charges^a for the three critical points found in the PES of Model2.

	M1	TSC	M2
Cys-73 S γ -H	2.30 Å	1.46 Å	1.34 Å
Cys-217 S γ -H	1.35 Å	1.35 Å	2.39 Å
Substrate C α -C	1.53 Å	1.43 Å	1.53 Å
Substrate C α -N	1.50 Å	1.48 Å	1.50 Å
Substrate C α -C β	1.53 Å	1.51 Å	1.54 Å
Substrate C-O1	1.26 Å	1.29 Å	1.27 Å
Substrate C-O2	1.28 Å	1.30 Å	1.28 Å
Charge on Cys-73 S γ -	-1.017	-0.640	-0.427
Charge on Cys-217 S γ -	-0.493	-0.451	-1.032
Charge on substrate C α	0.335	0.036	0.155
Total charge on substrate -NH ₃ group	0.535	0.557	0.567
Charge on substrate C	0.614	0.703	0.687
Charge on substrate O1	-0.761	-0.904	-0.842
Charge on substrate O2	-0.988	-1.120	-0.969
Total charge on substrate -CO1O2 group	-1.135	-1.321	-1.124

a) The reported charge values are computed with the CHELPG¹²⁷ method.

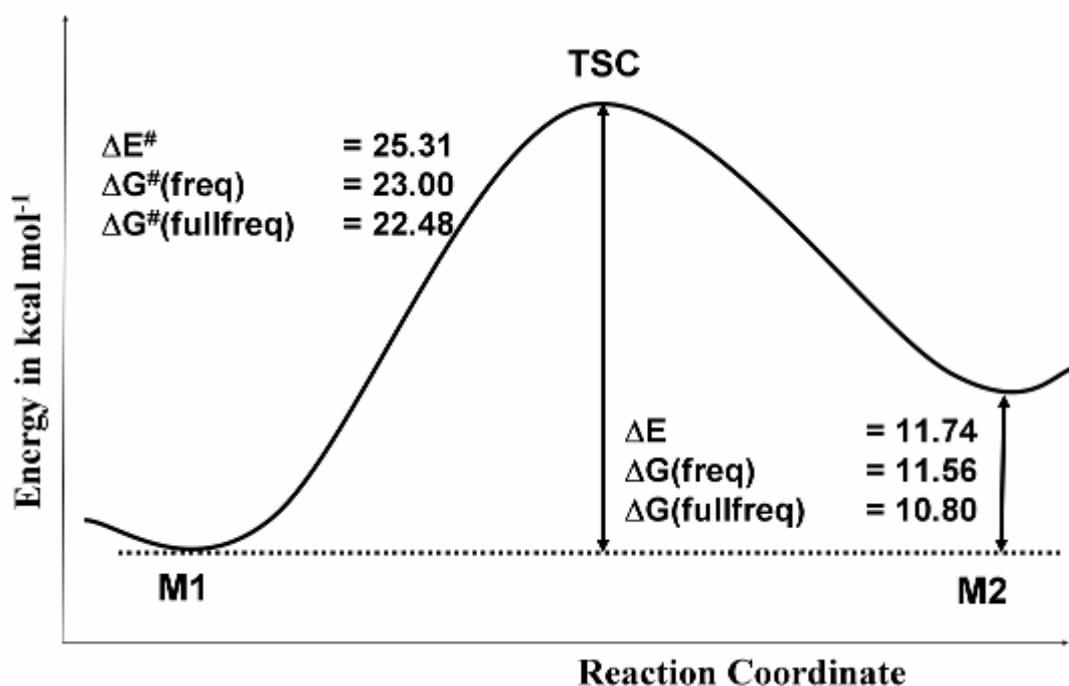


Figure 6.11 Reaction profile obtained for Model2.

Table 6.10 Energy difference between minima and reaction barrier for Model2.

ΔE	11.74 kcal mol ⁻¹
ΔE^\ddagger	25.31 kcal mol ⁻¹
ΔG (freq corrections)	11.56 kcal mol ⁻¹
ΔG^\ddagger (freq corrections)	23.00 kcal mol ⁻¹
ΔG (fullfreq corrections)	10.80 kcal mol ⁻¹
ΔG^\ddagger (fullfreq corrections)	22.48 kcal mol ⁻¹

Table 6.11 Electrostatic interaction energy data for Model2

$\Delta E_0(\text{M1}, \text{M2})^a$	7.17 kcal mol ⁻¹
$\Delta E_0(\text{M1}, \text{TSC})^a$	37.81 kcal mol ⁻¹
$\Delta E_{\text{qm}}(\text{M1}, \text{M2})^b$	10.20 kcal mol ⁻¹
$\Delta E_{\text{qm}}(\text{M1}, \text{TSC})^b$	24.45 kcal mol ⁻¹

a) See Equation 3.7

b) See Equation 3.8

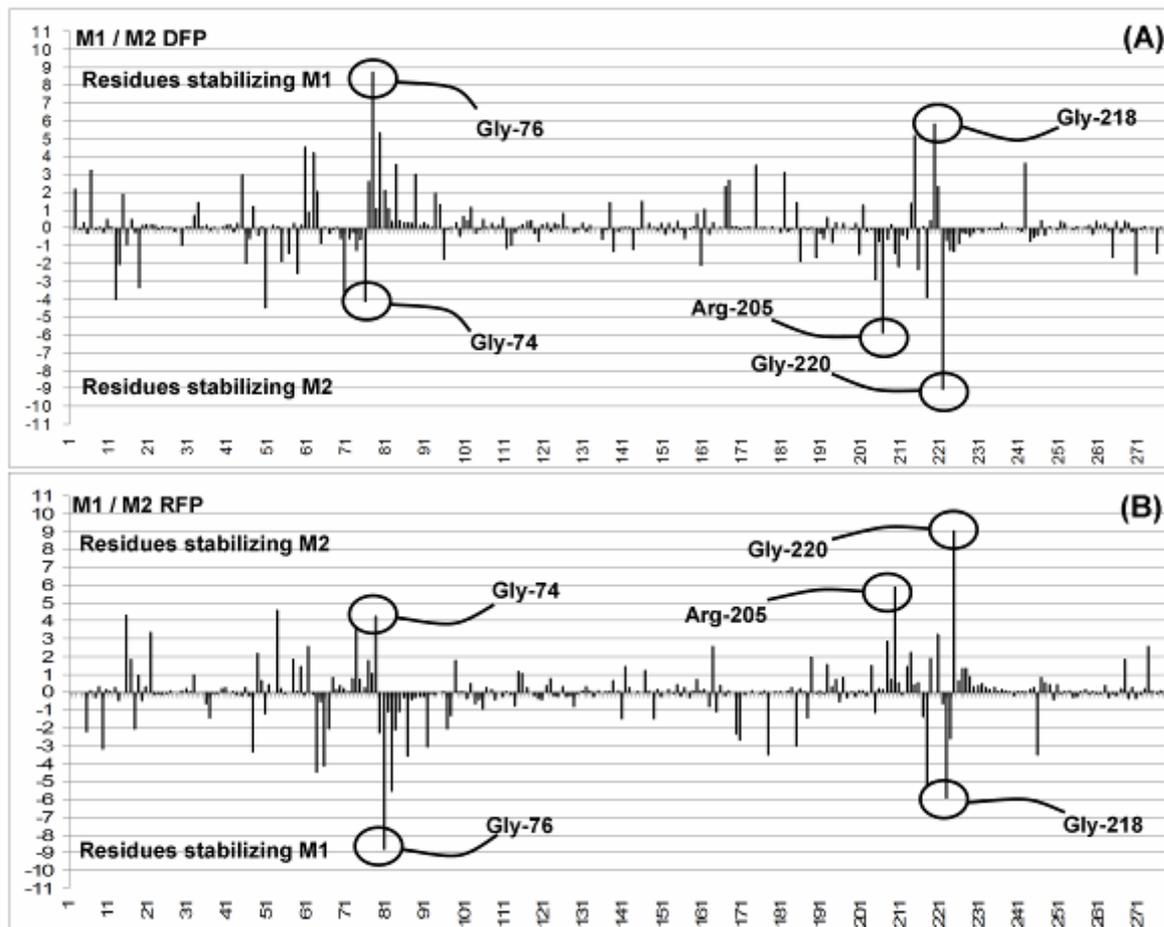


Figure 6.12 DFP and RFP diagrams for the M1/M2 transition.

Table 6.12 Stabilizing/destabilizing factor for the couple M1/M2 in Model2; values from selected residues are reported.

	Residue	S from DFP	D from RFP
Residues stabilizing M1	Lys-5	3.23	-2.22
	Hip-59	4.51	-4.46
	Arg-61	4.17	-4.16
	Gly-76	8.77	-8.81
	Arg-78	5.33	-5.57
	Arg-82	3.55	-3.56
	Glu-173	3.50	-3.54
	Glu-213	5.22	-5.28
	Gly-218	5.80	-5.95
	Asp-241	3.62	-3.53
Residues stabilizing M2	Asn-11	-4.00	4.28
	Asp-17	-3.32	3.36
	Glu-49	-4.45	4.56
	Glu-69	-3.63	3.69
	Gly-74	-4.13	4.24
	Lvs-203	-2.91	2.84
	Arg-205	-5.89	5.87
	Ala-216	-3.89	3.25
	Gly-220	-9.06	9.05

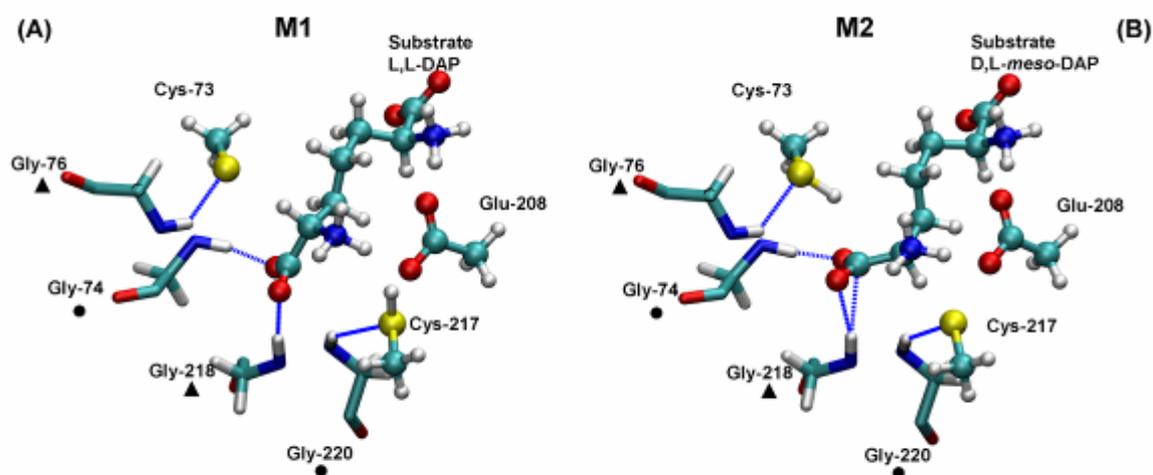


Figure 6.13 Schematic representation of the most important residues in the stabilisation/destabilisation of the M1/M2 couple

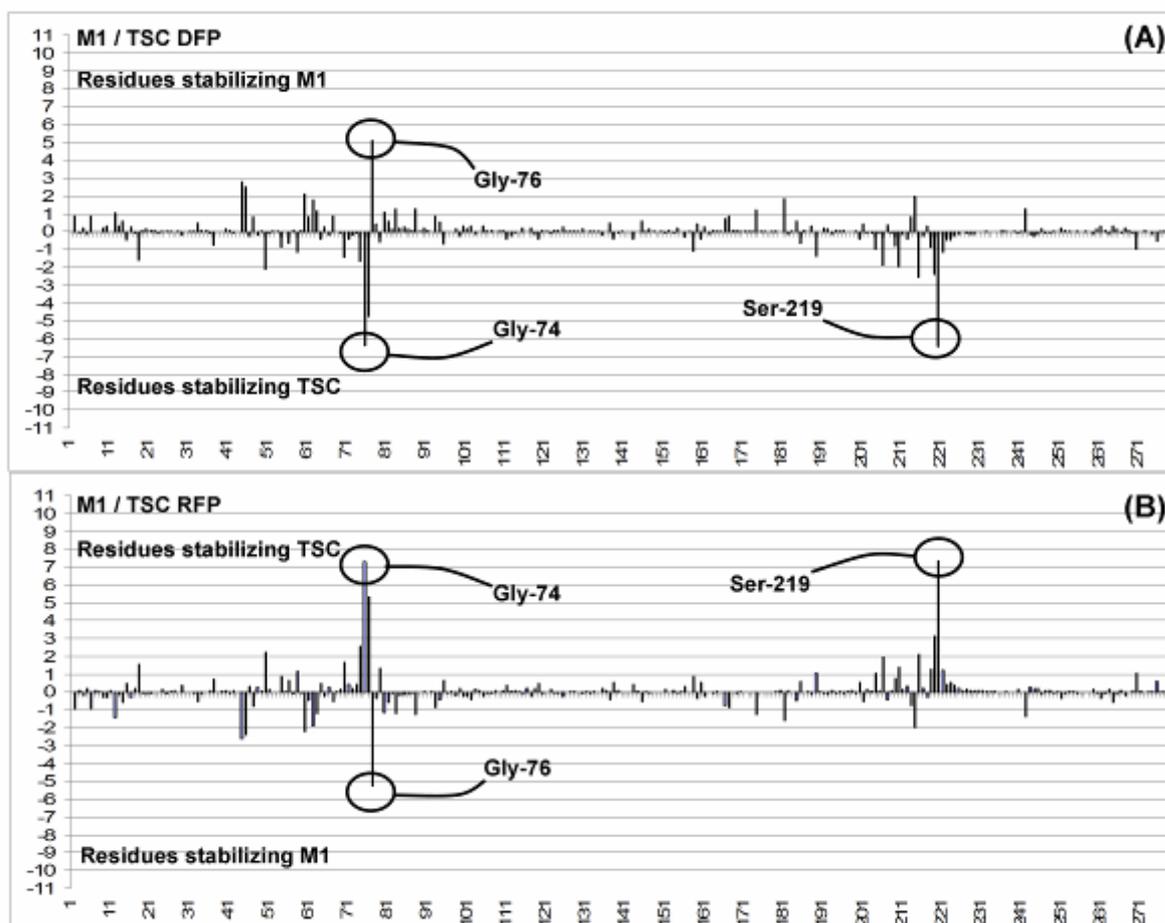


Figure 6.14 DFP and RFP diagrams for the M1/TSC transition.

Table 6.13 Stabilizing/destabilizing factor for the couple M1/TSC in Model2; values from selected residues are reported.

	Residue	S from DFP	D from RFP
Residues stabilizing M1	Asp-43	2.81	-2.62
	Gln-44	2.52	-2.40
	Gly-76	5.08	-5.28
Residues stabilizing TSC	Glu-49	-2.11	2.21
	Gly-74	-6.36	7.29
	Asn-75	-4.73	5.29
	Thr-214	-2.54	2.13
	Gly-218	-2.38	3.14
	Ser-219	-6.47	7.34

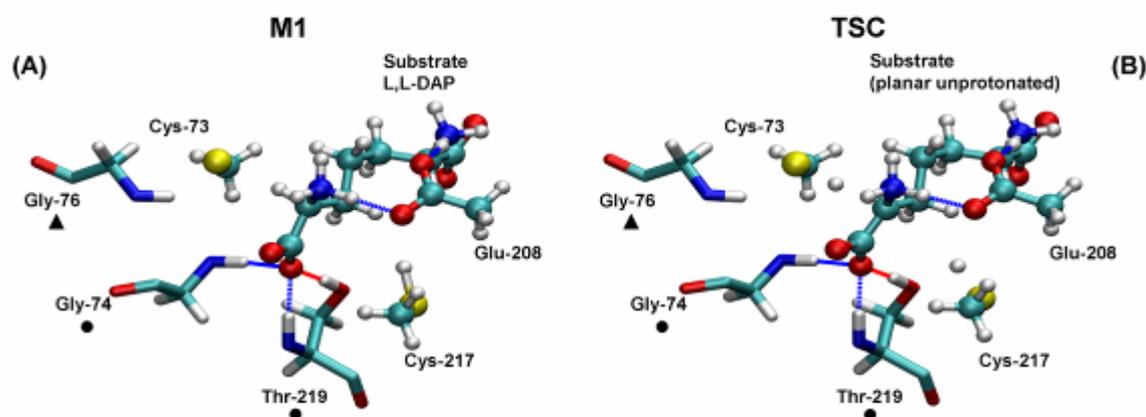


Figure 6.15 Schematic representation of the most important residues in the stabilisation/destabilisation of the M1/TSC couple

6.6 Conclusions

In this section an accurate computational study of the *Haemophilus influenzae* diaminopimelate epimerase have been described. The set-up procedure have been described in details, showing different approaches to obtain a suitable enzyme-substrate complex. The enzyme reaction mechanism have been described using an hybrid QM/MM potential, used to explore the PES associated to the stereoinversion of the C^α of the substrate (a doubly zwitterionic diaminopimelate molecule). Two different model systems have been taken into account, differing for the presence (**Model2**) or the absence (**Model1**) of a residue (Glu-208) in the QM layer. In both cases the calculations accounts for a concerted (but highly asynchronous) process of double proton transfer between the substrate as a couple of catalytically relevant cysteine residues (Cys couple). The energetic of the system have been analyzed to understand the role of the surrounding enzyme in the catalytic process. Two computational techniques, **DFP** and **RFP**, have been used to point out the role of each residue in the catalytic stabilisation of the transition structure. The results of both procedures suggest that a strong electrostatic effect helps in lowering the pK_a of the C^α of the substrate, by stabilizing the upcoming negative charge on the substrate itself (when passing from reactant to transition state). This charge accumulates on the carboxylic group of the diaminopimelate molecule and is stabilized by H-bond interaction with some residues near the reacting core and exposed to the active site. These residues have been successfully identified by both **DFP** and **RFP** and their effect was ranked.

Part III:

**Using Standard Computational
Approaches to Study Organic and
Bio-Organic Systems**

7 Investigations on the Electrochemical Behaviour of Some Ortho-Substituted (Aryl)(2-nitrobenzo[*b*]thiophen-3-yl)amines.

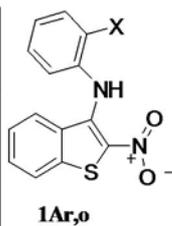
7.1 Introduction

The reduction potential values of the title compounds (**1Ar,o**, Scheme 7.1) have been evaluated²³² by cyclic voltammetry at a platinum electrode in the presence of 0.1 M N(Et)₄BF₄ in DMSO. Compounds **1Ar,o** give reversible reduction peaks. The ortho-substituent affects their values basically by electronic effects (good LFER have been observed). The peculiar behaviour of (2'-hydroxyphenyl)(2-nitrobenzo[*b*]thiophen-3-yl)amine **1Ar,o(i)** (showing two quasi reversible systems of peaks) has been explained on the basis of the special effect of the 2'-hydroxy substituent, which is able to assist a proton transfer through a network of hydrogen bonds involving the amine nitrogen atom (proton shuttle). This behaviour is unexpected in anhydrous DMSO and reminds of the reaction features observed in water. The above hypothesis has been confirmed by quantum-mechanical DFT calculations carried out on **1Ar,o(i)**. In this theoretical investigation all the possible species, that can form as a consequence of two mono-electron reduction processes and the relevant various chemical transformations involving the intermediate species, have been carefully investigated.

7.2 Electrochemical behaviour of **1Ar,o**

For compounds **1Ar,o(a-h)** the CVs were characterized by a cathodic-anodic peak system. At sufficiently high potential scan rate (> 200 mV/s), ΔE_p was about 59 mV and the ratio of the cathodic to anodic peak current was not significantly different from unity, confirming a quasi-reversible process. Furthermore, from the plot E vs $\log[(i_{pc}-i)/i]$, we argued that the process is mono-electronic. As an example, the voltammogram obtained for compound **1Ar,o(e)** is shown in Figure 7.1.

Compounds 1Ar,o	X
<i>a</i>	H
<i>b</i>	<i>o</i> -CH ₃
<i>c</i>	<i>o</i> -CH ₂ CH ₃
<i>d</i>	<i>o</i> -F
<i>e</i>	<i>o</i> -Cl
<i>f</i>	<i>o</i> -Br
<i>g</i>	<i>o</i> -OCH ₃
<i>h</i>	<i>o</i> -NH ₂
<i>i</i>	<i>o</i> -OH



Scheme 7.1 Compounds of the **1Ar,o** series.

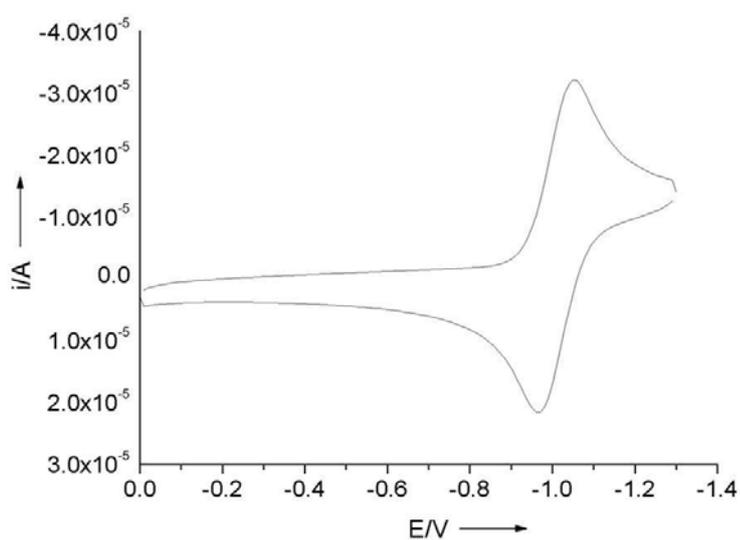


Figure 7.1 Cyclic voltammogram obtained for **1Ar,o(e)** (3 mM) in DMSO solution with 0.1 M N(Et)₄BF₄ as supporting electrolyte; $\nu = 200$ mV/s.

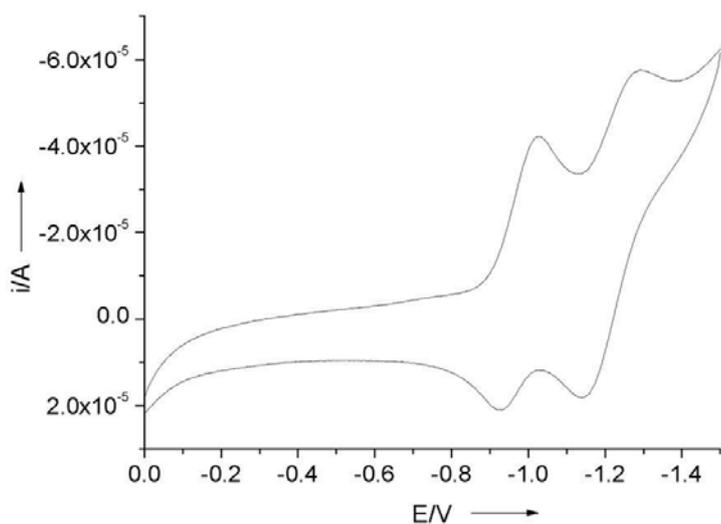


Figure 7.2 Cyclic voltammogram obtained for **1Ar,o(i)** (3 mM) in DMSO solution with 0.1 M N(Et)₄BF₄ as supporting electrolyte; $\nu = 200$ mV/s.

The cathodic peak can be attributed to the reduction of the nitro group to radical anion, and the anodic one to the reversible reaction of re-oxidation, *i.e.* these compounds behave in a way similar to nitrobenzenes in aprotic media.

The two methods²³² used to calculate the $E_{1/2}$ quantity, gave slightly different results, but the trend of the reduction half-wave potentials as a function of the benzene substituent, was the same (see values reported in Table 7.1). In **1Ar,o** the electronic nature of the benzene substituent strongly affects the driving force for the reduction. For instance, when X is an electron-withdrawing group (F, Cl, Br) the value of $E_{1/2}$ is more positive than *ca.* 50 mV, with respect to the potential value of the parent compound. On the other hand, when X is an electron-donating group (CH₃, CH₂CH₃, OCH₃, OH, NH₂), the $E_{1/2}$ values become more and more negative.

Table 7.1 Values of $E_{1/2}$ obtained by methods (i) and (ii) from the CVs recorded at $v = 20$ mV/s.

Compounds 1Ar,o	X(σ_p)	$E_{1/2}/V$	$E_{1/2calc}/V$
<i>a</i>	H (0.00)	-0.92	-0.89
<i>b</i>	<i>o</i> -CH ₃ (-0.14)	-0.94	-0.91
<i>c</i>	<i>o</i> -CH ₂ CH ₃ (-0.13)	-0.95	-0.91
<i>d</i>	<i>o</i> -F (0.06)	-0.89	-0.85
<i>e</i>	<i>o</i> -Cl (0.22)	-0.86	-0.82
<i>f</i>	<i>o</i> -Br (0.22)	-0.87	-0.82
<i>g</i>	<i>o</i> -OCH ₃ (-0.28)	-0.96	-0.94
<i>h</i>	<i>o</i> -NH ₂ (-0.57)	-1.01	-0.97
<i>i</i>	<i>o</i> -OH (-0.38)	-0.98 ^a	-0.94 ^a
		-1.22 ^b	-1.28 ^b

a) First peak

b) Second peak.

The electrochemical behaviour of (*ortho*-hydroxyphenyl)(2-nitrobenzo[*b*]thiophen-3-yl)amine [**1Ar,o(i)**] is different from the other *ortho*-substituted nitrocompounds **1Ar,o(a-h)**. In particular, the voltammogram is characterized by two quasi-reversible systems of peaks, as reported in Figure 2. The values of $E_{1/2}$ were calculated for both peaks and resulted -0.98 and -1.22 V, respectively when using the half-sum method, and -0.94 and -1.28 V when using the plot E vs $\log[(i_{pc}-i)/i]$. Furthermore, from the slope values of the plots we could deduce that both processes are mono-electronic.

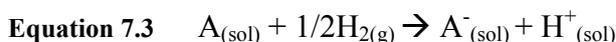
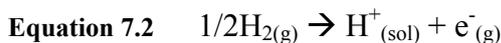
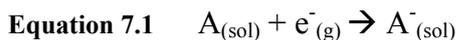
This different behaviour can be understood if we accept the hypothesis that the radical anion of compound **1Ar,o(i)** can be further reduced to a relatively stable dianion. The observed multi-step reduction (entirely unexpected in an anhydrous solvent) reminds of the behaviour of nitrocompounds in aqueous solvents. It is reasonable to believe that the presence of the *ortho*-hydroxy group in **1Ar,o(i)** can give rise to an intramolecular proton transfer *via* a network of hydrogen bonds with the amine nitrogen atom behaving as a proton shuttle. Since

it is rather unlikely, the possibility of an intermolecular proton transfer has not been considered. For a more complete discussion about the experimental procedures and the techniques used to determine the values reported in this section we refer to the original paper²³² and to its rich reference section. In this section we fully report the computational part of the study under examination.

7.3 Computational investigation of the behaviour of 1Ar,o(i)

In this section we discuss the results obtained in a DFT⁷⁵ computational study of the processes involving the molecular species **1Ar,o(i)**. All computations have been carried out with the Gaussian 03¹³¹ series of programs, using the B3LYP¹⁵¹ functional in the unrestricted form. The structures of the various critical points have been fully optimized with the gradient method and the nature of each critical point has been characterized by computing the harmonic vibrational frequencies. In all cases the geometry optimization has been carried out in the presence of solvent effects.

A locally dense basis set (LDBS)^{233,234} approach has been adopted. According to this approach, the system has been partitioned into two different regions, which were assigned basis sets of different accuracy. One region contains the atoms directly involved in the reaction (electron transfer and proton transfer processes) or in the formation of hydrogen bonds. For these atoms the 6-31++G(2p,d) basis has been used. This basis has the required accuracy to provide a reliable description of negatively charged atoms such as those involved in the process. Also, it is particularly suitable to describe the various hydrogen bonds occurring in the molecular system investigated here. All the remaining atoms are included in the second region and are described with the DZVP²²³ basis, which is a Local Spin Density (LSD)-optimized basis set of double-zeta quality. The two regions are marked by the contour lines shown in Figure 7.3, where the numbering of the various atomic centres for the molecular species **1Ar,o(i)** is also given. We briefly describe the method²³⁵ used to compute the reduction potential of the various Red/Ox couples considered in the present study. For the generalized Red/Ox couple A/A⁻ referred to the normal hydrogen electrode (NHE) we must consider the coupled reactions (Equation 7.1 and Equation 7.2) using the thermal-electron standard state²³⁵ and the resulting reaction (Equation 7.3), where A_(sol), A⁻_(sol), H⁺_(sol) refer to the solvated species and (g) indicates the gas-phase. The corresponding potential has been computed using the Nernst equation (Equation 7.4): here F is the Faraday Constant, n is number of involved electrons and $\Delta G^{\circ}_{\text{NHE,A/A}^-}$ is the Gibbs free energy variation (referred to the NHE) for the mono-electronic reduction of A_(sol) to A⁻_(sol) described by Equation 7.3. $\Delta G^{\circ}_{\text{NHE,A/A}^-}$ has been computed using Equation 7.5.



Equation 7.4 $E^\circ_{NHE,A/A^-} = -(\Delta G^\circ_{NHE,A/A^-})/nF$

Equation 7.5 $\Delta G^\circ_{NHE,A/A^-} = \Delta G_{A/A^-} + \Delta G_{NHE} = G(A^-_{(sol)}) - G(A_{(sol)}) + 426,8 \text{ kJ mol}^{-1}$

Equation 7.6 $E^\circ_{SCE,A/A^-} = E^\circ_{NHE,A/A^-} - 0.241$

The model system that we have considered is represented in Figure 7.3. We rationalize the macroscopic electrochemical results observed for **1Ar,o(i)** during the cyclic voltammetry²³² by examining the microscopic behaviour of the isolated **1Ar,o(i)** species by means of quantum-mechanical computations. The low concentration (3mM in DMSO) of the solute **1Ar,o(i)** has made possible to neglect in our model-system the formation of solute-solute interacting species such as dimers or clusters. To take into account the solvent effects, a Polarized Continuum Model approach has been used (SCRF-CPCM method¹⁵²). The emulated solvent was DMSO (dielectric constant $\epsilon = 46.7$).

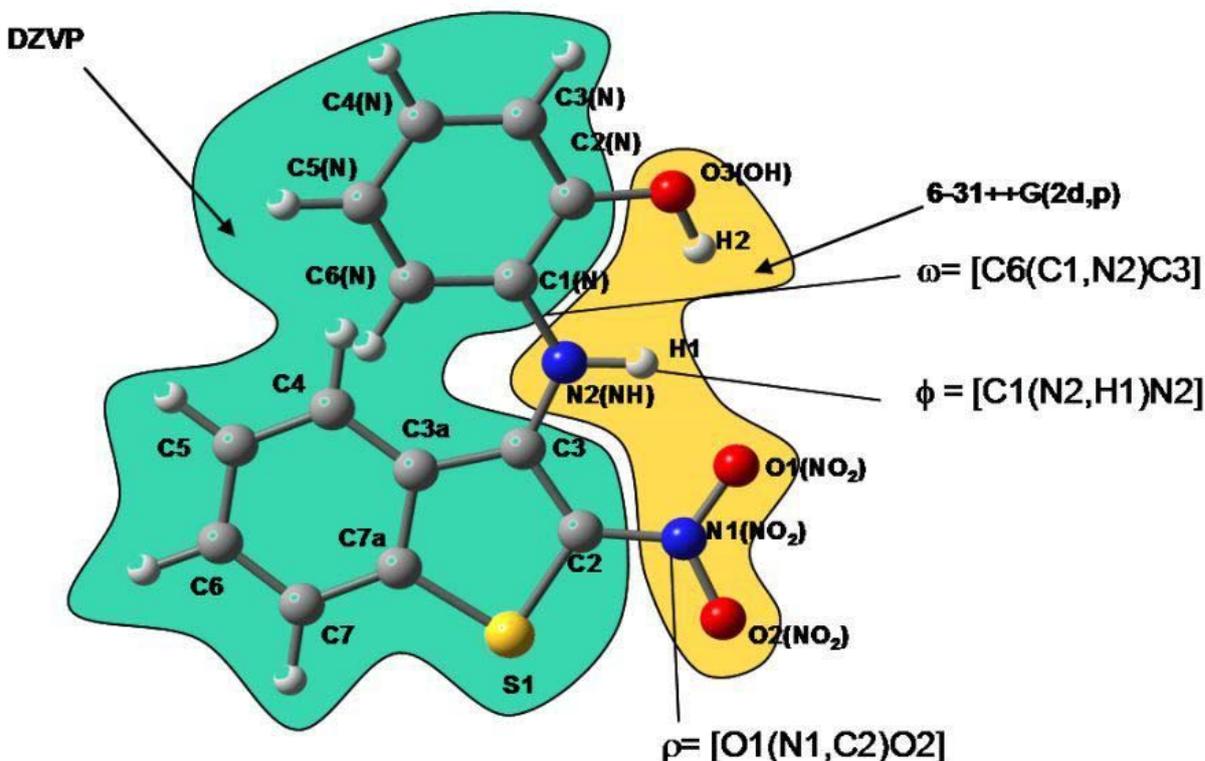
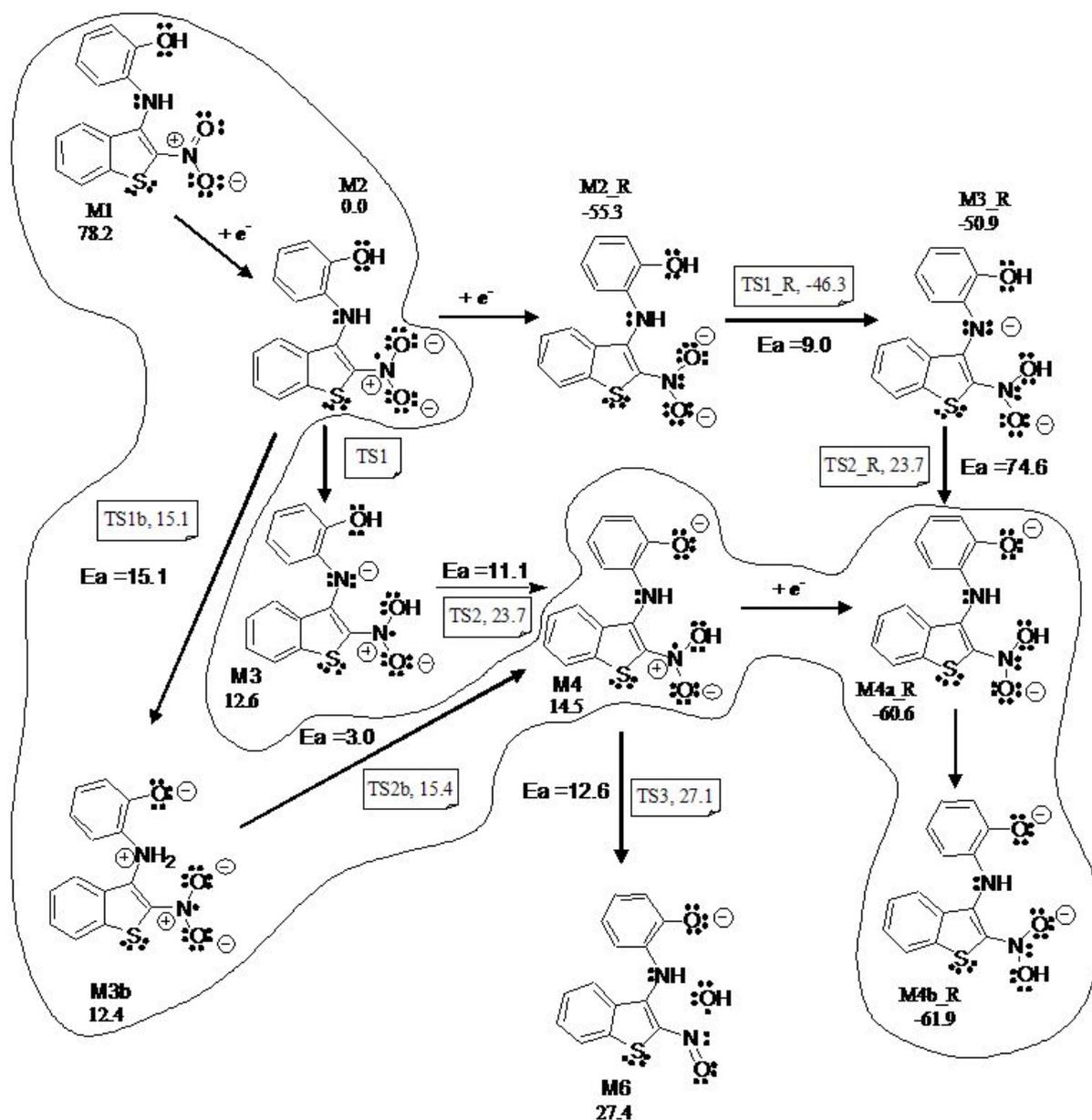


Figure 7.3 Schematic representation of the model-system used in the DFT computations. The two regions marked by the contour lines are described by the 6-31++G(2p,d) and DZVP basis set, respectively.

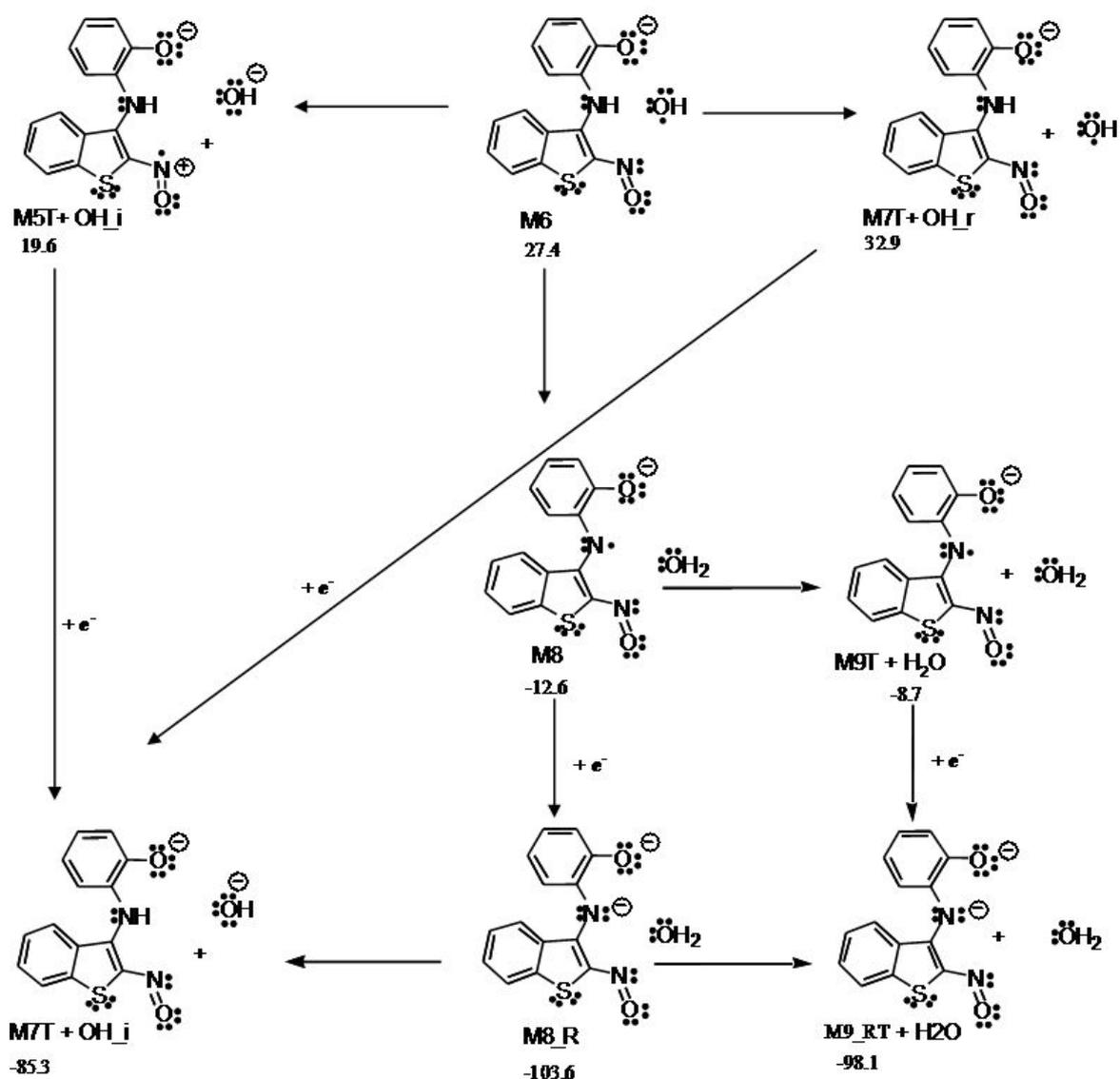


Scheme 7.2 Possible reaction paths originating from the starting species M1. The favored path is highlighted by the contour line.

In our investigation we have considered various possible routes that, in principle, could explain the experimental results. These routes are schematically represented in Scheme 7.2 and in Scheme 7.3. The structures corresponding to the most significant critical points of the reaction surface are shown in Figure 7.4, Figure 7.5 and Figure 7.6 with the values of the most important geometrical parameters. The computed total energy and free energy values of intermediates and transition states and the activation barriers for the various reaction steps are reported.

The starting molecular system M1 (see Figure 7.4) has a main reference plane that contains the two fused rings (thiophene and benzene) and the NO₂ group. The nitrogen atom N2 of the N2-H1 amino group has a partial sp³ character, the dihedral angle $\phi = [C1(N2,H1)C3]$ being

162.1° (this angle is a measure of the degree of pyramidalization of the nitrogen atom and is 180° for the sp^2 hybridization). The phenyl group is rotated with respect to the reference molecular plane and is approximately orthogonal to it. Two hydrogen bonds can be detected in M1. One involves the O1(NO₂) oxygen and the amino N2-H1 group. This O1...H1-N2 interaction is rather strong as indicated by the O1...H1 distance of 1.926 Å. The other hydrogen bond (O3-H2...N2) forms between the nitrogen N2 and the OH substituent in *ortho* position on the phenyl group. This bond is weaker than the previous one, the H2...N2 distance being 2.413 Å.



Scheme 7.3 Possible reaction paths originating from the hypothetical species M6.

The first electrochemical reduction leads to the radical anion system M2. Its energy is 78.3 kcal mol⁻¹ lower than that of M1. The corresponding computed reduction potential is -1.21 V (see Table 7.2), which is in rather good agreement with the experimental values of -0.98 and -0.94 V, obtained by methods (i) and (ii) (see Table 7.1). After reduction the O1...H1-N2

hydrogen bond remains almost unchanged, while the strength of the O3-H2 \cdots N2 interaction slightly increases (the H2 \cdots N2 distance becomes 2.289 Å). More significant geometrical changes are observed in the molecular skeleton, mainly in the region that comprehends the thiophene ring, the NO₂ group and the amine group NH. The most important changes are related to the C3-N2, N1-O1 and N1-O2 bond lengths that vary from 1.351, 1.258 and 1.249 Å in M1 to 1.404, 1.307 and 1.295 Å in M2, respectively. The augmented sp³ character of N2 and the consequent less effective conjugation of the ring π system with the nitrogen lone-pair, is evident from the dihedral angle value $\phi = [\text{C1}(\text{N2},\text{H1})\text{C3}]$, which is now 136.7°. A non-negligible shortening is observed for the C2-N1 bond that varies from 1.378 (M1) to 1.362 Å (M2).

Table 7.2 Computed Reduction Potentials for the various Redox Processes reported in Scheme 7.2 and in Scheme 7.3.

Redox couple A ^a /A ^{-b}	E° _{SCE} ^c (V)
M1/M2	-1,21
M2/M2 _R	-2,22
M3/M3 _R	-1,91
M4/M4a _R	-1,41
M5T + OH _i /M7T + OH _i	-0,11
M7T + OH _r /M7T + OH _i	-3,50
M8/M8 _R	-0,78
M9T + H ₂ O/M9 _{RT} + H ₂ O	-0,79

^a Oxidized species.

^b Reduced species.

^c Reduction potential referred to SCE (Saturated Calomel Electrode).

The shape of the SOMO and the computed spin density values indicate that the unpaired electron is delocalized on the following atomic centers: N1, O1, O2, C3, C7a, N2 and S (spin density on these atomic centers: 0.32, 0.16, 0.18, 0.29, 0.08, 0.05 and 0.01).

M2 opens the way to different reaction paths. M2 could undergo a second one-electron reduction to afford the M2_R dianion. However, the computed standard reduction potential for the Redox couple M2/M2_R is -2.22V (see Table 7.2), a too negative value when compared to the experimental second reduction potential of -1.22 or -1.28 V (obtained by methods (i) or (ii), respectively). This result suggests that the experimental reduction potential refers to a species different from M2 and that the system could be involved in some chemical transformation before undergoing the second reduction.

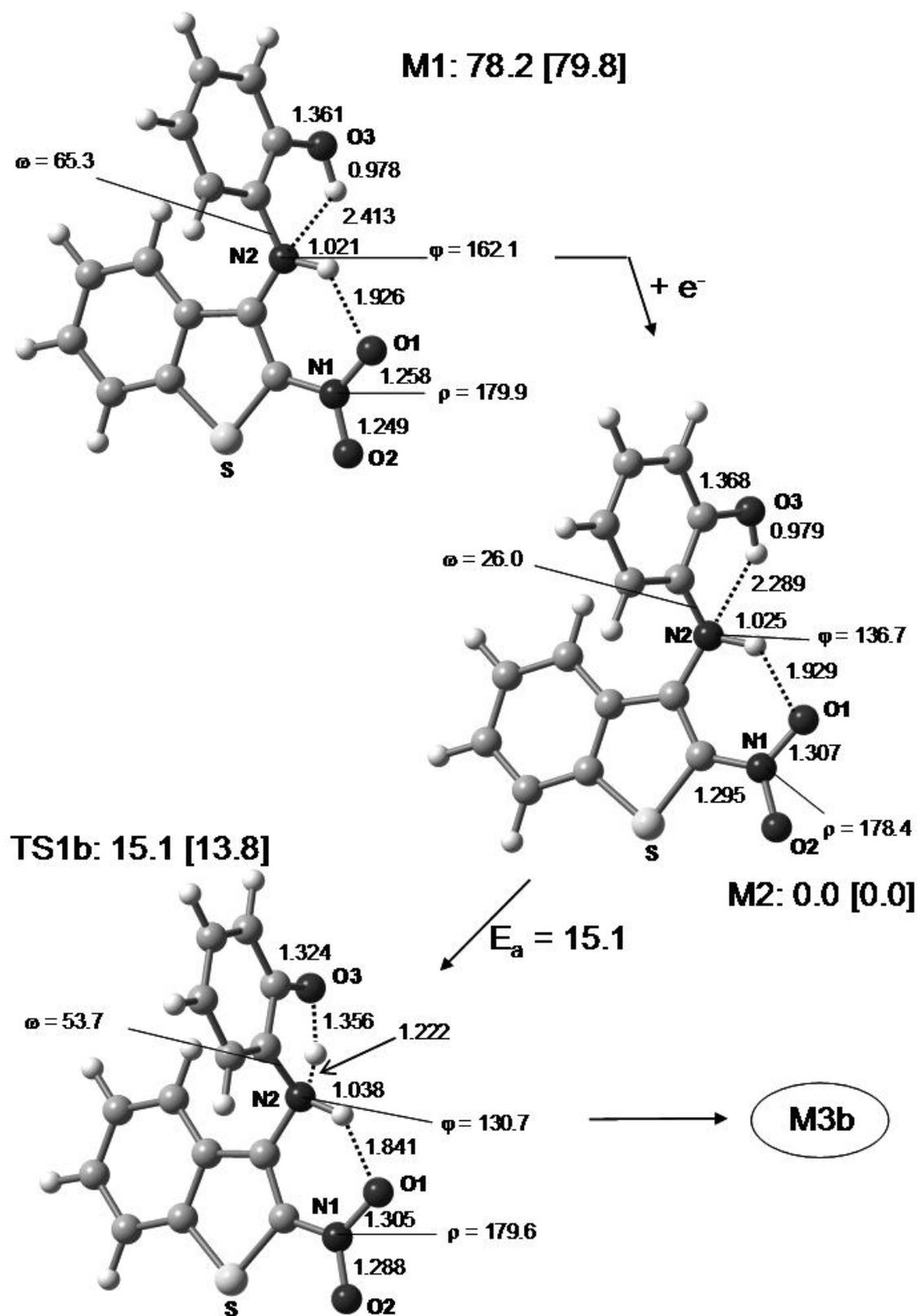


Figure 7.4 Schematic representation of the structure of M1, M2 and TS1b. The energies and Gibbs energies (values in brackets) are relative to the M2 species (kcal mol^{-1}). E_a = activation barriers. Bond lengths are in Ångstroms and angles in degrees.

An alternative reaction path is represented by a sequence of two proton transfers leading to a new intermediate M4. This proton transfer occurs in two steps (a first transfer from O3 to N2 and, then, a second one from N2 to O1) and any attempts to find a concerted transition state failed. However, even if the two steps are topologically distinct, the second one (M3b \rightarrow TS2b \rightarrow M4) has a small barrier (only 3.0 kcal mol⁻¹) compared to that of the first step (M2 \rightarrow TS1b \rightarrow M3b), which is 15.1 kcal mol⁻¹. These features make the whole transformation closely resemble a two-phase concerted process where the two proton transfers are highly asynchronous, but belong to the same kinetic step. The final effect of this double proton transfer is the shift of one proton from the O3-H2 group of the phenyl ring to the NO₂ group, with the amine fragment acting as a proton shuttle. This proton shuttle maintains along the reaction pathway the pair of hydrogen bonds that stabilize the systems and helps the proton transfer. The most significant structural variations on passing from M2 to M4 are observed for the C3-N2, N1-O1, N1-O2 and C2(N)-O3 bond lengths. The change of the C3-N2 bond (from 1.404 to 1.427 Å) can be explained by the new orientation of the N2-H bond, which is almost orthogonal to the molecular plane in M4. The consequent orientation of the nitrogen lone-pair in M4 eliminates the conjugation with the thiophene π system and causes a lengthening of the C3-N2 bond. Two new hydrogen bonds (O3 \cdots H2-N2 and N2 \cdots H1-O1), similar to those previously discussed for M2, can be recognized in M4. The second one is rather strong, the N2 \cdots H1 distance being 1.615 Å.

The double proton transfer leading from M2 to M4 can follow an alternative two-step path. In the first step a proton moves from N2 to O1 (transition state TS1) affording the M3 intermediate where the amine group is negatively charged. M3 is 12.6 kcal mol⁻¹ higher than M2. In the second step a proton is transferred from O3 to the negative nitrogen N2 (transition state TS2). For this step we have computed an activation barrier of 11.1 kcal mol⁻¹. Thus, since TS2 is 8.3 kcal mol⁻¹ higher than TS2b (i.e. the highest in energy point along the profile M2 \rightarrow M3b \rightarrow M4), this pathway leading to M4 is certainly less likely with respect to that previously discussed.

M4 could now accept a second electron or, alternatively, could eliminate a hydroxyl radical to give the nitroso derivative M6. However, since M6 is 12.9 kcal mol⁻¹ higher than M4, this process is probably much slower than a second electron transfer to give M4a_R, the final product of the electrochemical process. Two aspects supporting this hypothesis must be stressed. (a) The computed standard reduction potential for the Redox couple M4/M4a_R is -1.41 V, a value which compares rather well with the experimental second reduction potential of -1.28 V. On the contrary, the agreement between this experimental value and the computed

standard reduction potentials for the one-electron reductions that could occur after the formation of M6, is not satisfactory. This is evident from Table 7.2, where we have collected the computed reduction potentials for the various Redox processes reported in Scheme 7.2 and Scheme 7.3. (b) If we consider the pathway leading to the formation of a hydroxyl radical and the nitroso derivative, the whole process has definitely an irreversible character in contrast with the experimental evidence.

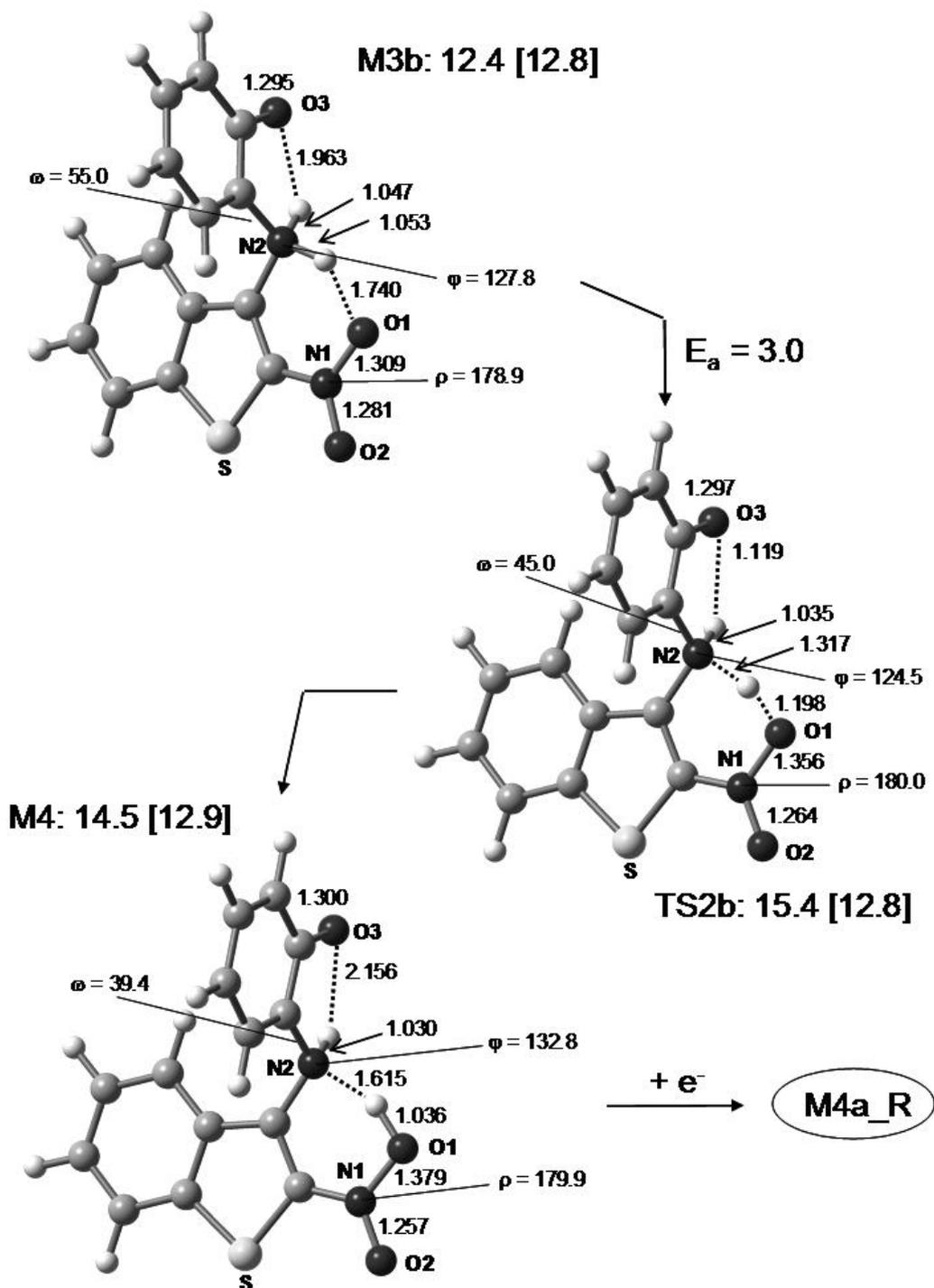


Figure 7.5 Schematic representation of the structure of M3b, TS2b and M4. The energies and Gibbs energies (values in brackets) are relative to the M2 species (kcal mol^{-1}). E_a = activation barriers. Bond lengths are in Ångströms and angles in degrees.

The structure of M4a_R is reported in Figure 7.6. The computations show that the double negative charge of the dianion is mainly localized on O2, O3 and N2 (the Mulliken charges are -0.73, -0.88 and -0.60, respectively). M4a_R exists in a conformational equilibrium with M4b_R, more stable by 1.3 kcal mol⁻¹. A rotation around the C2-N1 bond connects the two conformers that show very similar geometrical parameters. A representation of the M4b_R structure is given in Figure 7.6.

In conclusion, the satisfactory agreement of the computed values of the first and second reduction potentials with the experimental ones, the energy barriers of the double proton transfer and the irreversibility that characterizes the process leading to the formation of M6, suggest that the transformation M1 → M2 → TS1b → M3b → TS2b → M4 → M4a_R is a likely description of the whole electrochemical process. This path is highlighted by a contour line in Scheme 7.2.

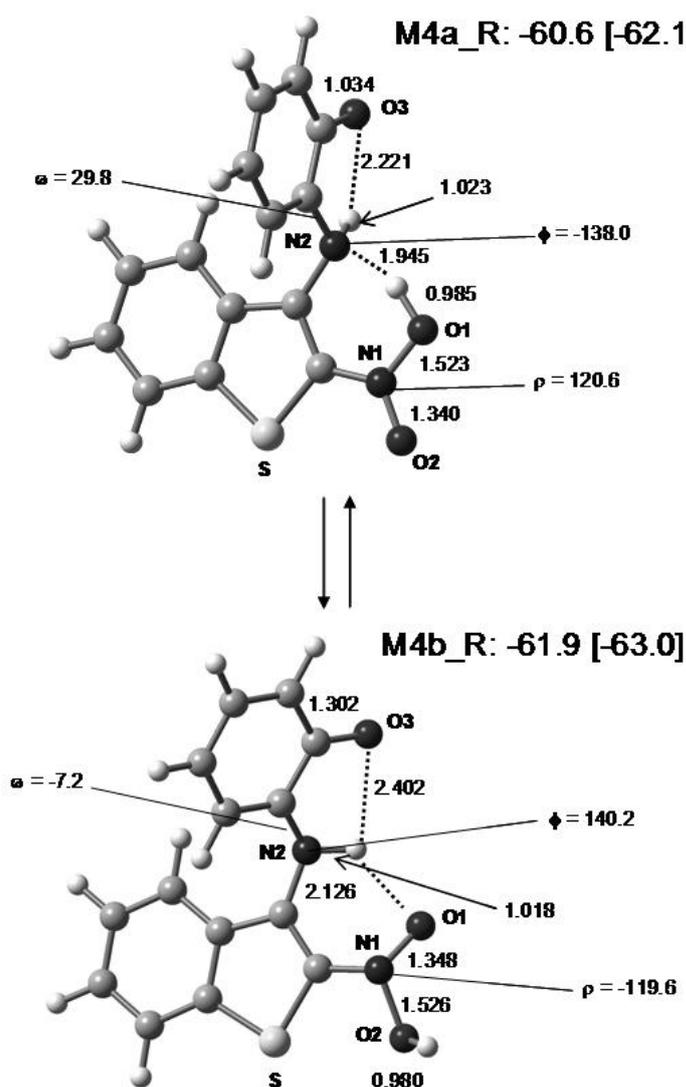


Figure 7.6 Schematic representation of the structure of M4a_R and M4b_R. The energies and Gibbs energies (values in brackets) are relative to the M2 species (kcal mol⁻¹). E_a = activation barriers. Bond lengths are in Ångstroms and angles in degrees.

7.4 Conclusions

The electrochemical behaviour of compounds **1Ar,o(a-i)** in anhydrous DMSO solution has been studied by cyclic voltammetry: the obtained experimental results give rather good ($r = 0.984$, $n = 9$) LFER *versus* Hammett constants (σ_p) with high susceptibility constants ($\rho = 182$). Compounds **1Ar,o(a-h)** showed a monoelectronic *quasi* reversible process.

An interesting electrochemical behaviour has been evidenced in the case of **1Ar,o(i)**. For this species the experimental data show a reversible electrochemical process characterized by two mono-electronic reduction steps. This process should lead to a rather stable dianion that does not easily evolve to a new more stable species. This hypothesis has been supported by accurate DFT computations on **1Ar,o(i)**. These computations have demonstrated that, after the first reduction, the resulting radical anion can undergo rather easily a double proton transfer. A key-role in determining the height of the activation barriers for this proton shift, is played by the network of hydrogen bonds that involve the NH group bridging the two ring systems. After the proton transfer the new species (M4) can undergo the second reduction or, alternatively, can eliminate a hydroxyl radical affording a nitroso derivative. Since the nitroso compound is rather high in energy (12.9 and 27.4 kcal mol⁻¹ higher than M4 and M2, respectively), the most likely event is a second electron transfer to give the dianion M4a_R. This hypothesis is supported by the good agreement between the computed and experimental half reduction potentials.

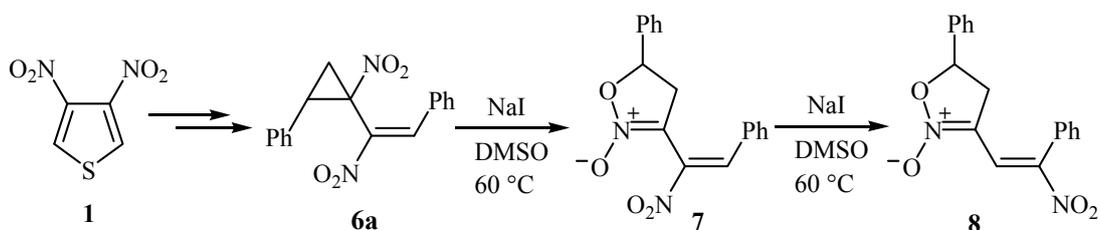
NOTE:

The experimental section²³² was carried out by the group of Professor D. Tonelli; Professor A. Bottoni must be acknowledged for the supervision to the computational section. The joint project was lead by Professor D. Spinelli.

8 Unprecedented 1,2-Migration of a Nitro Group within an α -Aryl- β -nitroethenyl Moiety

8.1 Introduction

In the last decades the ring-opening of nitrothiophenes has opened the way to the synthesis of new building-blocks from which both linear and homo- or hetero-cyclic compounds can be synthesized.²³⁶ The conversion of **1** to **6a** has been previously reported.²³⁷⁻²³⁹



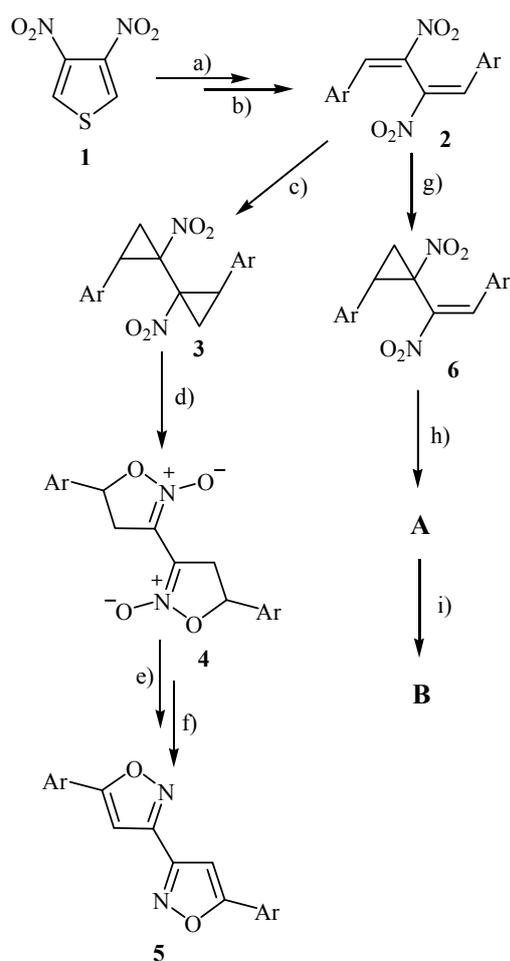
Scheme 8.1 Reaction scheme of the examined chemical transformation.

When heated at 60 °C in DMSO in the presence of NaI, **6a** isomerises to the isoxazoline *N*-oxide **7** (isolable in proper experimental conditions), which more slowly quantitatively evolves towards **8**. The structure of both **7** and **8** has been ascertained on the basis of ¹H and ¹³C NMR as well as MS and HRMS results, and definitively confirmed through an X-ray analysis. The **7** to **8** rearrangement represents the first example of isomerisation of an α -aryl- β -nitroethenyl into a α -aryl- α -nitroethenyl moiety. A rationale for such a surprising nitrogroup 1,2-shift is supported by DFT⁷⁵ calculations and some mechanistic details are provided.

8.2 Experimental section

All the experimental details and observed data are reported in the original paper,²⁴⁰ where a full list of references is also shown. When subjected to the experimental conditions previously employed for the 3 to 4 isomerisation above (see Scheme 8.2), the representative nitrovinylcyclopropane **6a** (Ar = Ph) has interestingly revealed (¹H NMR analysis of test runs carried out in DMSO-*d*₆) the relatively fast accumulation of a transient A which slowly decays to the final product B. In preparative runs, the intermediate A can be isolated by

work-up of the reaction after a suitable time (1 h): it is stable in DMSO at 60 °C but smoothly yields B quantitatively after addition of NaI, i.e. in experimental conditions similar to those applied to the precursor 6a. A single-crystal X-ray structural determination of A and B (see the ORTEP²⁴¹ drawings in Figure 8.1) has allowed to assign to such compounds the structures 7 (i.e. the expected outcome of a nitrocyclopropane to isoxazoline N-oxide isomerisation on 6a) and 8, respectively: a result which unveils an interesting and unprecedented 1,2-migration of a nitrogroup in relatively mild experimental conditions (Scheme 8.2). As a matter of fact, the 1,2-migration of a nitro group within a nitroethenyl moiety is a well-known example²⁴² of a more general isomerisation process undergone by unsaturated systems characterized by the presence of a heteroatom (Hal, O, S, Se, N) bonded to a Csp² atom (Scheme 8.3). In particular, in the case of nitroolefins proofs for the intramolecular base-catalyzed mechanism of Scheme 8.3 have been reached by means of isotopic labeling of either the substrate (sizeable kinetic deuterium isotope effect) or added nitrite (absence of ¹⁵N uptake into the product).



Reagents and conditions:

- (a) Et₂NH (excess), EtOH, 0 °C, overnight;
- (b) ArMgBr (2.2 mol equiv.), THF, 0 °C, 15'-45'; then acidic (NH₄Cl) quenching;
- (c) CH₂N₂ (excess), Et₂O, 0 °C to rt, overnight;
- (d) NaI (2 mol equiv.), DMSO, 60 °C, 4-22 h;
- (e) P(OMe)₃, (40 mol equiv.), anhydrous dioxane, reflux under argon, 17-24 h;
- (f) DDQ (4-6 mol equiv.), dry toluene, reflux under argon;
- (g) CH₂N₂ (1.2 mol equiv.), Et₂O, 0 °C to rt, overnight;
- (h) and
- (i) NaI, DMSO, 60 °C, 1-5 h (A) or 24-48 h (B).

Scheme 8.2 Experimental procedure diagram.

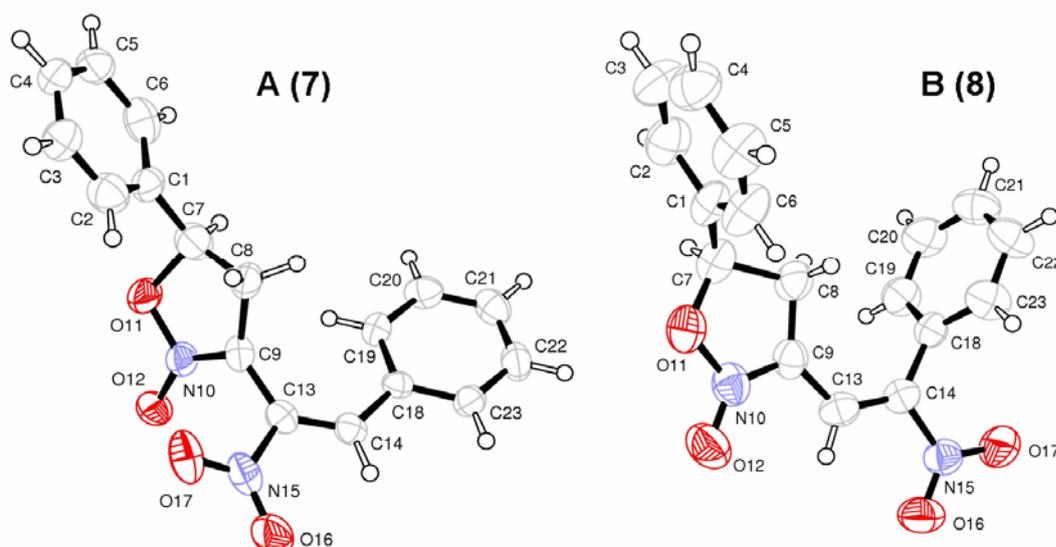
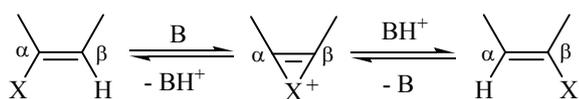
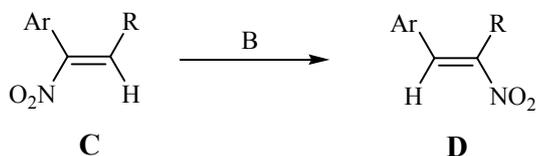


Figure 8.1 ORTEP representation of the crystal structures of **A (7)** and **B (8)**..



Scheme 8.3 Schematic representation of the hypothesised olephine rearrangement.



Scheme 8.4 Shift of a nitrogroup on a double bond.

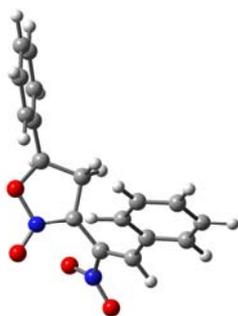
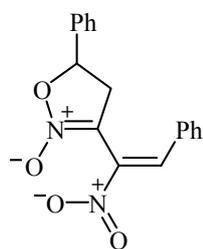
At the best of our knowledge, the literature provides exclusively examples of an irreversible base-catalyzed rearrangement towards the isomer (*i.e.* **D** in Scheme 8.4) which takes thermodynamic advantage of the through conjugation between the aryl and the nitrogroup. The results herein, though, feature a nitrogroup migration which is in the opposite direction with respect to that of Scheme 8.4: a behaviour which could most likely find its driving force in the release of the conceivable repulsive interactions of steric and/or stereoelectronic nature between the cyclic nitronate moiety and the nitrogroup in **A**. On the other hand, the loss of the existing through-conjugation between the aryl and the nitrogroup in **A** should be counterbalanced by a new extended conjugation between the nitrogroup and the cyclic nitronate in **B**.

8.3 Computational Section

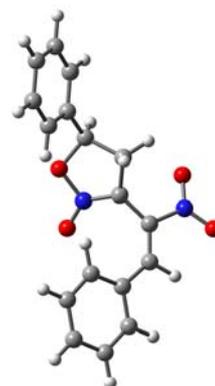
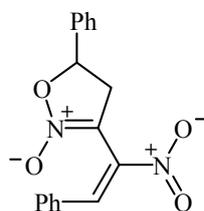
All calculations have been carried out using the Gaussian 03¹³¹ series of programs. The structures of the various minima points have been fully optimized with the gradient method. After the gas-phase optimization, single point calculations in solution were performed. The correctness of such a methodology was checked by means of a full-optimization calculation considering the solvent. In each case the solvent (ϵ 46.7) was treated with the SCRF-CPCM¹⁵² method.⁸ The calculations were performed using the DFT⁷⁵ method and the B3LYP¹⁵¹ functional, available in Gaussian 03. All atoms were assigned the double- ζ DZVP²²³ basis-set. The data obtained from calculations, carried out both on isolated molecules and in solution, provide definite thermodynamic support to the **A** to **B** isomerisation, enlightening a significant free-energy stabilisation due to the release of unfavourable interactions between the nitronate moiety and the nitro- (*s-trans-A*) or the phenyl group (*s-cis-A*) [Figure 8.2, where calculated relative energies in DMSO as well as in the gas phase (in parentheses) for the optimized geometries are reported].

The latter effect still negatively contributes (at least in the gas phase) to the energy of *s-cis-B*, furthermore most likely preventing coplanarity (and therefore full conjugation) within the unsaturated system. On the other hand, the *s-trans* conformer of the rearranged nitrovinyl moiety (*i.e.* *s-trans-B*) takes definite sizeable advantage (amounting to 5.6 and 3.9 Kcal/mol with respect to *s-trans-A* and *s-cis-A*, respectively) of an effective coupling between minimization of negative steric interactions and the ensuing full coplanarity within the extended unsaturated system.

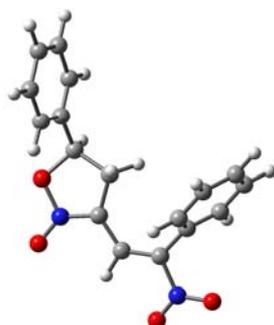
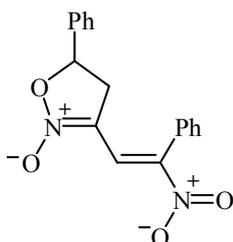
Studies are in progress in order (a) to gain more detailed mechanistic information and, at the same time, (b) to verify the practical limits of a rearrangement which undoubtedly unmask yet new appealing synthetic applications for the butadiene building-blocks originating from the ring-opening of 3,4-dinitrothiophene.



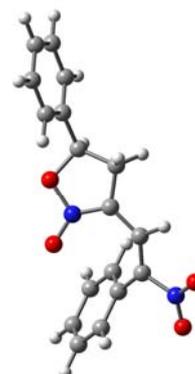
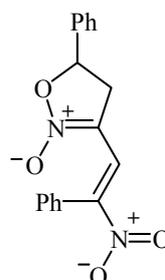
s trans A
 $E = 0.0 \text{ kcal/mol}$
 (reference)



s cis A
 $E = -1.64 \text{ kcal/mol}$
 (-2.25 kcal/mol)



s trans B
 $E = -5.57 \text{ kcal/mol}$
 (-5.00 kcal/mol)



s cis B
 $E = -1.92 \text{ kcal/mol}$
 (-0.33 kcal/mol)

Figure 8.2 Optimised structures of *s-trans* and *s-cis* isomers of both **A** (7) and **B** (8)

NOTE:

The experimental section²⁴⁰ was carried out by the group of Professor G. Petrillo; The joint project was lead by Professor D. Spinelli.

9 Rationalisation of complex molecular rearrangement

9.1 Introduction

The Mukaiyama–Michael-type addition of various silyl ketene acetals or silyl enol ethers on some 1,2-diaza 1,3-butadienes proceeds at room temperature in the presence of catalytic amounts of Lewis acid affording by heterocyclization 1-aminopyrrol-2-ones and 1-aminopyrroles, respectively (Figure 9.1, Scheme 9.1). 1- Aminoindoles have been also obtained by the same addition of 2-(trimethylsilyloxy)-1,3-cyclohexadiene on some 1,2-diaza-1,3-butadienes and subsequent aromatization. Mechanistic investigations indicate the coordination by Lewis acid of the enolsilyl derivative and its 1,4-addition on the azo-ene system of 1,2-diaza-1,3-butadienes. The migration of the silyl group from a hydrazone to an amidic nitrogen, its acidic cleavage and the final internal heterocyclization give the final products. Based on NMR studies and DFT calculations, a plausible explanation for the migration of the silyl protecting group is presented.

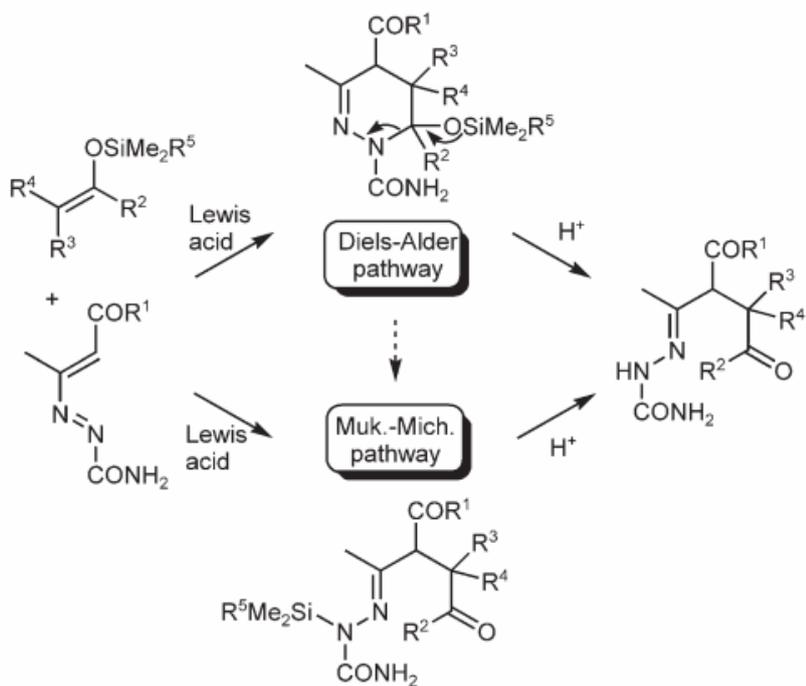
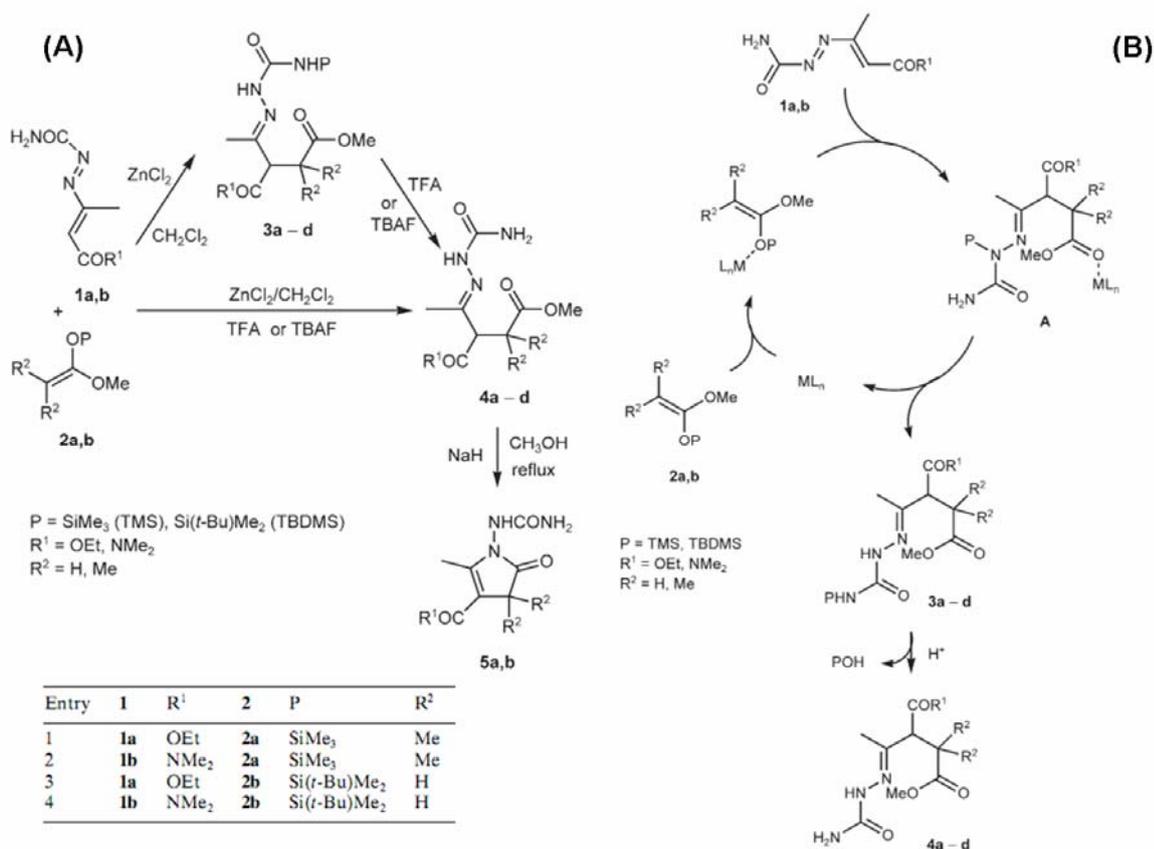


Figure 9.1 Possible reaction pathways for the addition of enolsilyl derivatives on 1,2-diaza-1,3-butadienes.

9.2 Discussion



Scheme 9.1 A) Mukaiyama-Michael-type addition/heterocyclization reaction of silyl ketene acetals **2a,b** on 1,2-diaza-1,3-butadienes **1a,b**; B) Proposed mechanism for the Lewis acid-catalyzed Mukaiyama-Michael-type addition of enolsilyl derivatives **2a,b** on 1,2-diaza-1,3-butadienes **1a,b**.

The experimental section with all the procedures and the characterisation of products and intermediates is fully reported in the original paper²⁴³, as well as a complete list of references. ¹H-NMR analysis²⁴³ of the reaction mixtures seems able to indicate that the addition of silyl ketene acetals **2a-b** to 1,2-diaza-1,3-butadienes in the presence of $ZnCl_2$ occurs via some intermediate (transient) products. As a matter of fact the presence of signals at 1.87, 3.78 and 4.88 ppm disappearing with the time seems a clear indication of the formation of short-lived intermediates **A**, able to give **3a-d** and then **4a-d** with high yields. The intermediates would have the structure (**A**) of the expected product of 1,4-addition of **2a-b** to **1a-b**.

In this section we will make some comments on the hypothesized rearrangement of **A** intermediates into **3a-d**, examining the possible reasons for the transformation, advancing some hypotheses on its occurrence, and then possibly offering a first proposal on its mechanism. It is well known that a “spontaneous” rearrangement occurring during the course of a reaction and/or during its work-up “must” be thermodynamically driven. That is, the “less stable” firstly formed product “can” (“can” means that this fact occurs only if the activation energy of the process is quite low considering the reaction experimental conditions) give the “more stable” one. We performed some calculations on simplified model systems to give a qualitative confirmation to our hypotheses. All calculations were performed at the DFT⁷⁵ level with the Gaussian 03¹³¹ series of programs using the B3LYP¹⁵¹ functional; in all calculations the solvent (CH_2Cl_2) was simulated using the SCRF-CPCM method. A locally dense basis set(LDBS)²⁴⁴ approach has been adopted. Preliminary calculations at the STO-3G level allowed a fast and rough exploration of PES to identify the most stable conformers; the so-obtained candidate geometries have been optimized at the DZVP²²³ level, thus obtaining results reported in Table 9.1. NOE experiments²⁴³ indicated an *E*-relationship for **3c**; to gain support for this result we calculated the stability of **A** and of **3** using the simplified models reported in Scheme 9.2 and indicated as *E-X* and *E-Y*, respectively.

Table 9.1 Energies of stationary points.

molecule	Energy	
	hartree	kcal mol ⁻¹
<i>E-X</i>	-958.713673	0.0
<i>E-Y</i>	-958.724789	-6.98
<i>Z-X</i>	-958.712949	0.45
<i>Z-Y</i>	-958.721609	-4.98

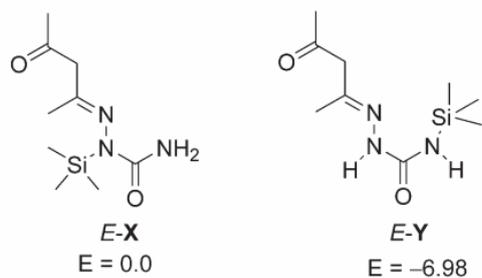


figure 3

Scheme 9.2 Simplified models for *E*-isomers of **A** and **3**: *E-X* and *E-Y* (calculated differences in energy contents is reported).

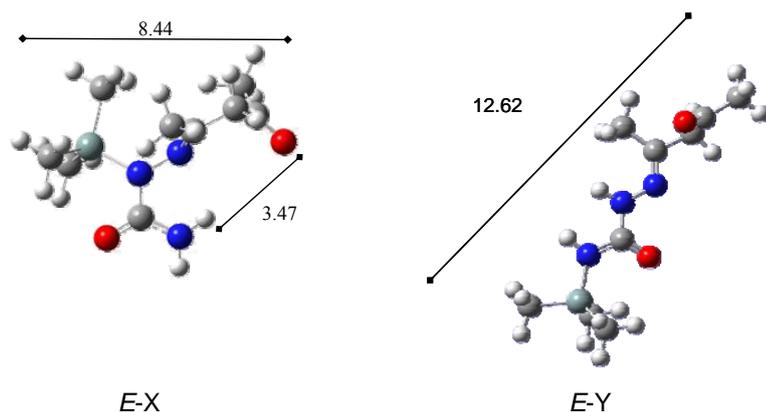
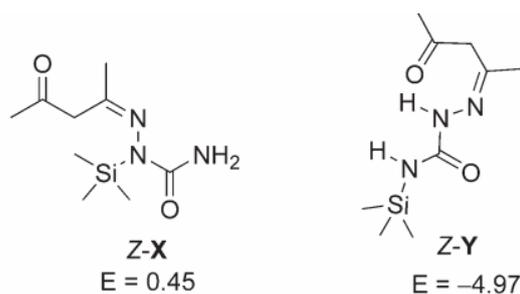


Figure 9.2 Three dimensional representations for *E-X* and *E-Y*. Reported distances are in Å.

In line with previsions, *E-Y* resulted to be much more stable than *E-X* ($\Delta E = \text{ca. } 7 \text{ kcal mol}^{-1}$). This result supports the reasonable hypothesis that the spontaneous rearrangement is thermodynamically driven. It is also interesting to point out that if an equilibrium between *E-X* and *E-Y* were present, the stability gap would ensure an almost complete shift towards *E-Y*, since the equilibrium constant, K_{eq} , is as large as 10^5 at room temperature. Having ascertained the significant difference in the energy content between *E-X* and *E-Y*, now by a deeper analysis of models it should be possible to understand the reason for this fact. We ascribe the lesser stability of *E-X* in respect to *E-Y* to an higher overcrowding in the first regioisomer, in fact the two structures do extend, respectively, in 8.44 and 12.62 Å (Figure 9.2). This fact causes in *E-X* strong steric interactions between one of the methyls of the trimethylsilyl group and that at the old C-3 of the 1,2-diaza-1,3-butadiene, that cannot be released by rotation along the single N-N bond because this would result in a larger overlap between bulkier groups: accordingly we failed in all attempts to find other minima when rotating along this bond. We have also examined the structures and the relative stability of the *Z*-isomers to gain information useful for the understanding of the reasons by which only the formation of the *E*-isomers (*E-X* and *E-Y*) has been observed. This calculation could be easily carried out with the relevant models of *Z-Y* and *Z-X*, respectively (Scheme 9.3).



Scheme 9.3 Simplified models for *Z*-isomers of **A** and **3**: *Z-X* and *Z-Y* (calculated differences in energy contents is reported).

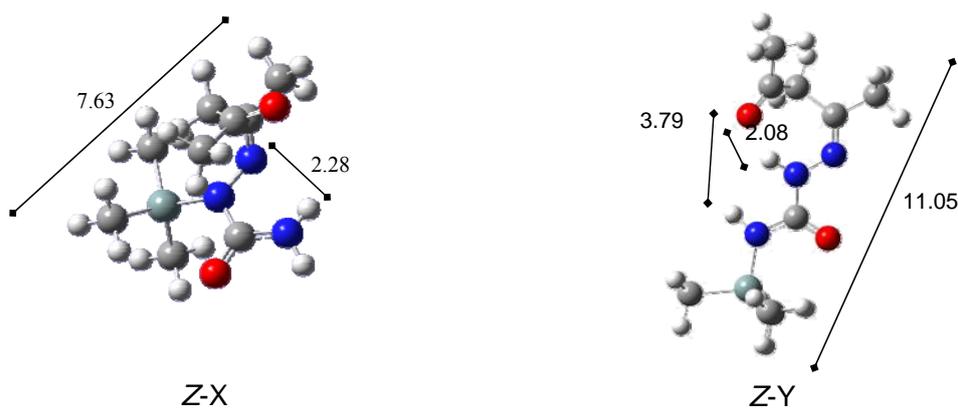


Figure 9.3 Three dimensional representations for *Z-X* and *Z-Y*. Reported distances are in Å.

The *Z*-isomers are less stable than *E*-isomers by 2 and 0.5 kcal mol⁻¹, respectively, that is, values well in line with the usual observed differences between *E* and *Z* stereoisomers. Of course comments as above made on *E-X* and *E-Y* can be attached to *Z-X* and *Z-Y* models. Interestingly an examination of **Z-Y** (Figure 9.3) evidences the presence of a feeble hydrogen bond (2.08 Å) between the carbonyl group and the hydrazonic proton.

Concerning the possible mechanism of the rearrangement, two pathways appear, in principle, to be possible: an intramolecular path, occurring via acyclic four-membered transition state, and an intermolecular one, occurring via a presumably more stable cyclic eight-membered transition state. The results of cross-over experiments²⁴³ seem able to support the idea that an intermolecular process occurs. Research in progress is devoted to a deeper understanding of this point.

NOTE:

The experimental section²⁴³ was carried out by the group of Professor O. Attanasi; The joint project was lead by O. Attanasi and Professor D. Spinelli.

10 The Cycloaddition Reaction between Vinylketenes and Imines

10.1 Introduction

Ketenes exhibit a very peculiar cycloaddition chemistry because of their structural and electronic properties.²⁴⁵ One of the most valuable and exploited pseudo-pericyclic reaction of ketenes is certainly the reaction with imines to afford β -lactams, discovered by Staudinger at the beginning of the 20th century.²⁴⁶ In view of the importance of β -lactams in medicinal chemistry as antibacterial agents²⁴⁷⁻²⁵¹ and as enzymatic inhibitors²⁵²⁻²⁵⁵, Staudinger reaction has been extensively studied, both experimentally²⁵⁶⁻²⁶⁸ and computationally.²⁶⁹⁻²⁷⁹

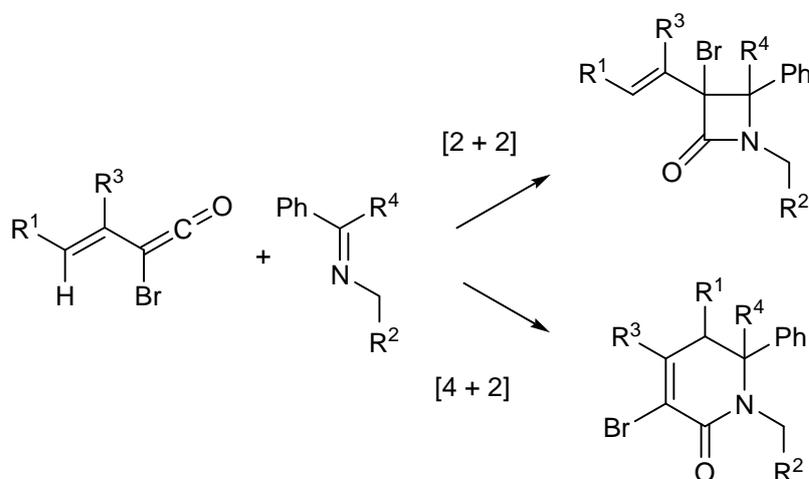
An interesting class of ketenes are the vinylketenes, that have proven to be versatile building blocks in cycloaddition reactions with various double bonds.²⁸⁰⁻²⁸⁴ They are reported to behave as electron-deficient dienophiles in [2+2] cycloadditions with electron-rich partners and as diene component in [4+2] cycloadditions with electron-poor species.^{285,286} Therefore, vinylketenes usually react with imines *via* Staudinger reaction. At the best of our knowledge, only silyl-vinylketenes have demonstrated to undergo [4+2] cycloaddition with imines.²⁸⁷⁻²⁹⁰

We carried out a theoretical investigation on the reactivity of a new class of vinylketenes, the α -bromo-vinylketenes,²⁹¹ with imines. The aim of this theoretical work is to rationalize recent experimental results obtained on this subject by a synthetic group working at the Dipartimento di Chimica "G. Ciamician" of the University of Bologna. In this work successful synthetic routes toward α -bromo substituted 3-alkenyl-azetidin-2-ones and 4-alkyl-5,6-dihdropyridin-2-ones have been described (Scheme 10.1). The halo substituent^{265,292-298} dramatically affects the α -bromo-vinylketene's reactivity in the cycloaddition with an imine, promoting an unusual diene behaviour. Furthermore, a fine-tuning of the substituents on both vinylketene and imine moieties allows to selectively obtain [2+2] or [4+2] products. Moreover, these heterocycles are suitable for further elaborations, *via* the substitution of the halogen atom and the transformation of the double bond.²⁹⁹⁻³⁰⁵ The derivatives of 3-bromo-3-alkenyl-azetidin-2-ones and 3-bromo-4-alkyl-5,6-dihdropyridin-2-ones have shown

inhibition of ACAT enzyme and antagonism of $\alpha_v\beta_3$ integrin respectively, thus demonstrating the versatility of these building blocks in the synthesis of biologically active compounds.³⁰⁶⁻³⁰⁹

A new synthetic approach was developed to obtain four and six membered lactams *via* cycloaddition reactions between α -bromo vinylketenes and imines and a theoretical investigation on the mechanism and on the stereoselectivity of these reactions, providing also a rationalisation for the unprecedented behaviour of this class of vinylketenes.

In the following sections we shall describe in details the experimental results since they are essential to understand the subsequent theoretical work.



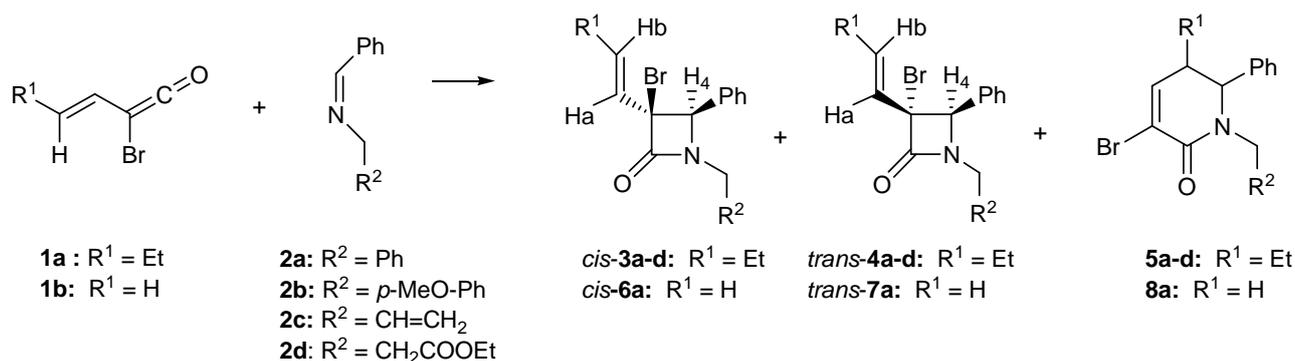
Scheme 10.1 Synthesis of 3-bromo-3-alkenyl-azetidin-2-ones and 3-bromo-4-alkyl-5,6-dihydropyridin-2-ones.

10.2 Experimental section

10.2.1 Reaction of linear α -bromo vinylketenes with imines.

Recently, we have investigated³¹⁰ the straightforward synthesis of 3-bromo-3-alkenyl-azetidin-2-ones *via* Staudinger reaction between α -bromo vinylketenes and an imine. The α -bromo vinylketenes **1a** and **1b** were prepared *in situ* starting from α -bromo hexenoyl chloride and α -bromo crotonyl chloride respectively in the presence of triethylamine.

The cycloaddition reaction of α -bromo vinylketenes **1a-b** with imines **2a-e** afforded, as major products, the *cis* β -lactams **3a-e** and **6a** but, unexpectedly, the formation of the six membered 5,6-dihydropyridin-2-ones **5a-e** and **8a** could not be avoided, even under a variety of experimental conditions (Scheme 10.1).



Scheme 10.2 Reaction of **1a** and **1b** with imines **2a-d**.

The reaction of vinylketenes with imines, that affords 3-alkenyl-azetidin-2-ones, was reported in the past by Bose and Manhas,³¹¹⁻³¹⁴ but no traces of the six membered lactams was observed by the authors. This result prompted us to investigate the effect of the halogen on the reactivity of the intermediate ketene.

The cycloaddition reaction was performed on **1a** and **1b** with the imines derived from benzaldehyde and benzylamine (**2a**), *p*-methoxy-benzylamine (**2b**), allylamine (**2c**), β-alanine ethyl ester (**2d**) and (*S*)-1-phenylethylamine (**2e**).

The α-bromo vinylketene **1a** and the imine **2a** were reacted under several different conditions in order to increase yield and selectivity in the formation of 3-Br-3-alkenyl-β-lactam. The detailed investigation of the reaction conditions showed that the best results in diastereoselectivity could be obtained in CH₂Cl₂ at reflux by adding the proper acyl chloride to a hot solution of imine **2a** and TEA. Following this procedure, the β-lactams **3a-d** and **4a-d** and **6a-7a** were obtained in good yield and high selectivity in favour of the *cis* isomer, accompanied by a significant amount of **5a-d** and **8a** (Table 10.1). Product distributions were determined by ¹H NMR integration of distinctive key signals and by quantitation of the individual isomers obtained after chromatographic separation on silica gel, eluting with benzene.

Table 10.1 Reaction of Vinylketenes **1a-b** with imines **2b-d**.

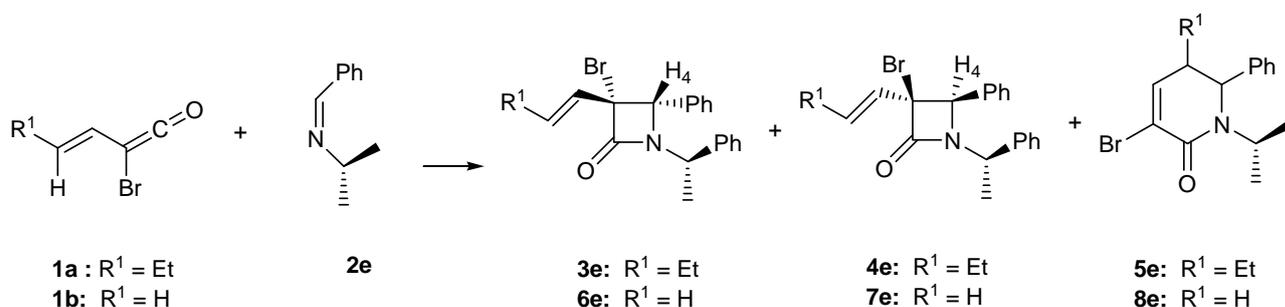
Entry	Ketene	Imine	3 + 4 (%) ^a	3/4 (%)	5 (%) ^a
1	1a	2a	57	95 : 5	22
1	1a	2b	60	80 : 20	27
2	1a	2c	50	80 : 20	23
3	1a	2d	55	93 : 7	14
4	1b	2a	30	90 : 10	15

^a) Reported yields refer to isolated products. A small amount of amide was observed in all the reactions.

Compounds **3b-d** and **4b-d** were obtained with exclusive E configuration of the double bond in the side chain, as shown by the coupling constant ($J = 15.6$ Hz). The *cis* configuration of **3** was established by NOE experiments using DPGFSE pulse sequence.

The isomer **3** exhibited a strong NOE between H₄ and the protons of the double bond on C₃, thus suggesting a *cis* relationship between these two moieties. The same experiment performed on the minor isomer **4** did not afford any NOE effect, thus suggesting a *trans* geometry between H₄ and the alkenyl group.

The Staudinger reaction carried out with α -bromo vinylketenes **1a-b** and the imine **2e**, obtained starting from benzaldehyde and (*S*)-1-phenylethylamine, gave the β -lactams in enantiomerically pure form (Scheme 10.3). Although four stereoisomers could possibly arise from this reaction, complete *cis* diastereoselectivity was observed and mixtures of *cis* β -lactams, together with a significant amount δ -lactams, were isolated both from the reaction of **1a** and **1b**.



Scheme 10.3 Reaction of **1a** and **1b** with imine **2e**.

The mixtures were easily separated by flash chromatography on silica gel eluting with cyclohexane/diethyl ether. The major isomer **3e** was isolated as a gum, while the minor isomer **4e** is a solid that was crystallized from methanol. The DPGFSE-NOE experiments, performed on **3e** and **4e**, indicated for both compounds a 3,4-*cis* configuration, since a strong NOE effect was recorded between H₄ and the double bond protons.

The E configuration of the side chain was determined from the double bond coupling constant ($J = 15.4$ Hz). Furthermore, the (1*S'*,3*S*,4*R*) absolute configuration of **4e** was established by X-ray diffraction (see Figure 1.1). On the basis of these considerations, the assignment of the (1*S'*, 3*R*,4*S*) absolute configuration to **3e** could be made. The comparison of the ¹H NMR data for **3e** and **4e** and the data for **6e** and **7e**, allowed us to find regularities when considering H_{1'} and CH_{3'} of the phenylethyl moiety. The signal of H_{1'}, indeed, occurs at 5.10 ppm and 5.09 ppm for (1*S'*, 3*R*,4*S*)-**3e** and for **6e** respectively, while it is observed at 4.33 ppm and 4.35 ppm for (1*S'*, 3*S*,4*R*)-**4e** and for **7e** respectively. In a similar way, the signal of CH_{3'}, occurs at

1.50 ppm and 1.52 ppm for (1*S'*, 3*R*,4*S*)-**3e** and for **6e** respectively, and at 1.95 ppm both for (1*S'*,3*S*,4*R*)-**4e** and **7e**.

On the basis of these observations, we attributed the (1*S'*,3*R*,4*S*) configuration to **6e** and the (1*S'*,3*S*,4*R*) configuration to **7e**.

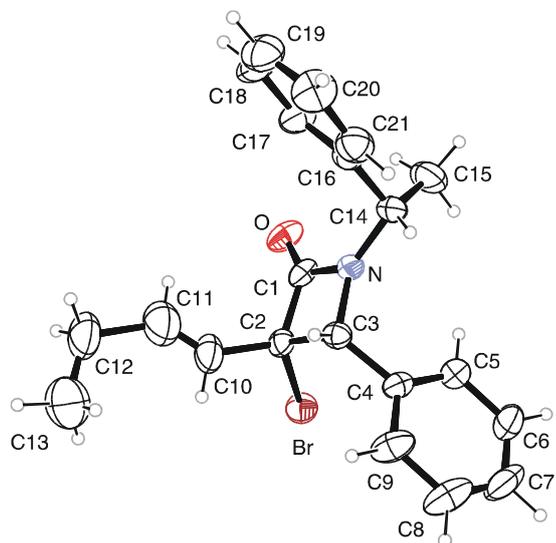
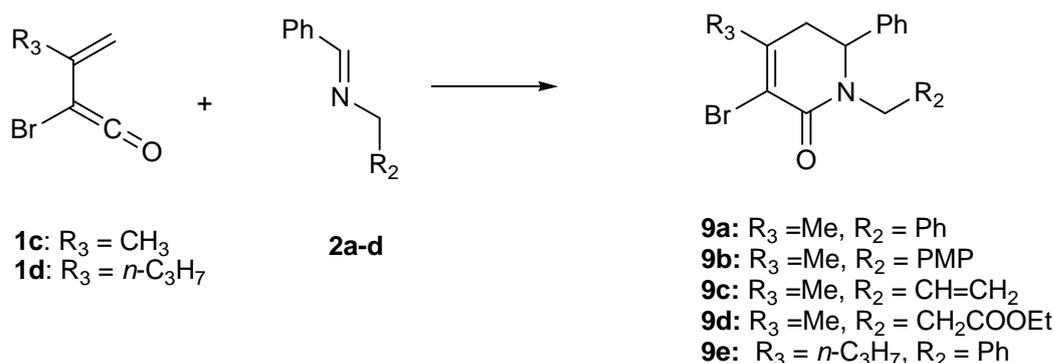


Figure 10.1 ORTEP diagram of **4e**.

10.2.2 Reaction of γ,γ -disubstituted- α -bromo vinylketenes with imines

Performing a modulation of the substituents of α -bromo vinylketenes, we discovered that their diene-behaviour could be enhanced with the introduction of a branch in β position. Actually, exclusively the [4+2] pathway is followed in the reactions of **1c-d** with imines **2a-d**, and the dihydropyridinones **9a-e** could be obtained in high yields, as previously reported.²⁹⁹ Indeed, α -bromo vinylketene **1d** is prepared with excellent regioselectivity treating 2-bromo-3-methyl-2-hexenoyl chloride with 2 eq. of TEA, despite the possibility of deprotonation of the methylene group to give an isomeric vinylketene.³¹⁵

No trace of the β -lactam was detected in the HPLC analysis and in the ¹H NMR spectra of the crude reaction (Scheme 10.4, Table 10.2).



Scheme 10.4 Reaction of γ,γ -disubstituted- α -bromo vinylketenes **1c-d** with imines **2a-d**.

The detailed investigation of the reaction conditions showed that the best results could be obtained when **1** and **2** were refluxed in CH₂Cl₂. Under these reaction conditions, **9a-e** were obtained in excellent yields (92-96%). Only the cycloaddition of **1c** and β-alanine derivative **2d** gave a lower yield, **9d** being isolated in 64% yield (Table 10.2, entry 4).

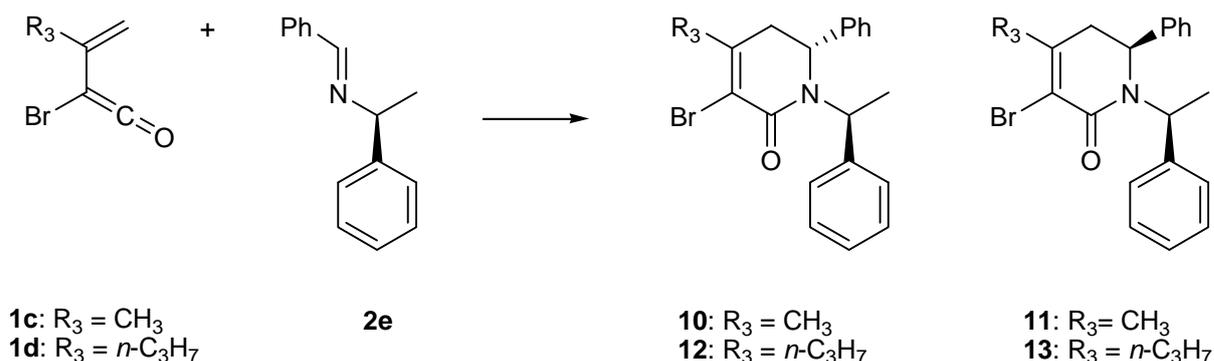
Table 10.2 Formation of 3-bromo-4-alkyl-5,6-dihydropyridin-2-one **3** via ketene-imine cyclisation.

Entry ^[a]	Ketene	Imine	Product	Yield (%) ^[b]
1	1c	2a	9a	98
2	1c	2b	9b	92
3	1c	2c	9c	96
4	1c	2d	9d	64
5	1d	2e	9e	55

^[a]Reactions were performed in CH₂Cl₂.

^[b]Yields correspond to the compounds purified by flash chromatography on silica gel.

Good yields and moderate diastereoselectivities were observed in the reactions of **1c** and **1d** with the chiral imine **2e** (98% yield and 62/38 d.r. for the reaction of **1c**, 55% yield and 68/32 d.r. for **1d**). The dihydropyridinones **10/11** and **12/13** were easily separated and fully characterised by NMR spectroscopy and LC-MS analysis. The (6*R*) absolute configuration of the newly created stereogenic centre in **12** was established by X-ray diffraction²⁹⁹, and the complete regularity of the ¹H NMR chemical shifts allowed us to confidently assign the stereochemistry to the other compounds of the class.

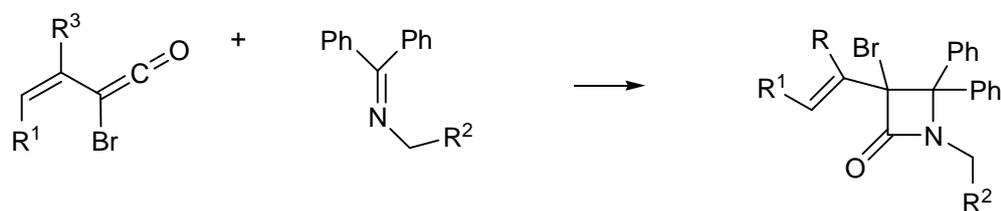


Scheme 10.5 Reaction of **1c** and **1d** with the chiral Schiff base **2e**.

10.2.3 Reaction of α-bromo vinylketenes with ketimines

The results reported above suggest that linear α-bromo vinylketenes **1a** and **1b** react with imines preferentially giving β-lactams, while γ,γ-disubstituted-α-bromo vinylketenes **1c** and **1d** afford exclusively six membered rings. In order to complete our investigation, we studied

the influence of the imine substitution on the reactivity, treating α -bromo vinylketenes **1a-c** with the ketimines **2d** and **2f**, derived from benzophenone (Scheme 10.6).



1a: R³ = H, R¹ = Et
1b: R³ = H, R¹ = H
1c: R³ = CH₃, R¹ = H

2d: R² = CH₂COOEt
2f: R² = COOEt

14d: R³ = H, R¹ = Et, R² = CH₂COOEt
14f: R³ = H, R¹ = Et, R² = COOEt
15d: R³ = H, R¹ = H, R² = CH₂COOEt
15f: R³ = H, R¹ = H, R² = COOEt
16d: R³ = CH₃, R¹ = H, R² = CH₂COOEt
16f: R³ = CH₃, R¹ = H, R² = COOEt

Scheme 10.6 Reaction of α -bromo vinylketenes with ketimines **2d** and **2f**.

The experimental results demonstrate that the substitution of the imine hydrogen with a phenyl group dramatically suppressed the strong preference of **1c** for [4+2] reaction, selectively leading to 3-bromo-3-alkenyl-azetidin-2-ones (Table 10.3).

Table 10.3 Formation of 3-bromo-3-alkenyl-azetidin-2-ones via ketene-ketimine cyclisation.

Entry ^[a]	Ketene	Imine	Product	Yield (%) ^[b]
1	1a	2f	14f	97
2	1a	2d	14d	84
3	1b	2f	15f	93
4	1b	2d	15d	90
5	1c	2f	16f	80
6	1c	2d	16d	87

^[a]Reactions were performed in CH₂Cl₂. ^[b]Yields correspond to the compounds purified by flash chromatography on silica gel.

Therefore, in the cycloadditions between α -bromo vinylketenes and imines, variations of the substituents can be exploited to drive the reaction toward the [2+2] or the [4+2] path, in order to selectively obtain β -lactams or δ -lactams.

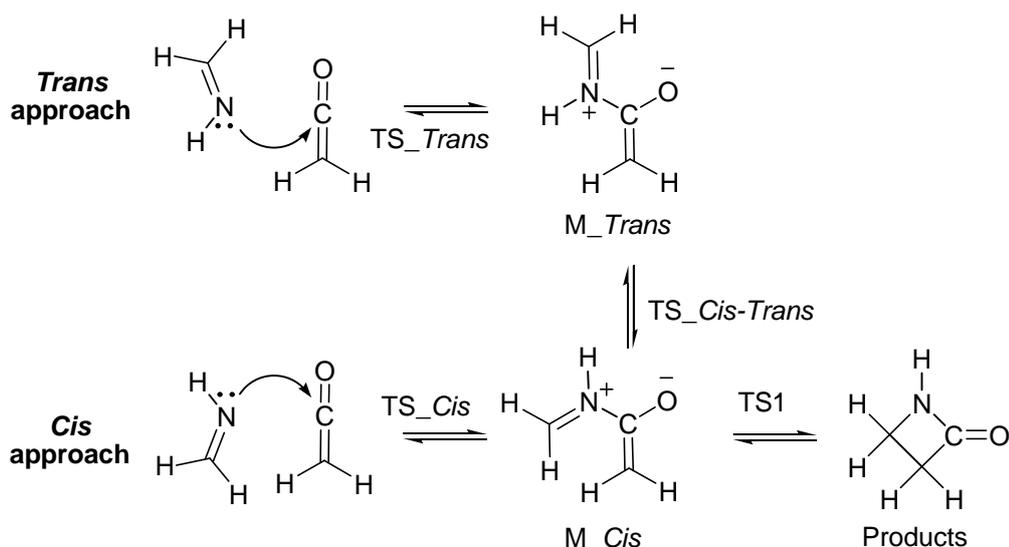
10.3 Computational Section

10.3.1 Choice of the model system and description of the computational method

To provide a rationale for the observed periselectivity, geometry optimisations of selected molecules in the presence of the solvent have been carried out. All computations have been

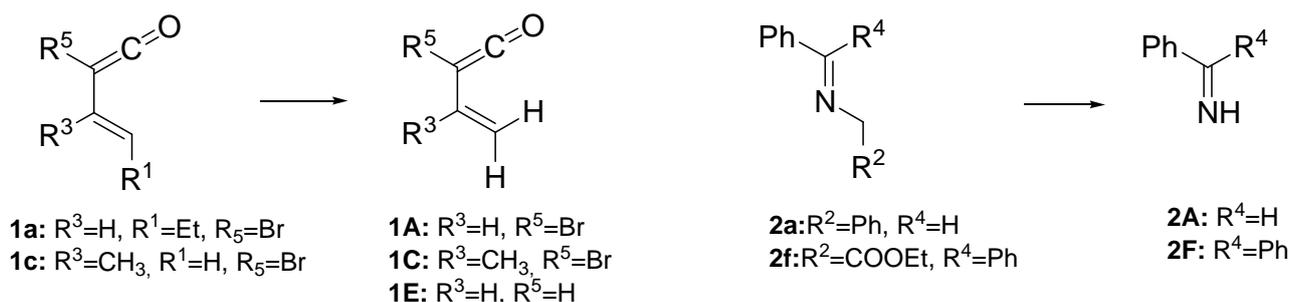
performed at the DFT/B3LYP/DZVP^{75,151,223} level, using Gaussian 03 package³¹⁶; the solvent (dichloromethane) has been simulated by means of SCRF-CPCM^{145,152,317,318} method. To validate this choice, the same level of calculation was used to study the addition of an unsubstituted imine to an unsubstituted ketene, in order to compare our results to those obtained by Venturini at CASSCF-CASPT2 level.²⁷¹ The results obtained with the DFT method in gas phase are comparable to the CASSCF ones, both in terms of energy and geometry (results not reported here). Besides, the inclusion of the solvent (dichloromethane) effects *via* the SCRF method gave a quantitatively different pathway, according to Venturini.²⁷¹ The solvent helps stabilizing the structures with a charge separation (see Scheme 10.7), and the transition state TS_Cis_Trans connecting the minima M_Cis and M_Trans (and thus Cis and Trans pathways) could be easily located. Moreover, no influence of unpaired spin states have been found; single point calculation on all the found critical points, with both restricted and unrestricted methods (B3LYP and UB3LYP), have proved to give identical results and stable wavefunctions. For these reasons, a restricted DFT formulation was employed. The use of the hybrid B3LYP functional accounts for correlation effects with a lower computational demand with respect to CASSCF-CASPT2 or MP2 approaches. Furthermore, the inclusion of the solvent is of paramount importance to correctly reproduce the Potential Energy Surface (TS) related to the reactions under examination, due to the presence of critical points with consistent charge separation.

The geometries were optimised using redundant internal coordinates²⁰⁸ and the nature of all the found critical points has been ascertained by means of frequency calculations, to check the presence of the correct number of negative eigenvalues. All optimisations were carried out in the presence of the solvent (see above) and the reported energy values include all the contribution due to solvent (both electrostatic and not).



Scheme 10.7 Pathways for the reaction of ketene and imine.

In order to decrease computational time, calculations have been carried out on simplified models of molecules, confidently assuming that these approximations do not affect the ability to reproduce experimental results (Scheme 10.8). The vinylketenes have been approximated with **1A** (α -bromo-substituted), **1C** (both α -bromo- β -methyl-substituted) and **1E** (unsubstituted). Two molecules have been used to describe imines and ketimines (**2A** and **2F**, respectively). The nature of N-substituent has been considered irrelevant since experimentally it does not affect the reactions' outcome, therefore, it has been approximated with an H atom in all the calculations.



Scheme 10.8 Model system used in the computational study.

10.4 Results and discussion

The [2+2] reaction between a ketene and an imine leading to β -lactams is unanimously considered a stepwise process; the first step is the nucleophilic attack of the imine nitrogen to the *sp*-carbon of the ketene, leading to a zwitterionic intermediate, that then undergoes a conrotatory electrocyclic ring-closure to give the final product. Theoretical studies have

provided an insight into the origins of the stereoselectivity,^{274,276,279} which is without question the most intriguing problem; in particular, a detailed study recently reported by Xu and co-workers²⁶⁹ clarified the role of the substituents in driving the stereoselectivity.

The conformations of the vinylketenes have been computationally studied;^{285,286} the equilibrium from *s*-E-conformation to *s*-Z-conformation is of particular interest since it may be crucial for the discrimination between the [2+2] and the [4+2] pathway. The *s*-E-conformation is more stable than *s*-Z-conformation for all the vinylketenes (Table 10.4); however, the energies of the conformational TS are markedly lower for **1A** and **1C** than for **1E**, thus suggesting a role of the bromo substituent in the stabilisation of the unfavoured conformation. On the basis of these results it is clear that all the examined ketenes adopt a *s*-E-conformation at the equilibrium, but in RT conditions the rotation around the single bond is possible and is faster for **1C**.

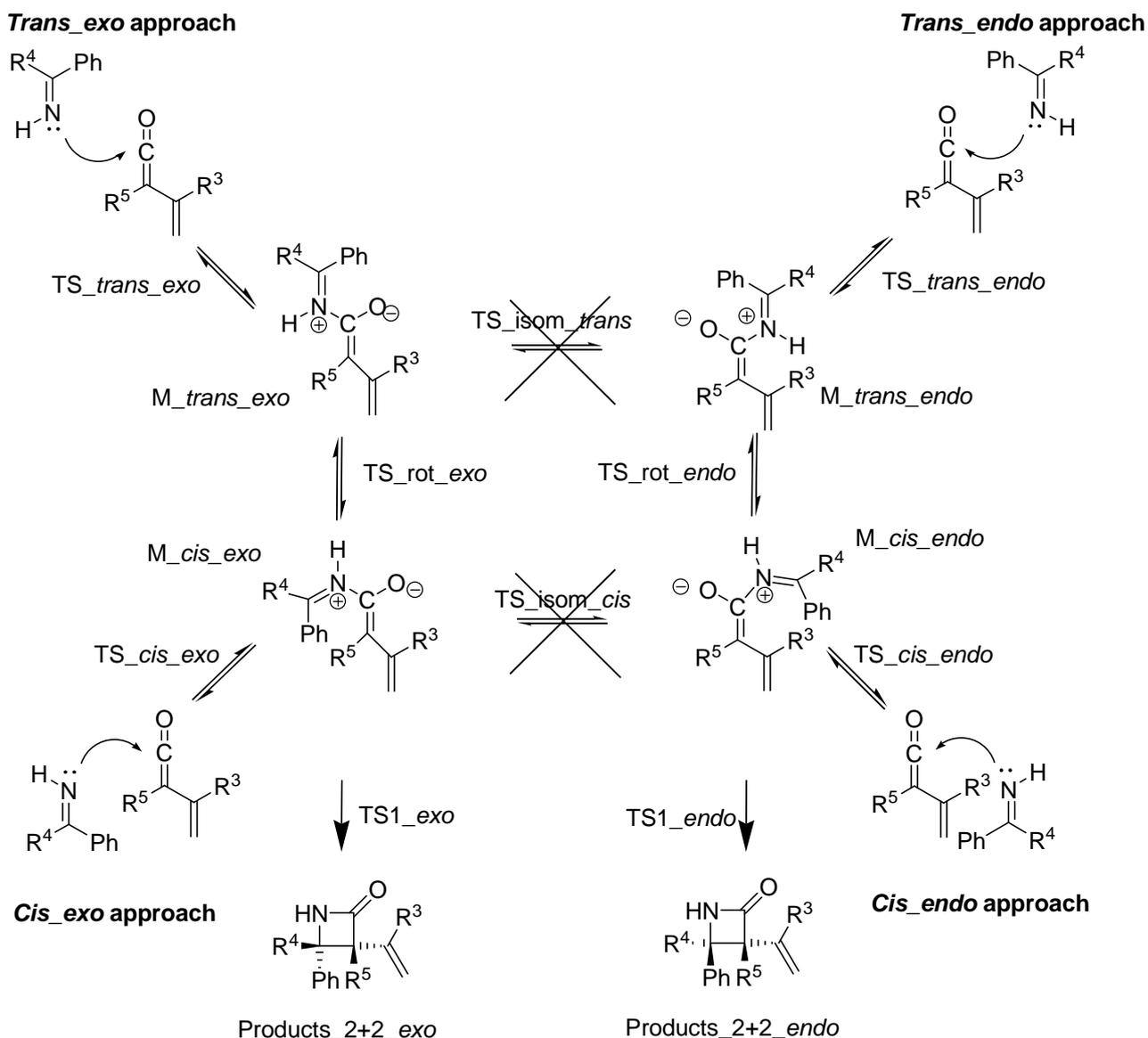
Table 10.4 Energies of the conformational equilibrium of the ketenes.

molecule	E (kcal mol ⁻¹) ^a	Dihedral angle ϕ (°)
1A <i>s</i> -Z	3,09	0
1A_TS	5,81	108
1A <i>s</i> -E	0,00	180
1C <i>s</i> -Z	2,09	-43,39
1C_TS	5,07	-112,39
1C <i>s</i> -E	0,00	180
1E <i>s</i> -Z	2,18	0
1E_TS	7,01	98,24
1E <i>s</i> -E	0,00	180

The energies are referred to the *s*-E conformer

The possibility of an isomerisation from an (E) imine to a (Z) imine has been taken into account,²⁶⁹ and finally the model imine **2A** was considered to be in the E configuration on the basis of both experimental and computational evidences. Our data suggest a barrier of about 30 kcal mol⁻¹ for the E/Z isomerisation of **2A**.

As previously reported in the literature^{271,277}, there are different approaches leading to β -lactams: two *endo* approaches, called *cis_endo* and *trans_endo*, and two *exo* approaches, called *cis_exo* and *trans_exo* (Scheme 10.9). In the following discussion, the *endo* suffix will be used for the imine approach on the vinyl group side of the ketene, while the *exo* suffix will be used for the imine approach on the R⁵ group side of the ketene. As showed before the *cis* and *trans* approaches in both *exo* and *endo* pathways are connected by a rotational transition state between the minima M_*trans* and M_*cis*.



Scheme 10.9 General scheme for the [2+2] pathways.

The energy of the critical points found on the PES for the [2+2] reaction between ketenes **1A**, **1C** and **1E** and imine **2A** and ketimine **2F** are reported in table 5. In the four fully examined pathways (**1E+2A** *endo/exo* and **1A+2A** *endo/exo*), the rate determining step coincides with the electrocyclic ring closure of the M_cis minimum to give the β -lactam. The comparison of the barriers associated to TS1 in the *endo* and *exo* approaches (TS1_endo and TS1_exo) accounts for the preference of the system to give different β -lactam's diastereoisomers. A qualitative relation was observed between the $\Delta E^{\text{TS1}}_{(\text{exo-endo})}$ values (Table 5) and the corresponding experimental ratio of products. In the case of the vinylketene **1E** ($R^3=H$) the greatest barrier (see Table 10.5 and Figure 10.2) is associated to the *endo* approach (TS1_endo = 15.75 kcal mol⁻¹), while the rate determining step of the *exo* approach was found to be 5.60 kcal mol⁻¹ lower in energy (TS1_exo = 10.16 kcal mol⁻¹). This result is in good

agreement with the observed experimental data³¹³, since *exo* approach leads to *cis* β -lactams. The opposite behaviour was observed in the reaction between α -bromo vinylketene **1A** and aldimine **2A**, since the [2+2] *endo* pathway, leading to the (Br-Ph)-*cis* β -lactam, is favoured over the *exo* of 4.46 kcal mol⁻¹ (see Table 10.5 and experimental products ratio in Table 10.1). On the basis of the reported results, the analysis of the [2+2] pathway preference was extended to other reacting couples (**1C+2A**, **1E+2F**, **1A+2F**, **1C+2F**), performing the calculations only on the most relevant critical point (TS1), confidently assuming the qualitative similarity of all the PES under examination.

Table 10.5 Energetics of the [2+2] pathways.

	1E+2A	1A+2A	1C+2A	1E+2F	1A+2F	1C+2F
	Energy	Energy	Energy	Energy	Energy	Energy
	(kcal mol ⁻¹) ^a	(kcal mol ⁻¹) ^b	(kcal mol ⁻¹) ^b	(kcal mol ⁻¹) ^a	(kcal mol ⁻¹) ^b	(kcal mol ⁻¹) ^b
TS_trans_endo	## ^c	## ^c				
M_trans_endo	-6.05	-12.62				
TS_rot_endo	-4.05	-11.13				
TS_cis_endo	3.95	2.01				
M_cis_endo	-4.19	-11.26				
TS1_endo	15.75	6.72	10.83	11.01	2.95	9.56
Products_endo	-28.05	-29.72				
TS_trans_exo	3.26	3.95				
M_trans_exo	-9.64	-13.21				
TS_Rot_exo	-4.94	-10.94				
TS_cis_exo	## ^c	## ^c				
M_cis_exo	-5.40	-10.61				
TS1_exo	10.16	11.18	13.80	5.45	6.10	8.11
Products_exo	-26.89	-29.86				
$\Delta E_{(exo-endo)}^{TS1}$	-5.60	+4.46	+2.97	-5.57	3.14	-1.45

a) Energies are referred to the sum of the energies of the reactants (1E and 2A)

b) Energies are referred to the sum of the energies of the reactants (1A and 2A)

c) The geometry obtained after the optimization was found to be a saddle of order higher than 1 and so was considered irrelevant for a chemical point of view.

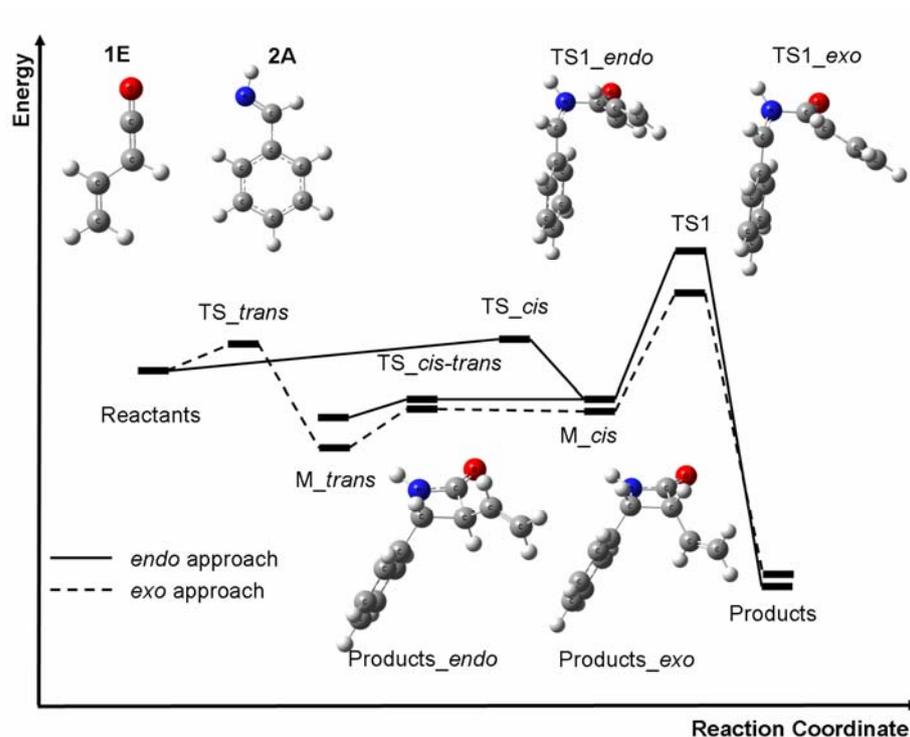


Figure 10.2 1E + 2A [2+2] reaction profile.

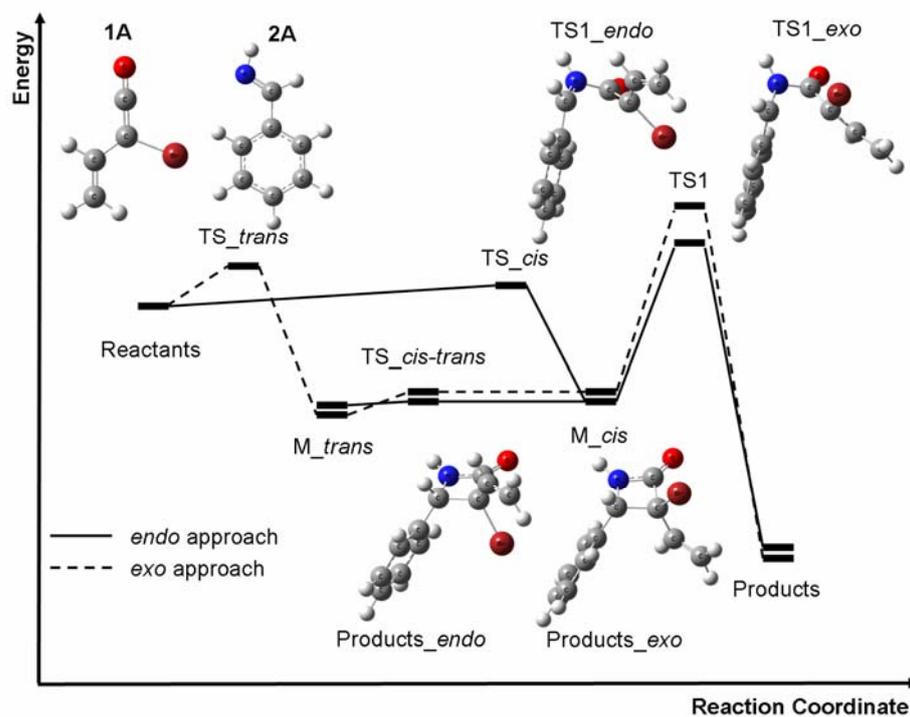
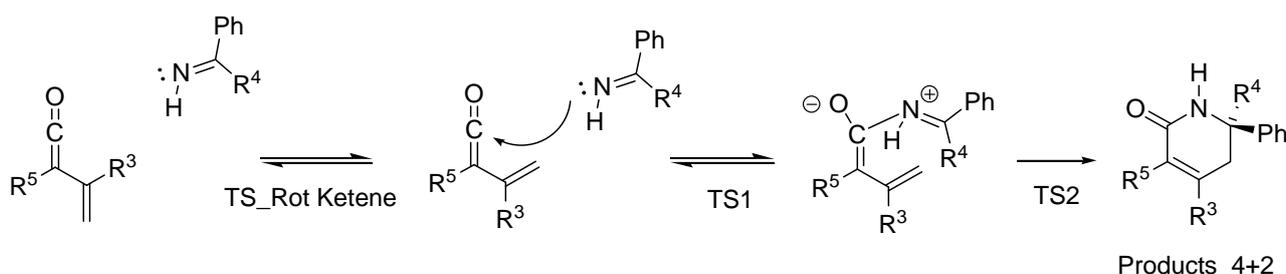


Figure 10.3 1A + 2A [2+2] reaction's profile.

We examined the isomerisation of the zwitterionic intermediates (see Scheme 10.9) from the more stable (E) imine geometry to the (Z); this pathway, if active, would connect the *endo* and *exo* pathways and result in an opposite diastereoselectivity. Even if some authors reported the occurrence of this mechanism in the case of benzaldehyde imines³¹⁹, no connection in between *endo* and *exo* pathways was found, since the location of the Transition States

connecting M_{cis_endo} and M_{cis_exo} (TS_isom_cis) or M_{trans_endo} and M_{trans_exo} (TS_isom_trans) failed. Thus, we excluded the possibility of this isomerisation process in the case under examination.

To understand the peculiar diene-behaviour of α -bromo vinylketenes in depth, [4+2] pathways have been afterwards studied. First, on the basis of computational evidences, we excluded that the zwitterionic intermediates (M_{cis_endo} and M_{cis_exo}) of [2+2] pathway could give the six-membered product by way of rotation of vinyl group and subsequent ring-closure. In fact the Minima and the Transition States related to this mechanism were found in none of the examined PES. Instead, an independent two-step mechanism starting from the *s*-Z-vinylketene has been discovered in [4+2] cycloaddition; indeed, the capability of bromo substituent to stabilise *s*-Z-conformation of α -bromo vinylketenes enable a new reaction's pathway (Scheme 10.10).



Scheme 10.10 General scheme for the [4+2] pathway.

Table 10.6 Energetics of the [4+2] pathways.

	1E+2A	1A+2A	1C+2A	1C+2F
Reactants	0.00	0.00	0.00	0.00
TS_rot_ketene	7.01	5.81	5.07	5.07
M0	2.18	3.09	2.09	2.09
TS1	10.63	8.26	7.27	8.68
M1	-1.30	-8.08	-8.11	-2.95
TS2	7.66	2.20	2.78	3.43
M2	-49.68	-53.93	-54.80	-46.76
TS3	-48.05	-52.02	-52.93	-43.43
M3	-51.51	-55.78	-56.72	-47.76

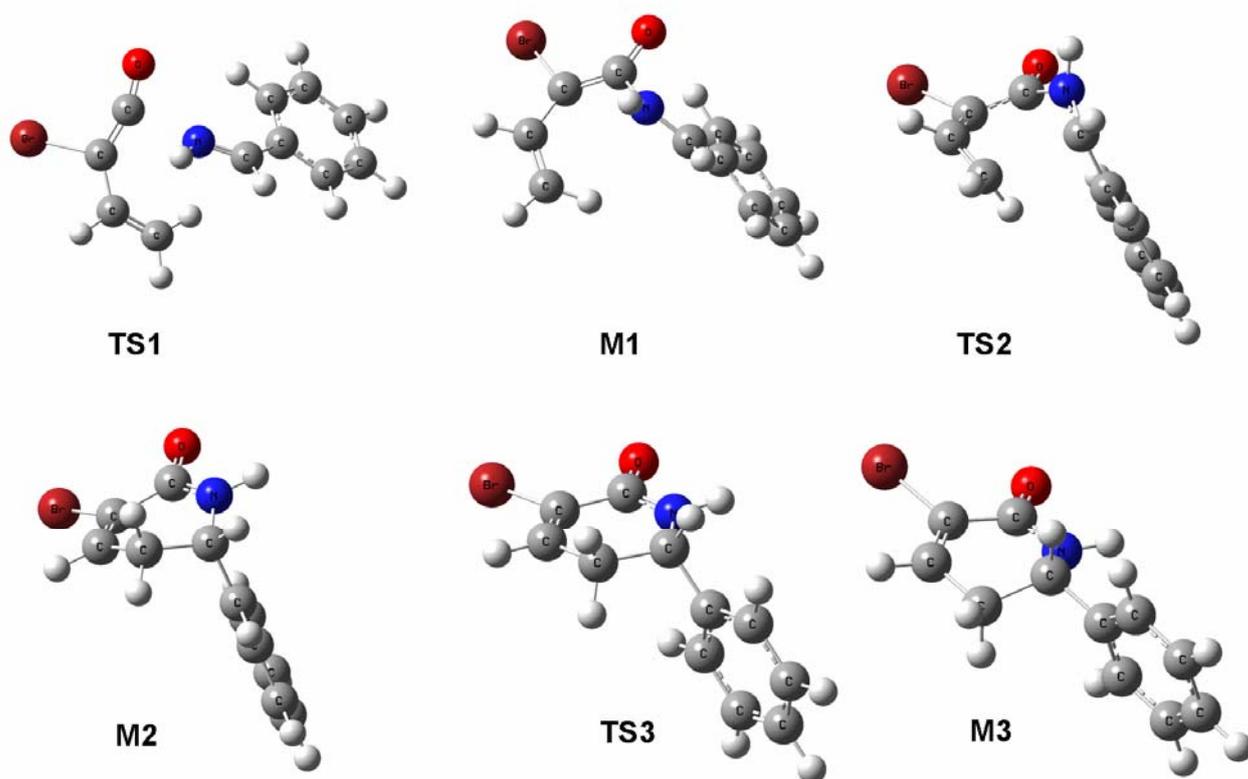


Figure 10.4 Geometries for [4+2] pathway (**1E**+**2A**)

The nucleophilic attack of the imine nitrogen to the *sp*-carbon of the α -bromo vinylketene in the *s*-*Z*-conformation is the first step of the reaction and it is also associated with the highest barrier (TS1);²⁸⁵ the ring closure that follows from M1, affords through the transition state TS2 the final six-membered product (M2) with the phenyl group in axial position (M3); a low barrier isomerisation (TS3) gives the stable conformer with the phenyl group in equatorial position. A “ball & stick” representation of the [4+2] pathway for the **1A**+**2A** reaction is reported in Figure 4. In Figures 5-8, the energy diagrams for the [4+2] reaction of **2A** with **1E**, **1A** and **1C** are reported, compared with the highest barriers of the competitive [2+2] pathways. It is worthwhile to note that the diastereoselectivity predicted by the proposed [4+2] mechanism is consistent with the observed stereochemistry as previously reported.²⁹⁹

In the cycloaddition between **1E** and **2A**, the energies of [4+2] TS1 and [2+2] TS1_{exo} are similar; the exclusive formation of the β -lactam may be ascribed to the more favoured attack of the imine (TS_{trans}_{exo} = 3.26 kcal mol⁻¹) on the *s*-*E*-vinylketene upon its isomerisation to give the *s*-*Z*-vinylketene (TS_{rot}_{ketene} = 7.01 kcal mol⁻¹). In the reaction between **1A** and **2A**, the computed barriers for [2+2] (TS1_{exo} = 11.18 kcal mol⁻¹ and TS1_{endo} = 6.72 kcal mol⁻¹) and [4+2] (TS1 = 8.26 kcal mol⁻¹) pathways are comparable; the experimental products ratio (Table 10.1) is in good qualitative agreement with the order of magnitude of the

corresponding rate-determining barriers. Concerning the cycloaddition of **1C** with **2A**, the preference for the [4+2] product is fully explained by the difference between TS1 (7.27 kcal mol⁻¹) and the higher barriers TS1_*exo* (13.80 kcal mol⁻¹) and TS1_*endo* (10.83 kcal mol⁻¹). Generally, the [4+2] pathway can be followed only by those vinylketenes that can easily populate the *s-Z*-conformation and it is unapproachable for other vinylketenes; in this sense, the unusual behaviour of the α -bromo vinylketenes in respect to unsubstituted vinylketenes may be rationalised.

However, when ketimine **2F** (R⁴=Ph, Scheme 10.8) is involved, TS1 energy increases (8.68 kcal mol⁻¹) and the [4+2] pathway becomes therefore unfavoured compared to [2+2] pathway (TS1_*exo* = 8.11 kcal mol⁻¹ and TS1_*endo* = 9.56 kcal mol⁻¹, Table 10.5), as represented in Figure 10.7.

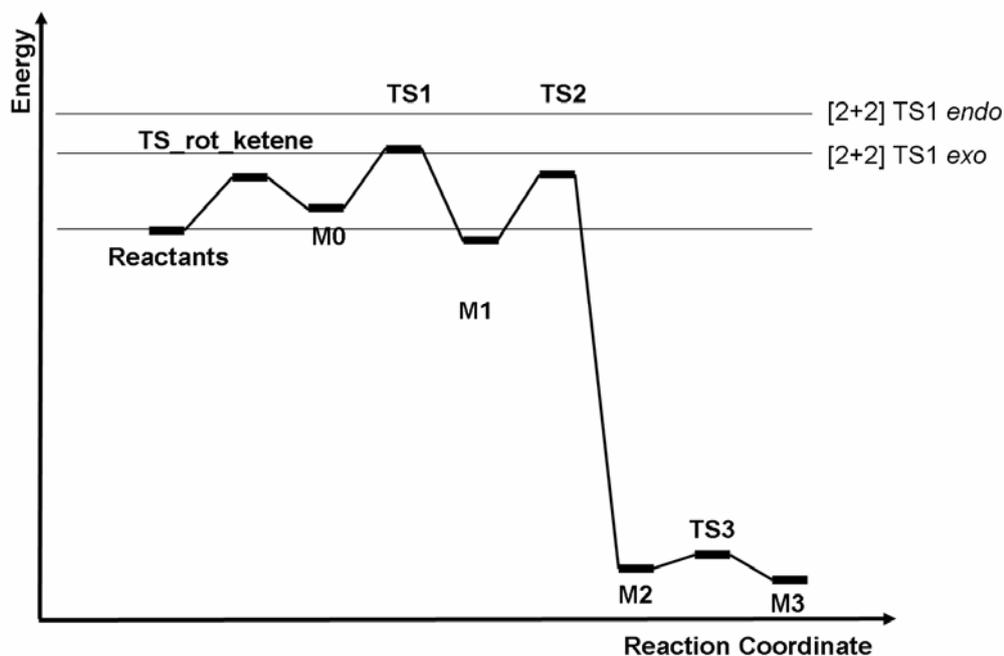


Figure 10.5 1E+2A [4+2] reaction profile.

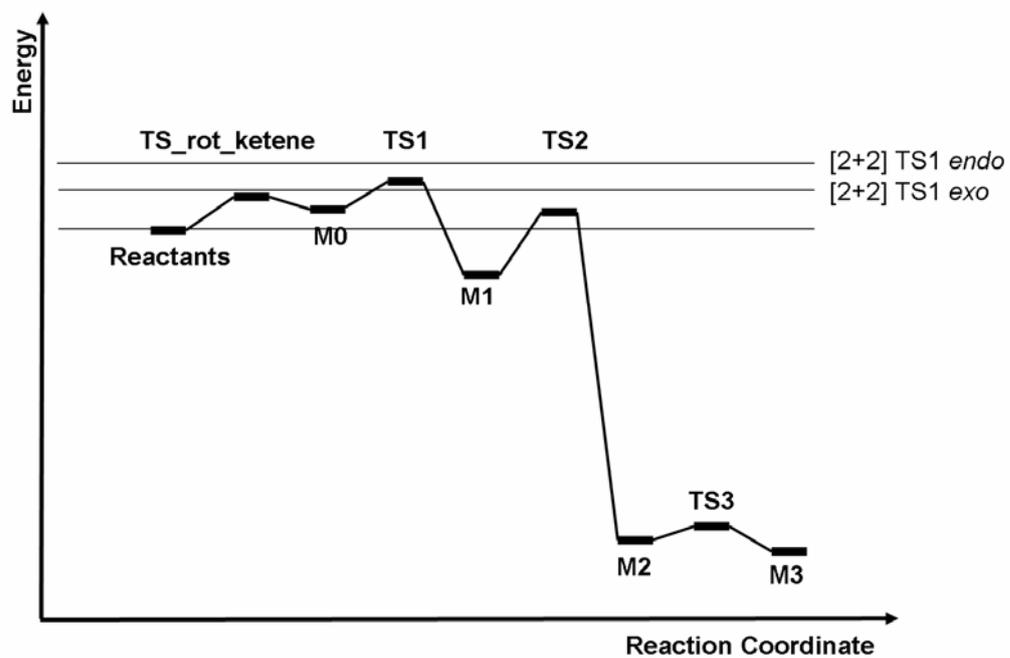


Figure 10.6 $1A+2A$ [4+2] reaction profile.

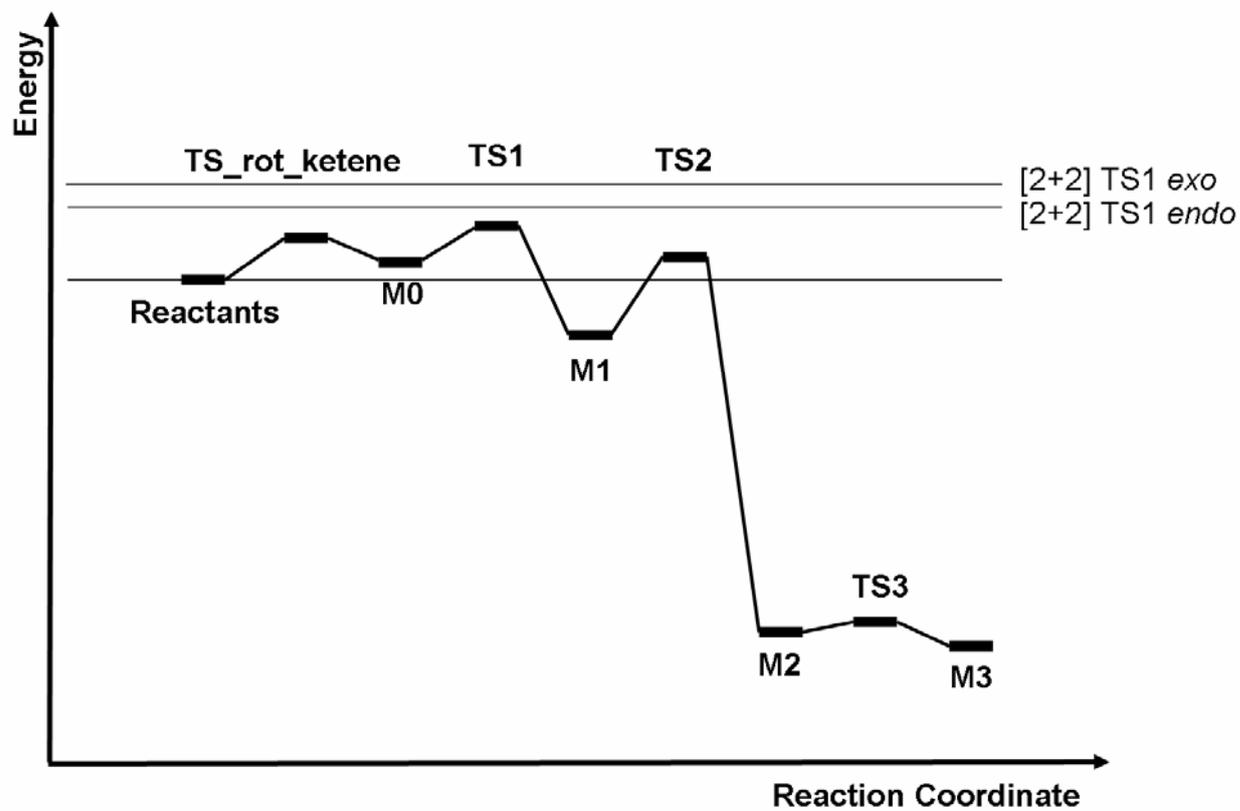


Figure 10.7 $1C+2A$ [4+2] reaction profile.

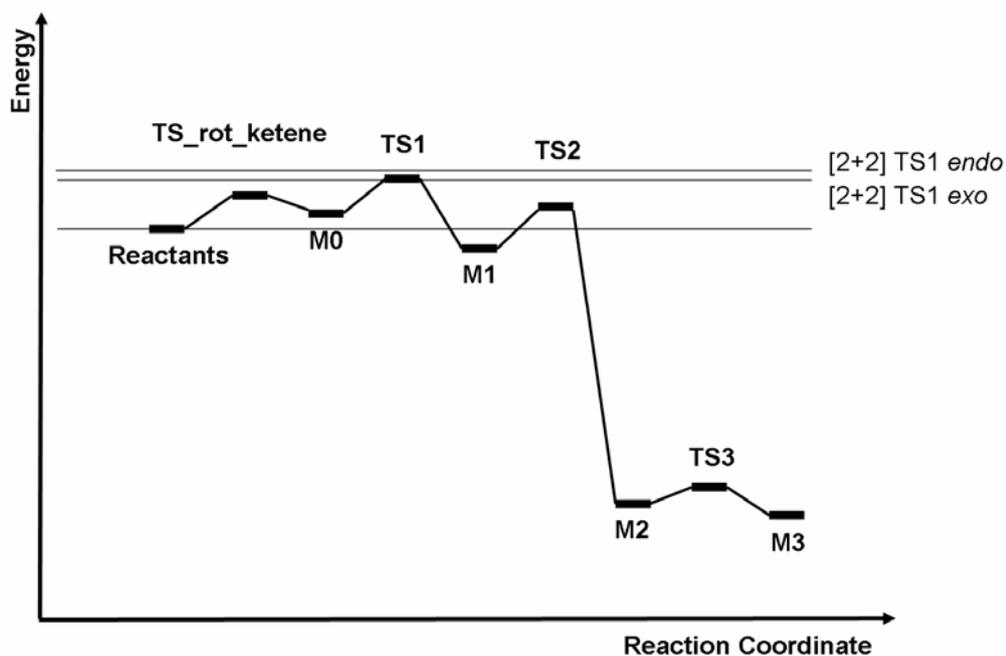


Figure 10.8 1C+2F [4+2] reaction profile.

10.5 Conclusions

In summary, we have documented the unprecedented behaviour of α -bromo vinylketenes in cycloadditions with imines; the singularity of this class of vinylketenes relies on the fact that they can act as dienophiles in [2+2] reactions, but they can also display an unusual diene reactivity in [4+2] reactions. Thus, α -bromo vinylketenes are versatile building blocks in the synthesis of highly functionalised heterocycles, since α -bromo substituted 3-alkenyl-azetidin-2-ones and 4-alkyl-5,6-dihdropyridin-2-ones can be easily obtained from [2+2] and [4+2] cycloadditions with imines respectively. Interestingly, the reactivity of α -bromo vinylketenes can be modulated *via* a fine tuning of the substituents; the introduction of a methyl group in β -position, for example, completely suppress the dienophile reactivity and allow to obtain the [4+2] cycloaddition products exclusively. To switch again to the [2+2] cycloaddition, it is enough to employ a more hindered ketimine instead of an imine: both β -methyl substituted and unsubstituted α -bromo vinylketenes give azetidinones when reacted with ketimines.

The theoretical investigation carried out in the present thesis in order to rationalise these observations has provided interesting computational results that are in good qualitative agreement with the experimental ones.

Two pathways have been considered for the [2+2] reaction with an imine i.e. the *endo* and the *exo* paths. The former (*endo*) was found to be favoured for α -bromo vinylketenes, while the

latter (*exo*) was preferred in the case of unsubstituted vinylketenes. The bromo substituent is crucial for the discrimination between [2+2] and [4+2] mechanisms, since [4+2] cycloaddition occurs starting from a vinylketene in the *s-Z*-conformation. For the unsubstituted vinylketene, the barrier from the *s-E*-conformation to the *s-Z*-conformation is too high and, therefore, the favourite pathways are the [2+2] ones. For α -bromo vinylketenes, and especially for the β -methyl substituted compounds, this barrier can be overcome. In this way the [4+2] pathway becomes accessible and, in the latter case, preferred. In the reactions involving ketimines, the presence of another phenyl group especially affects the formation of the zwitterionic intermediate in the [4+2] pathway, in such a way that the [2+2] pathways become favoured with all α -bromo vinylketenes.

NOTE:

The experimental section was carried out by the group of Professor G. Cardillo; Professor A. Bottoni must be acknowledged for the supervision of the computational section. The collaboration was mainly encouraged by the work of Dr. F. Benfatti.

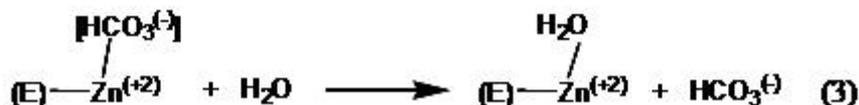
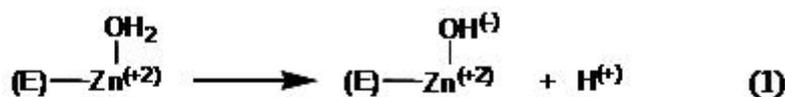
In this section was fully reported the experimental details of the project because the data, not been yet published, are of fundamental importance to understand the computational experiments and the obtained results.

11 New Computational Evidence for the Catalytic Mechanism of Carbonic Anhydrase

11.1 Introduction

Human carbonic anhydrase II (HCA II) is a zinc-based metalloenzyme that catalyzes the reversible hydration of CO₂ to bicarbonate. This enzyme is a single polypeptide chain of 260 aminoacids that catalyzes the reaction with rates up to 10⁷ times higher than in the uncatalyzed case.³²⁰⁻³²² Thus the reaction has the features of diffusion-controlled processes. As shown by X-ray diffraction,³²³⁻³²⁶ the active site is formed by a zinc cation Zn⁽⁺²⁾ in a tetrahedral coordination geometry. The metal is bound to three rigid imidazole groups belonging to three histidine residues (His94, His96 and His119) and a water molecule. One region of the active site (the so-called hydrophilic pocket) contains several water molecules and the His64 residue. This residue is thought to behave as an intra-molecular proton acceptor in the transfer of a proton from the zinc-bound water to an external buffer. Another region (hydrophobic pocket) is characterized by the presence of binding sites involved in the CO₂ transport and the so-called “deep water”. In the approaching process to the zinc ion, CO₂ probably displaces this water molecule, which is about 3.2 Å away from the metal.

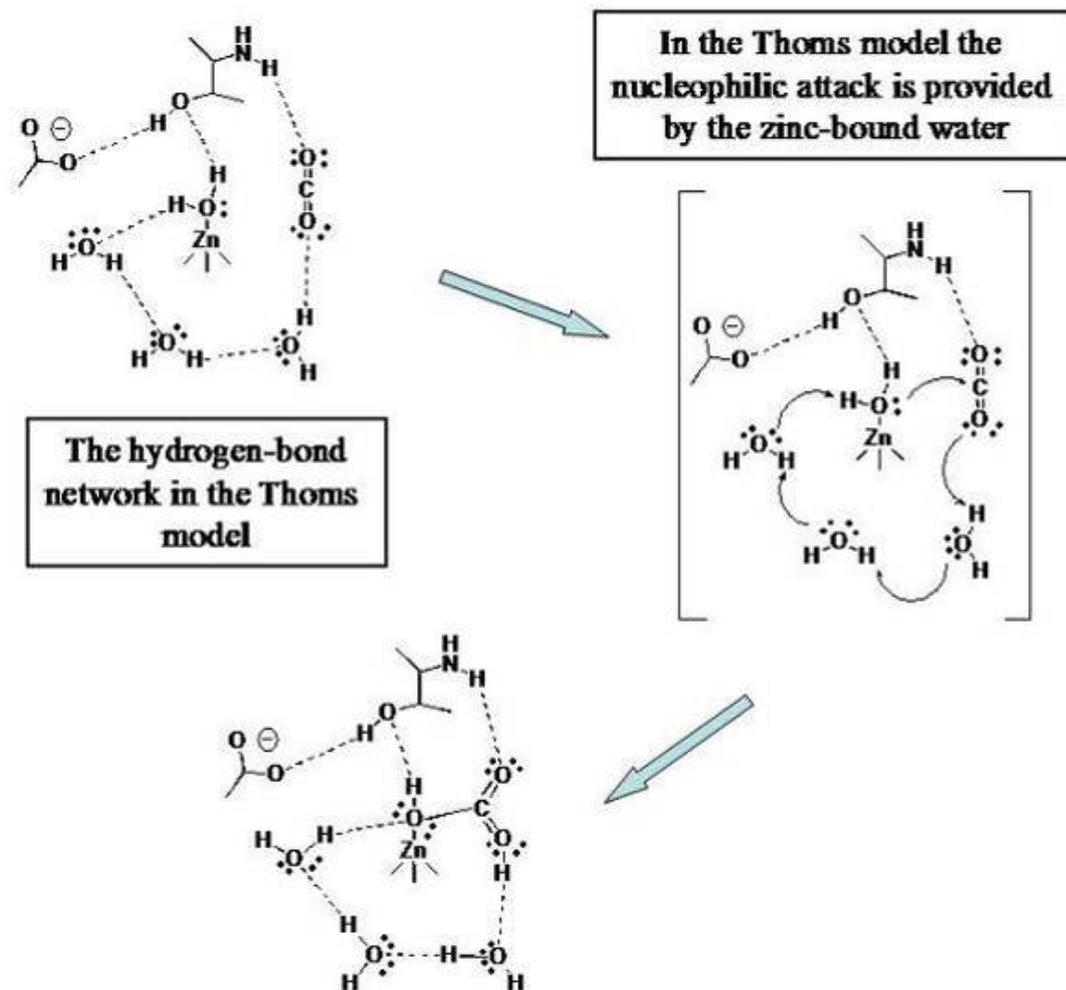
The almost universally accepted mechanism of HCAII (see Scheme 11.2) is based on a large number of experimental³²⁰⁻³³³ and theoretical^{191,334-350} investigations carried out during the last three decades. It consists of three main steps. The first step involves the proton release from the Zn-bound water to form a Zn-bound hydroxide (see step 1 of Scheme 11.1 where E indicates the enzyme). In the second step the zinc-bound hydroxide undergoes a nucleophilic attack on the CO₂ carbon to form bicarbonate (see step 2 of Scheme 11.1). In step 3 an external water molecule replaces the zinc-bound bicarbonate, which is expelled from the metal coordination sphere (see step 3 of Scheme 11.1) and closes the catalytic cycle.



Scheme 11.1 The fundamental equations of the HCA II enzyme.

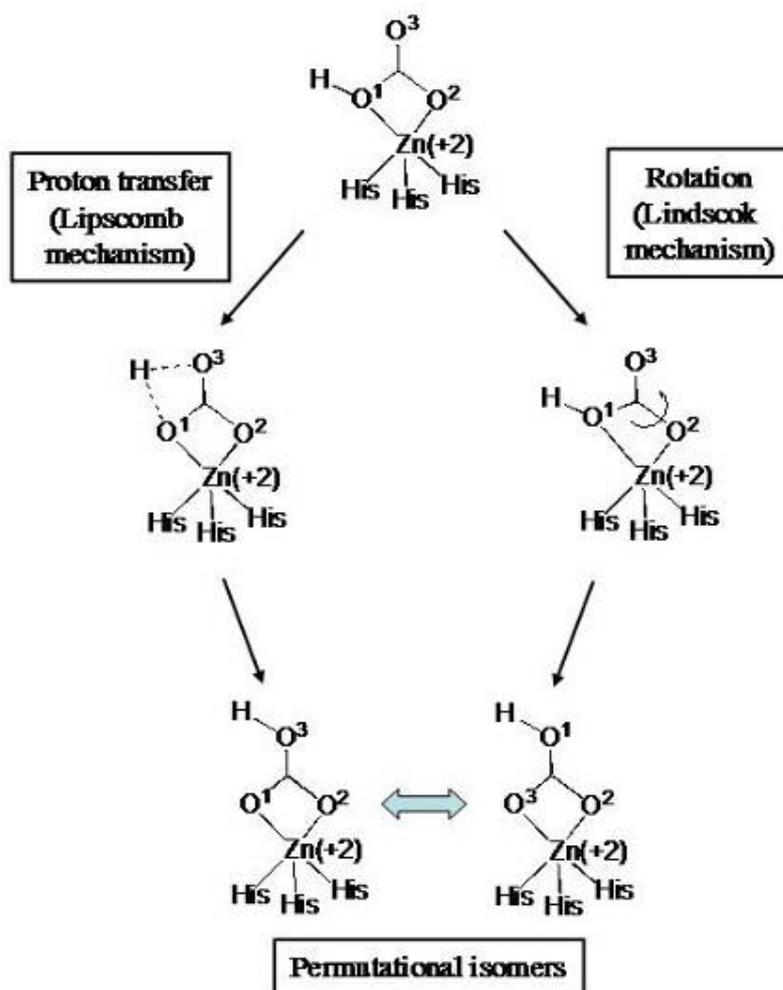
Thoms proposed a different mechanism, which is schematically represented in Scheme 11.2.³³³ This author, on the basis of the crystallographic structure of the hydrogen-bond network within the catalytic site, suggested that the zinc-bound water and not the zinc-bound hydroxide, carries the nucleophilic attack on CO₂. Within this mechanism the formation of a low-barrier hydrogen-bond network (LBHB) involving CO₂, would enhance the electrophilic character of the carbon dioxide and lower the energy of the transition state for the C-O bond formation. A partial activation of the zinc-bound water (increase of its nucleophilic character) should be determined during the nucleophilic attack by a simultaneous transfer of a proton from this water molecule to another water of the network, as indicated in Scheme 11.2.

Recently, we carried out a theoretical DFT investigation on the second and third step of the carbonic anhydrase catalytic cycle reported in Scheme 11.2.¹⁹¹ We used a model-system including the Glu106 and Thr199 residues, the “deep water” molecule and a Zn(+2) cation coordinated to an hydroxide ion and three imidazole rings. We found that the nucleophilic attack of the zinc-bound hydroxide on CO₂ has a negligible barrier, suggesting that this step cannot be the rate-determining step of the process. Also, we examined the mechanism of the internal rearrangement of the zinc-bicarbonate complex (see Scheme 11.3). We found that the direct intra-molecular proton transfer from the zinc-bound oxygen to another oxygen of the bicarbonate moiety (Lipscomb mechanism) can effectively compete with a rotational mechanism (Lindskog mechanism).



Scheme 11.2 Thoms model.

These transformations lead to identical permutational isomers. We demonstrated that the proton transfer is a two-step process, which is assisted by a complex network of hydrogen bonds involving Glu106 and Thr199 where the threonine residue acts as a proton shuttle. In the absence of Glu106 the barrier for the proton transfer significantly increases and the rotational (Lindskog) mechanism becomes definitely favored. In that paper we also considered the attack of the water on the zinc-bicarbonate complex (step 3) leading to a penta-coordinate intermediate, but we did not investigate in details the mechanism of the bicarbonate expulsion.



Scheme 11.3 Rearrangement of the product.

Our results, indicating a very low barrier for the nucleophilic attack, were in contrast with those obtained by Anders and co-workers who found a barrier of $5.7 \text{ kcal mol}^{-1}$ at the B3LYP level using a smaller model-system (formed by the $[(\text{NH}_3)_3\text{Zn}(\text{OH})]^+$ complex reacting with CO_2) but a more accurate basis set.³⁴⁹ On the other hand the investigation carried out by these authors on the internal bicarbonate rearrangement were in rather good agreement with our model. Their results pointed to the Lindskog-type mechanism (internal rotation) as the most favoured path and not to a direct proton shift, characterized by a significantly higher activation energy (about 28 kcal mol^{-1}). Interestingly, this value is rather close to the proton transfer activation barrier that we found in the absence of the Glu106 residue (32 kcal mol^{-1}). We must also outline that the results of our first study¹⁹¹ on HCA are not in contrast with other investigations, based on solvent isotope effects, which suggest that the rate-determining step of the reaction could be the enzyme activation, i.e. the proton transfer from the zinc-bound water to a proton acceptor (the imidazole ring of the His64 residue) contained in the

active site.^{321,336,350} The energetics of this proton transfer was accurately examined in a recent paper by Cui and Karplus.³⁵⁰ They considered different model-systems involving two, three and four water molecules acting as carriers. They demonstrated that the proton transfer process is fully concerted with two water molecules and becomes partially concerted and stepwise with three and four water molecules, respectively. Also, the barrier height increases as the number of water molecules increases (a value of 0.6, 3.6 and about 6 kcal mol⁻¹ was calculated for the three different models, respectively). These authors pointed out that a model based on three or four water molecules, provides results that are consistent with the experimental kinetic observations.

In the present paper we again investigate the HCA mechanism using a model-system based on Thoms' hypothesis. This has been obtained by adding to our first model¹⁹¹ the three water molecules involved in the low-barrier hydrogen-bond network evidenced in Thoms' theory. In particular we examine here again (i) the nucleophilic attack using as a nucleophile either a zinc-bound water (as suggested by Thoms) or a zinc-bound hydroxide (the usual active form of the enzyme), (ii) the internal bicarbonate arrangement and (iii) the final attack of a water molecule that should lead to the expulsion of the bicarbonate fragment from the metal coordination sphere. A detailed analysis of the effect of the hydrogen-bond network on the various reaction steps is given.

11.2 Computational details and choice of the model.

The model-system used here (see Figure 11.1) has been assembled using the crystallographic structure available in literature.³²⁶ This model includes: (i) a Zn(+2) cation bonded to a HO(-) group (or, alternatively, a water molecule) and three imidazole rings belonging to the three histidine residues His94, His96 and His119; (ii) the Glu106 residue; (iii) the Thr199 residue; (iv) the four water molecules W318, W338, W359 and W386; (v) a CO₂ molecule. To reduce the size of the model an acetate fragment has been used to emulate the Glu106 residue. Also, we have replaced the threonine with a serine and we have cut the protein backbone along the bonds adjacent to the carbonyl groups.

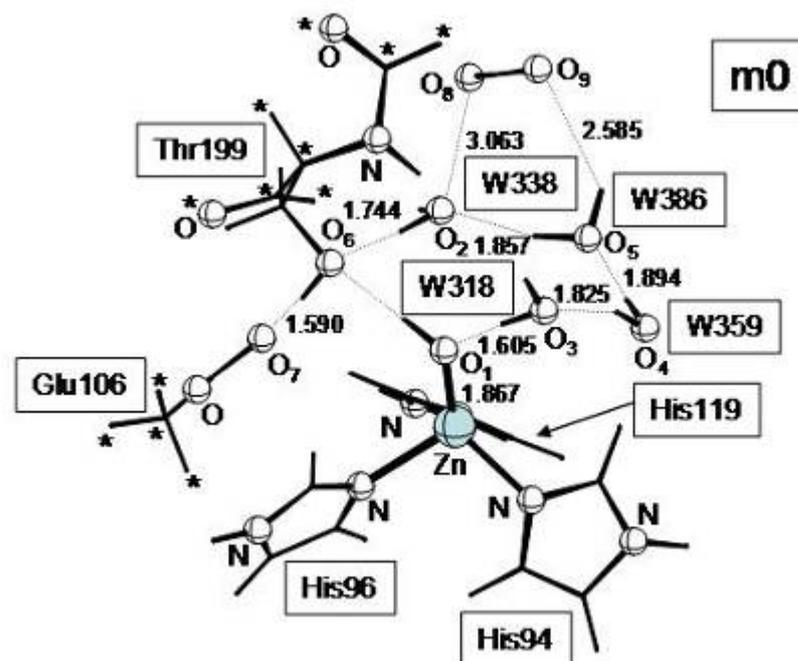


Figure 11.1 A schematic representation of the model-system used in this paper (bond lengths in Ångströms). The reported structure corresponds to that of the preliminary complex **m0** (absolute energy = -3694.03761 a.u.). The asterisks indicate the atoms (in addition to those of the imidazole rings) that have been kept frozen during the geometry optimization).

As already described in ref. 32, to emulate the partially constraining effect of the protein environment, during the geometry optimization procedure we have not optimized the positions of the atoms not directly involved in the reaction or in hydrogen bond formation. This approach, where a number of appropriately chosen atoms have been anchored to their crystallographic coordinates, preserves the geometry of the active-site cavity. The “frozen” atoms include the three imidazole rings and all the atoms marked by an asterisk in Figure 11.1. For the serine (threonine) residue we have not locked the whole CH₂-OH fragment. In this way the OH group should approximately have the same freedom it has in the real enzyme to adjust its position and form effective hydrogen bonds with the neighboring groups.

All the reported DFT computations have been carried out with the Gaussian 03 series of programs³³ using the B3LYP¹⁵¹ functional and the DZVP basis set.²²³ The B3LYP functional has been demonstrated to provide reliable description of systems including transition metals and involving hydrogen bond interactions.^{73,75,191,351,352} The DZVP basis is a Local Spin Density (LSD)-optimized basis set of double-zeta quality that includes polarization functions and is suitable to describe weak hydrogen interactions such as those occurring in the system investigated in this paper. The transition vector of the various transition states has been analyzed by means of frequency computations.

The effect of the whole protein environment has been evaluated with the solvent continuous model approach COSMO³⁵³ as implemented in the Turbomole package.¹³² The dielectric constant of nitromethane ($\epsilon = 38.2$) was used. This value should take into account the simultaneous presence of hydrophilic and hydrophobic groups around the active site. A value of about 40 was suggested elsewhere to describe the effect of charge-charge interactions in proteins.³⁵⁴ Several paper available in literature report the results of the COSMO method in the calculations on enzymatic models and systems involving hydrogen bonds and proton transfers [32, 43-46].^{32,43-46}

11.3 Results and discussion

In this section we examine in detail the singlet potential energy surface that describes the catalytic cycle of HCA. We consider either the case of a zinc-bound hydroxide or that of a zinc-bound water (Thoms' hypothesis) attacking CO₂. The corresponding energy profile is reported in Figure 11.2, while the structures of the various critical points are represented in Figure 11.1 and in Figure 11.3 to Figure 11.9.

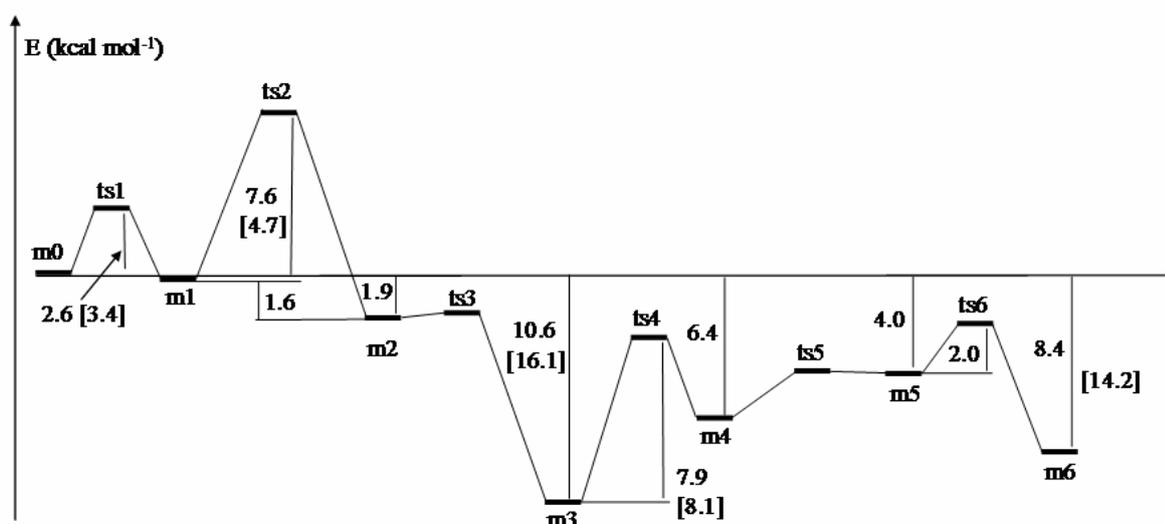


Figure 11.2 Energy profile obtained for the HCA catalytic process.

11.3.1 The starting complex, the nucleophilic attack and the origin of the barrier

The preliminary complex **m0** is depicted in Figure 11.1. This provides a schematic representation of the active site with the CO₂ molecule still far away from the reactive center. A chain of hydrogen bonds involves the four water molecules W318, W338, W359 and W386, the zinc-bound hydroxide and the Thr199 residue. All these molecules and groups are arranged in a ring-like structure that shields the zinc-bound hydroxide with respect to the

approaching carbon dioxide CO₂. Only a weak hydrogen interaction (O₉···H) between CO₂ and W386 can be recognized (O₉···H(O₅) distance = 2.585 Å). An additional hydrogen bond involves the oxygen atom O₇ of the Glu106 residue and the threonine hydroxide (O₇···H (O₆) distance = 1.590 Å).

A transition state **ts1** connects **m0** to a new complex **m1** (almost degenerate to **m0**) where CO₂ is still rather far from the zinc-bound hydroxide, but is now oriented in a suitable way for the nucleophilic attack (see Figure 11.3). To undertake the transformation **m0** → **m1** a small barrier of 2.6 kcal mol⁻¹ must be overcome. The new position of CO₂ partially breaks the original ring-like structure and W338 is now far-away from W386, the O₂···H(O₅) distance being 4.123 Å. In this new structural arrangement it is possible to recognize a smaller ring-like structure formed by the zinc-bound hydroxide and the three water molecules W318, W359 and W386. Rather strong hydrogen bonds characterize this circular frame around the hydroxide (O₁···H(O₃) = 1.702 Å, O₃···H(O₄) = 2.035 Å, O₅···H(O₄) = 1.998 Å, O₁···H(O₅) = 1.670 Å). CO₂ is partially bound to W386 and Thr199 by two weak hydrogen bonds involving the same oxygen atom O₈: O₅-H···O₈ interaction (H···O₈ distance = 2.496 Å) and N-H···O₈ interaction (H···O₈ distance = 2.778 Å).

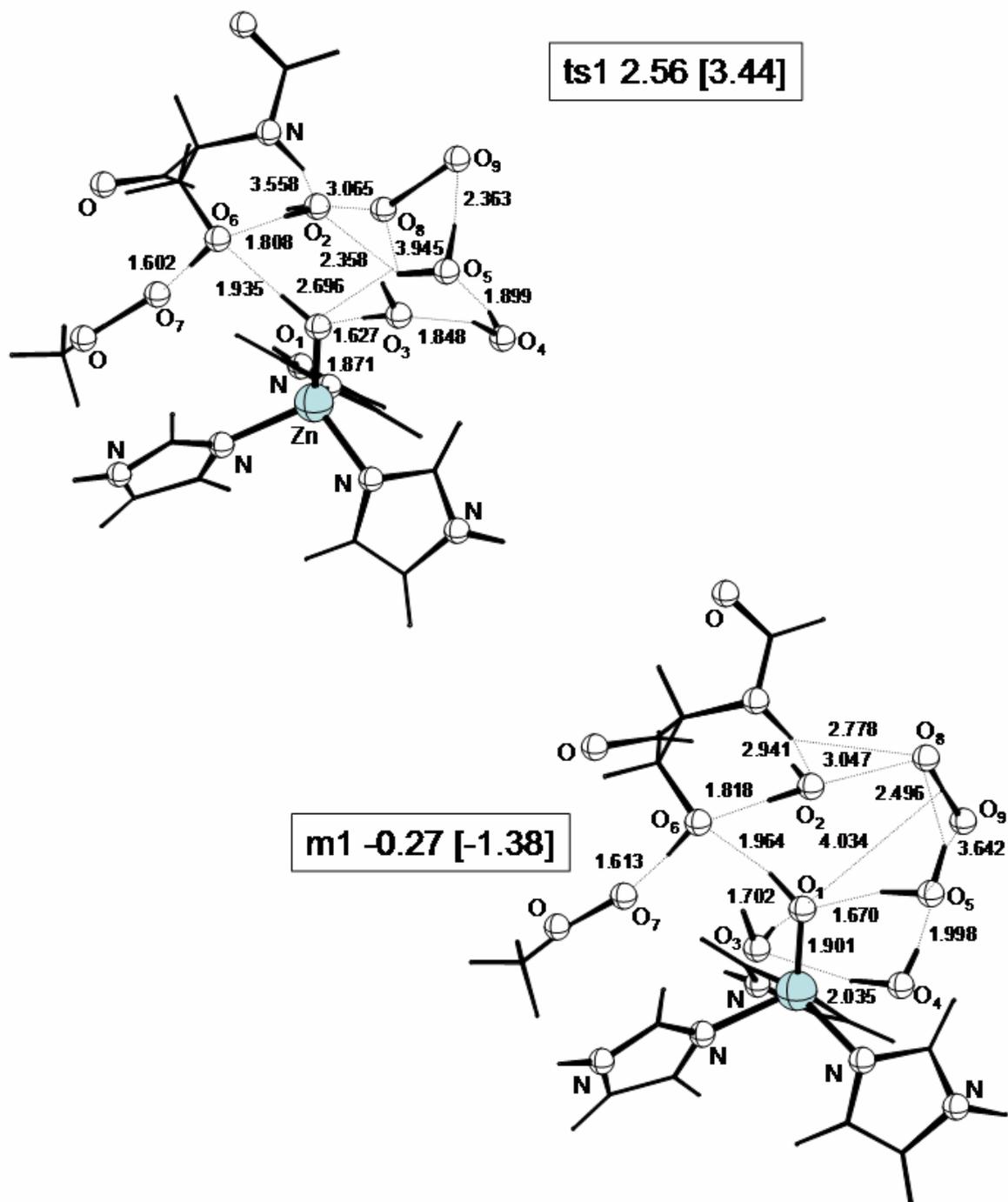


Figure 11.3 Schematic representation of the structure of the critical points **ts1** and **m1** (bond lengths are in ångströms). The energy values (kcal mol^{-1}) are relative to **m0**. Values in brackets have been obtained with the COSMO method.

In the subsequent step (**m1** \rightarrow **ts2** \rightarrow **m2**) the zinc-bound hydroxide (nucleophile) attacks CO_2 . To move closer to the nucleophile the carbon dioxide molecule breaks the network of hydrogen bonds involving the three water molecules W318, W359 and W386 that “protect” the hydroxide. In the transition state **ts2** (see Figure 11.4) the $\text{O}_5 \cdots \text{H}(\text{O}_4)$ interaction becomes significantly weaker (the $\text{O}_5 \cdots \text{H}$ distance changes from 1.998 in **m1** to 2.337 Å in **ts2**) and the

$O_1 \cdots H(O_5)$ hydrogen bond is destroyed by the approaching CO_2 . The new forming carbon-oxygen bond $O_1-C(CO_2)$ is 2.030 Å, but CO_2 remains strongly anchored to W386 by the $O_8 \cdots H(O_5)$ hydrogen interaction ($O_8 \cdots H$ distance = 2.005 Å). The shorter distance between the substrate and the hydroxide has the effect of bending the linear carbon dioxide molecule, the O_8CO_9 angle being now 155.7°.

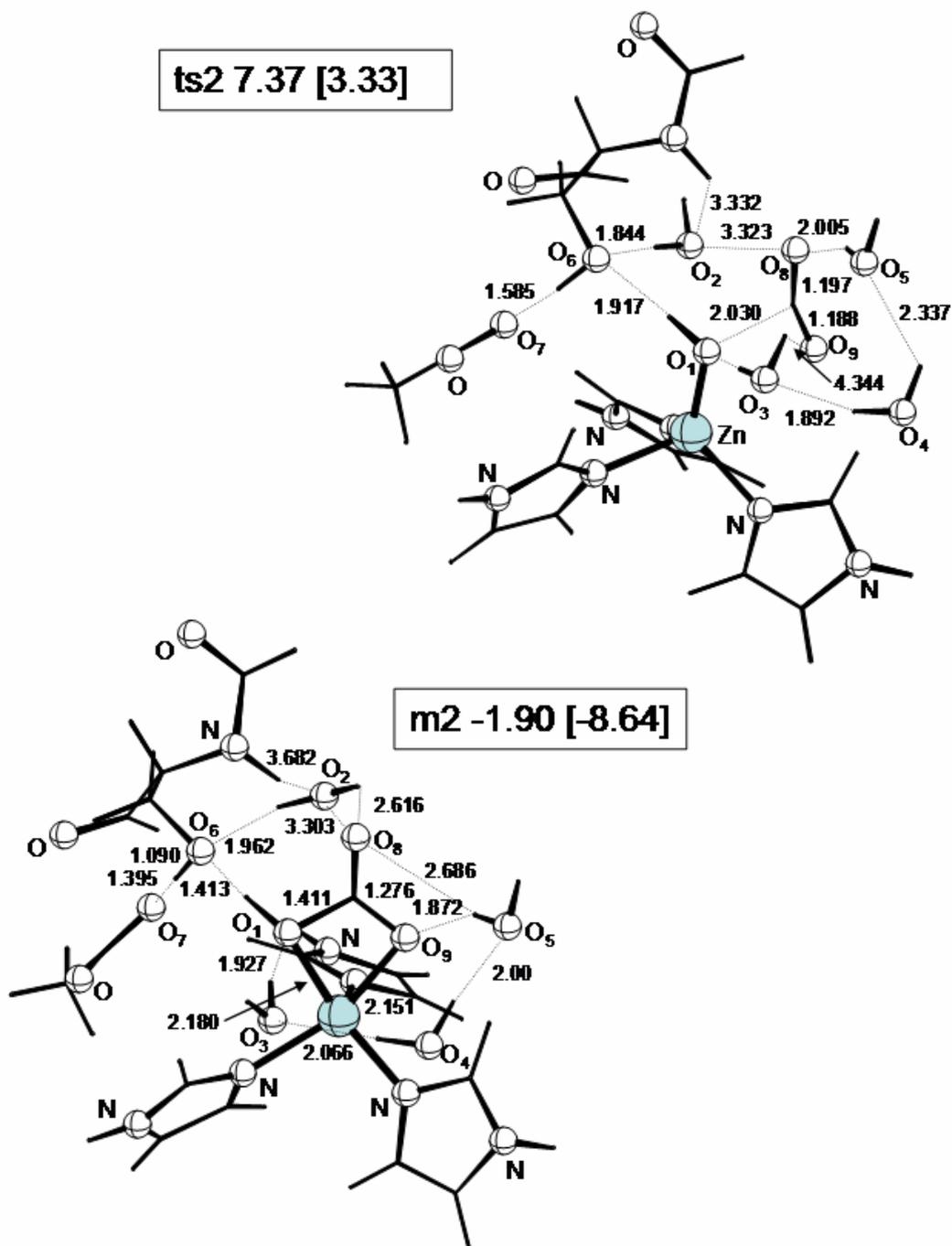


Figure 11.4 Schematic representation of the structure of the critical points **ts2** and **m2** (bond lengths are in ångstroms). The energy values (kcal mol^{-1}) are relative to **m0**. Values in brackets have been obtained with the COSMO method

The energy required to break the wall of water molecules around the zinc-bound hydroxide is responsible for the activation barrier computed for **ts2** (7.6 kcal mol⁻¹) and explains why an almost negligible barrier (about 1 kcal mol⁻¹) is found when the three water molecules W318, W359 and W386 are missed in the model-system. The new energy barrier value found here is in rather good agreement with that found by Anders and co-workers (5.7 kcal mol⁻¹).

The resulting complex **m2** is 1.6 kcal mol⁻¹ lower in energy than **m1**. The stabilization of **m2** is determined by the restoring of the hydrogen bond network after the nucleophilic attack. These hydrogen bonds form a new round structure now including the just formed bicarbonate fragment bonded to the metal. The HCO₃(-) unit chelates the metal giving rise to a penta-coordinated zinc complex which is rather similar to that already described in ref. 32 (the O₁-Zn and O₉-Zn distances are 2.180 and 2.151 Å, respectively). As already observed in our previous study, a strong hydrogen-bond anchors the bicarbonate to the Thr199 hydroxide fragment (O₆···H(O₁) distance = 1.413 Å).

Furthermore, to verify the Thoms' mechanistic hypothesis we have considered a model-system with a water molecule bonded to the metal in the place of the hydroxide. We have recomputed the structure of the starting complex (**m1'**), which is depicted in Figure 11.5. **m1'** is rather similar to **m1**. The four water molecules form again a ring-like structure involving CO₂, which interacts with one water molecule (W386) and the N-H bond of the Thr199 residue, as suggested in Thoms' model. The major difference between **m1** and **m1'** is the distance between the CO₂ carbon and the nucleophilic oxygen O₁. In the former case this distance is 4.034 Å, while in **m1'** it is much shorter i.e. 2.877 Å. We have investigated for **m1'** the reaction channel for the nucleophilic attack. However, in spite of extensive search on the potential surface we could not find any transition state for the formation of the new O-C bond. We observed a rather rapid energy increase when the oxygen of the zinc-bound water was approaching the carbon dioxide. All attempts to locate this critical point drove the search algorithm back to the starting complex with the zinc-bound water far away from CO₂. These results indicate that, even in the presence of the hydrogen-bond network, the electrophilic character of the CO₂ carbon is not enhanced enough to make possible the nucleophilic attack by the metal-bound water. Also, no simultaneous activation of this water molecule via a partial proton transfer was observed. This evidence makes Thoms' hypothesis quite unlikely.

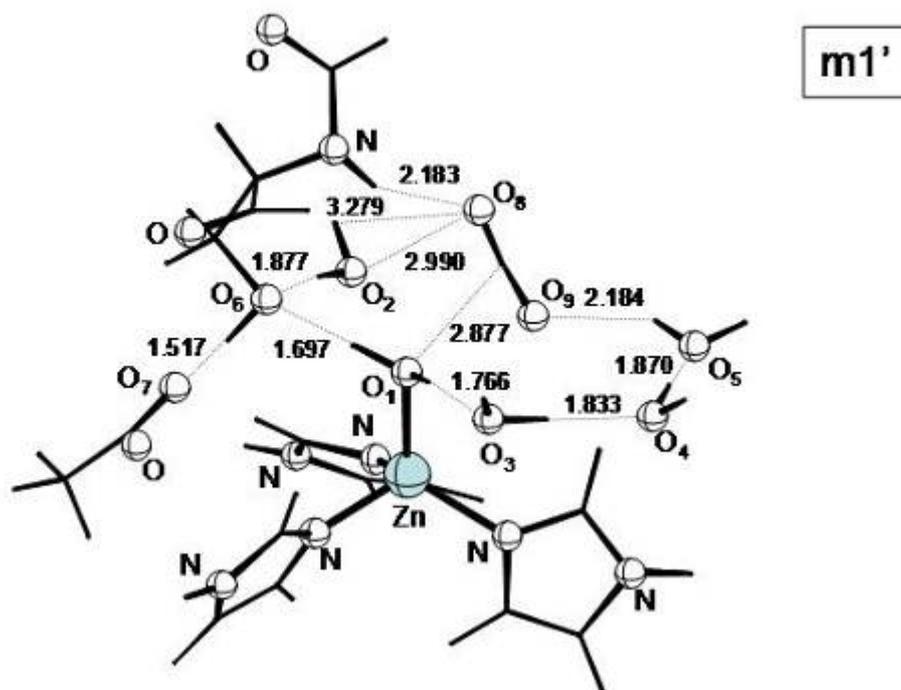


Figure 11.5 Schematic representation of the structure of the critical point **m1'** (bond lengths are in ångstroms).

11.3.2 The rearrangement of the zinc-bicarbonate complex

We have examined again the rearrangement mechanism of the zinc-bicarbonate complex in the presence of the three additional water molecules considered in the Thoms' model. In spite of extensive search, we did not locate any transition state for the Lindskog mechanism (rotation around the Zn-O₉ bond and simultaneous breaking of the Zn-O₁ bond) and all our attempts to follow this pathway on the potential surface led to a strong energy increase. This finding can be easily understood if we take into account the network of hydrogen bonds that involve the four water molecules and the bicarbonate fragment. Inspection of the structural features of the **m2** intermediate clearly shows that the rotation required by the Lindskog mechanism inevitably breaks this network and raises the energy.

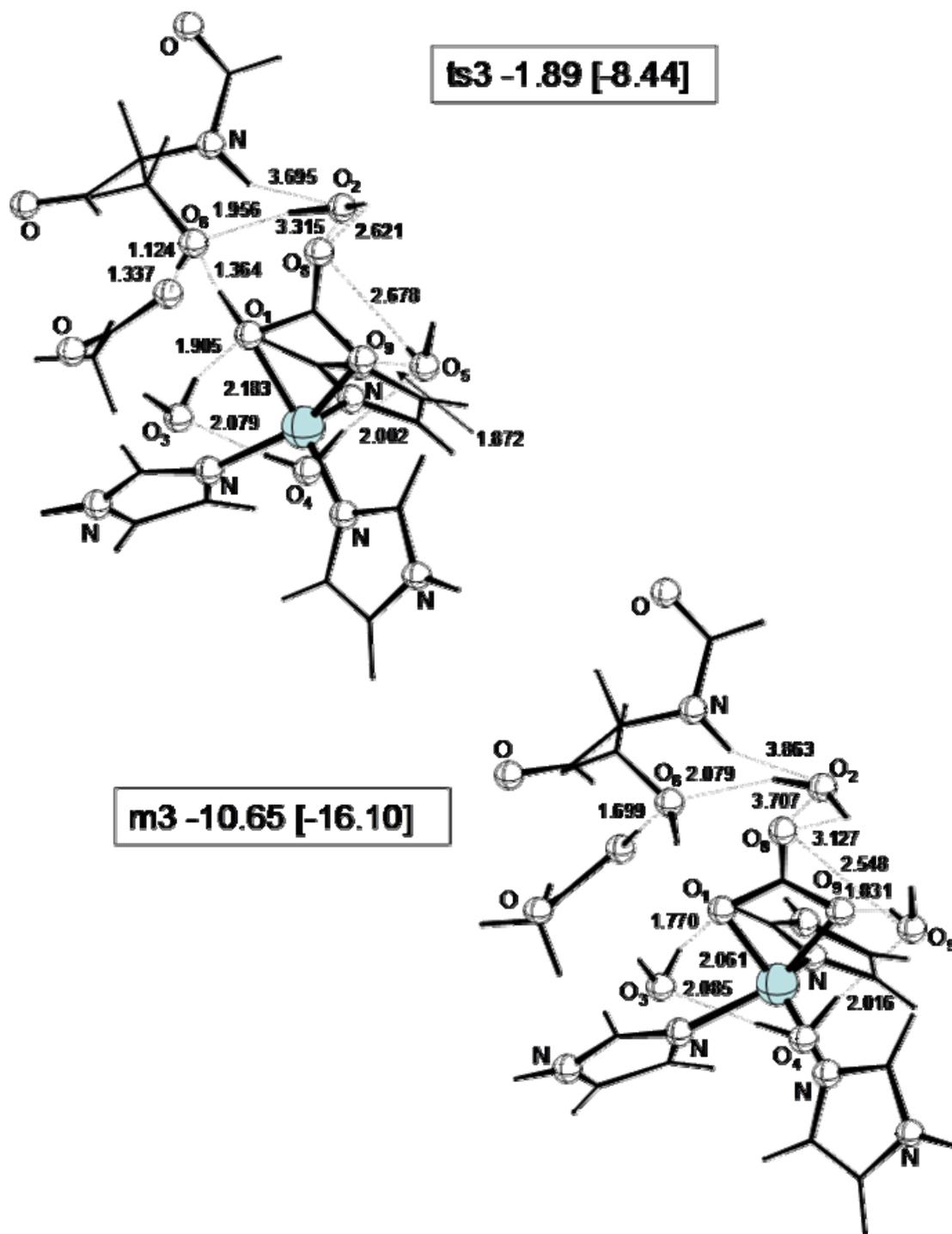


Figure 11.6 Schematic representation of the structure of the critical points **ts3** and **m3** (bond lengths are in ångstroms). The energy values (kcal mol^{-1}) are relative to **m0**. Values in brackets have been obtained with the COSMO method.

The rearrangement can occur rather easily via a double proton transfer mechanism (Lipscomb mechanism). This is very similar to that already discussed in our previous work and requires two subsequent steps where the threonine residue behaves as a proton shuttle. In the first step (**m2** \rightarrow **ts3** \rightarrow **m3**) a proton is transferred from O₁ (the original hydroxide oxygen) to O₆ of

the Thr199 group. Simultaneously the threonine proton is transferred to the Glu106 residue. This transformation leaves the frame of hydrogen bonds roughly unchanged. More precisely, since during the transformation some hydrogen bonds become stronger, **ts3** is stabilized and the corresponding barrier becomes negligible (for instance, on passing from **m2** to **ts3** the $O_1 \cdots H(O_3)$ distance changes from 1.927 to 1.905 Å, $O_7 \cdots H(O_6)$ from 1.395 to 1.337 Å and $O_6 \cdots H(O_1)$ from 1.413 to 1.364 Å, see Figure 11.6). The second step (**m3** \rightarrow **ts4** \rightarrow **m4**) corresponds to a double proton transfer in the opposite direction: from Glu106 to Thr199 and from Thr199 to the bicarbonate fragment (**ts4** and **m4** are depicted in Figure 11.7). Since we have a simultaneous breaking of the Zn-O₉ bond and a consequent reorientation of the bicarbonate, the proton is transferred to O₈. Thus, the final effect of the **m2** \rightarrow **m4** transformation is an internal proton transfer from O₁ to O₈. The barrier for this second proton transfer is 7.9 kcal mol⁻¹, a value lower than that found in our previous investigation (12.3 kcal mol⁻¹) where we carried out single-point computations, with the same basis set used here, on structures optimized at a lower computational level. The difference can be due to the loss of geometric re-optimization in our previous calculations and to the presence of additional hydrogen bonds in the present model. Thus, these results show that the hypothesis of a Lindskog mechanism is highly unlikely (it is impossible to locate the corresponding pathway on the potential surface since the energy raises very rapidly) and the discussion concerning the competition between the two possible rearrangement mechanisms (Lindskog vs Lipscomb) is only an academic issue.

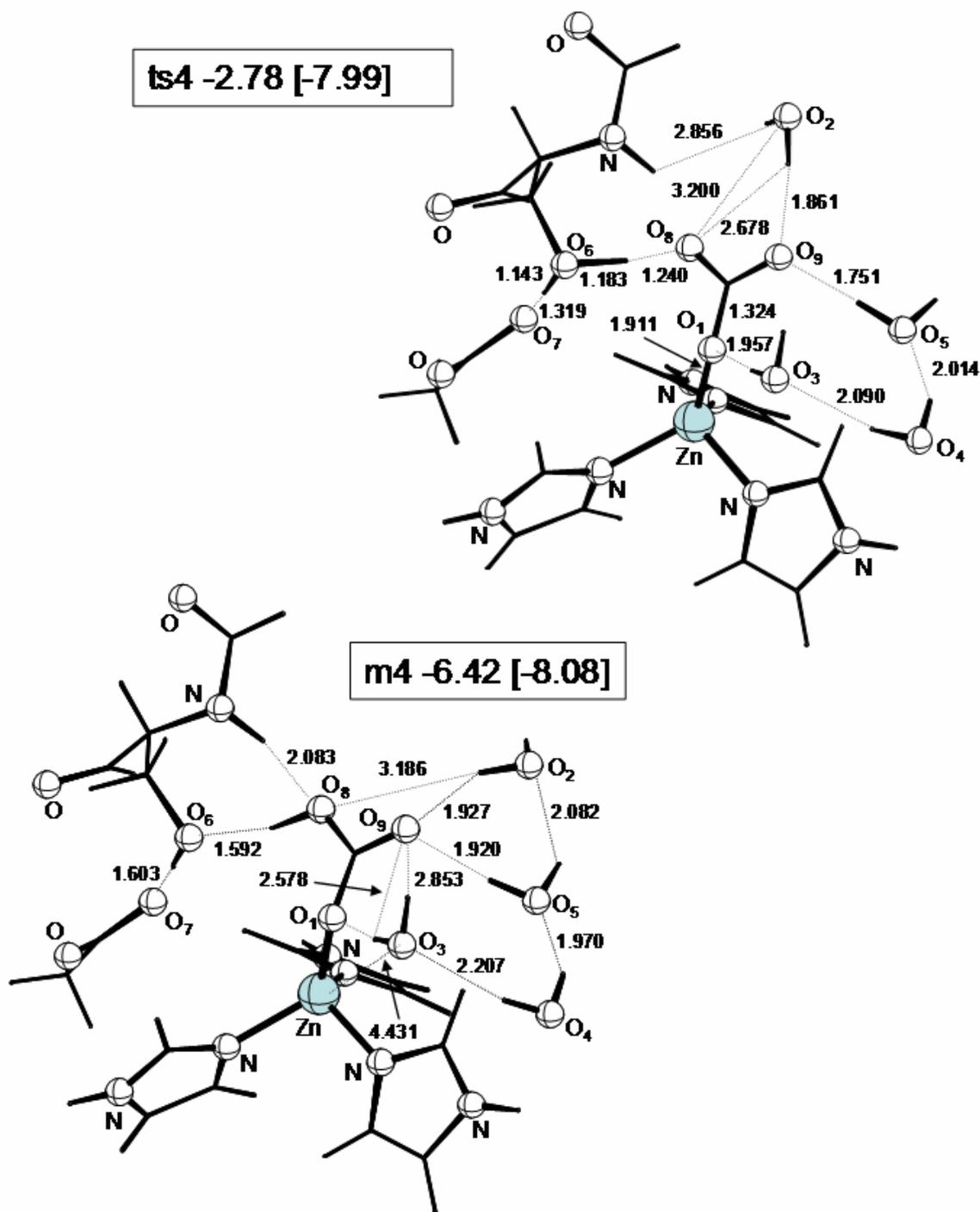


Figure 11.7 Schematic representation of the structure of the critical points **ts4** and **m4** (bond lengths are in ångstroms). The energy values (kcal mol^{-1}) are relative to **m0**. Values in brackets have been obtained with the COSMO method.

11.3.3 The attack of the water molecule and the formation of a penta-coordinated zinc complex.

It is interesting to note that the **m4** intermediate is now a four-coordinated complex where one water molecule is correctly oriented to undergo a nucleophilic attack on the metal atom (the

distance between the zinc atom and the water oxygen O₃ is 4.431 Å). All the attempts to locate the corresponding transition state (**ts5**) showed that the transition region is very flat indeed and the search algorithm led in all cases to the penta-coordinated intermediate **m5** that results from the water attack (see Figure 11.8). Even if it has not been located, **ts5** is indicated in the diagram of Figure 11.2 and is approximately degenerate to **m5**. Thus, we again observe that the attack of the water on the metal does not cause the simultaneous expulsion of the bicarbonate fragment but the formation of a new intermediate where both the bicarbonate and the water are firmly bonded to the zinc atom (the two Zn-O₁ and Zn-O₃ bond lengths are 2.030 and 2.298 Å, respectively). The complex **m5** is 2.4 kcal mol⁻¹ higher in energy than **m4**. This destabilization (in rather good agreement with the results of ref. 29) is probably due to the weakening of some hydrogen bonds on passing from **m4** to **m5** (for instance, O₃···H-O₄). Further investigation of the potential energy provides information very similar to that obtained in our first study. The expulsion is a rather complicated process as indicated by the subsequent transformation (**m5** → **ts6** → **m6**) characterized by a small barrier of 2.0 kcal mol⁻¹. In the resulting intermediate **m6** (see Figure 11.9) the bicarbonate-zinc bond is weakened and the water-zinc bond is enforced. More interesting the bicarbonate hydroxide has changed orientation and points now in the opposite direction. As a consequence the O₆···H(O₈) hydrogen bond is broken, the bicarbonate fragment is less firmly bound to the Thr199 residue and it should be easier for this fragment to abandon the active site. Thus, the intermediate **m6** seems to prepare and organize the real expulsion of the bicarbonate fragment. Since the model-system is still rather small and other residues can play, in principle, an important role in the expulsion process (for instance Thr200 that can anchor the leaving bicarbonate), we did not further investigate the potential surface. Another possibility, which is worth to explore, is the formation of carbonic acid (see ref. 29) via a proton transfer from O₃ to O₉. This alternative reaction path would have the advantage of leading directly to the enzyme reactivation (formation of the nucleophilic zinc-bound hydroxide). Also, H₂CO₃ could leave the field of the Zn²⁺ cation without significant problems.

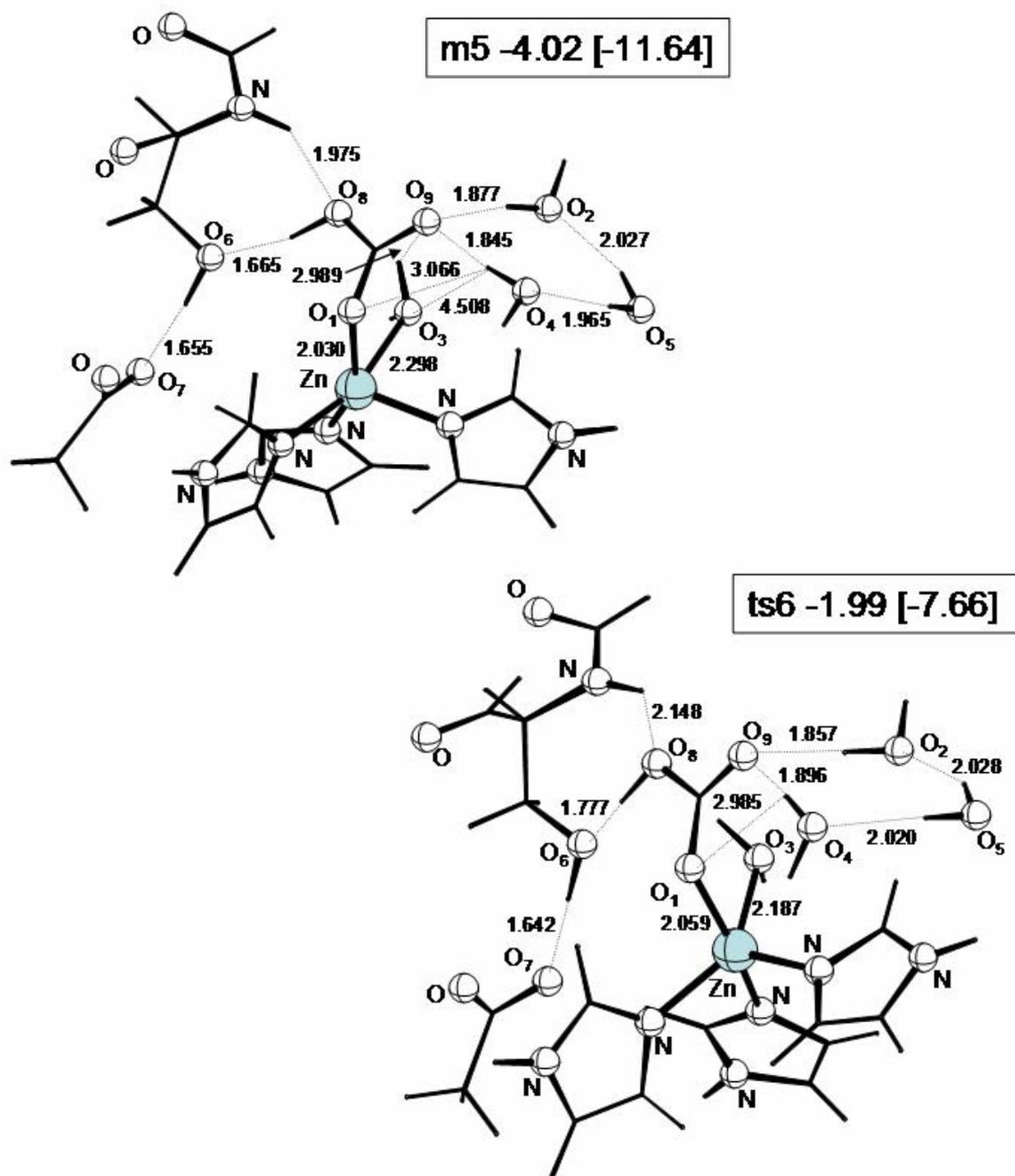


Figure 11.8 Schematic representation of the structure of the critical points **m5** and **ts6** (bond lengths are in ångstroms). The energy values (kcal mol^{-1}) are relative to **m0**. Values in brackets have been obtained with the COSMO method.

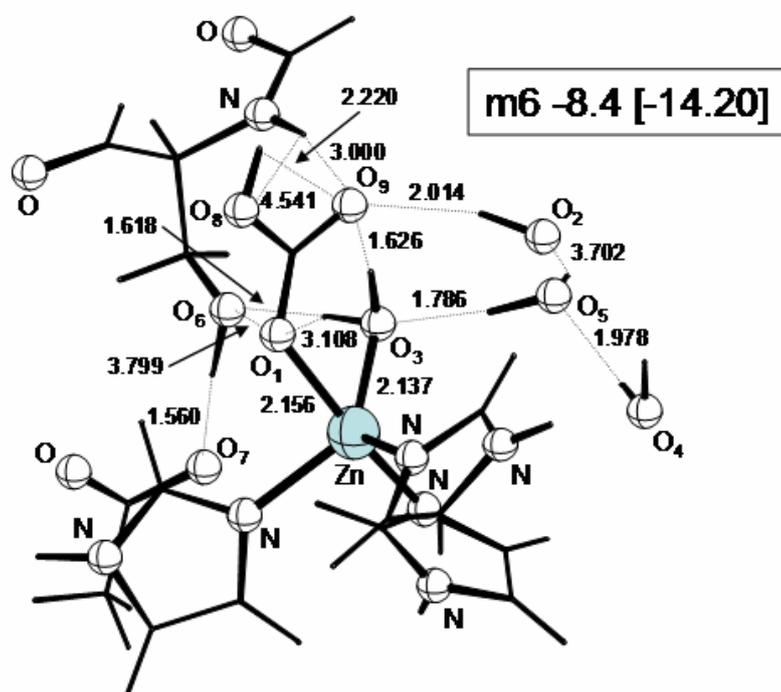


Figure 11.9 Schematic representation of the structure of the critical point **m6** (bond lengths are in Ångströms). The energy value (kcal mol^{-1}) is relative to **m0**. Values in brackets have been obtained with the COSMO method.

11.3.4 The effect of the protein environment.

The values of the two largest activation energies (for **ts1**, **ts2** and **ts4**) and the energy values of the **m3** and **m6** intermediates obtained in the presence of solvent effects are reported in brackets in Figure 1.1. Also, the energy corrected by solvents effects for all the critical points located on the surface are reported in square brackets in the figures. The barrier for the nucleophilic attack (**m1** \rightarrow **ts2**) decreases (from 7.6 to 4.7 kcal mol^{-1}) but this step remains the rate-determining step of the process. As a matter of fact the two transition states **ts1** and **ts2** become almost degenerate (3.4 and 3.3 kcal mol^{-1} above the starting complex **m0**, respectively), which suggests that in the presence of the protein environment the reorientation of the CO_2 molecule within the active site together with the nucleophilic attack on CO_2 , represent the real determining step of the reaction. Interestingly, even if the barrier for **ts4** remains almost the same (it changes from 7.9 to 8,1 kcal mol^{-1}), the energy of **m3** and **m6** become much lower. These two critical points are now 16.1 and 14.2 kcal mol^{-1} lower than **m0**, respectively. All these results (i.e. the lower barrier for the rate determining step and the larger exothermicity) indicate that the protein environment makes the reaction even faster. This finding is in agreement with the high reaction rate experimentally observed.

11.4 Conclusions

In this paper some aspects of the catalytic mechanism of HCA have been again investigated using a model-system based on the mechanistic hypothesis proposed by Thoms.³³³ This model-system includes three additional water molecules with respect to that examined in ref. 32. Following this scheme a zinc-bound water and not a zinc-bound hydroxide would carry out a nucleophilic attack on CO₂. The most significant results can be summarized as follows:

(i) Our computational evidence does not support Thoms' hypothesis. We could not locate any transition state for the nucleophilic attack when we have considered as nucleophile a zinc-bound water (as suggested by Thoms) in the place of a zinc-bound hydroxide. Our computations indicate that the presence of the hydrogen bond network is not enough either to enhance the electrophilic character of the CO₂ carbon or the nucleophilic character of the metal-bound water.

(ii) A transition state for the nucleophilic attack (with a barrier of 7.6 kcal mol⁻¹) has been located when an activated water (zinc-bound hydroxide) was used as a nucleophile. This activation energy is much larger than that found in our previous study and is determined by the breaking of the hydrogen-bond network when the CO₂ molecule approaches the reaction center.

(iii) No ambiguity exists about the mechanism for the bicarbonate rearrangement. The rotation pathway (Lindskog mechanism) is too energy-demanding since it causes the breaking of the hydrogen-bond network around the bicarbonate. The only possible rearrangement mechanism is a proton transfer (Lipscomb) that occurs in two steps (each step corresponding to a double proton transfer) and involves Thr199 as a proton shuttle. This mechanism is almost identical to that described in ref. 32 in the absence of the three additional water molecules used here.

(iv) The attack of the water on the metal in the zinc-bicarbonate complex does not cause the simultaneous expulsion of the bicarbonate fragment. On the contrary, we have observed the formation of a penta-coordinated zinc complex where both the bicarbonate and the water are firmly bonded to the metal. Further computations indicate that the expulsion of the bicarbonate is a rather complicated process where other residues that are missed in the present model, could play a key-role in stabilizing the leaving group.

(v) The barrier for the nucleophilic attack is comparable to that computed by Karplus³⁵⁰ for the activation step of the zinc-bound water (a value between 3.6 and 6 kcal mol⁻¹). The accuracy of the computational level used by Karplus (B3LYP/6-31G(d)) is similar to that employed in the present study. Thus, our results suggest that the nucleophilic attack could be

the rate-determining step of the catalytic cycle in agreement with the conclusions reached by Anders and co-workers.

(vi) The reaction is exothermic by $8.4 \text{ kcal mol}^{-1}$. This exothermicity value and the low barrier for the nucleophilic attack are consistent with the high reaction speed observed for this catalytic process.

(vii) The effect of the protein environment (emulated by solvent continuous model computations) does not change the mechanistic scenario. However, the barrier for the rate-determining step becomes lower and the reaction much more exothermic. This indicates an even faster reaction in agreement with the high reaction rate experimentally observed.

Conclusions

A versatile approach for QM/MM computations (concerning geometry optimisations, frequency computations and molecular dynamics) has been illustrated. This method is characterized by a partitioning of the system in two or three layers, each corresponding to a different level of accuracy. The present QM/MM implementation represents a general hybrid approach with a modular structure that is able to integrate some commercially available specialized programs, thus increasing the flexibility/efficiency of QM, MM and QM/MM computations at the same time. The flexibility and reliability of this implementation has been supported by several test examples, which span over rather different areas of chemistry including either ground or excited state topics. In particular, organic, bio-organic, photobiological and super/supra-molecular applications have been presented and discussed.

Moreover standard QM methods have been successfully employed in the study of organic reactions and to simulate the behaviour of two bimolecular systems, the HIV1-Pr and the HCA II enzymes. The performed accurate calculation helped to rationalize the experimental data and to investigate various mechanistic hypothesis.

Appendices

(1) Handling the charges in the “real” system

A particular care is needed in redistributing charges if the boundary region cuts one or more covalent bonds. Specifically, MM computations need a topology file for the whole system (*real*) that contains the connectivity data and the atomic point charges assigned on the basis of the adopted force field. The sum of all the atomic charges gives the total charge of the *real* system (Q^{real}), which is necessarily a net charge and must be a constant value all through QM/MM calculations. Thus, before starting calculations, atomic point charges for the whole system atoms have to be assigned and their sum must be a net constant value; but while the charges for the atoms in the MM region will remain constant throughout the course of all QM/MM calculations (at least this is our case, since we don't use a polarizable force field), charges for the QM atoms are generated at each step from the wave function and may change accordingly, which may lead to a variable fractional total charge for the *model* system (note that when the *real* system is split in *model* and *pod* (see **Figure 3.4**), the corresponding total charges, $Q^{model'}$ and $Q^{pod'}$, are not necessarily net constant values although it is their sum $Q^{real} = Q^{model'} + Q^{pod'}$). QM calculations are performed on *model-H*, whose net charge $Q^{model-H}$ and multiplicity are imposed by the user. To prepare a suitable topology file and grant invariability of the system total net charge, in the current implementation we force the total charge of *model* (Q^{model}) to be equal to that of *model-H* by redistributing the charge difference $\Delta Q = Q^{model-H} - Q^{model'}$ over the QM atoms (we do that proportionally to the magnitude of each QM atom charge, but alternative strategies are also possible). Thus:

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

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

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

For *pod* (i.e. the MM atoms) this is done only once, i.e. for the starting geometry during the initial set up stage and, as stated above, these charges do not change any more. This charge redistribution process may be accomplished according to different strategies since the choice is not unique (mostly depending by the system under study). Here, we can draw some general guidelines on the basis of our experience. When studying an enzymatic system the procedure is quite complex, because all the residues have a net charge. In this case, $-\Delta Q$ is equally redistributed only on the MM atoms of *pod* belonging to the residues crossed by the boundary line (thus leaving a net charge

on each one) and these new atomic point charges of *pod* will not change anymore during QM/MM calculations (i.e. we don't use a polarizable force field); then, *emb* is generated from *pod* at each step (discussion in section Chapter 3), to be included in the QM calculation according to the Electrostatic Embedding scheme¹⁰⁹.

Another crucial point is how to handle charges in the QM region. In many cases the charges of this region are not correctly parameterized in the adopted force field, or they are completely unavailable. In these cases it is necessary to get a guess of the charges to perform the first calculation step and different approaches are possible for accomplishing this task; for example, we found that computing the wave function of *model-H* at the AM1-BCC^{161,162} (or a low *ab initio*) level gives a good starting point. Anyway, during the QM/MM calculation, the wave function of *model-H* is computed at the higher level (including the effect of the Electrostatic Embedding) and its charges re-evaluated. Thus, according to the flowchart of our code (**Scheme 3.1**), at the end of each cycle the complete set of charges of *real* is reassembled using for *model* the charges computed at the high QM level. It is worth repeating that, QM calculations are performed on *model-H* and the obtained charge includes H link atoms. According to the procedure described above, at every cycle we redistribute the charges of the H link atoms (ΔQ) on the *model* (QM) atoms. Thus, by adopting all these precautions, the total charge of the system (Q^{real}) is preserved, as well as the total charge of *model* ($Q^{model} = Q^{model-H}$) and *pod* (Q^{pod}). The atomic point charges of *pod* are preserved all over the QM/MM calculations, while the charges of *model* do change at every step, according to the computed QM electronic distribution.

References

- [1] Leach, A. R. *Molecular Modelling: Principles and Applications*; second ed.; Pearson Education EMA: UK, 2001.
- [2] Cramer, C. J. *Essentials of Computational Chemistry: Theories and Models*; Second ed., 2004.
- [3] Jensen, F. *Introduction to Computational Chemistry*; first ed.; John Wiley & Sons: UK, 1999.
- [4] Born, M.; Oppenheimer, R. *Ann. Physic.* **1927**, *84*, 457-484.
- [5] Broyden, C. G. *Journal of the Institute of Mathematics and Its Applications* **1970**, *6*, 76-90.
- [6] Fletcher, R. *Computer Journal* **1970**, *13*, 317-322.
- [7] Goldfarb, D. *Math. Comput* **1970**, *24*, 23-26.
- [8] Shanno, D. F. *Math. Comput* **1970**, *24*, 647-656.
- [9] Nocedal, J. *Math. Comput* **1980**, *35*, 773-782.
- [10] Liu, D. C.; Nocedal, J. *Math. Prog. B* **1989**, *45*, 503-528.
- [11] Powell, M. J. D. *A new algorithm for unconstrained optimization*; Academic Press: New York, 1970.
- [12] Henkelman, G.; Jonsson, H. *J. Chem. Phys.* **1999**, *111*, 7010-7022.
- [13] Henkelman, G.; Jonsson, H. *J. Chem. Phys.* **2000**, *113*, 9978-9985.
- [14] Henkelman, G.; Uberuaga, B. P.; Jonsson, H. *J. Chem. Phys.* **2000**, *113*, 9901-9904.
- [15] Trygubenko, S. A.; Wales, D. J. *J. Chem. Phys.* **2004**, *120*, 2082-2094.
- [16] Carr, J., M.; Trygubenko, S. A.; Wales, D. J. *J. Chem. Phys.* **2005**, *122*, 234903.
- [17] Young, D. C. *Computational Chemistry: A Practical Guide for Applying Techniques to Real-World Problems*; Wiley: New York, 2001.
- [18] Dykstra, C. E. *Quantum Chemistry & Molecular Spectroscopy*; Prentice Hall: Englewood Cliffs, 1992.
- [19] McQuarrie, D. A. *Quantum Chemistry*; University Science Books: Mill Valley, 1983.
- [20] Atkins, P. W.; Friedman, R. S. *Molecular Quantum Mechanics*; Oxford: Oxford, 1997.
- [21] Haile, J. M. *Molecular dynamics simulation : elementary methods*; Wiley: New York 1997.
- [22] Rapaport, D. C. *The art of molecular dynamics simulation*; Cambridge University Press Cambridge 2001.
- [23] Becker, O. M. *Computational biochemistry and biophysics*; M. Dekker: New York, 2001.
- [24] Andrew, M.; Harvey, S. C. *Dynamics of proteins and nucleic acids*; Cambridge University Press: Cambridge, 1987.
- [25] Frenkel, D.; Smit, B. *Understanding molecular simulation : from algorithms to applications*; Academic Press: San Diego, 2002.
- [26] Allen, M. P.; Tildesley, D. J. *Computer simulation of liquids*; Oxford University Press: Oxford, 1989.
- [27] Sadus, R. J. *Molecular simulation of fluids: theory, algorithms, and object-orientation*; Elsevier: New York, 2002.
- [28] Marx, D.; Hutter, J. *Ab Initio Molecular Dynamics: Theory and Implementation* Second ed., 2000; Vol. 3.
- [29] Billeter, S. R.; Webb, S. P.; Agarwal, P. K.; Iordanov, T.; Hammes-Schiffer, S. *J. Am. Chem. Soc.* **2001**, *123*, 11262-11272.
- [30] Armunanto, R.; Schwenk, C. F.; Bambang Setiaji, A. H.; Rode, B. M. *Chem. Phys.* **2003**, *295*, 63-70.

- [31] Verlet, L. *Phys. Rev.* **1967**, *159*, 98-103.
- [32] Hockney, R. W. *Methods Comp. Phys.* **1970**, *9*, 136-211.
- [33] William, C. S.; Hans, C. A.; Peter, H. B.; Kent, R. W. *J. Chem. Phys.* **1982**, *76*, 637-649.
- [34] Beeman, D. *J. Comput. Phys.* **1976**, *20*, 130-139.
- [35] Ryckaert, J.-P.; Ciccotti, G.; Berendsen, H. J. C. *J. Comput. Phys.* **1977**, *23*, 327-341.
- [36] Ciccotti, G.; Ryckaert, J. P. *Computer Physics reports* **1986**, *4*, 346-392.
- [37] Procacci, P.; Darden, T.; Marchi, M. *J. Phys. Chem.* **1996**, *100*, 10464-10468.
- [38] Tuckerman, M.; Berne, B. J.; Martyna, G. J. *J. Chem. Phys.* **1992**, *97*, 1990-2001.
- [39] Francisco, F.; Ronald, M. L.; Ruhong, Z.; Berne, B. J. *J. Chem. Phys.* **1997**, *106*, 9835-9849.
- [40] Procacci, P.; Marchi, M. *J. Chem. Phys.* **1996**, *104*, 3003-3012.
- [41] Procacci, P.; Marchi, M.; Martyna, G. J. *J. Chem. Phys.* **1998**, *108*, 8799-8803.
- [42] Duan, Y.; Wu, C.; Chowdhury, S.; Lee, M. C.; Xiong, G.; Zhang, W.; Yang, R.; Cieplak, P.; Luo, R.; Lee, T.; Caldwell, J.; Wang, J.; Kollman, P. *J. Comput. Chem.* **2003**, *24*, 1999-2012.
- [43] Bayly, C. I.; Cieplak, P.; Cornell, W.; Kollman, P. A. *J. Phys. Chem.* **1993**, *97*, 10269-10280.
- [44] Cieplak, P.; Cornell, W. D.; Bayly, C.; Kollman, P. A. *J. Comput. Chem.* **1995**, *16*, 1357-1377.
- [45] Case, D. A.; Cheatham, T. E.; Darden, T.; Gohlke, H.; Luo, R.; Merz, K. M.; Onufriev, A.; Simmerling, C.; Wang, B.; Woods, R. J. *J. Comput. Chem.* **2005**, *26*, 1668-1688.
- [46] Ponder, J. W.; Case, D. A.; Valerie, D. In *Advances in Protein Chemistry*; Academic Press: 2003; Vol. Volume 66, p 27-85.
- [47] Dudek, M. J.; Ponder, J. W. *J. Comput. Chem.* **1995**, *16*, 791-816.
- [48] Cheatham, T. E. I.; Young, M. A. *Biopolymers* **2000**, *56*, 232-256.
- [49] Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. *J. Comput. Chem.* **2004**, *25*, 1157-1174.
- [50] Van Der Spoel, D.; Lindahl, E.; Hess, B.; Groenhof, G.; Mark, A. E.; Berendsen, H. J. C. *J. Comput. Chem.* **2005**, *26*, 1701-1718.
- [51] Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. *J. Comput. Chem.* **1983**, *4*, 187-217.
- [52] MacKerell, A. D.; Brooks, B.; Brooks, C. L.; Nilsson, L.; Roux, B.; Won, Y.; Karplus, M. In *The Encyclopedia of Computational Chemistry*; al., P. v. R. S. e., Ed.; John Wiley & Sons: Chichester, 1998, p 271-277.
- [53] Boys, S. F. *Proc. R. Soc. London* **1950**, *200*, 542.
- [54] Szabo, A.; Ostlund, N. S. *Modern Quantum Chemistry*; Dover Publications, INC: New York, 1996.
- [55] Roothaan, C. C. J. *Reviews of Modern Physics* **1951**, *23*, 69-89.
- [56] Møller, C.; Plesset, M. S. *Phys. Rev.* **1934**, *46*, 618-622.
- [57] Kohn, W.; Sham, L. J. *Phys. Rev.* **1965**, *140*, A1133.
- [58] Slater, J. C. *Phys. Rev.* **1951**, *81*, 385.
- [59] Becke, A. D. *Phys. Rev. A* **1988**, *38*, 3098.
- [60] Lee, C.; Yang, W.; Parr, R. G. *Physical Review B* **1988**, *37*, 785.
- [61] Tao, J.; Perdew, J. P.; Staroverov, V. N.; Scuseria, G. E. *Phys. Rev. Lett.* **2003**, *91*, 146401.
- [62] Becke, A., D. *J. Chem. Phys.* **1993**, *98*, 5648-5652.
- [63] Perdew, J. P.; Chevary, J. A.; Vosko, S. H.; Jackson, K. A.; Pederson, M. R.; Singh, D. J.; Fiolhais, C. *Physical Review B* **1992**, *46*, 6671.
- [64] Grimme, S. *J. Comput. Chem.* **2004**, *25*, 1463-1473.

- [65] Grimme, S. *J. Comput. Chem.* **2006**, *27*, 1787-1799.
- [66] Harvey, J. N. *Annu. Rep. Prog. Chem., Sect. C: Phys. Chem.* **2006**, *102*, 203-226.
- [67] Grimme, S. *J. Chem. Phys.* **2006**, *124*, 034108.
- [68] De Proft, F.; Geerlings, P. *Chem. Rev.* **2001**, *101*, 1451-1464.
- [69] Schwabe, T.; Grimme, S. *Phys. Chem. Chem. Phys.* **2007**, *9*, 3397-3406.
- [70] Grimme, S.; Neese, F. *J. Chem. Phys.* **2007**, *127*, 154116.
- [71] Neese, F.; Schwabe, T.; Grimme, S. *J. Chem. Phys.* **2007**, *126*, 124115.
- [72] Schwabe, T.; Grimme, S. *Phys. Chem. Chem. Phys.* **2006**, *8*, 4398-4401.
- [73] Ziegler, T. *Chem. Rev.* **1991**, *91*, 651-667.
- [74] Bernardi, F.; Bottoni, A.; Garavelli, M. *Quant. Struct. - Act. Rel.* **2002**, *21*, 128-148.
- [75] Geerlings, P.; De Proft, F.; Langenaeker, W. *Chemical Reviews* **2003**, *103*, 1793-1873.
- [76] Elstner, M.; Porezag, D.; Jungnickel, G.; Elsner, J.; Haugk, M.; Frauenheim, T.; Suhai, S.; Seifert, G. *Physical Review B* **1998**, *58*, 7260.
- [77] Seifert, G. *J. Phys. Chem. A* **2007**, *111*, 5609-5613.
- [78] Otte, N.; Scholten, M.; Thiel, W. *J. Phys. Chem. A* **2007**, *111*, 5751-5755.
- [79] Pu, J.; Gao, J.; Truhlar, D. G. *J. Phys. Chem. A* **2004**, *108*, 5454-5463.
- [80] Cai, Z. L.; Lopez, P.; Reimers, J. R.; Cui, Q.; Elstner, M. *J. Phys. Chem. A* **2007**, *111*, 5743-5750.
- [81] Zheng, G.; Witek, H. A.; Bobadova-Parvanova, P.; Irle, S.; Musaev, D. G.; Prabhakar, R.; Morokuma, K.; Lundberg, M.; Elstner, M.; Kohler, C.; Frauenheim, T. *J. Chem. Theory Comput.* **2007**, *3*, 1349-1367.
- [82] Heine, T.; DosSantos, H. F.; Patchkovskii, S.; Duarte, H. A. *J. Phys. Chem. A* **2007**, *111*, 5648-5654.
- [83] Hu, H.; Lu, Z.; Elstner, M.; Hermans, J.; Yang, W. *J. Phys. Chem. A* **2007**, *111*, 5685-5691.
- [84] Reimers, J. R.; Solomon, G. C.; Gagliardi, A.; Bilic, A.; Hush, N. S.; Frauenheim, T.; DiCarlo, A.; Pecchia, A. *J. Phys. Chem. A* **2007**, *111*, 5692-5702.
- [85] Frauenheim, T.; Seifert, G.; Elstner, M.; Niehaus, T.; hler, C.; Amkreutz, M.; Sternberg, M.; Hajnal, Z.; Carlo, A. D.; Suhai, S.; ndor *J. Phys.: Condens. Matter* **2002**, *14*, 3015-3047.
- [86] Bhattacharyya, S.; Stankovich, M. T.; Truhlar, D. G.; Gao, J. *J. Phys. Chem. A* **2007**, *111*, 5729-5742.
- [87] Kubar, T.; Jurecka, P.; Cerny, J.; Rezac, J.; Otyepka, M.; Valdes, H.; Hobza, P. *J. Phys. Chem. A* **2007**, *111*, 5642-5647.
- [88] Riccardi, D.; Cui, Q. *J. Phys. Chem. A* **2007**, *111*, 5703-5711.
- [89] Riccardi, D.; Li, G.; Cui, Q. *J. Phys. Chem. B* **2004**, *108*, 6467-6478.
- [90] Seabra, G. M.; Walker, R. C.; Elstner, M.; Case, D. A.; Roitberg, A. E. *J. Phys. Chem. A* **2007**, *111*, 5655-5664.
- [91] Woodcock, H. L.; Hodoscek, M.; Brooks, B. R. *J. Phys. Chem. A* **2007**, *111*, 5720-5728.
- [92] Xu, D.; Guo, H.; Cui, Q. *J. Phys. Chem. A* **2007**, *111*, 5630-5636.
- [93] Whitten, J. L. *J. Chem. Phys.* **1973**, *58*, 4496-4501.
- [94] Dunlap, B. I.; Connolly, J. W. D.; Sabin, J. R. *J. Chem. Phys.* **1979**, *71*, 3396-3402.
- [95] Vahtras, O.; Almlof, J.; Feyereisen, M. W. *Chem. Phys. Lett.* **1993**, *213*, 514-518.
- [96] Weigend, F. *Phys. Chem. Chem. Phys.* **2002**, *4*, 4285-4291.
- [97] Feyereisen, M.; Fitzgerald, G.; Komornicki, A. *Chem. Phys. Lett.* **1993**, *208*, 359-363.
- [98] Bernholdt, D. E.; Harrison, R. J. *Chem. Phys. Lett.* **1996**, *250*, 477-484.
- [99] Weigend, F.; Haser, M.; Patzelt, H.; Ahlrichs, R. *Chem. Phys. Lett.* **1998**, *294*, 143-152.
- [100] Eichkorn, K.; Treutler, O.; Ohm, H.; Haser, M.; Ahlrichs, R. *Chem. Phys. Lett.* **1995**, *240*, 283-289.

- [101] Eichkorn, K.; Weigend, F.; Treutler, O.; Ahlrichs, R. *Theor. Chem. Acc.* **1997**, *97*, 119-124.
- [102] Skylaris, C. K.; Gagliardi, L.; Handy, N. C.; Ioannou, A. G.; Spencer, S.; Willetts, A. *Journal of Molecular Structure: THEOCHEM* **2000**, *501-502*, 229-239.
- [103] Sierka, M.; Hogekamp, A.; Ahlrichs, R. *J. Chem. Phys.* **2003**, *118*, 9136-9148.
- [104] Altoè, P.; Stenta, M.; Bottoni, A.; Garavelli, M. *Theor. Chem. Acc.* **2007**, *118*, 219-240.
- [105] Bakowies, D.; Thiel, W. *J. Phys. Chem.* **1996**, *100*, 10580-10594.
- [106] Sherwood, P. *NIC series* **2000**, *3*, 285-305.
- [107] Warshel, A.; Levitt, M. *J. Mol. Biol.* **1976**, *103*, 227-249.
- [108] Gao, J. In *Reviews in Computational Chemistry*; Lipkowitz, K. B., Boyd, D. B., Eds.; VHC Publishers: New York, 1995; Vol. 7, p 119-185.
- [109] Lin, H.; Truhlar, D. G. *Theor. Chem. Acc.* **2006**, *117*, 185-199.
- [110] Gao, J.; Truhlar, D. G. *Annu. Rev. Phys. Chem.* **2002**, *53*, 467-505.
- [111] Vreven, T.; Mennucci, B.; da Silva, C. O.; Morokuma, K.; Tomasi, J. *J. Chem. Phys.* **2001**, *115*, 62-72.
- [112] Svensson, M.; Humbel, S.; Froese, R. D. J.; Matsubara, T.; Sieber, S.; Morokuma, K. *J. Phys. Chem.* **1996**, *100*, 19357-19363.
- [113] Maseras, F.; Morokuma, K. *J. Comput. Chem.* **1995**, *16*, 1170-1179.
- [114] Sherwood, P.; de Vries, A. H.; Guest, M. F.; Schreckenbach, G.; Catlow, C. R. A.; French, S. A.; Sokol, A. A.; Bromley, S. T.; Thiel, W.; Turner, A. J. *J. Mol. Struct.* **2003**, *632*, 1-28.
- [115] Peng, C.; Ayala, P. Y. S., H. Bernhard; Frisch, M. J. *J. Comput. Chem.* **1996**, *17*, 49-56.
- [116] Vreven, T.; Morokuma, K.; Farkas, Ö.; Schlegel, H. B.; Frisch, M. J. *J. Comput. Chem.* **2003**, *24*, 760-769.
- [117] Prat-Resina, X.; Bofill, J. M.; Gonzalez-Lafont, A.; Lluch, J. M. *Int. J. Quantum Chem.* **2004**, *98*, 367-377.
- [118] Prat-Resina, X.; Garcia-Viloca, M.; Gonzalez-Lafont, A.; Lluch, J. M. *Phys. Chem. Chem. Phys.* **2002**, *4*, 5365-5371.
- [119] Klahn, M.; Braun-Sand, S.; Rosta, E.; Warshel, A. *J. Phys. Chem. B* **2005**, *109*, 15645-15650.
- [120] Field, M. J.; Bash, P. A.; Karplus, M. *J. Comput. Chem.* **1990**, *11*, 700-733.
- [121] Singh, U. C.; Kollman, P. A. *J. Comput. Chem.* **1986**, *7*, 718-730.
- [122] Ferre, N.; Olivucci, M. *J. Mol. Struct.* **2003**, *632*, 71-82.
- [123] Pu, J.; Gao, J.; Truhlar, D. G. *J. Phys. Chem. A* **2004**, *108*, 632-650.
- [124] Gao, J.; Amara, P.; Alhambra, C.; Field, M. J. *J. Phys. Chem. A* **1998**, *102*, 4714-4721.
- [125] Théry, V.; Rinaldi, D.; Rivail, J.-L.; Maignet, B.; Ferenczy, G. G. *J. Comput. Chem.* **1994**, *15*, 269-282.
- [126] Kastner, J.; Thiel, S.; Senn, H. M.; Sherwood, P.; Thiel, W. *J. Chem. Theory Comput.* **2007**, *3*, 1064-1072.
- [127] Breneman, C. M.; Wiberg, K. B. *J. Comput. Chem.* **1990**, *11*, 361-373.
- [128] Singh, U. C.; Kollman, P. A. *J. Comput. Chem.* **1984**, *5*, 129-145.
- [129] Besler, B. H.; Jr, M. K. M.; Kollman, P. A. *J. Comput. Chem.* **1990**, *11*, 431-439.
- [130] Karlstrom, G.; Lindh, R.; Malmqvist, P.-A.; Roos, B. O.; Ryde, U.; Veryazov, V.; Widmark, P.-O.; Cossi, M.; Schimmelpfennig, B.; Neogrady, P.; Seijo, L. *Computational Materials Science* **2003**, *28*, 222-239.
- [131] Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, J., J. A.; ; 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. In *Gaussian 03, Revision C.02*; Gaussian, Inc., Wallingford CT 2004.
- [132] Ahlrichs, R.; Bar, M.; Haser, M.; Horn, H.; Kolmel, C. *Chem. Phys. Lett.* **1989**, *162*, 165-169.
- [133] Neese, F. In *ORCA An ab initio, DFT and semiempirical SCF-MO package* 2006.
- [134] Sinnecker, S.; Neese, F. *J. Comput. Chem.* **2006**, *27*, 1463-1475.
- [135] Wales, D. J. *Energy Landscapes: Applications to Clusters, Biomolecules and Glasses*; Cambridge University Press: Cambridge, 2003.
- [136] van Rossum, G.; Drake, F. L. In *Python Reference Manual, PythonLabs, Virginia, USA, 2001*. Available at <http://www.python.org>.
- [137] Cui, Q.; Karplus, M. *J. Am. Chem. Soc.* **2001**, *123*, 2284-2290.
- [138] Larsson, P. E.; Marti, S.; Moliner, V.; Andres, J. *Abstracts of Papers of the American Chemical Society* **2004**, 228, U291-U291.
- [139] Marti, S.; Moliner, V. *J. Chem. Theory Comput.* **2005**, *1*, 1008-1016.
- [140] Roca, M.; Marti, S.; Andres, J.; Moliner, V.; Tunon, I.; Bertran, J.; Williams, I. H. *J. Am. Chem. Soc.* **2003**, *125*, 7726-7737.
- [141] Soriano, A.; Silla, E.; Tuñón, I.; Martí, S.; Moliner, V.; Bertrán, J. *Theor. Chem. Acc.* **2004**, *V112*, 327-334.
- [142] Gordon, J. C.; Myers, J. B.; Folta, T.; Shoja, V.; Heath, L. S.; Onufriev, A. *Nucleic Acids Res.* **2005**, *33*, 368-371.
- [143] Gatti, F. G.; Leon, S.; Wong, J. K. Y.; Bottari, G.; Altieri, A.; Morales, M. A. F.; Teat, S. J.; Frochot, C.; Leigh, D. A.; Brouwer, A. M.; Zerbetto, F. *Proc. Natl. Acad. Sci. U. S. A.* **2003**, *100*, 10-14.
- [144] Allinger, N. L.; Yuh, Y. H.; Lii, J. H. *J. Am. Chem. Soc.* **1989**, *111*, 8551-8556.
- [145] Cossi, M.; Barone, V. *J. Chem. Phys.* **1998**, *109*, 6246-6254.
- [146] Hawkins, G. D.; Cramer, C. J.; Truhlar, D. G. *J. Phys. Chem.* **1996**, *100*, 19824-19839.
- [147] Corchado, J. C.; Sanchez, M. L.; Aguilar, M. A. *J. Am. Chem. Soc.* **2004**, *126*, 7311-7319.
- [148] Miura, N.; Taniguchi, T.; Monde, K.; Nishimura, S.-I. *Chem. Phys. Lett.* **2006**, *419*, 326-332.
- [149] Momany, F. A.; Appell, M.; Willett, J. L.; Bosma, W. B. *Carbohydr. Res.* **2005**, *340*, 1638-1655.
- [150] Momany, F. A.; Appell, M.; Willett, J. L.; Schnupf, U.; Bosma, W. B. *Carbohydr. Res.* **2006**, *341*, 525-537.
- [151] Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 1372-1377.
- [152] Cossi, M.; Rega, N.; Scalmani, G.; Barone, V. *J. Comput. Chem.* **2003**, *24*, 669-681.
- [153] Woods, R. J.; Dwek, R. A.; Edge, C.; Fraser-Reid, J. B. *J. Phys. Chem.* **1995**, *99*, 3832-3846.
- [154] He, X.; Bell, A. F.; Tonge, P. J. *J. Phys. Chem. B* **2002**, *106*, 6056-6066.

- [155] Nielsen, S. B.; Lapiere, A.; Andersen, J. U.; Pedersen, U. V.; Tomita, S.; Andersen, L. H. *Phys. Rev. Lett.* **2001**, *87*, 228102.
- [156] Tsien, R. Y. *Annu. Rev. Biochem.* **1998**, *67*, 509-544.
- [157] Andersen, L. H.; Nielsen, I. B.; Kristensen, M. B.; ElGhazaly, M. O. A.; Haacke, S.; Nielsen, M. B.; Petersen, M. A. *J. Am. Chem. Soc.* **2005**, *127*, 12347-12350.
- [158] Freedman, K. A.; Becker, R. S. *J. Am. Chem. Soc.* **1986**, *108*, 1245-1251.
- [159] Sinicropi, A.; Andruniow, T.; Ferre, N.; Basosi, R.; Olivucci, M. *J. Am. Chem. Soc.* **2005**, *127*, 11534-11535.
- [160] Altoe, P.; Bernardi, F.; Garavelli, M.; Orlandi, G.; Negri, F. *J. Am. Chem. Soc.* **2005**, *127*, 3952-3963.
- [161] Jakalian, A.; Bush, B. L.; Jack, D. B.; Bayly, C. I. *J. Comput. Chem.* **2000**, *21*, 132-146.
- [162] Jakalian, A.; Jack, D. B.; Bayly, C. I. *J. Comput. Chem.* **2002**, *23*, 1623-1641.
- [163] Martin, M. E.; Negri, F.; Olivucci, M. *J. Am. Chem. Soc.* **2004**, *126*, 5452-5464.
- [164] Schellenberg, P.; Johnson, E.; Esposito, A. P.; Reid, P. J.; Parson, W. W. *J. Phys. Chem. B* **2001**, *105*, 5316-5322.
- [165] Creemers, T. M. H.; Lock, A. J.; Subramaniam, V.; Jovin, T. M.; Volker, S. *Nat. Struct. Mol. Biol.* **1999**, *6*, 557-560.
- [166] Feese, M. D.; Faber, H. R.; Bystrom, C. E.; Pettigrew, D. W.; Remington, S. J. *Structure* **1998**, *6*, 1407-1418.
- [167] Cembran, A.; Bernardi, F.; Olivucci, M.; Garavelli, M. *J. Am. Chem. Soc.* **2004**, *126*, 16018-16037.
- [168] Vreven, T.; Bernardi, F.; Garavelli, M.; Olivucci, M.; Robb, M. A.; Schlegel, H. B. *J. Am. Chem. Soc.* **1997**, *119*, 12687-12688.
- [169] Cembran, A.; Gonzalez-Luque, R.; Altoe, P.; Merchan, M.; Bernardi, F.; Olivucci, M.; Garavelli, M. *J. Phys. Chem. A* **2005**, *109*, 6597-6605.
- [170] Andruniow, T.; Ferre, N.; Olivucci, M. *Proc. Natl. Acad. Sci. U. S. A.* **2004**, *101*, 17908-17913.
- [171] Cembran, A.; Gonzalez-Luque, R.; Altoe, P.; Merchan, M.; Bernardi, F.; Olivucci, M.; Garavelli, M. *J. Phys. Chem. A* **2005**, *109*, 6597-6605.
- [172] Cembran, A.; Bernardi, F.; Olivucci, M.; Garavelli, M. *Proc. Natl. Acad. Sci. U. S. A.* **2005**, *102*, 6255-6260.
- [173] Houjou, H.; Inoue, Y.; Sakurai, M. *J. Am. Chem. Soc.* **1998**, *120*, 4459-4470.
- [174] Houjou, H.; Sakurai, M.; Inoue, Y. *Chem. Lett.* **1996**, 1075-1076.
- [175] Prabu-Jeyabalan, M.; Nalivaika, E.; Schiffer, C. A. *J. Mol. Biol.* **2000**, *301*, 1207-1220.
- [176] Davies, D. R. *ANNU. REV. BIOPHYS. BIOPHYS. CHEM.* **1990**, *19*, 189-215.
- [177] Dreyer, G. B.; Metcalf, B. W.; Tomaszek, T. A.; Carr, T. J.; Chandler, A. C.; Hyland, L.; Fakhoury, S. A.; Magaard, V. W.; Moore, M. L.; Strickler, J. E.; Debouck, C.; Meek, T. D. *Proc. Natl. Acad. Sci. U. S. A.* **1989**, *86*, 9752-9756.
- [178] Piana, S.; Bucher, D.; Carloni, P.; Rothlisberger, U. *J. Phys. Chem. B* **2004**, *108*, 11139-11149.
- [179] Hensen, C.; Hermann, J. C.; Nam, K.; Ma, S.; Gao, J.; Holtje, H. D. *J. Med. Chem.* **2004**, *47*, 6673-6680.
- [180] Cecconi, F.; Micheletti, C.; Carloni, P.; Maritan, A. *Proteins: Struct., Funct., Genet.* **2001**, *43*, 365-372.
- [181] Warshel, A. *J. Biol. Chem.* **1998**, *273*, 27035-27038.
- [182] Liu, H.; Muller-Plathe, F.; van Gunsteren, W. F. *J. Mol. Biol.* **1996**, *261*, 454-469.
- [183] Lee, H.; Darden, T. A.; Pedersen, L. G. *J. Am. Chem. Soc.* **1996**, *118*, 3946-3950.
- [184] Piana, S.; Sebastiani, D.; Carloni, P.; Parrinello, M. *J. Am. Chem. Soc.* **2001**, *123*, 8730-8737.

- [185] Harte, W. E. B., Jr. David L. *J. Am. Chem. Soc.* **1993**, *115*, 3883-3886.
- [186] Rodriguez, E. J.; Angeles, T. S.; Meek, T. D. *Biochemistry* **1993**, *32*, 12380-12385.
- [187] Hyland, L. J.; Tomaszek, T. A.; Roberts, J. G. D.; Carr, S. A.; Magaard, V. W.; Bryan, H. L.; Fakhoury, S. A.; Moore, M. L.; Minnich, M. D. *Biochemistry* **1991**, *30*, 8441-8453.
- [188] Hyland, L. J.; Tomaszek, T. A.; Meek, J. T. D. *Biochemistry* **1991**, *30*, 8454-8463.
- [189] Cavalli, A.; Carloni, P.; Recanatini, M. *Chem. Rev.* **2006**, *106*, 3497-3519.
- [190] Pillai, B.; K., K. K.; Hosur, V., M. *Proteins: Struct., Funct., Genet.* **2001**, *43*, 57-64.
- [191] Bottoni, A.; Lanza, C. Z.; Miscione, G. P.; Spinelli, D. *J. Am. Chem. Soc.* **2004**, *126*, 1542-1550.
- [192] Bottoni, A.; Miscione, G. P.; De Vivo, M. *Proteins: Struct., Funct., Bioinf.* **2005**, *59*, 118-130.
- [193] Scott, W. R. P.; Schiffer, C. A. *Structure* **2000**, *8*, 1259-1265.
- [194] Piana, S.; Carloni, P.; Parrinello, M. *J. Mol. Biol.* **2002**, *319*, 567-583.
- [195] Basolo, F.; Hayes, J. C.; Neumann, H. M. *J. Am. Chem. Soc.* **1953**, *75*, 5102-5106.
- [196] Basolo, F.; Hayes, J. C.; Neumann, H. M. *J. Am. Chem. Soc.* **1954**, *76*, 3807-3809.
- [197] Dickens, J. E.; Basolo, F.; Neumann, H. M. *J. Am. Chem. Soc.* **1957**, *79*, 1286-1290.
- [198] Dickens, J. E.; Basolo, F.; Neumann, H. M. *J. Am. Chem. Soc.* **1957**, *79*, 6579-6579.
- [199] Jensen, A.; Basolo, F.; Neumann, H. M. *J. Am. Chem. Soc.* **1958**, *80*, 2354-2358.
- [200] Seiden, L.; Basolo, F.; Neumann, H. M. *J. Am. Chem. Soc.* **1959**, *81*, 3809-3813.
- [201] Van Meter, F. M.; Neumann, H. M. *J. Am. Chem. Soc.* **1976**, *98*, 1388-1394.
- [202] Amis, E. S. *Solvent Effects on Reaction Rates and Mechanisms*; Academic Press: New York, 1966.
- [203] Yamagishi, A. *Inorg. Chem.* **1986**, *25*, 55-57.
- [204] Rodger, A.; Schipper, P. E. *J. Phys. Chem.* **1987**, *91*, 189-195.
- [205] Rodger, A.; Johnson, B. F. G. *Polyhedron* **1988**, *7*, 1107-1120.
- [206] Rodger, A.; Schipper, P. E. *Inorg. Chem.* **1988**, *27*, 458-466.
- [207] Rodger, A.; Johnson, B. F. G. *Inorg. Chem.* **1988**, *27*, 3061-3062.
- [208] Peng, C.; Ayala, P. Y.; Schlegel, H. B.; Frisch, M. J. *J. Comput. Chem.* **1996**, *17*, 49-56.
- [209] Perdew, J. P. *Physical Review B* **1986**, *33*, 8822.
- [210] Treutler, O.; Ahlrichs, R. *J. Chem. Phys.* **1995**, *102*, 346-354.
- [211] Weigend, F.; Ahlrichs, R. *Phys. Chem. Chem. Phys.* **2005**, *7*, 3297-3305.
- [212] Weigend, F. *Phys. Chem. Chem. Phys.* **2006**, *8*, 1057-1065.
- [213] Yoshimura, T.; Esak, N. *J. Biosci. Bioeng.* **2003**, *96*, 103-109.
- [214] Amadasi, A.; Bertoldi, M.; Contestabile, R.; Bettati, S.; Cellini, B.; Luigi di Salvo, M.; Borri-Voltattorni, C.; Bossa, F.; Mozzarelli, A. *Curr. Med. Chem.* **2007**, *14*, 1291-1324.
- [215] Stadtman, T. C.; Elliott, P. *J. Biol. Chem.* **1957**, *228*, 983-997.
- [216] Cardinale, G. J.; Abeles, R. H. *Biochemistry* **1968**, *7*, 3970-3978.
- [217] Rudnick, G.; Abeles, R. H. *Biochemistry* **1975**, *14*, 4515-4522.
- [218] Chamond, N.; Gregoire, C.; Coatnoan, N.; Rougeot, C.; Freitas, L. H.; da Silveira, J. F.; Degrave, W. M.; Minoprio, P. *J. Biol. Chem.* **2003**, *278*, 15484-15494.
- [219] Chamond, N.; Goytia, M.; Coatnoan, N.; Barale, J. C.; Cosson, A.; Degrave, W. M.; Minoprio, P. *Mol. Microbiol.* **2005**, *58*, 46-60.
- [220] Tonelli, R. R.; Silber, A. M.; Almeida-de-Faria, M.; Hirata, I. Y.; Colli, W.; Alves, M. J. M. *Cell. Microbiol.* **2004**, *6*, 733-741.
- [221] Buschiazio, A.; Goytia, M.; Schaeffer, F.; Degrave, W.; Shepard, W.; Gregoire, C.; Chamond, N.; Cosson, A.; Berneman, A.; Coatnoan, N.; Alzari, P. M.; Minoprio, P. *Proc. Natl. Acad. Sci. U. S. A.* **2006**, *103*, 1705-1710.

- [222] Feig, M.; Karanicolas, J.; Brooks, C. L. In *MMTSB Tool Set, The Scripps Research Institute*. 2001.
- [223] Godbout, N.; Salahub, D. R.; Andzelm, J.; Wimmer, E. *Can. J. Chem.* **1992**, *70*, 560-571.
- [224] Antoniou, D.; Basner, J.; Nunez, S.; Schwartz, S. D. *Chem. Rev.* **2006**, *106*, 3170-3187.
- [225] Warshel, A.; Sharma, P. K.; Kato, M.; Xiang, Y.; Liu, H.; Olsson, M. H. M. *Chem. Rev.* **2006**, *106*, 3210-3235.
- [226] Cirilli, M.; Zheng, R. J.; Scapin, G.; Blanchard, J. S. *Biochemistry* **1998**, *37*, 16452-16458.
- [227] Koo, C. W.; Blanchard, J. S. *Biochemistry* **1999**, *38*, 4416-4422.
- [228] Pillai, B.; Cherney, M. M.; Diaper, C. M.; Sutherland, A.; Blanchard, J. S.; Vederas, J. C.; James, M. N. G. *Proc. Natl. Acad. Sci. U. S. A.* **2006**, *103*, 8668-8673.
- [229] Diaper, C. M.; Sutherland, A.; Pillai, B.; James, M. N. G.; Semchuk, P.; Blanchard, J. S.; Vederas, J. C. *Org. Biomol. Chem.* **2005**, *3*, 4402-4411.
- [230] Kuntz, I. D.; Blaney, J. M.; Oatley, S. J.; Langridge, R.; Ferrin, T. E. *J. Mol. Biol.* **1982**, *161*, 269-288.
- [231] Moustakas, D. T.; Lang, P. T.; Pegg, S.; Pettersen, E.; Kuntz, I. D.; Brooijmans, N.; Rizzo, R. C. *J. Comput.-Aided Mol. Des.* **2006**, *20*, 601-619.
- [232] Bottoni, A.; Cosimelli, B.; Scavetta, E.; Spinelli, D.; Spisani, R.; Stenta, M.; Tonelli, D. *Mol. Phys.* **2006**, *104*, 2961-2982.
- [233] DiLabio, G. A.; Wright, J. S. *Chem. Phys. Lett.* **1998**, *297*, 181-186.
- [234] Wright, J. S.; Johnson, E. R.; DiLabio, G. A. *J. Am. Chem. Soc.* **2001**, *123*, 1173-1183.
- [235] Lewis, A.; Bumpus, J. A.; Truhlar, D. G.; Cramer, C. J. *J. Chem. Educ.* **2004**, *81*, 596-604.
- [236] Bianchi, L.; Dell'Erba, C.; Maccagno, M.; Morganti, S.; Petrillo, G.; Rizzato, E.; Sancassan, F.; Severi, E.; Spinelli, D.; Tavani, C. *Arkivoc* **2006**, *vii*, 169-185.
- [237] Dell'Erba, C.; Mele, A.; Novi, M.; Petrillo, G.; Stagnaro, P. *Tetrahedron* **1992**, *48*, 4407-4418.
- [238] Armaroli, T.; Dell'Erba, C.; Gabellini, A.; Gasparrini, F.; Mugnoli, A.; Novi, M.; Petrillo, G. T., C. *Eur. J. Org. Chem.* **2002**, 1284-1291.
- [239] Bianchi, L.; Dell'Erba, C.; Gasparrini, F.; Novi, M.; Petrillo, G.; Sancassan, F.; Tavani, C. *Arkivoc* **2002**, *xi*, 142-158.
- [240] Bianchi, L.; Giorgi, G.; Maccagno, M.; Petrillo, G.; Rizzato, E.; Spinelli, D.; Stenta, M.; Tavani, C. *Lett. Org. Chem.* **2007**, *4*, 268-272.
- [241] Johnson, C. K.; ORTEP-II: A FORTRAN Thermal-Ellipsoid Plot Program for Crystal Structure Illustrations: Oak Ridge National Laboratory Report ORNL-5138, 1976.
- [242] Kral, V.; Leseticky, L. *Collect. Czech. Chem. Commun.* **1975**, *40*, 2816-2825.
- [243] Attanasi, O. A.; Filippone, G. F. P.; Lillini, S.; Mantellini, F.; Spinelli, D.; Stenta, M. *Adv. Synth. Catal.* **2007**, *349*, 907-915.
- [244] Thom H. Dunning, Jr. *J. Chem. Phys.* **1989**, *90*, 1007-1023.
- [245] Tidwell, T. T. *Ketenes*; Wiley: New York.
- [246] Staudinger, H. *Justus Liebigs Ann. Chem.* **1907**, *356*, 51.
- [247] Walsh, C. T.; Wright, G. D. *Chem. Rev.* **2005**, *105*, 391-394.
- [248] Southgate, R.; Branch, C.; Coulton, S.; Hunt, E. In *Recent progress in the Chemical Synthesis of Antibiotics and Related Microbial Products*; Luckacs, G., Ed.; Springer-Verlag: Berlin, 1993; Vol. 2, p 621.
- [249] Buynak, J. D. *Curr. Med. Chem.* **2004**, *11*, 1951-1964.
- [250] Lysek, R.; Borsuk, K.; Furman, B.; Kałuzza, Z.; Kazimierski, A.; Chmielewski, M. *Curr. Med. Chem.* **2004**, *11*, 1813-1835.

- [251] Fisher, J. F.; Meroueh, S. O.; Mobashery, S. *Chem. Rev* **2005**, *105*, 395-424.
- [252] Clader, J. W. *J. Med. Chem.* **2004**, *47*, 1-9.
- [253] Burnett, D. A. *Curr. Med. Chem.* **2004**, *11*, 1873-1887.
- [254] Wilmouth, R. C.; Westwood, N. J.; Anderson, K.; Brownlee, W.; Claridge, T. D. W.; Clifton, I. J.; Pritchard, G. J.; Aplin, R. T.; Schofield, C. J. *Biochemistry* **1998**, *37*, 17506-17513.
- [255] Bonneau, P. R.; Hasani, F.; Plouffe, C.; Malenfant, E.; LaPlante, S. R.; Guse, I.; Ogilvie, W. W.; Plante, R.; Davidson, W. C.; Hopkins, J. L.; Morelock, M. M.; Cordingley, M. G.; Deziel, R. *J. Am. Chem. Soc.* **1999**, *121*, 2965-2973.
- [256] Bharatam, P. V.; Kumar, R. S.; Mahajan, M. P. *Org. Lett.* **2000**, *2*, 2725-2728.
- [257] Komatsu, M.; Ogawa, H.; Mohri, M.; Ohshiro, Y. *Tetrahedron Lett.* **1990**, *31*, 3627-3630.
- [258] Kirmse, W.; Rondan, N. G.; Houk, K. N. *J. Am. Chem. Soc.* **1984**, *106*, 7989-7991.
- [259] Georg, G. I.; He, P.; Kant, J.; Wu, Z. J. *J. Org. Chem.* **1993**, *58*, 5771-5778.
- [260] Georg, G. I. *The Organic Chemistry of β -Lactams*; Verlag Chemie: New York, 1993.
- [261] Palomo, C.; Aizpurua, J. M.; Ganboa, I.; Oiarbide, M. *Eur. J. Org. Chem.* **1999**, *1999*, 3223-3235.
- [262] Singh, G. S. *Tetrahedron* **2003**, *59*, 7631-7649.
- [263] Benaglia, M.; Cinquini, M.; Cozzi, F. *Eur. J. Org. Chem.* **2000**, *2000*, 563-572.
- [264] Hart, D. J.; Ha, D. C. *Chem. Rev* **1989**, *89*, 1447-1465.
- [265] Moore, H. W.; Hernandez, L.; Sing, A. *J. Am. Chem. Soc.* **1976**, *98*, 3728-3730.
- [266] Alcaide, B.; Almendros, P. *Curr. Med. Chem.* **2004**, *11*, 1921-1949.
- [267] Deshmukh, A. R. A. S.; Bhawal, B. M.; Krishnaswamy, D.; Govande, V. V.; Shinkre, B. A.; Jayanthi, A. *Curr. Med. Chem.* **2004**, *11*, 1889-1920.
- [268] Palomo, C.; Aizpurua, J. M.; Ganboa, I.; Oiarbide, M. *Curr. Med. Chem.* **2004**, *11*, 1837-1872.
- [269] Jiao, L.; Liang, Y.; Xu, J. *J. Am. Chem. Soc.* **2006**, *128*, 6060-6069.
- [270] Sordo, J. A.; Gonzalez, J.; Sordo, T. L. *J. Am. Chem. Soc.* **1992**, *114*, 6249-6251.
- [271] Venturini, A.; Gonzalez, J. *J. Org. Chem.* **2002**, *67*, 9089-9092.
- [272] Macias, A.; Alonso, E.; delPozo, C.; Venturini, A.; Gonzalez, J. *J. Org. Chem.* **2004**, *69*, 7004-7012.
- [273] Bongini, A.; Panunzio, M.; Piersanti, G.; Bandini, E.; Martelli, G.; Spunta, G.; Venturini, A. *Eur. J. Org. Chem.* **2000**, *2000*, 2379-2390.
- [274] Alajarín, M.; Vidal, A.; Tovar, F.; Arrieta, A.; Lecea, B.; Cossío, F. P. *Chemistry - A European Journal* **1999**, *5*, 1106-1117.
- [275] Arrastia, I.; Arrieta, A.; Ugalde, J. M.; Cossio, F. P.; Lecea, B. *Tetrahedron Lett.* **1994**, *35*, 7825-7828.
- [276] Arrieta, A.; Cossio, F. P.; Lecea, B. *J. Org. Chem.* **2000**, *65*, 8458-8464.
- [277] Cossio, F. P.; Ugalde, J. M.; Lopez, X.; Lecea, B.; Palomo, C. *J. Am. Chem. Soc.* **1993**, *115*, 995-1004.
- [278] Lecea, B.; Arrastia, I.; Arrieta, A.; Roa, G.; Lopez, X.; Arriortua, M. I.; Ugalde, J. M.; Cossio, F. P. *J. Org. Chem.* **1996**, *61*, 3070-3079.
- [279] Martín-Zamora, E.; Ferrete, A.; Llera, J. M.; Muñoz, J. M.; Pappalardo, R. R.; Fernández, R.; Lassaletta, J. M. *Chemistry - A European Journal* **2004**, *10*, 6111-6129.
- [280] Lee, S. Y.; Kulkarni, Y. S.; Burbaum, B. W.; Johnston, M. I.; Snider, B. B. *J. Org. Chem.* **1988**, *53*, 1848-1855.
- [281] Nguyen Minh, T.; Ha, T. K.; More O'Ferrall, R. A. *J. Org. Chem.* **1990**, *55*, 3251-3256.
- [282] Allen, A. D.; Gong, L.; Tidwell, T. T. *J. Am. Chem. Soc.* **1990**, *112*, 6396-6397.

- [283] Danheiser, R. L.; Brisbois, R. G.; Kowalczyk, J. J.; Miller, R. F. *J. Am. Chem. Soc.* **1990**, *112*, 3093-3100.
- [284] Birchler, A. G.; Liu, F.; Liebeskind, L. S. *J. Org. Chem.* **1994**, *59*, 7737-7745.
- [285] Zhou, C.; Birney, D. M. *J. Am. Chem. Soc.* **2002**, *124*, 5231-5241.
- [286] Barbaro, G.; Battaglia, A.; Giorgianni, P. *J. Org. Chem.* **1987**, *52*, 3289-3296.
- [287] Loebach, J. L.; Bennett, D. M.; Danheiser, R. L. *J. Org. Chem.* **1998**, *63*, 8380-8389.
- [288] Bennett, D. M.; Okamoto, I.; Danheiser, R. L. *Org. Lett.* **1999**, *1*, 641-644.
- [289] Brandstadter, S. M.; Ojima, I.; Herai, K. *Tetrahedron Lett.* **1987**, *28*, 613-616.
- [290] Zhang, S.; Liebeskind, L. S. *J. Org. Chem.* **1999**, *64*, 4042-4049.
- [291] Cardillo, G.; Fabbroni, S.; Gentilucci, L.; Perciaccante, R.; Tolomelli, A. *Tetrahedron: Asymmetry* **2004**, *15*, 593-601.
- [292] Snider, B. B.; Walner, M. *Tetrahedron* **1989**, *45*, 3171-3182.
- [293] Brady, W. T.; Owens, R. A. *Tetrahedron Lett.* **1976**, *17*, 1553-1556.
- [294] Moore, H. W.; Hughes, G. *Tetrahedron Lett.* **1982**, *23*, 4003-4006.
- [295] Araki, K.; Wichtowski, J. A.; Welch, J. T. *Tetrahedron Lett.* **1991**, *32*, 5461-5464.
- [296] Dey, P. D.; Sharma, A. K.; Bharatam, P. V.; Mahajan, M. P. *Tetrahedron* **1997**, *53*, 13829-13840.
- [297] Sharma, A. K.; Mahajan, M. P. *Tetrahedron* **1997**, *53*, 13841-13854.
- [298] Buynak, J. D.; Mathew, J.; Narayana Rao, M.; Haley, E.; George, C.; Siriwardane, U. *J. Chem. Soc., Chem. Commun.* **1987**, 735-737.
- [299] Cardillo, G.; Fabbroni, S.; Gentilucci, L.; Perciaccante, R.; Piccinelli, F.; Tolomelli, A. *Tetrahedron* **2004**, *60*, 5031-5040.
- [300] Benfatti, F.; Cardillo, G.; Fabbroni, S.; Gentilucci, L.; Perciaccante, R.; Piccinelli, F.; Tolomelli, A. *Synthesis* **2005**, 61-70.
- [301] Benfatti, F.; Cardillo, G.; Fabbroni, S.; Gentilucci, L.; Perciaccante, R.; Tolomelli, A. *Arkivoc* **2005**, *vi*, 136-152.
- [302] Cardillo, G.; Fabbroni, S.; Gentilucci, L.; Perciaccante, R.; Piccinelli, F.; Tolomelli, A. *Org. Lett.* **2005**, *7*, 533-536.
- [303] Cardillo, G.; Fabbroni, S.; Gentilucci, L.; Perciaccante, R.; Tolomelli, A. *Adv. Synth. Catal.* **2005**, *347*, 833-838.
- [304] Benfatti, F.; Cardillo, G.; Gentilucci, L.; Tolomelli, A. *Eur. J. Org. Chem.* **2007**, *2007*, 3199-3205.
- [305] Benfatti, F.; Cardillo, G.; Gentilucci, L.; Tolomelli, A.; Monari, M.; Piccinelli, F. *Adv. Synth. Catal.* **2007**, *349*, 1256-1264.
- [306] Benfatti, F.; Cardillo, G.; Fabbroni, S.; Gentilucci, L.; Perciaccante, R.; Tolomelli, A.; Baiula, M.; Spampinato, S. *Tetrahedron: Asymmetry* **2006**, *17*, 167-170.
- [307] Benfatti, F.; Cardillo, G.; Gentilucci, L.; Perciaccante, R.; Tolomelli, A.; Catapano, A. *J. Org. Chem.* **2006**, *71*, 9229-9232.
- [308] Benfatti, F.; Cardillo, G.; Fabbroni, S.; Galzerano, P.; Gentilucci, L.; Juris, R.; Tolomelli, A.; Baiula, M.; Sparta, A.; Spampinato, S. *Bioorg. Med. Chem.* **2007**, *15*, 7380-7390.
- [309] Benfatti, F.; Cardillo, G.; Gentilucci, L.; Tolomelli, A. *Bioorg. Med. Chem. Lett.* **2007**, *17*, 1946-1950.
- [310] Benfatti, F.; Cardillo, G.; Gentilucci, L.; Perciaccante, R.; Tolomelli, A. *Synlett* **2005**, 2204-2208.
- [311] Bose, A. K.; Krishnan, L.; Wagle, D. R.; Manhas, M. S. *Tetrahedron Lett.* **1986**, *27*, 5955-5958.
- [312] Bose, A. K.; Manhas, M. S.; Ghosh, M.; Shah, M.; Raju, V. S.; Bari, S. S.; Newaz, S. N.; Banik, B. K.; Chaudhary, A. G.; Barakat, K. J. *J. Org. Chem.* **1991**, *56*, 6968-6970.
- [313] Manhas, M. S.; Ghosh, M.; Bose, A. K. *J. Org. Chem.* **1990**, *55*, 575-580.

- [314] Firestone, R. A.; Barker, P. L.; Pisano, J. M.; Ashe, B. M.; Dahlgren, M. E. *Tetrahedron* **1990**, *46*, 2255-2262.
- [315] Snider, B. B.; Ron, E.; Burbaum, B. W. *J. Org. Chem.* **1987**, *52*, 5413-5419.
- [316] Frisch, M. J. 2004.
- [317] Barone, V.; Cossi, M. *J. Phys. Chem. A* **1998**, *102*, 1995-2001.
- [318] Barone, V.; Cossi, M.; Tomasi, J. *J. Comput. Chem.* **1998**, *19*, 404-417.
- [319] Hegedus, L. S.; Montgomery, J.; Narukawa, Y.; Snustad, D. C. *J. Am. Chem. Soc.* **1991**, *113*, 5784-5791.
- [320] Bertini, I.; Luchinat, C. *Acc. Chem. Res.* **1983**, *16*, 272-279.
- [321] Silverman, D. N.; Lindskog, S. *Acc. Chem. Res.* **1988**, *21*, 30-36.
- [322] Christianson, D. W.; Fierke, C. A. *Acc. Chem. Res.* **1996**, *29*, 331-339.
- [323] Woolley, P. *Nature* **1975**, *258*, 677-682
- [324] Liljas, A.; Kannan, K. K.; Bergsten, P. C.; Waara, I.; Friberg, K.; Strandberg, B.; Carlbom, U.; Jarup, L.; Lovgren, S.; Petef, M. *Nature New Biology* **1972**, *235*, 131-137.
- [325] Eriksson, E. A.; Jones, T. A.; Liljas, A. *Zinc Enzymes*; Birkhäuser: Boston, 1986.
- [326] Hakansson, K.; Carlsson, M.; Svensson, L. A.; Liljas, A. *J. Mol. Biol.* **1992**, *227*, 1192-1204.
- [327] Kimura, E. *Acc. Chem. Res.* **2001**, *34*, 171-179.
- [328] Huang, C.; Lesburg, C. A.; Kiefer, L. L.; Fierke, C. A.; Christianson, D. W. *Biochemistry* **1996**, *35*, 3439-3446.
- [329] Lesburg, C. A.; Huang, C. C.; Christianson, D. W.; Fierke, C. A. *Biochemistry* **1997**, *36*, 15780-15791.
- [330] Zhang, X.; Hubbard, C. D.; van Eldik, R. *J. Phys. Chem.* **1996**, *100*, 9161-9171.
- [331] Nair, S. K.; Calderone, T. L.; Christianson, D. W.; Fierke, C. A. *J. Biol. Chem.* **1991**, *266*, 17320-17325.
- [332] Tu, C.; Tripp, B. C.; Ferry, J. G.; Silverman, D. N. *J. Am. Chem. Soc.* **2001**, *123*, 5861-5866.
- [333] Thoms, S. *J. Theor. Biol.* **2002**, *215*, 399-404.
- [334] Liang, J. Y.; Lipscomb, W. N. *J. Am. Chem. Soc.* **1986**, *108*, 5051-5058.
- [335] Liang, J. Y.; Lipscomb, W. N. *Biochemistry* **1987**, *26*, 5293-5301.
- [336] Merz, K. M.; Hoffmann, R.; Dewar, M. J. S. *J. Am. Chem. Soc.* **1989**, *111*, 5636-5649.
- [337] Liang, J.-Y.; Lipscomb, W. N. *Int. J. Quantum Chem.* **1989**, *36*, 299-312.
- [338] Jacob, O.; Cardenas, R.; Tapia, O. *J. Am. Chem. Soc.* **1990**, *112*, 8692-8705.
- [339] Krauss, M.; Garmer, D. R. *J. Am. Chem. Soc.* **1991**, *113*, 6426-6435.
- [340] Zheng, Y. J.; Merz, K. M. *J. Am. Chem. Soc.* **1992**, *114*, 10498-10507.
- [341] Sola, M.; Lledos, A.; Duran, M.; Bertran, J. *J. Am. Chem. Soc.* **1992**, *114*, 869-877.
- [342] Sakurai, M.; Furuki, T.; Inoue, Y. *J. Phys. Chem.* **1995**, *99*, 17789-17794.
- [343] Murcko, M. A. *Theor. Chem. Acc.* **1997**, *96*, 56-60.
- [344] Merz, K. M.; Banci, L. *J. Am. Chem. Soc.* **1997**, *119*, 863-871.
- [345] Lu, D.; Voth, G. A. *J. Am. Chem. Soc.* **1998**, *120*, 4006-4014.
- [346] Toba, S.; Colombo, G.; Merz, K. M. *J. Am. Chem. Soc.* **1999**, *121*, 2290-2302.
- [347] Denisov, V. P.; Jonsson, B. H.; Halle, B. *J. Am. Chem. Soc.* **1999**, *121*, 2327-2328.
- [348] Mauksch, M.; Bräuer, M.; Weston, J.; Anders, E. *ChemBioChem* **2001**, *2*, 190-198.
- [349] Brauer, M.; Perez-Lustres, J. L.; Weston, J.; Anders, E. *Inorg. Chem.* **2002**, *41*, 1454-1463.
- [350] Cui, Q.; Karplus, M. *J. Phys. Chem. B* **2003**, *107*, 1071-1078.
- [351] Bernardi, F.; Bottoni, A.; Miscione, G. P. *Organometallics* **2001**, *20*, 2751-2758.
- [352] Bottoni, A.; Higuieruelo, A. P.; Miscione, G. P. *J. Am. Chem. Soc.* **2002**, *124*, 5506-5513.
- [353] Klamt, A. *J. Phys. Chem.* **1995**, *99*, 2224-2235.

[354] Warshel, A.; Naray-Szabo, G.; Sussman, F.; Hwang, J. K. *Biochemistry* **1989**, *28*, 3629-3637.

List of Figures

Figure 1.1 a) Graphical representation of a three-dimensional energy surface; b) bi-dimensional cross-section. The MEP (Minimum Energy Path) connecting the two critical points M1 and M2 is showed.....	13
Figure 1.2 Schematic representation of the NEB approach.....	18
Figure 1.3 Schematic representation of a harmonic potential energy function and first vibrational levels (ν_0 to ν_2). The Zero Point Energy (ZPE) is indicated.	19
Figure 3.1 Schematic representation of the hybrid QM/MM approach.	35
Figure 3.2 Graphical representation of the tree-layer scheme adopted in our code.....	37
Figure 3.3 Boundary region crossing a covalent bond.....	40
Figure 3.4 The atom-link approach implies a change in the original charges (pod) of the MM region to be introduced in the QM calculation, according to the Electrostatic Embedding scheme. The procedure to obtain the new set of charges (emb) is described in the text.	40
Figure 3.5 Schematic representation of the N (total number of residues) SP calculations performed in a) DFP and b) RFP analysis.....	47
Figure 4.1 Rotaxane ground state structure. a) The dimethyl-fumaramide (ball&stick) is treated at the QM level while the rest (tube) is calculated at the MM level. b) Hydrogen bonds (dashed lines) linking the ring to the thread: QM/MM (red, up), mm3 (blue, centre) and experimental X-ray (black, low) values.	55
Figure 4.2 (a) Haworth and (b) ball & stick representations of the α - and β -D-glucopyranose.....	57
Figure 4.3 Schematic representation of the model system used to study the (α - and β -) D-glucopyranose (solute) in solution. The solvent molecules (water) are partitioned between shell1 (free to move) and shell2 (kept frozen to the initial geometry).	58
Figure 4.4 Molecular Dynamics results from (a) the MM system (ML calculation type) and (b) the QM-MM system (HML calculation type). The behavior of the α - (red line) and β - (black line) anomers is reported, highlighting the production part of each run. The difference between the averaged energies of the two anomers is also reported.....	61
Figure 4.5 Chemical structure of the a) GFP chromophore and its isolated model (HBDI) and b) the protonated Schiff base of 11-cis retinal.	62
Figure 4.6 Bond distances values calculated for the GFP chromophore in vacuo, solvent (PCM and QM/MM) and protein: a) QM/MM values ¹⁵⁹ and b) our QM/MM results.....	64
Figure 4.7 Model system used for the QM/MM study of the anionic GFP. Ball&Stick is the QM part, the tube is the movable medium layer, while the remaining atoms are kept frozen.....	65
Figure 4.8 Bond distances computed for RPSB in vacuo (red) and in protein (black).....	66
Figure 4.9 a) Model system used for the QM study; the • symbol is used to highlight the atoms kept frozen to their initial positions. b) The basis set adopted for the atoms of the QM system is shown using different colors. The model-H region used in QM/MM computations do coincide with the QM model system and the basis set adopted is the same.	69
Figure 4.10 a) Pure QM and b) QM/MM reaction profiles. c) Schematic representation of the M2 and M3 species; the two Asp residues of the aspartic dyad (Asp25 and Asp25') belong to two identical chains.....	69
Figure 4.11 Model system used to study the QM/MM reaction profile of the enzyme HIV1-Pr.....	71
Figure 4.12 The chemical structure of the 2,2-bypiridine (bipy) and 1,10-phenanthroline (phen) ligands.	72
Figure 4.13 Structural representation of the two optical isomers of the octahedral complexes of a generic metal (M^{n+}) with a) 2,2-bypiridine and b) 1,10-phenanthroline.....	73
Figure 4.14 The two enantiomers of the $[Fe(phen)_3]^{2+}$ and the two most plausible enantiomerisation TS.....	74
Figure 4.15 Ball & stick representation of the coordination complexes of Fe and Zn discussed here.....	75
Figure 4.16 Bond distance differences for the $[Fe(bipy)_3]^{2+}$ between a) b3lyp and bp86, b) dftb and b3lyp, dftb and bp86 optimised geometries. The difference values are reported in Å for all the bonds.....	78
Figure 4.17 Bond distance differences for the $[Fe(phen)_3]^{2+}$ between a) b3lyp and bp86, b) dftb and b3lyp, dftb and bp86 optimised geometries. The difference values are reported in Å for all the bonds.....	79
Figure 4.18 Bond distance differences for the $[Zn(bipy)_3]^{2+}$ between a) b3lyp and bp86, b) dftb and b3lyp, dftb and bp86 optimised geometries. The difference values are reported in Å for all the bonds.....	80
Figure 4.19 Bond distance differences for the $[Zn(phen)_3]^{2+}$ between a) b3lyp and bp86, b) dftb and b3lyp, dftb and bp86 optimised geometries. The difference values are reported in Å for all the bonds.....	81
Figure 5.1 Different representation of the QM sub-region (high layer): a) ball & stick model within the whole system; b) yellow surface within the whole system; c) ball & stick model-H stand-alone.	87
Figure 5.2 The model-H sub-system.....	87
Figure 5.3 Different representations of the medium layer: a) ball & stick model (around the yellow surface representing the high layer) within the whole system; b) blue surface within the whole system; c) blue ball & stick model (around the yellow ball & stick model representing the high layer).	88
Figure 5.4 The low layer is all but the blue surface. The frozen low region is represented in 'cartoon', while the free low region is represented using: a) ball & stick model around the blue surface corresponding to the medium layer within the whole system; b) a red surface within the whole system; c) red ball & stick model around the blue and yellow ball & stick models corresponding to the medium and high layers, respectively.	88

Figure 5.5 The residues belonging to the low layer and exposed to solvent are kept frozen during the calculations. The free residues of the low layer are represented by the red surface, while the frozen ones are represented by the cyan surface. Two views (a and b) of the enzyme show that the free low region is almost completely buried in the system.	89
Figure 5.6 Distribution of the added point charges (purple spheres) around the TcPRAC enzyme. Negatively and positively charged residues are represented respectively with red and green surfaces. In Mod0 the total charge of the enzyme is not counterbalanced by any added charge.	91
Figure 5.7 QM portion (i.e. model-H) of the QM/MM model system used for the study of TcPRAC.	91
Figure 5.8 Bar diagram representing the energy difference between the M2a and the M1a minima for all the models studied. The contribution of added point charges is reported in each case as a red portion of the whole bar.	93
Figure 5.9 Gibbs Free Energy profile for the enzymatic reaction.	94
Figure 5.10 Representation of the geometry of model-H subsystem from QM/MM optimisations. In the lower part is reported a superposition with the corresponding atoms as taken from X-ray structure.	95
Figure 5.11 Atomic point charges of the QM region. H atoms are omitted for clarity.	95
Figure 5.12 Schematic representation of the residues interacting with the QM sub-region (shown in green).	96
Figure 5.13 The stability parameter S (kcal mol ⁻¹) as a function of the residue number.	97
Figure 5.14 Residues having a significant effect on electrostatic stabilization/destabilization of TSCa with respect to M1a: a) charged long-range interacting residues (smaller effect, see Table 5.6); b) uncharged residues in close contact with the QM sub-region (higher effect, see Table 5.7).	97
Figure 5.15 Asn133 stabilizes M1a with respect to TSCa. In TSCa the charge on the S atom of Cys130 is less negative than in M1a. The Asn133-N-H - - - S-Cys130 H-bond is weaker in TSCa with respect to M1a.	98
Figure 5.16 Asp296 stabilizes M1a with respect to TSCa. In TSCa the negative charge of the S atom in Cys300 increases. The Asp296-CO1-O2 - - - S-Cys300 distance increases because of the unfavorable interaction.	99
Figure 5.17 Gly131 stabilizes TSCa with respect to M1a. In TSCa the total negative charge of COO ⁻ of the Pro residue increases. The Gly131-N-H - - - O2-CO1-Pro H-bond is stronger in TSCa with respect to M1a.	100
Figure 5.18 Gly301 stabilizes TSCa with respect to M1a. In TSCa the total negative charge of COO ⁻ of the Pro residue increases. The Gly301-N-H - - - O1-CO2-Pro H-bond is stronger in TSCa with respect to M1a.	100
Figure 5.19 Thr302 stabilizes TSCa with respect to M1a. In TSCa the total negative charge of COO ⁻ of the Pro residue increases; the Thr302-O-H - - - O2-CO1-Pro H-bond is stronger in TSCa with respect to M1a.	101
Figure 5.20 Gly303 stabilizes TSCa with respect to M1a. In TSCa a more negative charge accumulates on the S atom; the Gly303-N-H - - - S-Cys300 H-bond is stronger in TSCa with respect to M1a.	101
Figure 5.21 The stability parameter S (in kcal mol ⁻¹) is plotted as a function of the various residues.	102
Figure 5.22 Residues having a noticeable effect on electrostatic stabilization/destabilization of M2a in respect to M1a: a) charged long-range interacting residues (Table 5.8); b) uncharged residues in close contact with QM sub-region (Table 5.9).	102
Figure 5.23 Asn133 stabilizes M1a with respect to M2a. In M2a the charge on the S atom of Cys130 is less negative than in M1a. The Asn133-N-H - - - S-Cys130 H-bond is weaker in M1a with respect to M2a.	103
Figure 5.24 Asp296 stabilizes M1a with respect to M2a. In M2a the negative charge of the S atom in Cys300 increases. The Asp296-CO1-O2 - - - S-Cys300 distance increases for the unfavourable interaction.	104
Figure 5.25 Gly301 stabilizes M1a. In M1a the COO ⁻ group of the Pro residue is better oriented for the H-bond interaction with Gly301.	104
Figure 5.26 Gly131 stabilizes M2a. In M2a the geometry is more favorable for an effective Gly131-N-H - - - OOC-Pro H-bond interaction.	105
Figure 5.27 Gly303 stabilizes M2a. In M2a a more negative charge accumulates on the S atom. The Gly133-N-H - - - S-Cys300 H-bond is stronger in M2a with respect to M1a.	105
Figure 5.28 Representation of the residues stabilizing the transition state.	106
Figure 5.29 Representation of some π -symmetry molecular orbitals within the model-H system.	107
Figure 6.1 Sub-unit A of the diaminopimelate epimerase enzyme.	109
Figure 6.2 Partitioning scheme adopted to build the Model1 model system.	113
Figure 6.3 Representation of the model-H sub-system for the critical points found on the Model1 PES.	115
Figure 6.4 Reaction profile obtained for Model1.	116
Figure 6.5 DFP and RFP diagrams for the M1/M2 transition.	118
Figure 6.6 Schematic representation of the most important residues in the stabilisation/destabilisation of the M1/M2 couple.	119
Figure 6.7 DFP and RFP diagrams for the M1/TSC transition.	121
Figure 6.8 Schematic representation of the most important residues in the stabilisation/destabilisation of the M1/TSC couple.	122
Figure 6.9 Partitioning scheme adopted to build the Model2 model system.	124
Figure 6.10 Representation of the model-H sub-system for the critical points found on the Model2 PES.	125
Figure 6.11 Reaction profile obtained for Model2.	126
Figure 6.12 DFP and RFP diagrams for the M1/M2 transition.	127
Figure 6.13 Schematic representation of the most important residues in the stabilisation/destabilisation of the M1/M2 couple.	128

Figure 6.14 DFP and RFP diagrams for the M1/TSC transition.	129
Figure 6.15 Schematic representation of the most important residues in the stabilisation/destabilisation of the M1/TSC couple	130
Figure 7.1 Cyclic voltammogram obtained for 1Ar,o(e) (3 mM) in DMSO solution with 0.1 M N(Et) ₄ BF ₄ as supporting electrolyte; $\nu = 200$ mV/s.....	134
Figure 7.2 Cyclic voltammogram obtained for 1Ar,o(i) (3 mM) in DMSO solution with 0.1 M N(Et) ₄ BF ₄ as supporting electrolyte; $\nu = 200$ mV/s.....	134
Figure 7.3 Schematic representation of the model-system used in the DFT computations. The two regions marked by the contour lines are described by the 6-31++G(2p,d) and DZVP basis set, respectively.	137
Figure 7.4 Schematic representation of the structure of M1, M2 and TS1b. The energies and Gibbs energies (values in brackets) are relative to the M2 species (kcal mol ⁻¹). E _a = activation barriers. Bond lengths are in Ångstroms and angles in degrees.....	141
Figure 7.5 Schematic representation of the structure of M3b, TS2b and M4. The energies and Gibbs energies (values in brackets) are relative to the M2 species (kcal mol ⁻¹). E _a = activation barriers. Bond lengths are in Ångstroms and angles in degrees.	143
Figure 7.6 Schematic representation of the structure of M4a_R and M4b_R. The energies and Gibbs energies (values in brackets) are relative to the M2 species (kcal mol ⁻¹). E _a = activation barriers. Bond lengths are in Ångstroms and angles in degrees.	144
Figure 8.1 ORTEP representation of the crystal structures of A (7) and B (8)..	148
Figure 8.2 Optimised structures of s-trans and s-cis isomers of both A (7) and B (8).....	150
Figure 9.1 Possible reaction pathways for the addition of enolsilyl derivatives on 1,2-diaza-1,3-butadienes.	152
Figure 9.2 Three dimensional representations for E-X and E-Y. Reported distances are in Å.	154
Figure 9.3 Three dimensional representations for Z-X and Z-Y. Reported distances are in Å.....	155
Figure 10.1 ORTEP diagram of 4e.....	160
Figure 10.2 1E + 2A [2+2] reaction profile.	169
Figure 10.3 1A + 2A [2+2] reaction's profile.	169
Figure 10.4 Geometries for [4+2] pathway (1E+2A).....	171
Figure 10.5 1E+2A [4+2] reaction profile.	172
Figure 10.6 1A+2A [4+2] reaction profile.	173
Figure 10.7 1C+2A [4+2] reaction profile.	173
Figure 10.8 1C+2F [4+2] reaction profile.	174
Figure 11.1 A schematic representation of the model-system used in this paper (bond lengths in ångstroms). The reported structure corresponds to that of the preliminary complex m0 (absolute energy = -3694.03761 a.u.). The asterisks indicate the atoms (in addition to those of the imidazole rings) that have been kept frozen during the geometry optimization).....	181
Figure 11.2 Energy profile obtained for the HCA catalytic process.	182
Figure 11.3 Schematic representation of the structure of the critical points ts1 and m1 (bond lengths are in ångstroms). The energy values (kcal mol ⁻¹) are relative to m0. Values in brackets have been obtained with the COSMO method. .	184
Figure 11.4 Schematic representation of the structure of the critical points ts2 and m2 (bond lengths are in ångstroms). The energy values (kcal mol ⁻¹) are relative to m0. Values in brackets have been obtained with the COSMO method..	185
Figure 11.5 Schematic representation of the structure of the critical point m1' (bond lengths are in ångstroms).	187
Figure 11.6 Schematic representation of the structure of the critical points ts3 and m3 (bond lengths are in ångstroms). The energy values (kcal mol ⁻¹) are relative to m0. Values in brackets have been obtained with the COSMO method. .	188
Figure 11.7 Schematic representation of the structure of the critical points ts4 and m4 (bond lengths are in ångstroms). The energy values (kcal mol ⁻¹) are relative to m0. Values in brackets have been obtained with the COSMO method. .	190
Figure 11.8 Schematic representation of the structure of the critical points m5 and ts6 (bond lengths are in ångstroms). The energy values (kcal mol ⁻¹) are relative to m0. Values in brackets have been obtained with the COSMO method. .	192
Figure 11.9 Schematic representation of the structure of the critical point m6 (bond lengths are in ångstroms). The energy value (kcal mol ⁻¹) is relative to m0. Values in brackets have been obtained with the COSMO method.....	193

List of Schemes

Scheme 2.1	24
Scheme 2.2 HF-SCF approach	28
Scheme 3.1 QM/MM implementation flowchart.	44
Scheme 4.1 Schematic representation of the interaction between solvent, PSB and counter ion: a) tight bound ionic pair, b) dissociated couple.....	67
Scheme 7.1 Compounds of the 1Ar,o series.....	134
Scheme 7.2 Possible reaction paths originating from the starting species M1. The favored path is highlighted by the contour line.....	138
Scheme 7.3 Possible reaction paths originating from the hypothetical species M6.	139
Scheme 8.1 Reaction scheme of the examined chemical transformation.	146

Scheme 8.2 Experimental procedure diagram.....	147
Scheme 8.3 Schematic representation of the hypothesised olephine rearrangement.....	148
Scheme 8.4 Shift of a nitrogroup on a double bond.....	148
Scheme 9.1 A) Mukaiyama-Michael-type addition/heterocyclization reaction of silyl ketene acetals 2a,b on 1,2-diaza-1,3-butadienes 1a,b; B) Proposed mechanism for the Lewis acid-catalyzed Mukaiyama-Michael-type addition of enolsilyl derivatives 2a,b on 1,2-diaza-1,3-butadienes 1a,b.....	152
Scheme 9.2 Simplified models for E-isomers of A and 3: E-X and E-Y (calculated differences in energy contents is reported).....	154
Scheme 9.3 Simplified models for Z-isomers of A and 3: Z-X and Z-Y (calculated differences in energy contents is reported).....	155
Scheme 10.1 Synthesis of 3-bromo-3-alkenyl-azetid-2-ones and 3-bromo-4-alkyl-5,6-dihydropyridin-2-ones.....	157
Scheme 10.2 Reaction of 1a and 1b with imines 2a-d.....	158
Scheme 10.3 Reaction of 1a and 1b with imine 2e.....	159
Scheme 10.4 Reaction of γ,γ -disubstituted- α -bromo vinylketenes 1c-d with imines 2a-d.....	160
Scheme 10.5 Reaction of 1c and 1d with the chiral Schiff base 2e.....	161
Scheme 10.6 Reaction of α -bromo vinylketenes with ketimines 2d and 2f.....	162
Scheme 10.7 Pathways for the reaction of ketene and imine.....	164
Scheme 10.8 Model system used in the computational study.....	164
Scheme 10.9 General scheme for the [2+2] pathways.....	167
Scheme 10.10 General scheme for the [4+2] pathway.....	170
Scheme 11.1 The fundamental equations of the HCA II enzyme.....	177
Scheme 11.2 Thoms model.....	178
Scheme 11.3 Rearrangement of the product.....	179

List of Tables

Table 3.1 Calculation types.....	38
Table 4.1 Absorption maxima for the anionic GFP chromophore.....	63
Table 4.2 Absorption maxima for RPSB.....	66
Table 4.3 Relevant geometrical parameters and statistical analysis (in Å) for the $[\text{Fe}(\text{bipy})_3]^{2+}$ complex.....	78
Table 4.4 Relevant geometrical parameters and statistical analysis (in Å) for the $[\text{Fe}(\text{phen})_3]^{2+}$ complex.....	79
Table 4.5 Relevant geometrical parameters and statistical analysis (in Å) for the $[\text{Zn}(\text{bipy})_3]^{2+}$ complex.....	80
Table 4.6 Relevant geometrical parameters and statistical analysis (in Å) for the $[\text{Zn}(\text{phen})_3]^{2+}$ complex.....	81
Table 5.1 Residues in the medium region.....	88
Table 5.2 Residues in the moving low region.....	89
Table 5.3 Number and value of the point charges added to the model system used for the QM/MM study of TcPRAC.....	91
Table 5.4 Energy difference between the two minima: $\Delta E = E_{M2a} - E_{M1a}$; E_c represents the electrostatic contribution to ΔE due to the added external charges (ions or pseudo-ions); E_{Enz} represents the contribution from the enzyme itself.....	93
Table 5.5 QM/MM energy (E) and Gibbs Free Energy ($G = E + \text{Thermal Free Energies}$). The reported values (kcal mol^{-1}) are relative to M1a.....	94
Table 5.6 Long range interacting residues. If $S < 0$ the residue stabilizes TSCa.....	98
Table 5.7 Short range interacting residues. If $S < 0$ the residue stabilizes TSCa.....	98
Table 5.8 Long range interacting residues. If $S < 0$ the residue stabilizes M2a.....	103
Table 5.9 Short range interacting residues. If $S < 0$ the residue stabilizes M2a.....	103
Table 6.1 Partitioning of the whole system into layers.....	113
Table 6.2 Some relevant distances and atomic point charges ^a for the three critical points found in the PES of Model1.....	115
Table 6.3 Energy difference between minima and reaction barrier for Model1.....	116
Table 6.4 Electrostatic interaction energy data for Model1.....	118
Table 6.5 Stabilizing/destabilizing factor for the couple M1/M2 in Model1; values from selected residues are reported.....	119
Table 6.6 Stabilizing/destabilizing factor for the couple M1/TSC in Model1; values from selected residues are reported.....	121
Table 6.7 Stabilizing/destabilizing van der Waals factor for the couple M1/M2 in Model1; values from selected residues are reported.....	122
Table 6.8 Stabilizing/destabilizing van der Waals factor for the couple M1/TSC in Model1; values from selected residues are reported.....	122
Table 6.9 Some relevant distances and atomic point charges ^a for the three critical points found in the PES of Model2.....	126
Table 6.10 Energy difference between minima and reaction barrier for Model2.....	126

Table 6.11 Electrostatic interaction energy data for Model2.....	127
Table 6.12 Stabilizing/destabilizing factor for the couple M1/M2 in Model2; values from selected residues are reported.	128
Table 6.13 Stabilizing/destabilizing factor for the couple M1/TSC in Model2; values from selected residues are reported.	129
Table 7.1 Values of $E_{1/2}$ obtained by methods (i) and (ii) from the CVs recorded at $v= 20$ mV/s.	135
Table 7.2 Computed Reduction Potentials for the various Redox Processes reported in Scheme 7.2 and in Scheme 7.3.	140
Table 9.1 Energies of stationary points.	153
Table 10.1 Reaction of Vinylketenes 1a-b with imines 2b-d.	158
Table 10.2 Formation of 3-bromo-4-alkyl-5,6-dihydropyridin-2-one 3 via ketene-imine cyclisation.....	161
Table 10.3 Formation of 3-bromo-3-alkenyl-azetidin-2-ones via ketene-ketimine cyclisation.....	162
Table 10.4 Energies of the conformational equilibrium of the ketenes.	166
Table 10.5 Energetics of the [2+2] pathways.	168
Table 10.6 Energetics of the [4+2] pathways.	170

Publications

Some of the work presented here has been published in the following papers:

Chapter 3 and Chapter 4

Altoè, P.; Stenta, M.; Bottoni, A.; Garavelli, M. *Theor. Chem. Acc.* **2007**, *118*, 219-240.

Altoè, P.; Stenta, M.; Bottoni, A.; Garavelli, M. *AIP Conf. Proc.* **2007**, *963*, 491-505.

Chapter 5

Stenta, M.; Calvaresi, M.; Altoè, P.; Spinelli, D.; Garavelli, M.; Bottoni, A. *J. Phys. Chem. B* **2007**.
(published on-line; DOI =10.1021/jp7104105). Paper selected as cover for the 31 January, 2008, Issue (number 112) of *J. Phys. Chem. B*.

Chapter 7

Bottoni, A.; Cosimelli, B.; Scavetta, E.; Spinelli, D.; Spisani, R.; Stenta, M.; Tonelli, D. *Mol. Phys.* **2006**, *104*, 2961-2982.

Chapter 8

Attanasi, O. A.; Filippone, G. F. P.; Lillini, S.; Mantellini, F.; Spinelli, D.; Stenta, M. *Adv. Synth. Catal.* **2007**, *349*, 907-915.

Chapter 9

Bianchi, L.; Giorgi, G.; Maccagno, M.; Petrillo, G.; Rizzato, E.; Spinelli, D.; Stenta, M.; Tavani, C. *Lett. Org. Chem.* **2007**, *4*, 268-272.

Chapter 11

Miscione, G.; Stenta, M.; Spinelli, D.; Anders, E.; Bottoni, A. *Theor. Chem. Acc.* **2007**, *118*, 193-201.