



**HAL**  
open science

# Etudes de systèmes organométalliques et biologiques par des méthodes hybrides mécanique quantique/mécanique moléculaire

Marius Retegan

► **To cite this version:**

Marius Retegan. Etudes de systèmes organométalliques et biologiques par des méthodes hybrides mécanique quantique/mécanique moléculaire. Other. Université Joseph-Fourier - Grenoble I, 2009. English. NNT: . tel-00440656

**HAL Id: tel-00440656**

**<https://theses.hal.science/tel-00440656>**

Submitted on 11 Dec 2009

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# THÈSE

présentée par

**Marius RETEGAN**

En vue d'obtenir le grade de

**Docteur de l'Université Joseph Fourier**

Spécialité : Chimie Physique Moléculaire et Structurale

Ecole Doctorale de Chimie et Sciences du Vivant

**Etudes de systèmes organométalliques et  
biologiques par des méthodes hybrides  
mécanique quantique/mécanique  
moléculaire.**

soutenue le 27 Février 2009 devant le jury composé de :

Isabelle DEMACHY	Rapporteur
Manuel F. RUIZ-LÓPEZ	Rapporteur
Paul FLEURAT-LESSARD	Examineur
Martin J. FIELD	Examineur
Anne MILET	Directrice de thèse
Hélène JAMET	Directrice de thèse

Thèse préparée au Laboratoire Département de Chimie Moléculaire, Grenoble



*For my parents*



# Résumé

Ces dernières années, les méthodes hybrides QM/MM combinant la mécanique quantique (QM) et la mécanique moléculaire (MM) se sont révélées des méthodes de choix pour l'étude de systèmes chimiques et biochimiques contenant plus d'une centaine d'atomes. Nous avons mis en évidence les apports et difficultés liés à leur utilisation à travers des systèmes variés : modélisation de ligands phosphines, réactivité d'une protéine de type acide phosphatase pourpre, modélisation de l'interaction protéine-ligand. Le premier exemple nous a permis d'évaluer l'impact du choix de la frontière QM-MM sur la géométrie et structure électronique du système. Cet effet a été à nouveau étudié sur le deuxième exemple. L'influence de l'environnement de la protéine acide phosphatase pourpre sur sa réactivité a pu être étudiée par l'approche QM/MM. Enfin au cours de la dernière étude, nous avons été amené à adapter une méthode théorique originale appelée QM/MM-PB/SA permettant de décrire de façon réaliste les énergies d'interaction entre une protéine et un substrat. Cette dernière a été testée et validée sur différents inhibiteurs de la protéine casein kinase 2 et a permis de proposer une explication aux résultats obtenus lors d'un criblage de la chimiothèque de notre laboratoire. Les programmes mis en place nous permettent d'envisager une utilisation aisée sur d'autres systèmes originaux.



# Abstract

In the recent years the hybrid QM/MM methods, which combine the quantum mechanics (QM) with molecular mechanics (MM), have emerged as a promising route for studying complex biological and chemical systems containing more than a few hundreds atoms. In the present manuscript we showed their advantages and shortcomings in studying various problems, ranging from the modelization of phosphine ligands, enzymatic reactivity of the purple acid phosphatase and the modelization of protein-ligand interaction between the casein kinase 2 protein and its inhibitors.

The first study has allowed us to evaluate how the position of the QM-MM border influences the electronic structure and the geometrical parameters of the phosphine ligands commonly used in the Pauson-Khand reaction. Subsequently, the influence of the protein environment on the reactivity of purple acid phosphatase was studied by employing a QM/MM method. Finally, in the last study, we proposed an original methodology for description of the interaction energy between a protein and an inhibitor. The method was tested and validated on different inhibitors of the casein kinase 2 protein. This allowed us to explain the results obtained by a virtual screening of the DCM chemical library.





# Contents

<b>Contents</b>	<b>9</b>
<b>List of Tables</b>	<b>13</b>
<b>List of Figures</b>	<b>15</b>
<b>Introduction Générale</b>	<b>19</b>
<b>General Introduction</b>	<b>21</b>
<b>I Theoretical Background</b>	<b>25</b>
<b>1 The QM/MM methods</b>	<b>29</b>
1.1 Terminology . . . . .	29
1.2 QM/MM Energy Expression . . . . .	30
1.2.1 Subtractive Schemes . . . . .	30
1.2.2 Additive Schemes . . . . .	31
1.3 Bonded QM-MM Interactions . . . . .	32
1.4 Non-bonded Interaction Between Inner and Outer Subsystems . . . . .	33
1.4.1 van der Waals Interaction . . . . .	33
1.4.2 Electrostatic Interaction . . . . .	33
1.5 Bonds Across the Boundary . . . . .	35
1.5.1 Link Atom . . . . .	36
1.5.2 Frozen Localized Orbitals . . . . .	39
1.6 The ONIOM method . . . . .	40

<b>II Studied Systems</b>	<b>43</b>
<b>2 Modelization of Phosphine Ligands</b>	<b>47</b>
2.1 Introduction . . . . .	47
2.2 Enantioselective Pauson-Khand Reaction . . . . .	48
2.3 Methodology . . . . .	49
2.4 Results and Discussions . . . . .	51
2.5 Conclusions . . . . .	57
<b>3 Reactivity of Purple Acid Phosphatase</b>	<b>61</b>
3.1 Introduction . . . . .	61
3.2 Purple acid phosphatase . . . . .	62
3.3 Computational Details . . . . .	65
3.4 Results and Discussion . . . . .	66
3.4.1 DFT results . . . . .	66
3.4.2 ONIOM results . . . . .	69
3.4.3 Mechanistic Studies . . . . .	71
3.5 Conclusions and Perspectives . . . . .	75
<b>4 CK2 Protein and Its Inhibitors</b>	<b>81</b>
4.1 Introduction . . . . .	81
4.2 Free Energy Calculations and Ligand Binding . . . . .	82
4.2.1 Free energy perturbation . . . . .	82
4.2.2 Thermodynamic Integration . . . . .	84
4.2.3 Thermodynamic cycle . . . . .	85
4.2.4 End-Point Methods . . . . .	86
4.2.5 Improved Potential Energy Models . . . . .	90
4.3 The System . . . . .	92
4.4 Methodology . . . . .	94
4.4.1 Construction of the Initial Models . . . . .	94
4.4.2 Hybrid Potential Simulations . . . . .	94
4.4.3 QM/MM-PB/SA and Energy Decomposition . . . . .	96
4.5 Results and Discussion . . . . .	99
4.5.1 Structural Features . . . . .	99
4.5.2 Relative Binding Free Energies . . . . .	104

4.5.3	Electrostatic Energy Decomposition . . . . .	107
4.5.4	Energetic Contribution of Individual Amino Acids . . . . .	109
4.5.5	Application of the QM/MM-PB/SA Method to New Inhibitors	111
4.6	Conclusions . . . . .	115
<b>Conclusions Générales</b>		<b>117</b>
<b>Concluding Remarks and Perspectives</b>		<b>119</b>
<b>Bibliography</b>		<b>121</b>



# List of Tables

2.1	Calculated Bond Lengths (Å) for 1. . . . .	53
2.2	Calculated Bond Lengths (Å) for 2. . . . .	53
2.3	Calculated Bond Lengths (Å) for 3. . . . .	54
2.4	Calculated Bond Lengths (Å) for 4. . . . .	56
3.1	Selected active site distances (Å) for the DFT models . . . . .	68
3.2	Selected active site distances (Å) for the ONIOM models. . . . .	71
3.3	Main optimized distances (Å) for the reactant and transition state of mechanism B computed with the DFT and ONIOM models. . . . .	73
4.1	Inhibitor activity against the CK2 for the TBB derived inhibitors. . . . .	95
4.2	Average bond distances (Å) and their standard deviations between each inhibitor and selected amino acids from the active site of CK2. Values are in parenthesis represent crystallographic data. . . . .	100
4.3	Binding energies computed with the QM/MM-PB/SA method. . . . .	106
4.4	Components of the electrostatic interaction energy . . . . .	108
4.5	Individual amino acid residues contribution to the van der Waals interaction energy. . . . .	110
4.6	Individual amino acid contribution to the electrostatic interaction energy calculated using the gas-phase ( $\psi_0$ ) and the polarized ( $\psi$ ) semiempirical wavefunction for the description of the inhibitor. . . . .	112
4.7	Inhibitor activity against the CK2 for the <i>N</i> -hydroxyphthalimide derived inhibitors. . . . .	113
4.8	Binding energies and electrostatic interaction energy decomposition. . . . .	114



# List of Figures

1.1	Division of the system ( $S$ ) into inner ( $I$ ) and outer ( $O$ ) subsystems . . . . .	29
1.2	The link atom concept. The partitioning is across the M1-Q1 bond, with the link atom is denoted L. Atoms directly connected to M1 are labeled M2, those separated from M1 by two bonds are labeled M3, etc. The same naming conventions applies to QM atoms. . . . .	37
1.3	ONIOM terminology . . . . .	40
2.1	The catalytic version of the Pauson-Khand Reaction . . . . .	48
2.2	ONIOM partition scheme used in the calculations. Black: QM part, red MM part. . . . .	52
2.3	MESP isosurface (-18.82 kcal/mol) for compound 2. The dark dot in the surface is the $V_{\min}$ , and has a value of -21.60 kcal/mol for (a), -25.75 kcal/mol for (b), and -36.34 kcal/mol for (c). . . . .	54
2.4	Superposition of X-ray structure (yellow), full QM optimized (blue), ONIOM optimized models (red), and QM optimized models (green) for compound 3	55
2.5	MESP isosurface (-25.10 kcal/mol) for compound 3. The dark dot in the surface is the $V_{\min}$ , and has a value of -49.58 kcal/mol for (a), -49.33 kcal/mol for (b), and -49.86 kcal/mol for (c). . . . .	55
2.6	Superposition of X-ray structure (yellow), full QM optimized (blue), ONIOM optimized models (red), and QM optimized models (green) for compound 4	57
2.7	MESP isosurface (-18.82 kcal/mol) for compound 4. The dark dot in the surface is the $V_{\min}$ , and has a value of -22.65 kcal/mol for (a), -21.60 kcal/mol for (b), and -42.23 kcal/mol for (c). . . . .	57
3.1	Schematic representation of the active site of red kidney bean purple acid phosphatase (kbPAP) . . . . .	63



3.2	Proposed mechanisms for the hydrolysis of the phosphate monoester by PAPs, involving either a terminal (mechanism A) of a bridging (mechanism B) nucleophile . . . . .	64
3.3	Schematic representation of the DFT models used in this work. . . . .	67
3.4	Superposition of the optimized geometries for model 1DM (red) and 1DL (blue). . . . .	69
3.5	Cartoon representation of the crystal structure of kbPAP complex with phosphate (subunit D). The amino acids bound to the two metal ions and the two histidine residues found close to the substrate are respresented with a ball and stick model. . . . .	70
3.6	Superposition of the optimized geometries for model 1OM (red) and 1OL (blue). Hydrogens were removed for clarity. . . . .	72
3.7	Root mean square deviation (RMSD) of the protein backbone during molecular dynamics. . . . .	75
3.8	Results for the ONIOM PES scan for mechanism B. The reaction coordinate $\xi$ is defined as $\xi = d_{R1} - d_{R2}$ . Here $d_{R1}$ is the distance between the phosphorous atom and the oxygen atom of the phenolate and $d_{R2}$ is the distance between the oxygen atom of the nucleophilic hydroxide bridge and the phosphorous atom. The relative value of the QM energy for each point is plotted on the ordinate. . . . .	76
4.1	Thermodynamic cycle used in FEP/TI calculations of the relative binding free energies of two ligands, $L_1$ and $L_2$ , to a receptor molecule R. . . . .	86
4.2	Final snapshot from the MD trajectory of CK2 complexed with L4 in its neutral form (a), anionic form (b), and crystal structure of the complex (e). The inhibitor, residues K68, E114, V116 and all water molecules within 5 Å of the inhibitor are represented in ball-and-stick model. The protein is represented with grey cartoons. Hydrogens are removed for clarity. Measured distances are depicted as black dashed lines. Evolution of the selected distances between the neutral form and CK2, and the anionic form and CK2, are depicted in figures (c) and (d), respectively. . . . .	102

4.3	Final snapshot from the MD trajectory of CK2 complexed with L2 in simulation L2A (a), and simulation L2B (b). Representation is made as in Figure 4.2. Evolution of the selected distances between the inhibitor and CK2 in simulation L2A, and simulation L2B, are depicted in figure (c) and (d), respectively. . . . .	104
4.4	The X-ray structure of CK2 with AMPPNP (PDB code: 1DAW). The ANP is represented in ball-and-stick model, while the CK2 protein is represented with grey cartoons. The adenine pocket (blue) comprise the residues I66, V95, F113, E114, Y115, V116 and M163. The ribose region (green) comprise the residues V45, V53, N161 and I174. The triphosphate binding region (orange) comprise the residues G46, R47, G48, S51, K68 and D175. . . . .	109



# Introduction Générale

Si les méthodes de chimie quantique permettent de simuler les réactions chimiques et autres procédés complexes, elles restent par le coût du calcul engendré limiter à l'étude de petites molécules. Or les systèmes biologiques sont souvent composés d'un grand nombre d'atomes. Les simulations de mécanique moléculaire permettent de décrire ces systèmes tant qu'on ne cherche pas à décrire leur structure électronique. Dans ce cas, une solution naturelle qui vient à l'esprit est d'utiliser une description quantique pour la partie du système où la prise en compte de la structure électronique est nécessaire et une description classique en mécanique moléculaire pour le reste du système.

Cette approche couramment appelée méthode mixte QM/MM a vu le jour en 1976 avec l'étude de Warshel and Levit [WL76] sur la réaction enzymatique du lysosyme. L'énergie du système y est décomposée en deux parties, un terme incluant les termes classiques de mécanique moléculaire, et un autre incluant un Hamiltonien quantique semi-empirique prenant en compte la polarisation de la partie quantique par les charges ponctuelles de la partie classique. Les liaisons chimiques entre la partie quantique et classique sont coupées et des orbitales hybrides sont placées sur les atomes de la frontière. 10 ans plus tard, Singh et Kollman [SK86] ont été les premiers à publier la première optimisation de géométrie QM/MM avec une partie quantique ab-initio. La coupure entre la partie quantique et classique est traitée à l'aide d'atomes d'hydrogènes fictifs. En 1990, Field, Bash et Karplus [FBK90] ont décrit en détail le couplage entre un Hamiltonien quantique semi-empirique et le champs de force CHARMM. Là encore des atomes d'hydrogènes sont placés au niveau de la frontière QM/MM et la polarisation de la densité électronique quantique par les charges ponctuelles MM est prise en compte.

Dans cette thèse, nous avons appliqué ces méthodes QM/MM pour l'étude de différents problèmes, allant de la chimie organométallique, à la réactivité des métal-

loprotéines et au drug design. La première partie du manuscrit s'attache à décrire les fondements théoriques des méthodes QM/MM. La deuxième partie traite des systèmes étudiés. Le premier exemple nous a permis d'évaluer l'impact du choix de la frontière QM/MM sur la géométrie et structure électronique du système. Cet effet a été à nouveau étudié sur le deuxième exemple. L'influence de l'environnement d'une métalloprotéine sur sa réactivité a été mis en évidence par l'approche QM/MM. Enfin au cours de la dernière étude, nous avons été amené à adapter une méthode théorique originale appelée QM/MM-PB/SA permettant de décrire de façon réaliste les énergies d'interaction entre une protéine et un substrat aux dynamiques QM/MM. Cette dernière a été testée et validée sur différents inhibiteurs de la protéine CK2 et a permis de proposer une explication aux résultats obtenus lors d'un criblage de la chimiothèque de notre laboratoire.

# General Introduction

The modeling of complex chemical systems is still a difficult challenge. One could use elaborate first-principles methods for simulating reactions and other complex processes to achieve high accuracy, but these are limited by their computational cost to small molecules. The systems of chemical interest in computational biology and catalysis are often condensed phase systems with many thousands of participating atoms. The use of molecular mechanics, based on classical potentials, have been proven to provide an effective means for simulating these systems. On the other hand, the description of chemical reactions, like bond forming/breaking and other processes that involve changes in the electronic structure, such as charge transfer or electronic excitations can not be described by the molecular mechanics (MM) simple functions. A natural solution for this dilemma is to use a QM description for the chemically active region (e.g., substrates and cofactors in an enzymatic reaction) and a MM treatment for the rest of the system (e.g., full protein and solvent). The resulting schemes are commonly referred to as combined or hybrid QM/MM methods. By embedding a QM calculation in a classical MM model of the environment, the hybrid QM/MM scheme attempts to incorporate environmental effects at an atomistic level, including influences such as mechanical constraints, electrostatic perturbations and dielectric screening.

The seminal contribution in the field belongs to Warshel and Levitt [WL76], who in 1976, more than thirty years ago, presented a method that took into account all important aspects of the QM/MM approach and applied it to the lysozyme enzymatic reaction. The energy expression used in this work was remarkably complex and included the usual MM terms, a semi-empirical QM Hamiltonian that accounted for the polarization of the QM density by the MM charges as well as by induced dipoles placed on all MM atoms of the protein and by the permanent dipoles of the water molecule, the interaction of the point charges, induced dipoles, and permanent dipoles

with each other, and a classical bonded and van der Waals QM-MM coupling terms. Covalent bonds across the QM-MM boundary were saturated by single hybrid orbitals placed on the frontier MM atoms.

Ten years later, Singh and Kollman [SK86] were the first to report coupled ab initio QM/MM structure optimization. They used link atoms to cap the covalent bonds across the QM-MM boundary. They also allowed for the mutual polarization of the QM and MM regions.

In 1990 Field, Bash and Karplus [FBK90] described the coupling of a semi-empirical QM Hamiltonian with the CHARMM force field in considerable detail. They also used link atoms to cap dangling bonds in the QM region and accounted for the polarization of the QM electronic density by the MM point charges. They carefully evaluated the accuracy of the QM/MM treatment against ab initio and experimental data.

In the present manuscript we applied the QM/MM approach for the study of different problems, ranging from organometallic chemistry, reactivity of metalloproteins and drug design.

The first part provides a theoretical background of the QM/MM approach. It describes the two schemes used to evaluate the QM/MM energy expression (subtractive and additive) and takes a closer look at the methods used to treat the non bonded interaction between the two subsystems. Subsequently, the link atom methodology is presented, followed by an in depth description of the ONIOM method.

The second part consists of three chapters and focalizes on the studied systems. Chapter 2 discusses the treatment of phosphine ligands commonly used in the asymmetric Pauson-Khand reaction by means of QM/MM methods, more specifically ONIOM. The impact of different placements of the QM-MM border is evaluated by direct comparison of the ONIOM optimized geometries with full QM results and by means of MESP (molecular electrostatic potential) calculations. In Chapter 3, the importance of including the environment of the enzyme, in order to correctly model the hydrolysis mechanism of phosphorylated substrates by the red kidney bean purple acid phosphatase, is illustrated by a comparative DFT and ONIOM study. Finally in Chapter 4, after a short presentation of different methodologies used to access the free energy of binding, we present a detailed description of an adaptation of the MM-PB/SA method to trajectories obtained from QM/MM molecular dynamics. With a quantum mechanical description of the ligand, problems related to force field parameter development are circumvented. The QM/MM-PB/SA approach was tested on different

inhibitors of the casein kinase 2 protein. This allowed to propose an explanation for the experimental data obtained in the case of inhibitors discovered after a virtual screening of the DCM chemical library.





# Part I

## Theoretical Background



# Résumé

Dans ce premier chapitre, nous nous sommes attachés à décrire les fondements théoriques des méthodes QM/MM. Le premier point aborde l'expression de l'énergie avec les deux types de schémas additif ou soustractif. Le traitement de l'interaction entre les sous-systèmes quantique et classique est ensuite présenté avec notamment les différents niveaux de complexité de l'interaction électrostatique que sont les approches mechanical embedding, electrostatic embedding ou encore polarized embedding. Si la première approche ne permet pas de prendre en compte l'influence de l'environnement sur la fonction d'onde de la partie quantique, la seconde l'inclut directement dans l'hamiltonien quantique et décrit ainsi la polarisation de la partie quantique par les charges de la partie classique. La dernière approche rend ce processus self-consistent en introduisant la polarisation à l'aide de dipôles et non plus de charges ponctuelles.

Différentes modes de jonction entre la partie quantique et classique sont possibles. Une des méthodes les plus simples utilise un atome monovalent comme l'hydrogène pour saturer les liaisons coupées au niveau de la jonction. Le principe de cette méthode appelée link atom, son implémentation ainsi que le traitement des charges proches de la frontière sont présentés. Enfin l'approche ONIOM est plus précisément décrite. Cette méthode basée sur le calcul à bas et haut niveau de théorie d'un système modèle à l'intérieur d'un système réel calculé lui seulement à bas niveau a été utilisé pour modéliser les systèmes des Chapitres 2 et 3 de cette thèse.



# 1

## The QM/MM methods

### 1.1 Terminology

In a QM/MM approach the system ( $S$ ) is partitioned into the inner region ( $I$ ) to be treated quantum-mechanically and the outer region ( $O$ ) described by a force field (Figure 1.1).

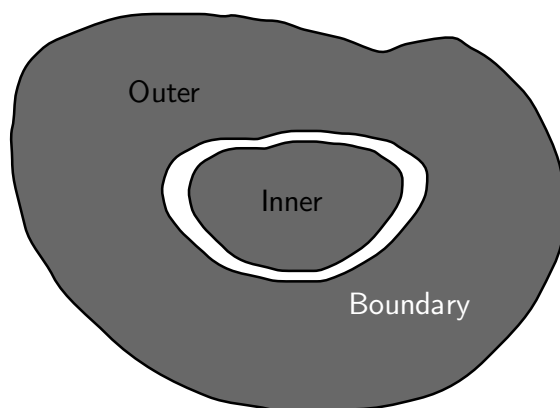


Figure 1.1: Division of the system ( $S$ ) into inner ( $I$ ) and outer ( $O$ ) subsystems

Inner and outer regions are therefore also frequently referred to as QM and MM regions, respectively. Each atom of the entire system is assigned to either of the subsystems. There are three types of interactions in a QM/MM model: interaction between the atoms in the MM subsystem, interactions between the particles (nuclei and explicit electrons) in the QM subsystem, and interactions between the two subsystems. Because the later type of interactions are generally strong, it is not possible to simply write down the total energy of the system as the sum of the energies of the subsystems. The boundary designates the region where the standard QM and MM procedure are modified or augmented in any way.

## 1.2 QM/MM Energy Expression

Based on the form for the energy expression, the QM/MM methods can be divided into two broad classes: subtractive and additive schemes.

### 1.2.1 Subtractive Schemes

In subtractive QM/MM schemes three independent calculations are performed: a calculation of the entire system  $\mathcal{S}$ , a QM calculation on the inner subsystem  $I$ , and a MM calculation of the inner subsystem. The total energy of the system is obtained by summing the first two terms and subtracting the last term to avoid double counting. From this point of view the energy expression for a subtractive scheme can be regarded as a correction (or extrapolation).

If we use the link atom implementation to saturate the free valences in the inner layer (Section 1.5.1) the total energy takes the following form:

$$E_{\text{QM/MM}}^{\text{subtractive}}(\mathcal{S}) = E_{\text{MM}}(\mathcal{S}) + E_{\text{QM}}(I + \mathcal{L}) - E_{\text{MM}}(I + \mathcal{L}) \quad (1.1)$$

Here the subscript denotes the level of calculation, while the system on which is performed is denoted in parenthesis.

Conceptually, the subtractive QM/MM scheme can be seen as an MM approach in which a certain region has been cut out and replaced by a higher level treatment. Its main advantage is the simplicity and ease of implementation. No explicit coupling term are necessary, avoiding modifications of the standard QM and MM procedures. The subtracted term corrects for any artifacts caused by the link atoms, provided that

the MM force field parameters referring to the link atoms reproduce the QM potential reasonably well.

As a disadvantage, in a subtractive scheme the coupling between the subsystems is treated entirely at the MM level. This can be especially problematic for electrostatic interactions, which are described using fixed atomic charges at the MM level. As a result, a strictly subtractive QM/MM method implies mechanical embedding, meaning that the QM density is not polarized by the environment (Section 1.4.2).

### 1.2.2 Additive Schemes

The energy expression for an additive QM/MM scheme is:

$$E_{\text{QM/MM}}^{\text{additive}}(S) = E_{\text{MM}}(O) + E_{\text{QM}}(I + L) + E_{\text{QM-MM}}(I, O) \quad (1.2)$$

In contrast to the subtractive scheme, the MM calculation is performed only on the outer system. The capped inner system is treated in the same manner as in Equation 1.1. The last term of the equation represents the coupling term between the two regions, and usually includes the bounded, van der Waals and electrostatic interactions between QM and MM atoms. The electrostatic coupling term (Section 1.4.2) has the largest impact and is also the most technically challenging.

If we assume a link atom scheme with mechanical embedding, we can derive the additive energy expression from the subtractive scheme [BT96]. For a classical force field the total energy of the system can be decomposed into contribution from exclusive set of atoms as following:

$$E_{\text{MM}}(S) = E_{\text{MM}}(O) + E_{\text{MM}}(I) + E_{\text{MM}}(I, O) \quad (1.3)$$

$$E_{\text{MM}}(I + L) = E_{\text{MM}}(I) + E_{\text{MM}}(L) + E_{\text{MM}}(I, L) \quad (1.4)$$

If we substitute these equations in Equation 1.1, we obtain the following equation for the QM/MM energy (after cancellation of  $E_{\text{MM}}(I)$ ):

$$E_{\text{QM/MM}}^{\text{additive}}(S) = E_{\text{MM}}(O) + E_{\text{QM}}(I + L) + E_{\text{MM}}(I, O) - [E_{\text{MM}}(L) + E_{\text{MM}}(I, L)] \quad (1.5)$$



Because we assumed a mechanical embedding formalism,  $E_{\text{MM}}(I, O)$  represents the coupling between the inner and the outer regions, and is equivalent to the last term from Equation 1.2. The term in square brackets of Equation 1.5 is referred as the “link atom correction”, and contains the interaction energy between the link atoms and the interaction between the link atoms and the atoms in the inner layer. Because these interactions are also included into the  $E_{\text{QM}}(I + \mathcal{L})$  term, this correction is formally justified. In practice the link atom correction is often omitted, and this has mostly to do with validity of using the MM level to correct the QM interactions.

The working equation adopted by the majority of the QM/MM scheme is thus Equation 1.2. What defines a particular QM/MM method is the exact description of the QM-MM coupling term,  $E_{\text{QM-MM}}(I, O)$ .

The interaction between QM and MM centers can be decomposed in accordance with the interactions considered in the force field:

$$E_{\text{QM-MM}}(I, O) = E_{\text{QM-MM}}^{\text{b}} + E_{\text{QM-MM}}^{\text{vdW}} + E_{\text{QM-MM}}^{\text{elec}} \quad (1.6)$$

The first term describes the QM-MM bonded interaction localized on the boundary (Section 1.3), while the last two terms described the QM-MM non-bonded interaction (Section 1.4).

### 1.3 Bonded QM-MM Interactions

Classical computation of selected QM/MM terms is required to maintain reasonable geometries. There is an operational difference between subtractive and additive schemes with respect to the treatment of link atoms, which leads to different terms being included in the final energy expression.

A subtractive scheme removes the QM-QM bonded interactions, *i.e.*, those involving atoms from the capped inner region  $I + \mathcal{L}$ , and retains all mixed QM-MM bonded terms. It thus implicitly corrects for the link atoms.

In the first computations using additive schemes and link atoms [SK86, FBK90], all bonded terms between at least one MM and at least one QM atoms were treated classically. In an approach employed by Eurenus *et al.* [ECBH96], this treatment is refined, avoiding duplication of the interactions computed quantum mechanically. They only include all the bond terms involving one QM atom and one MM atom,

all the angles and dihedral angles with one or two MM central atoms, and all the improper dihedral terms with the outer atoms in the MM region. All the other bonded interactions are assumed to be described by the QM Hamiltonian.

## 1.4 Non-bonded Interaction Between Inner and Outer Subsystems

### 1.4.1 van der Waals Interaction

The treatment of the van der Waals interaction is usually described by a Lennard-Jones potential and follows the model used in the MM calculation. The problem is that these parameters are only valid within the parametrization for which they were derived, that is, for MM-MM interactions. Fortunately the short-range nature of the van der Waals interaction makes this restriction less severe. While every atom in the inner region interacts with all the atoms from the MM region, only the closest ones contribute significantly, therefore unoptimized parameters only affect QM atoms close to the boundary. If one is concerned that this might influence the result, one solution would be to move the QM-MM boundary further away from incriminated QM atoms. Sometimes this strategy is not applicable and a re-fitting of the non-bonded parameters must be carried out, particularly in the case of solvation studies [GX92, TL99].

### 1.4.2 Electrostatic Interaction

The electrostatic interaction between the QM charge density and the charge model used for MM can be handled at different level of sophistication. Bakowies and Thiel [BT96] defined three ways for treating it: mechanical embedding, electrostatic embedding and polarized embedding.

#### **Mechanical Embedding**

In a mechanical embedding scheme the QM calculation is basically performed in gas phase, without the inclusion of the charges from the outer region into the QM Hamiltonian. This means that the QM density is not polarized. In this approach the electrostatic interaction between QM and MM region is either omitted or treated

using the MM code. If one chooses to treat this interaction, the MM electrostatic charges for the atoms in the inner region have to be derived when the system evolves along the reaction path. This can work, but in practice, it requires a large inner region so that the calculated charge distribution is converged with respect of the QM-MM boundary. Moreover, this procedure is often not trivial, and the resulting charges might be inconsistent with the charge model present in the force field.

### Electrostatic Embedding

The major shortcomings of mechanical embedding can be eliminated by performing the QM calculation in the presence of the charge distribution from the environment. The polarization of the QM region is done by incorporating the MM point charges as one electron terms in the QM Hamiltonian:

$$\mathcal{H}_{\text{QM-MM}}^{\text{el}} = - \sum_i^{\text{electrons}} \sum_{M \in O} \frac{q_M}{|\mathbf{r}_i - \mathbf{r}_M|} + \sum_{\alpha \in I+L} \sum_{M \in O} \frac{q_M Z_\alpha}{|\mathbf{r}_\alpha - \mathbf{r}_M|} \quad (1.7)$$

where  $q_M$  are the MM point charges and  $Z_\alpha$  the nuclear charge of the QM atoms; the index  $i$  runs over all electrons,  $M$  over the point charges, and  $\alpha$  over the QM nuclei.

With this approach, the electronic structure of the inner region can adapt to the charge distribution of the environment. The QM-MM electrostatic interaction is incorporated into the QM calculation, which makes it more accurate, but also increases the computational cost, especially for calculations of the electrostatic force due to the QM density acting on many MM point charges.

However some inconsistencies remain. The MM point charges model was not derived to provide a faithful representation of the real charge distribution, and its validity is interlink with the other force-field parameters. So, in principal, this kind of approach is not legitimate. Nevertheless, this has become standard procedure, and the experience shows that it generally yields reasonable results. Also special care should be taken at the QM-MM boundary, where close MM charges can overpolarize the QM electron density.

### Polarized Embedding

The electrostatic embedding scheme results in a asymmetric description of the electrostatic QM-MM interaction. While the polarization of inner region is admitted,

that of the outer region is not. In a first approximation, the rigid MM point charges are allowed to be polarized by the QM region, but the resulting field does not act back on the QM density. This procedure can be made fully self-consistent by including the resulting induced dipoles into the QM Hamiltonian.

A variety of models were devised to account for the classical polarization: (i) the polarized point dipoles model in which the polarizabilities assigned to the atoms interact with the electric field, thus inducing dipoles; (ii) the Drude oscillators model, which uses a spring to connect a mobile point charge of opposite sign to a charged site, hence the alternative name charge-on-spring model [LR03]; (iii) the fluctuating charges model based on the principle of electronegativity equalization, which means that the partial charges are optimized with respect to the total electrostatic energy.

Even though the first QM/MM approaches used the polarized embedding schemes [WL76, SK86], they remain limited in the present implementations. These has mostly to do with the lack of well established polarized MM force fields. As seen with electrostatic embedding, the increase level of complexity adds to the computational cost. In this case the addition of self consistency cycles for the QM charge density and MM polarizable model can lead to a computational overhead or convergence problems.

## 1.5 Bonds Across the Boundary

This section is concerned with the various approaches that have been devised to treat covalent bonds cut by the QM-MM boundary. The simplest solution to this problem is to define the subsystems in such a way that no covalent bonds are cut by the boundary. This is trivial for explicit solvation studies, where the solute can be described at a QM level, while the solvent by a MM force field. It is sometimes possible also for biomolecular system if the reactants are not covalently bound to the enzyme and no protein residue directly involved in the chemical transformation. Unfortunately this is not always feasible, and covalent bonds are cut by the QM-MM boundary. As a result one has to deal with two important issues resulting from cutting bonds at the QM-MM interface: the unsatisfied valences on the QM atoms and the overpolarization of the QM density by the close MM partial charges (Section 1.4).

Although many different models have been developed to deal with covalent bonds at the interface, they can be broadly classified into two groups.

1. methods based on **link atom** [SK86, FBK90] that introduce an extra atom, most often hydrogen, along the broken QM-MM bond at an appropriate distance from the QM atoms. The link atom is treated quantum mechanically, while the bond between QM and MM atoms is described molecular mechanically.
2. methods based on **frozen localized orbital** [TRR<sup>+</sup>94, GAAF98] that place a set of hybrid orbitals on the QM atom which has the broken bond. This approach is elegant because no unphysical atoms are introduced, but their formulation and implementation is more complex than for the link atom methods.

### 1.5.1 Link Atom

Even though the link atom methods have been criticized on a number of counts, like the necessity of introducing extra, unphysical atoms into the system, their conceptual simplicity remains a strong argument for using these methods.

A schematic link atom method is shown in Figure 1.2. The link atom is positioned at a certain distance along the M1-Q1 bond vector, but is completely free during the structure optimization. This can lead to inconsistencies because the Q1-L and Q1-M1 might deviate from collinearity. One way to avoid this problem is to introduce an artificial bend angle L-Q1-M1 with an equilibrium value of 0° [ECBH96].

The excess degrees of freedom introduced by the link atom can also be eliminated by constraining the M1 atom to lie along the Q1-L bond. This was first suggested by Maseras and Morokuma [MM95]. Because this approach removes four, rather than three degrees of freedom, the procedure was modified by making the coordinates of the link atom depend on the coordinates of atoms Q1 and M1 [DKB<sup>+</sup>99]:

$$\mathbf{r}_L = \mathbf{r}_{Q1} + g(\mathbf{r}_{M1} - \mathbf{r}_{Q1}) \quad (1.8)$$

where  $\mathbf{r}$  is a position vector. This definition places L along Q1-M1 bond, and relate the distance Q1-L to the distance Q1-M1 by the scaling factor  $g$ .

A suitable way to define the scaling factor  $g$  is the ratio between the equilibrium bond lengths Q1-L and Q1-M1 taken from the force field [DKB<sup>+</sup>99]. For example if we want to cut a C-C single bond and in its place use a C-H bond, a reasonable value for  $g$  would be a standard C-H bond length (1.084 Å) divided by a standard C-C bond length (1.528 Å), which results in a value of 0.709 for the scaling factor.

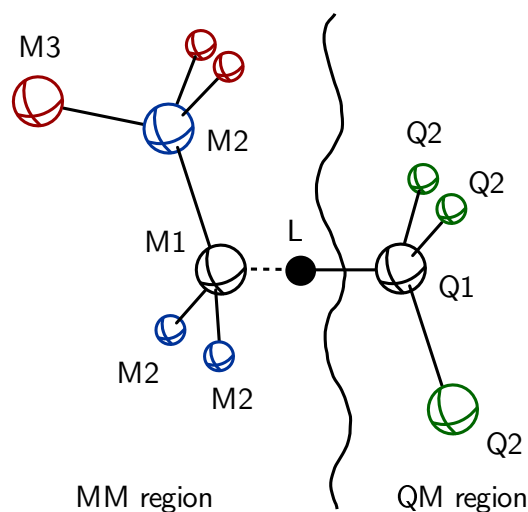


Figure 1.2: The link atom concept. The partitioning is across the M1-Q1 bond, with the link atom is denoted L. Atoms directly connected to M1 are labeled M2, those separated from M1 by two bonds are labeled M3, etc. The same naming conventions applies to QM atoms.

Another possibility [FAB<sup>+</sup>00, Swa03] is to define  $g$  as:

$$g = \frac{r_{Q1-L}}{|\mathbf{r}_{M1} - \mathbf{r}_{Q1}|} \quad (1.9)$$

where  $r_{Q1-L}$  is a predefined constant, which is the bond distance between the link atom and the atom Q1.

The scaled position link atom method (SPLAM) developed by Eichinger *et al.* [ETHP99] adds an extra term to Equation 1.9 reflecting the deviation of Q1-M1 from its equilibrium value. Their implementation also accounts for the stiffness of the Q1-M1 and Q1-L bonds.

### Forces on Link Atom

If the position of the link atom is expressed as a function of other atomic positions, its coordinates are eliminated from the set of coordinates used to describe the whole system. This way the atom becomes “invisible” to geometry optimizations and molecular dynamics algorithms which handle the entire coordinate set. However, since the QM algorithm treats the link atom the same manner as the other atoms, the forces acting on them are going to be non-zero.

In the IMOMM scheme of Maseras and Morokuma [MM95] the forces are relayed on the neighboring atoms by working in internal coordinates.

When working in pure cartesian the same effect can be realized by using the chain rule to establish the contribution of the forces acting on an atom  $K \in S$ :

$$\frac{\partial E}{\partial \mathbf{r}_K} = \frac{\partial E}{\partial \mathbf{r}_K} + \frac{\partial E}{\partial \mathbf{r}_L} \frac{\partial \mathbf{r}_L}{\partial \mathbf{r}_K} \quad (1.10)$$

The link atom contribution to the force is the second term of Equation 1.10, and vanishes if atom  $K$  is not involved in the definition of the link atom  $L$ .

### Electrostatic Interactions at the Boundary

As underlined previously the problem that has caused the most concern for the link atom method is that the electronic charge density of the QM system is poised to polarize non physically around the pointlike MM charges positioned in the immediate proximity. Various approaches have been employed to minimize this overpolarization effect, and they are outlined below.

The spurious electrostatic interaction between the QM density and the MM point charges is dominated by the one-electron Hamiltonian contributions involving the link atom basis functions and the full set of MM charges. By *selective deletion of the one-electron integral term* this interaction can be effectively removed [BT96, AT98, RDMK00]. While this approximation holds for semi-empirical methods or QM methods with small basis sets, it becomes problematic with larger basis sets because it leads to an unbalanced representation of the polarization and electrostatic potential of the QM density. Antes and Thiel defined a variant of this procedure in which all one-electron integrals are included with the exception of those that involve functions on the link atom and the charge on center M1.

Perhaps the simplest way to avoid overpolarization is to *delete point charges* close to the QM region. Different implementations of this idea have been proposed: (i) deletion of the M1 charge [WHGP90]; (ii) deletion of the charges M1 and M2 [LT05]; (iii) deletion of the charges M1, M2 and M3 [SK86, LT05]. The danger with these schemes is that simply deleting charges according to the connectivity will often result in the creation of a net partial charge near the QM region, which can lead to serious artifacts [LT05].

An alternative to charge deletion is *charge shifting* [dVSC<sup>+</sup>99]. In this scheme the M1 charge is shifted evenly onto the M2 atoms that are connected to M1. The resulting dipole is compensated by placing a pair of point charges near M2.

By replacing the M1 point charge by a *Gaussian blur* charge distribution the spurious overpolarization is greatly reduced. Thus the charge distribution can be written as:

$$\rho_M(\mathbf{r}) = q_M \exp \left[ - \left( \frac{\mathbf{r} - \mathbf{r}_M}{\sigma_M} \right)^2 \right] (\sqrt{\pi} \sigma_M)^{-3} \quad (1.11)$$

Such a scheme has been used by Eichinger *et al.* [ETHP99]. They chose a value of 0.8 Å for the  $\sigma_M$  parameter.

## 1.5.2 Frozen Localized Orbitals

Warshel and Levitt [WL76] were the first to suggest that a hybrid orbital could be used to treat the QM-MM bond, though the approach was not elaborated further. Two formalisms that have proved successful for treating the boundary region are the local self-consistent field (LSCF) [TRR<sup>+</sup>94, MLT<sup>+</sup>96, AR96, FAR02] and the generalized hybrid orbitals (GHO) [GAAF98, AFAG00].

The LSCF scheme developed by Rival and co-workers involves the primary calculation of the strictly localized bond orbital (SLBO), that could be described as a kind of frozen lone pair placed on QM boundary atom pointing towards the MM boundary atom. In the QM/MM calculation, it is excluded from the SCF optimization and does therefore not mix with other orbitals. A similar approach was presented by Friesner and co-workers [PF99, MPF00] that differs in some technical details from the original one.

The GHO method of Gao and co-workers differs from the LSCF method in orbital optimization and system partition. The set of localized hybrid orbitals are placed on the MM boundary atom. The orbital pointing towards the QM region is active and participates in the SCF iterations, while the remaining hybrids are kept frozen.

While these approaches promise to mimic the electronic structure of the QM region, they are technically considerably more complicated to implement when compared to link atoms. In addition, the localized orbitals themselves, must be determined beforehand, involving calculations on model compounds. These parameters are usually not



transferable and need to be recalculated upon changing the MM force field or the QM level of theory or basis set.

## 1.6 The ONIOM method

The ONIOM (our N-layered integrated molecular orbital + molecular mechanics) method developed in Morokuma and co-workers is an oniom-skin-like subtractive method, that uses link atoms to cap bonds severed by boundaries. The current version is the latest incarnation of a series of hybrid methods which also include IMOMM (integrated molecular orbital + molecular mechanics) [MM95] and IMOMO (integrated molecular orbital + molecular orbital) [SHF<sup>+</sup>96]. IMOMM and IMOMO should not be considered a subset of ONIOM method, since the placement of the link atom in ONIOM is obtained with a scale factor (Equation 1.8), while in IMOMM and IMOMO the link atom is placed at a specified distance from the atom to which it is connected.

In a two-layer ONIOM(QM:MM) the total energy of the system is obtained from three independent calculations:

$$E^{\text{ONIOM}} = E^{\text{real,MM}} + E^{\text{model,QM}} - E^{\text{model,MM}} \quad (1.12)$$

The *real* system contains all atoms and it is calculated only at the MM level. The *model* system contains the system that is treated at the QM level. Both QM and MM calculations need to be carried for the model system (Figure 1.3).

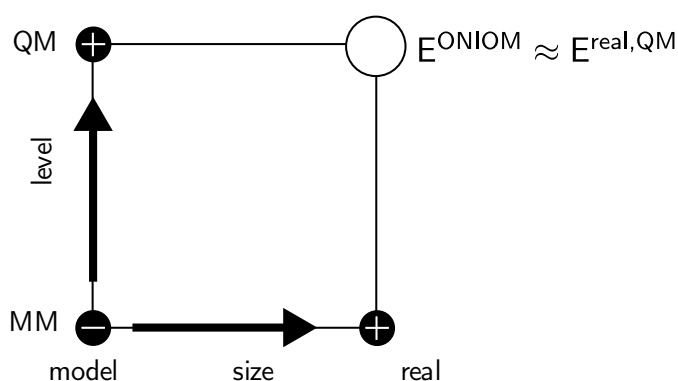


Figure 1.3: ONIOM terminology

In the ONIOM scheme, most of the MM terms in the model system exist in the real system as well and cancel in the full expression. The difference between the real

system and model system MM calculations, the  $S$ -value, describes the contribution from the MM region, which includes both the energy of the MM region as well as the interaction between the QM region and the MM region:

$$S^{\text{MM}} = E^{\text{real,MM}} - E^{\text{model,MM}} \quad (1.13)$$

$S^{\text{MM}}$  plays the same role as the  $E_{\text{MM}}(O) + E_{\text{QM-MM}}(I, O)$  for an additive scheme (Equation 1.2). When no bonded interactions are present between the regions,  $S^{\text{MM}}$  is identical to  $E_{\text{MM}}(O) + E_{\text{QM-MM}}(I, O)$ , and  $E^{\text{ONIOM}}$  becomes identical to  $E_{\text{QM/MM}}^{\text{additive}}$ .

When bonded interactions are present in the system, the functions for  $E^{\text{ONIOM}}$  and  $E_{\text{QM/MM}}^{\text{additive}}$  are not identical. As with any subtractive schemes, in ONIOM the correction for the link atoms is included in the last term of Equation 1.12. Because in its first implementation the position of the link atoms was assumed invariable in the two model calculation, this correction is less effective. One possibility to minimize this error is to allow different  $g$  scale factors for the QM model system and the MM model system.

As previously underlined the interaction between the QM layer and MM layer is always included at the lowest computational level, via the  $E^{\text{real,MM}}$  term. From the original formulation discussed above it follows that ONIOM uses mechanical embedding by default (ONIOM-ME). However, in many QM/MM methods, the electrostatic interaction between the layers is moved from the MM level to the QM level. Recently an electrostatic embedding scheme has been incorporated into the ONIOM framework [VBK<sup>+</sup>06, HPR<sup>+</sup>08] (ONIOM-EE). Because the authors wanted to stay true the extrapolation philosophy, they modified both the QM and MM model system calculations by adding the electrostatic interaction with the MM region:

$$E^{\text{ONIOM-EE}} = E^{\text{real,MM}} + E_v^{\text{model,QM}} - E_v^{\text{model,MM}} \quad (1.14)$$

where

$$E_v^{\text{model,MM}} = E^{\text{model,MM}} + \sum_J \sum_N \frac{s_{JN} q_J q_N}{\mathbf{r}_{JN}} \quad (1.15)$$

while the gas phase Hamiltonian corresponding to the energy  $E_v^{\text{model,QM}}$ , was augmented with a term similar to the one from Equation 1.7

$$\mathcal{H}^{\text{model}} = \mathcal{H}_0^{\text{model}} - \sum_i \sum_J \frac{q_J}{\mathbf{r}_{iJ}} + \sum_J \sum_N \frac{q_N Z_J}{\mathbf{r}_{JN}} \quad (1.16)$$

In the equations above the subscripts  $J$ ,  $N$ , and  $i$  refer to the atoms in the real system, atoms in the model system, and electrons, respectively, while  $s_{JN}$  is a charge scaling factor.

With this implementation the electrostatic interaction is evaluated three times, once in each term on the r.h.s of Equation 1.14 (the  $E^{\text{real,MM}}$  also includes the electrostatic interaction between the two layers at the MM level). Since many of the electrostatic terms from the first and third term cancel, the ONIOM-EE formalism can be regarded as more or less equivalent to the additive QM/MM electrostatic embedding. Again, if no bonded interactions are present between the layers, the two electrostatic embedding versions become identical. It must be noted that the link atoms from the ONIOM implementation “see” the surrounding charges. When the layers are covalently bound, ONIOM uses a charge scaling scheme to overcome the artifacts arising from close point charges. Typically, the interactions between the centers that are one or two bonds separated are scaled to zero, those who are three bonds separated are scaled by 0.5, and interactions between charges separated by more than three bonds are fully included.

The main advantage of the ONIOM method remains its flexibility, allowing the combination different levels of QM methods and an MM method in a single integrated calculation of the energy or other electronic properties. This aspect has contributed significantly to the popularity of the method within the community.

# **Part II**

## **Studied Systems**



## Résumé

Ce chapitre discute des différentes façons de modéliser des ligands phosphine. Ces ligands sont couramment utilisés dans la réaction de Pauson-Khan notamment dans sa version asymétrique en particuliers avec des substituants volumineux et souvent conjugués à l'origine de l'orientation de la réaction. Le calcul quantique de ces systèmes étant souvent trop coûteux, une solution est l'utilisation des méthodes mixtes et notamment l'approche ONIOM connue pour ces succès en chimie organométallique. Quatre complexes  $[W(CO)_5L]$  où L est un ligand phosphine avec des substituants encombrants ont été considérés. Différentes coupures entre les sous-systèmes quantique et classique des ligands ont été étudiées. Les coupures sont évaluées par deux approches.

La première consiste à comparer les géométries optimisées au niveau ONIOM et calcul DFT complet. Les calculs ont montré que la coupure de liaisons polarisées ou conjuguées n'induit pas de fortes perturbations des géométries si elle est restée suffisamment éloignée du phosphore. Dès qu'on se rapproche du phosphore, les géométries obtenues après optimisation diffèrent. Les mêmes résultats sont obtenus par la deuxième approche. Celle-ci compare le potentiel électrostatique autour de la paire libre du phosphore. La géométrie du ligand est la même pour tous les calculs. On a choisi de considérer celle du ligand dans le complexe optimisé au niveau DFT pur. On aurait également pu prendre une structure cristallographique. Le calcul du potentiel et le tracé des isosurfaces permet de comparer les différents calculs ONIOM. Ce calcul apparaît comme un mode d'évaluation rapide et fiable de l'effet de la position de la coupure sur la description de la structure électronique de ces ligands.



# 2

## Modelization of Phosphine Ligands

### 2.1 Introduction

Nowadays, theoretical chemistry is a common tool used to answer different questions in organometallic chemistry. Unfortunately, the influence of steric effects has been largely absent from this spectacular progress. This is mostly related to the fact that the accurate description of properties at metal centers leaves little room for the introduction of the bulky ligands responsible for steric effects. This situation is, however, currently changing, in part because of the introduction of hybrid QM/MM methods in this field of chemistry.

A fundamental question that arises after the decision of using QM/MM methods for a certain problem, is how to partition the system, which atoms are going to be included in the QM region and which atoms are going to be included in the MM region? Chemical knowledge is most of the time the guiding line for this choice, but tests should be run in case of uncertainty. These tests usually include the comparison of the geometries obtained with a hybrid QM/MM scheme and those from pure QM calculations. This would provide a more just criterion than comparison with



experiment, because there are a number of reasons for which QM/MM results are different from experimental data, and these might not be related to how the QM/MM partitioning was done. These include inaccuracy related to the QM method, packing effects or the presence of solvent molecules in the experimental data. Nevertheless comparison with experiment remains a useful criterion, especially when theoretical results converge towards them.

Comparison with structural experimental data is indeed important, but it might not give us sufficient information about how the QM-MM partition reproduces electronic effects compared to pure QM results. In this sense we decided to investigate how the placement of the QM-MM border influences the geometrical parameters and electronic structure for a series of phosphine ligands commonly used in Pauson-Khand reaction.

## 2.2 Enantioselective Pauson-Khand Reaction

The Pauson-Khand reaction has been since its discovery in 1971 [KKPW71, KKPW73, KKP<sup>+</sup>73] a powerful tool for synthesizing cyclopentanone derivatives or cyclopentane-containing natural products by cocyclization of alkynes, alkenes and carbon monoxide (Figure 2.1).

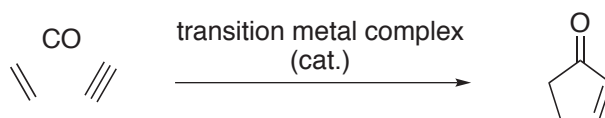


Figure 2.1: The catalytic version of the Pauson-Khand Reaction

In the recent years, there was an increasing interest in developing an enantioselective version of the Pauson-Khand reaction, while using a chiral metal catalyst. The first report was made by Buchwald and Hicks in 1996 [HB96]. They managed to induce asymmetry by using a chiral titanocene complex (*S,S*)-(EBTHI)Ti(CO)<sub>2</sub>. Hiroi *et al.* reported the first asymmetric cobalt catalyzed Pauson-Khand reaction [HWKA00b, HWKA00a]. They compared different chiral bidentate phosphine ligands, and found that BINAP ligands provide the highest level of enantioselectivity, but with moderate yields. In 2000, Jeong and coworkers reported an asymmetric Pauson-Khand cyclization. The catalyst was prepared *in situ* from [RhCl(CO)<sub>2</sub>]<sub>2</sub> and BINAP and AgOTf [JSC00]. Good to excellent selectivities were obtained for a small range of

intramolecular substrates. Independently, Shibata and Tagaki reported the catalysis by Ir-tolBINAP. The chiral catalyst was prepared *in situ* from  $[\text{IrCl}(\text{COD})]_2$  and (*S*)-tolBINAP. Again, excellent selectivities were obtained for the cyclization of a small range of enynes. Additionally, the authors reported the first asymmetric intermolecular cyclization between 1-phenyl-1-propyne and norbornene using the same catalyst and under the same reaction conditions. The success of their approach is partially due to the fact that the number of coordination sites on the Ir(I) catalyst for the organic partners was reduced to two after the coordination of the chiral ligand. The recent work of Gimbert *et al.* [MWGG08] shares the same fundamental idea. By using two bidentate phosphine ligands, the authors tried to limit the available positions on the monometallic center  $\text{M}(\text{CO})_6$  ( $\text{M}=\text{Mo}, \text{W}$ ).

From a theoretical point of view the Pauson-Khand reaction has been subject to a number of studies [dBMR<sup>+</sup>01, RMG<sup>+</sup>01, GLM<sup>+</sup>03, dBMGG04, dBMV<sup>+</sup>06], however the systems studied were limited to a relatively small number of atoms. If we want to envisage a theoretical study that takes into account the effects of large phosphine ligands on the Pauson-Khand reaction, DFT methods become rapidly unpractical.

## 2.3 Methodology

For this study four phosphine ligands were used, two monodentate and two bidentate. The bidentate ligands were recently employed in the study of Gimbert *et al.* [MWGG08]. X-ray structures of the ligands bound to the tungsten penta(tetra)carbonyl were available, and they were used as a starting point for all calculations. For the evaluation of the position of the QM-MM border, two approaches were used.

The first approach consisted in a series of geometry optimizations, followed by an analysis of the optimized structures. As discussed above, there are several inaccuracies that would result in different geometries between theoretical and experimental data, even in the case of a full QM geometry optimization. Surely with a hybrid representation of the system, we are going to add to these inaccuracies. So in order to separate the two problems, all structures were optimized using full QM and the ONIOM scheme. For the ONIOM scheme, different placements of the QM-MM border were used. In addition to these calculations, another set of optimizations was ran only on the QM part of the system including the link atoms (the inner layer),

without the inclusion of the MM atoms (outer layer). This calculations served to compare the ONIOM optimizations with the common strategy use in computational organometallic chemistry, which consists of substituting bulky groups with hydrogens atoms or methyl groups.

The second approach for the evaluation of the placement of the QM-MM border is based on the evaluation of the molecular electrostatic potential (MESP) for the ligands alone, without optimizing it. Suresh *et al.* quantified the electron donating power of phosphine ligands by calculating the minimum value of the MESP around the lone pair region of the phosphorus atom. Subsequently this methodology was used to unravel the stereoelectronic profile for the ligands [Sur06, MTS07]. For the separation of the electronic and steric effects, the authors coupled MESP methodology with an ONIOM representation of the system, in which only the phosphorus atom was included in the high level of calculation. The MESP of a molecule is a real physical property, that can be determined from X-ray diffraction experiments. This property can also be determined from the electron density,  $\rho(\mathbf{r})$ , by employing the following equation:

$$V(\mathbf{r}) = \sum_A^N \frac{Z_A}{|\mathbf{r} - \mathbf{R}_A|} - \int \frac{\rho(\mathbf{r}')d^3\mathbf{r}'}{|\mathbf{r} - \mathbf{r}'|} \quad (2.1)$$

where  $Z_A$  is the charge of the nucleus  $A$ , located at  $\mathbf{R}_A$ . Local minima of the MESP ( $V_{\min}$ ) are often observed at the lone pair region of a molecule, because the electronic term in Equation 2.1 has a larger value as compared to the nuclear term. At these inflection points,  $\nabla V$  will be zero, and the Hessian matrix of  $V(\mathbf{r})$  will have three positive eigenvalues [Sur06]. In our case, the evaluation of the MESP could give a rapid picture to how the electronic structure around the lone pair region of the phosphorous atom is affected by changing the atoms that are included in the QM and MM region.

Unconstrained geometry optimizations were carried out using the Gaussian 03 software package [FTS<sup>+</sup>04]. The full QM calculations were carried out within the framework of DFT using the B3PW91 hybrid functional [Bec93, PCV<sup>+</sup>92]. The tungsten atom was represented by the relativistic effective core potential (RECP) from the Los Alamos National Laboratory and the associated basis set (LANL2DZ) [HW85] augmented by an f polarization function ( $\alpha = 0.823$ ) [EBD<sup>+</sup>93]. The remaining atoms were represented by a 6-31G(d,p) basis set [HP73]. In the ONIOM calculations, the QM part was described using the same setup as above, while for the MM part the internally stored UFF parameters [RCC<sup>+</sup>92, CCR92, RCC93] were used. The initial

tests that were performed showed that the inclusion of point charges in the QM Hamiltonian (electronic embedding) had very little effect on the resulting geometries. As a result, all the ONIOM calculations were done with mechanical embedding. Considering the fact that by default the ONIOM algorithm in Gaussian 03 applies scaling factors to charges close to the QM region, these results were not surprising. This behavior can be modified by making these scaling parameters equal to unity, but that would lead to overpolarization of the wavefunction due to close charges. For the MESP evaluation the structure of the ligand alone was taken. For different QM-MM border placements, a cube file containing the MESP is generated for the inner layer atoms using a fine grid dimension  $100 \times 100 \times 100$  at the same theoretical level as described above. The MESP was visualized using the freely available PyMol program [DeL02]. From the visual inspection of the resulted cube file, the most negative value MESP ( $V_{\min}$ ) in lone pair region of the phosphorus atom can be obtained. If this coordinate is used as a guess along with `prop=(efg, opt)` keyword in Gaussian 03, the exact location of the  $V_{\min}$  can be obtained.

## 2.4 Results and Discussions

The different systems partitioning are presented in Figure 2.2. A single partitioning was used in the case of **1**, and two for the rest of the compounds. When using hybrid QM/MM methods to study biological systems where bonds have to be severed, it is advisable to place the border between the  $C_{\alpha}$  and  $C_{\beta}$  of the amino acids. In organometallic systems the search for reasonable location to separate the two regions, we can end up by treating the hole system at the QM level of theory. As a result, in the present study we were forced to place the QM-MM border in a way that it often cuts through polarized bonds and fused cycles, even though we are aware that this could lead to perturbation of the system.

For compound **1** the results of the geometry optimizations are present in Table 2.1. Interestingly, the full QM and ONIOM optimized geometries are in good agreement with a increase W-P distance of  $0.02 \text{ \AA}$  for the ONIOM structure. On the other hand, when only the atoms from the inner region are optimized at the QM level, the W-P distance is shorter by  $0.07 \text{ \AA}$  when compared to full QM geometry optimization. In the case of ONIOM optimized structure, the position of the link atoms in the inner

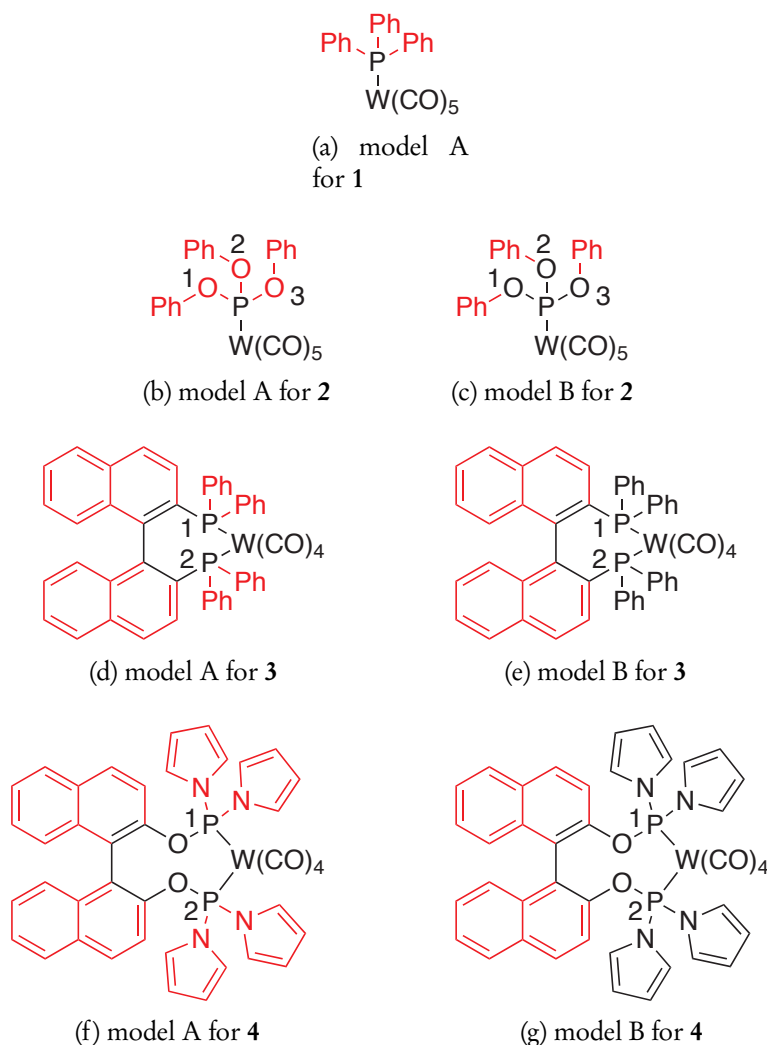


Figure 2.2: ONIOM partition scheme used in the calculations. Black: QM part, red MM part.

layer will be governed by the steric effects exerted between the phenyl groups. This leads to different values for the W-P-H angle of ONIOM optimized geometry and of the QM optimized inner layer geometry. Electronically, this means that the  $sp^3$  orbital on the phosphorous atom will have less p character in the later case, the final effect being the larger difference for the W-P distance observed above.

In the second compound (2), the phenyl groups on the phosphorus atom are substituted by phenol. Besides the changes in the electronic structure, this substitution also leads to an increased flexibility. The optimized geometries are presented in Table 2.2.

Table 2.1: Calculated Bond Lengths ( $\text{\AA}$ ) for 1.

	X-ray	full QM	model A ONIOM	model A QM
W – P	2.544	2.571	2.591	2.505
W – C <sub>ax</sub>	2.005	2.002	1.997	2.008
C <sub>ax</sub> – O <sub>ax</sub>	1.125	1.158	1.157	1.155

It can be seen that even at the full QM level the P-O bonds are  $\sim 0.04 \text{ \AA}$  longer when compared to the X-ray structure. It's very probable that these differences are due to the inaccuracies discussed in the introduction of this chapter. As a result, the optimized geometries for the two models will be compared with full QM results, rather than the experimental data.

Table 2.2: Calculated Bond Lengths ( $\text{\AA}$ ) for 2.

	X-ray	full QM	model B ONIOM	model B QM	model A ONIOM	model A QM
W – P	2.442	2.469	2.491	2.460	2.569	2.506
W – C <sub>ax</sub>	2.035	2.015	2.015	2.022	2.001	2.008
C <sub>ax</sub> – O <sub>ax</sub>	1.116	1.155	1.155	1.154	1.157	1.155
P – O1	1.600	1.634	1.627	1.633	–	–
P – O2	1.590	1.632	1.619	1.617	–	–
P – O3	1.609	1.654	1.641	1.641	–	–

As expected the different placement of the QM-MM border in the two models has a major influence on the resulting optimized geometries. For example the W-P distance is overestimated by more than  $0.1 \text{ \AA}$  in the case of model A, while for model B we find a difference of  $0.02 \text{ \AA}$ .

Subsequently we were interested to see if the evaluation of MESP could reproduce the results observed above. The method is very simple and inexpensive in term of computational costs, since no geometry optimizations are necessary. Figure 2.3 gives a visual representation of the MESP isosurface and the resulted values of  $V_{\min}$ . We can notice that with the increasing size of the outer region, we obtained a more negative value for the MESP. While the difference between full QM and model B is of  $\sim 4 \text{ kcal/mol}$ , for model A we obtain  $\sim 15 \text{ kcal/mol}$ . Even though a correlation between

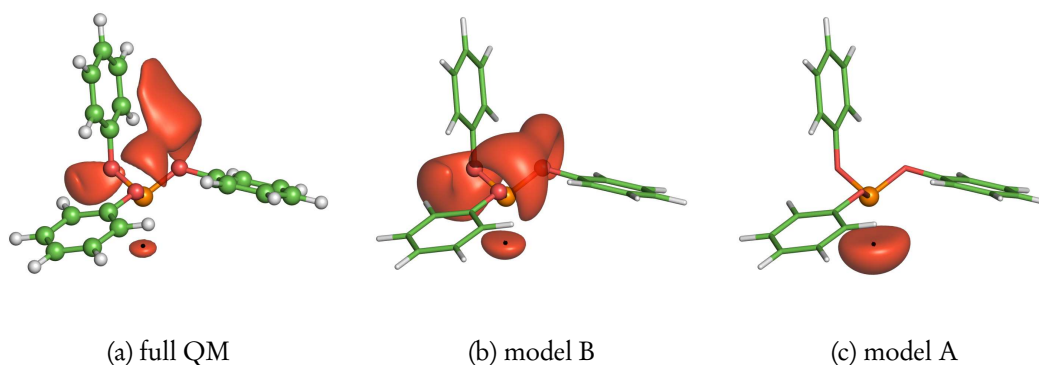


Figure 2.3: MESP isosurface (-18.82 kcal/mol) for compound **2**. The dark dot in the surface is the  $V_{\min}$ , and has a value of -21.60 kcal/mol for (a), -25.75 kcal/mol for (b), and -36.34 kcal/mol for (c).

the results obtained from geometry optimizations and MESP are hard to envisage, we notice that the same trends are followed, and that in the case of compound **2** at least the inclusion of the oxygen atoms in the inner layer is necessary.

The bidentate ligand in compound **3** is BINAP, a commonly used ligand in the Pauson-Khand reaction. As can be seen from Table 2.3, we obtain a good agreement between the full QM optimized geometries and the ONIOM optimized geometries of the two models. Another interesting result is that in the case of the optimization of the model A at the QM level, the W-P distances are underestimated. A similar situation was found for the QM optimized inner layer geometry of compound **1**, showing the need to include at least in the MM part the steric effects of the phenyl groups bound to the phosphorus atom.

Table 2.3: Calculated Bond Lengths (Å) for **3**.

	X-ray	full QM	model B ONIOM	model B QM	model A ONIOM	model A QM
W – P1	2.561	2.553	2.572	2.550	2.578	2.493
W – P2	2.523	2.588	2.603	2.555	2.606	2.493
W – C1 <sub>ax</sub>	1.936	1.988	1.990	1.987	1.983	1.998
W – C2 <sub>ax</sub>	2.009	1.983	1.980	1.988	1.978	1.998
C1 <sub>ax</sub> – O1 <sub>ax</sub>	1.161	1.163	1.161	1.164	1.163	1.160
C2 <sub>ax</sub> – O2 <sub>ax</sub>	1.121	1.164	1.164	1.163	1.163	1.160

Figure 2.4 presents a superposition of the X-ray and the optimized structures. Overall, the geometry of the compound is correctly described at all level of calculation. We notice that the phenyl groups bound to the phosphorus atom are very flexible, and even at the full QM level of theory their position deviates from the X-ray structure (Figure 2.4a). The cycle defined by the tungsten atom and the ligand atoms is correctly described in the case of QM optimized model B, while for QM optimized model A, a small deviation can be observed (Figure 2.4c).

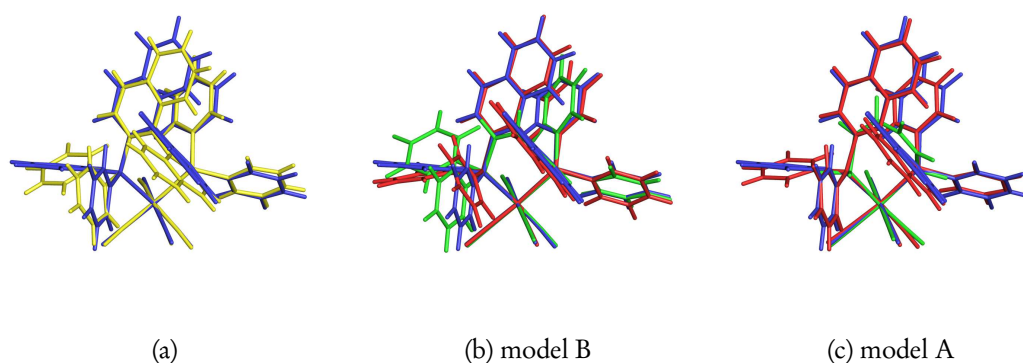


Figure 2.4: Superposition of X-ray structure (yellow), full QM optimized (blue), ONIOM optimized models (red), and QM optimized models (green) for compound **3**

The results obtained after the evaluation of the MESP in the case of compound **3**, are presented in Figure 2.5. There is practically no difference for the value of the  $V_{\min}$ . Nevertheless, they seem to be in accordance with the results obtained from geometrical optimizations, suggesting that even model A could provide a reliable representation for the present ligand.

The forth compound (**4**) differs from BINAP in that the phenyl groups bound to the phosphorus atom are substituted by pyrrole and the previous described cycle is now complemented with two oxygen atoms. This last aspect leads to an increased flexibility in that part of the molecule. Very good agreement is obtained between the full QM optimized geometry and the ONIOM optimized geometry of model B (Table 2.4). With the inclusion of the pyrrole groups in the outer layer (model A ONIOM) we notice a decrease of  $\sim 0.03$  Å for the W-P1 distances and an increase of  $\sim 0.04$  Å for the W-P2 distance.

The superposition of the X-ray and the optimized structures (Figure 2.6) show large deviations for the QM optimized model B, especially in the area of the cycle



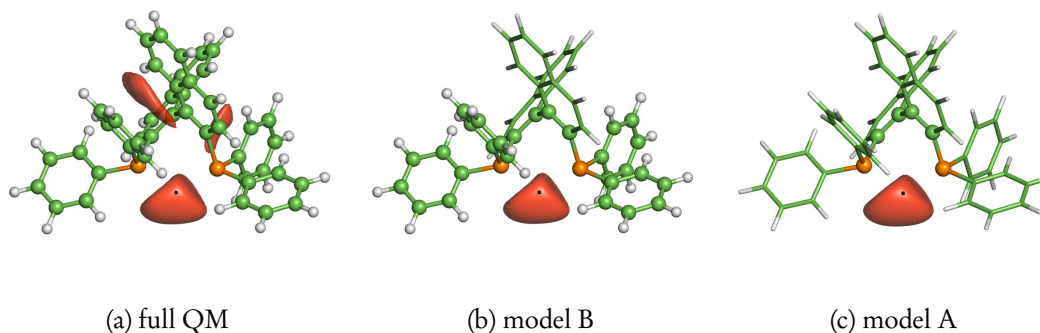


Figure 2.5: MESP isosurface (-25.10 kcal/mol) for compound 3. The dark dot in the surface is the  $V_{\min}$ , and has a value of -49.58 kcal/mol for (a), -49.33 kcal/mol for (b), and -49.86 kcal/mol for (c).

Table 2.4: Calculated Bond Lengths ( $\text{\AA}$ ) for 4.

	X-ray	full QM	model B ONIOM	model B QM	model A ONIOM	model A QM
W – P1	2.465	2.485	2.484	2.461	2.454	2.463
W – P2	2.468	2.484	2.485	2.457	2.521	2.464
W – C1 <sub>ax</sub>	2.020	2.004	2.004	2.009	1.996	2.007
W – C2 <sub>ax</sub>	2.009	2.004	2.004	2.015	2.000	2.007
C1 <sub>ax</sub> – O1 <sub>ax</sub>	1.134	1.158	1.158	1.157	1.160	1.158
C2 <sub>ax</sub> – O2 <sub>ax</sub>	1.170	1.158	1.158	1.155	1.160	1.158

defined by the tungsten atom and the atoms from the ligand. This is also reflected by the optimized distances presented in Table 2.4. From this point of view, ONIOM methodology provides a better description of the compound, because it helps keeping rigid this part of the molecule.

In the case of MESP, the same trends are observed as those resulted from geometry optimizations. The value of the  $V_{\min}$  in the case of model B is in good agreement with the one obtained if the entire ligand is treated at the QM level of theory (Figure 2.7). For model A,  $V_{\min}$  shows a large 20 kcal/mol deviation. We can conclude from the present results that, for a representation of this compound, model A is not correct and the inclusion of the pyrrol groups is crucial.

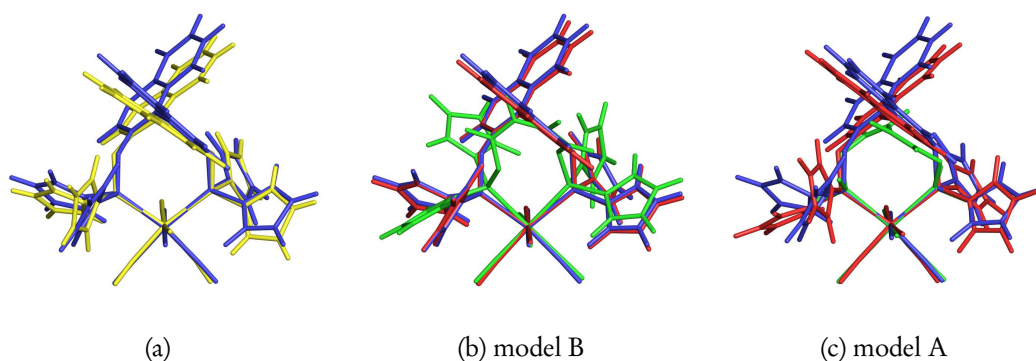


Figure 2.6: Superposition of X-ray structure (yellow), full QM optimized (blue), ONIOM optimized models (red), and QM optimized models (green) for compound 4

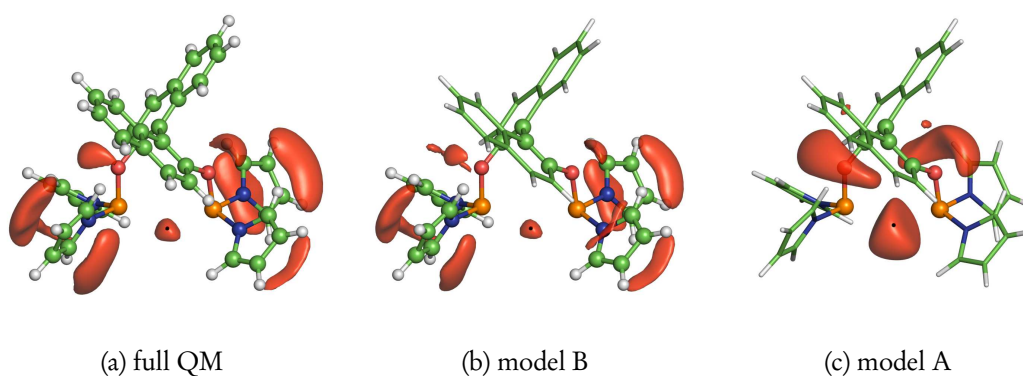


Figure 2.7: MESP isosurface (-18.82 kcal/mol) for compound 4. The dark dot in the surface is the  $V_{\min}$ , and has a value of -22.65 kcal/mol for (a), -21.60 kcal/mol for (b), and -42.23 kcal/mol for (c).

## 2.5 Conclusions

In the present section, we compare the results obtained from ONIOM method using different placements of the QM-MM border with those obtained from full quantum mechanical calculations. Two methodologies were employed: geometry optimizations of the ligand complexed to tungsten penta(tetra)carbonyl and evaluation of the MESP for the uncomplexed ligand. This later approach has the advantage of being fast and computationally inexpensive.

In the case of compound 2 we found that the inclusion of the oxygen atoms in the

inner layer is necessary to get reliable results. A more interesting situation was encountered in the study of the bidentate ligands. For compound **3** both ONIOM models provided an accurate enough representation. This was not the case for compound **4**, for which the model A proved to be a much to severe approximation.

We have underlined previously that, when using QM/MM methods for the study of organometallic systems, we are often forced to cut trough polarized bonds and fused cycles. Interestingly, even though the later rule is broken in the case of model B of compound **4**, the ONIOM optimized geometry and the MESP of the ligand is very good agreement with full QM calculations. We could conclude that if the QM-MM border is sufficiently far away from the center of interest, the errors due to it are going to be very small.

## Résumé

Pendant longtemps les calculs théoriques pour déterminer un mécanisme réactionnel enzymatique se sont limités à l'étude quantique sur de petits systèmes modèles autour du site actif. Si de tels calculs ont l'avantage de bien décrire le site actif, ils ne permettent pas de prendre en compte l'environnement de l'enzyme qui donne la spécificité de ces réactions.

Dans ce chapitre, ces différences sont illustrées à travers l'étude du mécanisme d'hydrolyse d'un phosphomonoester par la protéine purple acid phosphatase (PAP). Cette dernière forme une classe de métalloenzymes contenant dans son site actif le motif bimétallique Fe(III)-M(II) (M=Fe, Zn, ou Mn). Elle catalyse l'hydrolyse d'une large gamme de phosphoesters ou phosphoanhydres comme l'ATP sous des conditions acides. De nombreux travaux se sont attachés à décrire les structures des PAP et le rôle des métaux dans la réaction. L'hydrolyse enzymatique passe par un mécanisme de substitution nucléophile mais l'identité du nucléophile reste très controversée. Dans un des mécanismes proposés, il s'agirait d'un groupement hydroxyle pontant les deux centres métalliques. Afin de vérifier la validité de ce mécanisme, différents calculs ont été envisagés. L'étude DFT montre qu'une importante réorganisation du site actif accompagne l'attaque nucléophile et une forte énergie d'activation est obtenue pour ce mécanisme. Pour l'étude ONIOM, plusieurs calculs ont été réalisés en jouant sur la taille de la partie quantique. Tous donnent une forte diminution de l'énergie d'activation de la réaction rendant ce mécanisme compatible avec une activité enzymatique. L'introduction de résidue His202 liée par liaison hydrogène au substrat est proposée pour expliquer ce résultat.



# 3

## Reactivity of Purple Acid Phosphatase

### 3.1 Introduction

In the precedent study, we illustrated the application of the QM/MM methods in the field of organometallic chemistry. Another important application is the description of biochemical systems and more precisely theoretical studies of their reactivities. Determining a realistic enzymatic reaction pathway keep a challenge for the theoretical chemist. For a long time, most researchers confined the study of enzyme reactivity to models containing only a few representative atoms (i.e., those believed to contribute mostly to the reactivity). This drastic limitation in the size of model systems was mainly due to both the limited computing power available and the necessity of using quantum chemistry to access the making and breaking of bonds that usually appear in enzymatic reactions. If these models allowed a good understanding of the active site, they didn't take into account the specificity of enzymes and in particular the effect of the protein environment. The use of QM/MM methods has made it possible to include this effect.

In this chapter we present a study directed towards the reactivity of the purple acid

phosphatase protein from red kidney bean. The catalytic mechanism of this dinuclear Fe(III)-Zn(II) protein was studied by means of DFT and ONIOM calculations.

## 3.2 Purple acid phosphatase

Purple acid phosphatases (PAPs) belong to a growing class of metallohydrolases containing bimetallic sites that catalyze the hydrolysis of a broad range of phosphorylated substrates. Elevated PAP concentrations were observed in relationships with disease like osteoporosis [AEH<sup>+</sup>03, AWF<sup>+</sup>00, HJB<sup>+</sup>96], Morbus Gaucher [SRM91], hairy cell leukemia [HLY<sup>+</sup>97], and AIDS [SGM89]. The presumed role of these enzymes makes them viable targets for drug developing. In order to facilitate this process, the precise catalytic mechanism must be elucidated, including the function of each residue involved in the catalysis.

PAPs have been isolated from a variety of sources including mammals, plants and microorganisms. The animal enzymes are 35 kDa monomers with redox active Fe(III)-Fe(II/III) centers where only the heterovalent form is catalytically active [WHD<sup>+</sup>91, BSW04]. Plant PAPs are 110 kDa homodimers, containing Fe(III)-Zn(II) or Fe(III)-Mn(II) centers [BMS<sup>+</sup>86, SGC<sup>+</sup>99, DES<sup>+</sup>99, BdJZ<sup>+</sup>88, SBC<sup>+</sup>01], and a recombinant isoform from sweet potato has been shown to contain a di-iron center [WKSV06]. The characteristic purple color in concentrated solution is due to a charge transition between a tyrosine side-chain and the Fe(III) atom [AAL83, YMS97].

The crystal structure of the free red kidney bean PAP (kbPAP), complexes with phosphate (both a reaction product and substrate analog), tungstate (an inhibitor), sulfate (a pre-catalytic complex) and fluoride are available [SKT<sup>+</sup>95, KSF<sup>+</sup>96, SEL<sup>+</sup>08]. In addition, structures of phosphate-bound sweet potato PAP [SGC<sup>+</sup>05] and several mammalian PAPs (human, pig, rat) have been determined [GMH<sup>+</sup>99, LJK<sup>+</sup>99, ULS<sup>+</sup>99]. Their overall structures differers dependent on the phyla, but the amino acids of their active site are remarkably conserved, with seven invariant metal ligands (see Figure 3.1).

The equivalence of their active site structures has also been demonstrated by metal ion replacement studies. They have shown that the Zn(II)-derivatives of pig, rat and human PAPs are kinetically indistinguishable from native di-iron forms [BKJZ84, FLW<sup>+</sup>01, FKS<sup>+</sup>01, MA98, PMA99, FBA05], while the di-iron derivative of red kidney bean PAP has similar properties with the native Fe(III)-Zn(II) form [BMJZ88,

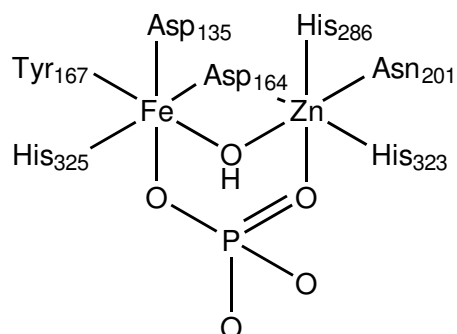


Figure 3.1: Schematic representation of the active site of red kidney bean purple acid phosphatase (kbPAP)

BdJZ<sup>+</sup>88]. <sup>1</sup>H NMR studies of the di-iron kbPAP and native mammalian enzymes also provide strong evidence for a similar coordination environment of the Fe(III) center [BDLW97].

The Fe(III) site in PAPs is coordinated to the oxygen atom of a deprotonated tyrosine (Tyr167), the nitrogen atom of a histidine (His325), and the oxygen atoms of two aspartic acid residues (Asp135, Asp164). This later bridges both metal ions (Figure 3.1). The divalent metal site (Zn(II), Fe(II), Mn(II)) is coordinated to the oxygen atom of the bridging aspartic acid (Asp164), the nitrogen atoms of two histidine residues (His286, His323), and an asparagine oxygen atom (Asn201). The X-ray structure of kbPAP shows that the uncoordinated oxygen atoms of the phosphate group form hydrogen bonds with His296 (3.16 Å) and His202 (3.25 Å). No large conformational changes are observed in the active site of kbPAP upon binding of phosphate. Superposition of the active sites of the resting and phosphate-bound forms reveal minor differences in the orientation of His295 and His296 (only the latest residue is conserved throughout PAPs). For kbPAP the crystallographic metal-metal distances are in the range 3.1-3.5 Å.

Despite the availability of structural informations, the catalytic mechanism of PAPs remains ambiguous. Experiments with bovine spleen PAP (bsPAP) and a substrate containing a chiral phosphate group showed that the hydrolysis occurs via nucleophile substitution mechanism with full inversion of configuration at the phosphorous center [MCAK93], ruling out the mechanism with a phosphoenzyme intermediate that has been proposed earlier [VCA91]. This supports a mechanism in which the substrate is directly attacked by water or by a hydroxide group. To date the na-



ture of this nucleophile has eluded identification. Two different proposals assign it to either a hydroxide ligand terminally bound to the Fe(III) center (Figure 3.2a) [DMSW91, ALS94, KSF<sup>+</sup>96, TS99, MPA99] or a bridging hydroxide (Figure 3.2b) [Kim00, WHWQ00, DFAC02, SQW<sup>+</sup>02, SEL<sup>+</sup>08]

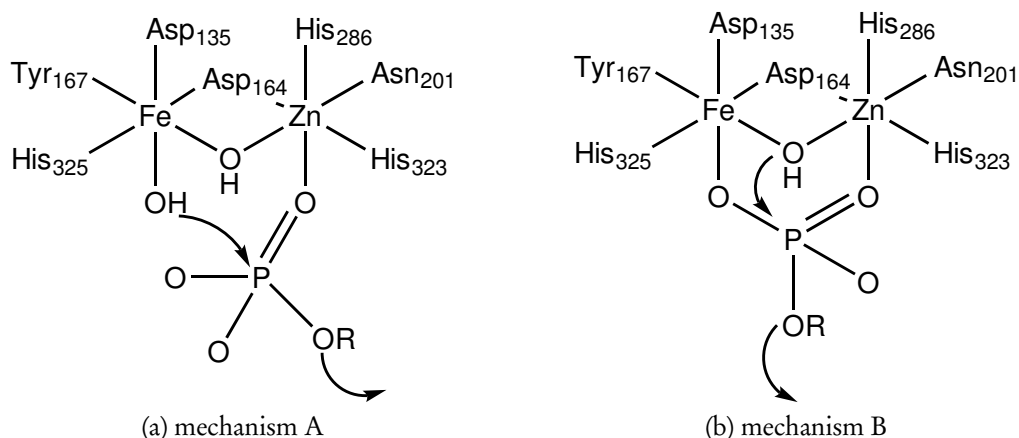


Figure 3.2: Proposed mechanisms for the hydrolysis of the phosphate monoester by PAPs, involving either a terminal (mechanism A) or a bridging (mechanism B) nucleophile

The initial structure of the resting kbPAP enzyme failed to identify any water molecules in the first coordination sphere of the two metal ions due to the limited resolution of the crystallographic data [SKT<sup>+</sup>95, KSF<sup>+</sup>96]. In conjunction with spectroscopic and kinetic studies the coordination environments of the Fe(III) and Zn(II) were completed by modeling a hydroxide ligand bound to the Fe(III), a H<sub>2</sub>O ligand bound to Zn(II), and a bridging hydroxide. The presence of a bridging hydroxide was observed in the electron density maps of pig, rat and a more recent structure of red kidney bean PAP complexed with sulfate [LJK<sup>+</sup>99, ULS<sup>+</sup>99, SEL<sup>+</sup>08]. ENDOR [SQW<sup>+</sup>02] and EXAFS [WQ98] studies of the Fe(III)-Fe(II) and Fe(III)-Zn(II) form of pig PAP have confirmed the presence of the bridging hydroxide ligand in the resting enzyme. Additionally, the ENDOR measurements didn't find terminal water/hydroxide bound to Fe(III). This observation is consistent with the X-ray structure of kbPAP-sulfate complex which mimics the pre-catalytic stage, in which the only solvent derived ligand is the  $\mu$ -hydroxide [SEL<sup>+</sup>08]. Que *et al.* [WQ98] showed that fluoride can replace the bridging hydroxide in PAP. Upon the addition of phosphate, a ternary PAP·F·PO<sub>4</sub> complex was formed, where phosphate binds to both metal

ions. Thus, for this case, the fluoride ion has been proposed to replace the bridging nucleophile. This hypothesis is in agreement with the crystal structure of the sweet potato PAP, which shows an unusual tripodal complex with one oxygen atom of the phosphate group replacing the hydroxide bridge [ULS<sup>+</sup>99], and with the structure of kbPAP complexed with fluoride [SEL<sup>+</sup>08].

Previous studies proposing a terminal Fe(III) bound hydroxide as a candidate nucleophile (mechanism A), argued that the nucleophilicity of a bridging hydroxide would be too low to assure efficient reactivity [ALS94, TS99, MPA99]. However, Smoukov *et al.* [SQW<sup>+</sup>02] pointed out that the reduced nucleophilicity of the bridging hydroxide can be compensated by an increased electrophilicity of the bridging substrate, which undergoes double Lewis acid activation when acting as bridge to the dimetal site. Furthermore, the binding of the substrate decreases the coupling interaction between the two metal centers, which may arise from a shift of the hydroxide bridge toward the divalent cation [YMS97].

Theoretical chemistry can give molecular-level insights into enzymatic reactivity through determinations of intermediates and transition states structures and energies. In the initial calculations we were interested to see how well the DFT and ONIOM results for the different models reproduce the active site structure of kbPAP (Section 3.4.1 and Section 3.4.2). Subsequently, we were interested to see if the reaction mechanism B is feasible. The obtained results show remarkable reduction of the height of the activation energy barrier when the amino acids close to the active site are included in the model (Section 3.4.3). The conclusion and perspectives that close this chapter are exposed in Section 3.5.

### 3.3 Computational Details

The initial geometries were constructed starting from the X-ray structure of kbPAP complexed with phosphate (PDB code: 4KBP) [KSF<sup>+</sup>96]. All DFT geometry optimizations we carried out with the spin-unrestrictive B3LYP functional (U-B3LYP) [Bec93, LYP88]. The atoms were represented with the LANL2DZ basis set [HW85], whose core parts in the case of iron, zinc and phosphorus were described by an effective core potentials (ECP). A set of polarization functions was added for the oxygen ( $\alpha = 0.961$ ), nitrogen ( $\alpha = 0.736$ ) and phosphorous ( $\alpha = 0.364$ ) atoms [CFB<sup>+</sup>01]. In

the ONIOM-ME (mechanical embedding) calculations the inner layer was described using the same level of theory as above, while the outer layer was describe with the UFF force field [RCC<sup>+</sup>92, CCR92, RCC93]. According to the EPR studies [DA82], high spin configuration was chosen for the iron. All geometries were fully optimized without any symmetry constraint and subsequently followed by frequency analysis to determine the nature of the stationary point, i.e, checked for the absence of imaginary frequencies in the minima and for the presence of one and only on imaginary frequency in the transition states.

## 3.4 Results and Discussion

The different labels used in this chapter follow the subsequent pattern: a number that indicates the study, a letter (D or O) for the level of calculation used (DFT or ONIOM), and a letter (S, M, L) to indicate the complexity of model (small, medium, large). In the case of the mechanistic study, the notation was complemented with an additional letter to differentiate the reactant(r), transition state (ts) and product (p).

### 3.4.1 DFT results

Due to the computational cost of the DFT calculations, we have chosen complex models including mainly the ligands of the first coordination sphere of the iron and zinc ions. In the three type of models used (1DS, 1DM and 1DL) (Figure 3.3), the amino acids were modeled as following: histidine residues were modeled as NH<sub>3</sub>, NHCH<sub>2</sub> and imidazole, the tyrosine was modeled as OH, OHCH<sub>2</sub> and O-C<sub>6</sub>H<sub>5</sub>, depending on the model, while the aspartic acid and asparagine were modeled as OCHO and OCHNH<sub>2</sub>, respectively, in all three models. To complete the coordination sphere, the  $\mu$ -hydroxide bridge and the phosphate substrate were added.

Experimentally, the X-ray structure of the kbPAP complexed with phosphate was determined at pH = 4.5. At this pH value, the phosphate is likely in its monoanionic (H<sub>2</sub>PO<sub>4</sub><sup>-</sup>) form. However, studies of a  $\mu$ -amido- $\mu$ -phosphato-bis[tetra-amminecobalt(III)] complex showed that, at this pH, the bridging phosphate is dianionic (HPO<sub>4</sub><sup>2-</sup>) [ES73]. Upon phosphate addition to the Zn(II)-derivate of pig PAP, David and Que noticed significant changes in the EPR and the visible spectra of the complex, indicating that the ferric ion is perturbed [DQ90]. The authors proposed a model

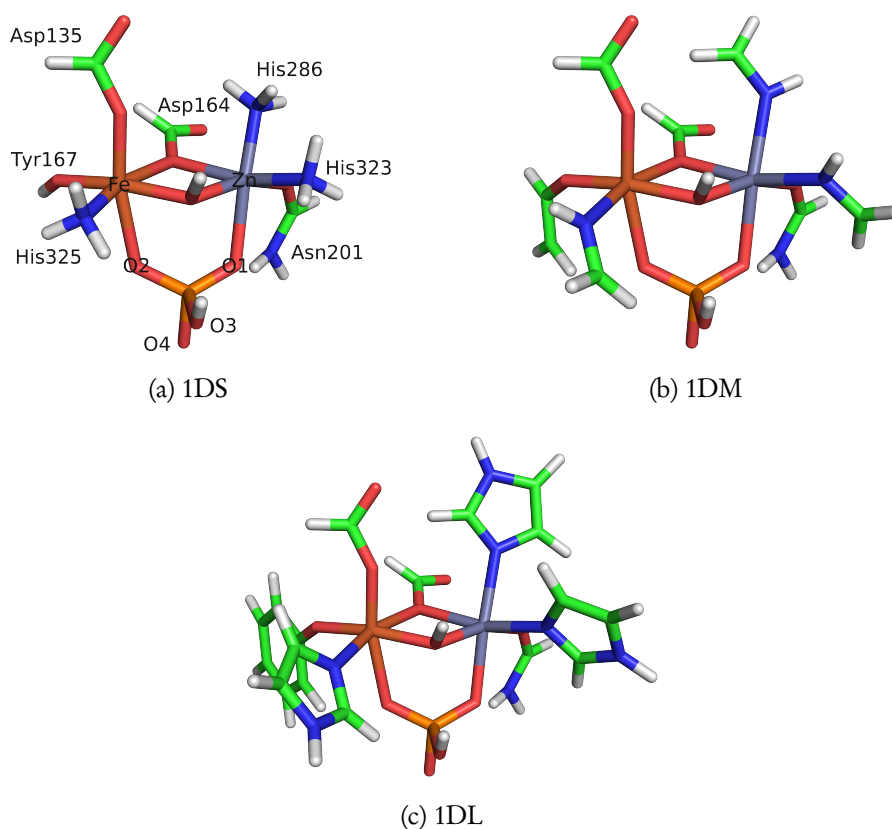


Figure 3.3: Schematic representation of the DFT models used in this work.

in which the proton of the phosphate can hydrogen bond to the  $\mu$ -hydroxide bridge to account for these experimental observations. In the following the phosphate is modeled accordingly.

Table 3.1 gathers the relevant optimized distances of the DFT models and the corresponding X-ray experimental data.

Significant deviations from the X-ray values are observed for all models. There are several causes for this. First of all in the DFT models the effect of the protein environment is neglected. It's very probable that strain imposed by the protein backbone or the interaction with other residues close to the active site might lead to different results. Secondly, the relatively low resolution of the X-ray structure (2.70 Å) raises some concerns about the experimental values. Finally, the theoretical level used to perform these computations, might influence the theoretical results. Regarding the last point, we were forced to chose a theoretical level while having in mind that a

Table 3.1: Selected active site distances (Å) for the DFT models

From	To	X-ray kbPAP	1DS	1DM	1DL
Zn	Asp164, O <sup>δ1</sup>	2.26	2.18	2.20	2.22
	Asn201, O <sup>δ1</sup>	2.16	2.29	2.22	2.20
	His286, N <sup>ε</sup>	2.05	2.19	2.23	2.24
	His323, N <sup>δ</sup>	2.14	2.27	2.26	2.23
	O1	1.98	2.10	2.06	2.08
	μ-OH, O	-	2.07	2.08	2.14
Fe	Asp135, O <sup>δ1</sup>	2.00	2.04	2.10	2.13
	Asp164, O <sup>δ1</sup>	2.31	2.16	2.10	2.09
	Tyr167, O	2.22	1.87	1.92	1.92
	His325, N <sup>ε</sup>	2.48	2.22	2.21	2.24
	O2	1.77	1.95	1.93	1.94
	μ-OH, O	-	2.19	2.11	2.09
P	μ-OH,	-	3.09	3.06	3.04

compromise between speed and accuracy has to be made.

While aware of all the possible source of errors presented above, we considered that a comparison between the different models is still meaningful. For the values presented in Table 3.1 we find good agreement between the two larger models 1DM and 1DL. This result is also reflected by the superposition of the two optimized structures presented in Figure 3.4. In both models, the phosphate group is bridged in a nearly symmetrical way, in a good position for the nucleophilic attack of the μ-hydroxide bridge.

In the case of model 1DS, we notice that the Fe-Tyr167(O) is underestimated when compared with models 1DM and 1DL. This effect is compensated by an elongation of the Fe-μ-OH(O) bond, which is approximately 0.1 Å longer. Additionally for this model we noticed a hydrogen bond between the His286 modeled as a NH<sub>3</sub> group and the fragment modeling the bridged Asp164. Together with the previous observation this clearly shows that the approximations made in representing the amino acids, make this model unsuitable to describe the active site of the protein.

Following these results, we decided that the intermediate model (1DM) is sufficient

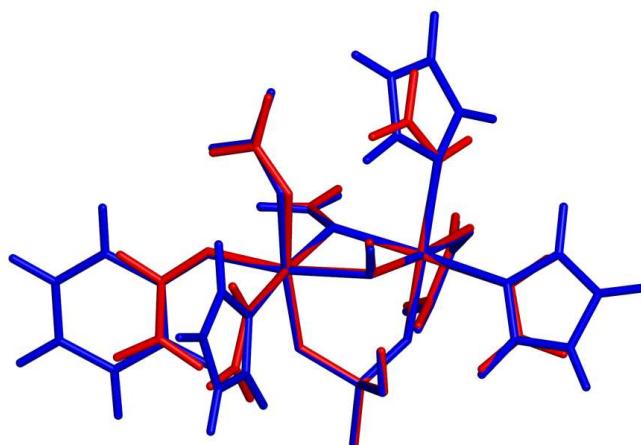


Figure 3.4: Superposition of the optimized geometries for model 1DM (red) and 1DL (blue).

for providing a good description of the active site, and was subsequently employed for the DFT mechanistic studies and as the QM part of the ONIOM calculations.

### 3.4.2 ONIOM results

In the X-ray structure of kbPAP complexed with phosphate [KSF<sup>+</sup>96] we can notice two histidine residues, His202 and His296, that help in orienting of the phosphate by hydrogen bond interactions (Figure 3.5).

In order to take into account these interactions, two ONIOM models were used, 1OM and 1OL. The inner layer of these calculations, which is modeled with the QM method, included all the atoms in model 1DM, with the additional inclusion of His202, His295 and His296 modeled as NHCH<sub>2</sub> groups. Since the complex was crystallized at a pH = 4.5 these amino acids were considered to be protonated.

While the two models share the same inner region (50 atoms, link atoms excluded), the outer region is different. The largest model considered, labeled 1OL, consists of a total of 328 atoms. The active site model includes not only all the amino acids linked to the two metal centers but also the backbone between these amino acids, comprising a total of 20 amino acids (from His286 to His296, His323 to His325, Asp164 to Tyr167, Asn201 to His202 and finally the complete Asp135 residue). In the smaller model (1OM), some of the linking amino acids were excluded, giving a total of 163 atoms. To ensure the environmental effect of the enzyme, the atoms from the peptide bond,

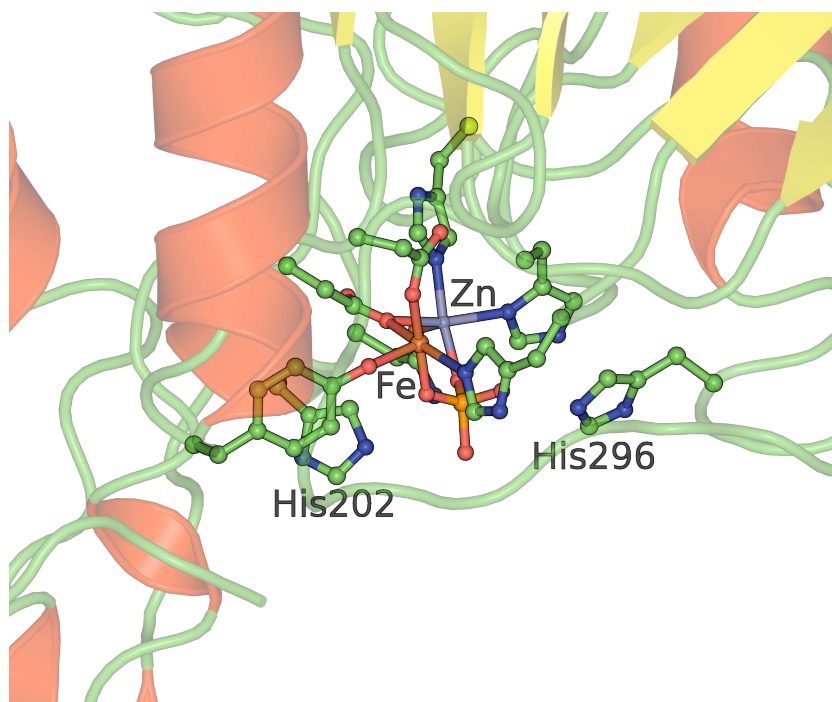


Figure 3.5: Cartoon representation of the crystal structure of kbPAP complex with phosphate (subunit D). The amino acids bound to the two metal ions and the two histidine residues found close to the substrate are represented with a ball and stick model.

which have been excluded, were frozen to their crystallographic positions.

The relevant distances for the optimized structures of the two models together with the X-ray data, are presented in Table 3.2.

As for the case of the DFT models, the results obtained for the ONIOM models show, in some cases, large deviations when compared to the X-ray structure, but are consistent with each other. For the amino acids that constitute the first coordination sphere of the two metal ions, the largest deviation is observed in the case of His323 and His325, for which we find a difference of  $\sim 0.25$  Å and  $\sim 0.24$  Å, respectively, when compared to the experimental data. Additionally, the distances between the metal ions and the oxygen atoms of the bridging phosphate are overestimated on average by  $\sim 0.25$  Å. Interestingly, even though in the DFT models the same distances are overestimated on average by  $\sim 0.14$  Å, which suggest a more tightly bound phosphate, the P-O( $\mu$ -OH) distance is larger in the case of the DFT models (3.06 Å and 3.04 Å in 1DM and 1DL, 2.83 Å and 2.82 Å in 1OM and 1OL). This can be regarded as an

Table 3.2: Selected active site distances (Å) for the ONIOM models.

From	To	X-ray kbPAP	1OM	1OL
Zn	Asp164, O <sup>δ1</sup>	2.26	2.22	2.19
	Asn201, O <sup>δ1</sup>	2.16	2.17	2.15
	His286, N <sup>ε</sup>	2.05	2.17	2.15
	His323, N <sup>δ</sup>	2.14	2.37	2.39
	O1	1.98	2.20	2.19
	μ-OH, O	-	2.16	2.16
Fe	Asp135, O <sup>δ1</sup>	2.00	1.90	1.88
	Asp164, O <sup>δ1</sup>	2.31	2.01	2.03
	Tyr167, O	2.22	2.23	2.25
	His325, N <sup>ε</sup>	2.48	2.24	2.25
	O2	1.77	2.09	2.11
	μ-OH, O	-	1.96	1.93
P	μ-OH, O	-	2.83	2.82

indication of the positioning role of the two histidine residues, His202 and His296.

The optimized geometries for the ONIOM models are in good agreement. A visual indication of this fact is provided in Figure 3.6. Thus, for the mechanistic studies, we chose to use model 1OM.

### 3.4.3 Mechanistic Studies

The following section presents the results obtained for the mechanism B. As underlined previously for the DFT study, the model 1DM was used, while in the case of the ONIOM study we choose model 1OM. The substrate was modeled as dianionic phenyl phosphate ( $\text{PO}_3(\text{O}^-\text{Ph})^{2-}$ ), bridging the two metal ion. In order to envisage a nucleophilic attack of the bridging water molecule, the phenyl group of the substrate has to be placed on the oxygen atom found in the axial position.

After the geometry optimization of the reactant (2rDM and 2rOM), we noticed that for both models the tilted orientation of the substrate was kept, and the P-O(μ-OH) distances presented in Table 3.3 are good agreement with those obtained for models 1DM (Table 3.1) and 1OM (Table 3.2). Again, we find that the two histidine



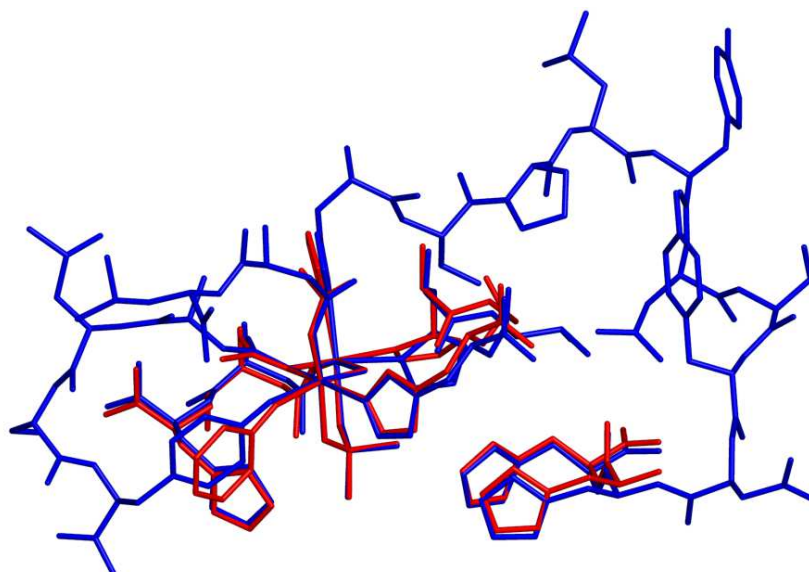


Figure 3.6: Superposition of the optimized geometries for model 1OM (red) and 1OL (blue). Hydrogens were removed for clarity.

residues (His202 and His296) which form hydrogen bonds with the substrate, aid in its orientation closer to the nucleophilic  $\mu$ -hydroxide bridge.

In the case of the ONIOM model, the transition state structure (2tsOM) was localized by scanning the distance P-O( $\mu$ -OH) in increments of 0.2 Å. The initial guess obtained from this PES scan was subsequently optimized, and in the following frequency calculation we obtained one imaginary value at  $-159\text{ cm}^{-1}$  which corresponds to the normal mode coupled with the attack of the phosphorus atom by the  $\mu$ -hydroxide group.

As expected, in the case of the transition state structures, the most notable differences are observed around the atoms directly implicated in the reaction. The distance P-O( $\mu$ -OH) decreases by 1.18 Å for the DFT model, while in the case of the ONIOM model we found a difference of 0.53 Å (see Table 3.3). A direct result of this modification is that the distances of the  $\mu$ -hydroxide bridge with the metal ions and the Fe-Zn distances are increased. In the ONIOM model the aforementioned values are closer those obtained for the reactant, most probably due to the stabilizing effect of the atoms from the outer layer.

Table 3.3: Main optimized distances (Å) for the reactant and transition state of mechanism B computed with the DFT and ONIOM models.

From	To	2rDM	2tsDM	2rOM	2tsOM
Zn	Asp164, O <sup>δ1</sup>	2.22	2.15	2.19	2.20
	Asn201, O <sup>δ1</sup>	2.21	2.14	2.17	2.16
	His286, N <sup>ε</sup>	2.24	2.17	2.15	2.15
	His323, N <sup>δ</sup>	2.26	2.24	2.34	2.34
	O1	2.08	2.05	2.37	2.25
	μ-OH, O	2.04	2.30	2.09	2.18
Fe	Asp135, O <sup>δ1</sup>	2.07	2.01	1.91	1.89
	Asp164, O <sup>δ1</sup>	2.11	2.16	2.05	2.04
	Tyr167, O	1.92	1.94	2.28	2.21
	His325, N <sup>ε</sup>	2.23	2.24	2.22	2.23
	O2	1.99	2.27	2.12	2.10
	μ-OH, O	2.04	2.27	1.89	1.96
Zn	Fe	3.17	3.45	3.23	3.32
P	μ-OH, O	3.08	1.90	2.81	2.27
P	O4	1.71	2.45	1.89	2.09

For the activation energy barrier we obtained a value of 151 kJ/mol for the DFT model, and a value of 53 kJ/mol for the ONIOM model. Thus, the inclusion in the ONIOM model of the amino acids not directly bound to the metal ions have a stabilizing role, thereby decreasing the energy required to initiate the reaction.

It is important to underline that in the case of all the ONIOM calculations presented until now we were forced to place the QM-MM bonded in a way that cuts through aromatic cycles in order to limit the system to a computationally manageable size. Nevertheless, we must ask ourselves what could be the influence of these approximations in the results obtained above. We have seen in the previous chapter that if the QM-MM border is placed sufficient far from the center of interest (e.g. reaction center), we obtain good agreement between ONIOM and DFT calculations on the same model. While this approximation holds true for the amino acids bound to the metal ions, the histidine residues close to the active site, which are modeled as NHCH<sub>2</sub>, might have influenced the height of the activation energy barrier. Thus, a more complete ONIOM model was elaborated which included the whole protein. Additionally, the position of

the QM-MM border was changed.

The initial model was prepared starting from the X-ray structure of kbPAP complexed with phosphate (PDB code: 4KBP) [KSF<sup>+</sup>96]. Since the four subunits presented in the crystallographic data are equivalent, only subunit D was included. The substrate was modeled as  $\text{PO}_3(\text{OPh})^{2-}$ . Hydrogen atoms were placed according to the predicted  $\text{pK}_a$  of the amino acids at the crystallization pH by employing the PROPKA 2.0 [LRJ05] software. The system was solvated in a cubic water box of about  $93 \times 93 \times 93 \text{ \AA}^3$  and was neutralized by randomly placing 13 chlorine ions, giving a total of 80,828 atoms. Prior to the molecular dynamics simulations, the system was energy minimized for 1000 steps using a steepest decent algorithm. During this minimization, the amino acids bound to the metal ions, the histidine His202, His295, His296, the two metal ions and the substrate were fixed to their crystallographic positions. Subsequently, a 100 ps NVT molecular dynamics simulation was performed, while keeping the same parts of the system fixed. A leap-frog stochastic dynamics integrator was used with a time constant  $\tau_i = 0.1 \text{ ps}$ . LINCS [HBBF97] was used to constrain all bonds with H-atoms. A  $1.4 \text{ \AA}$  cutoff was used for the Lennard-Jones interactions. Coulomb interactions within  $10 \text{ \AA}$  were computed each step, beyond this cutoff, the particle mesh Ewald method [DYP93] was used, with a reciprocal grid spacing of  $1.2 \text{ \AA}$ . All simulations were performed using the Gromacs simulation package [SLH<sup>+</sup>05] together with the AMBER-94 force field [CCB<sup>+</sup>95] for the protein and a TIP3P model [JCM<sup>+</sup>83] for water. The missing charges for the substrate were derived from the ESP by using the RESP methodology [BCCK93].

As can be seen from Figure 3.7, the system is stable during the equilibration, with an average RMSD, for the last 20 ps of the simulation, of  $0.8 \text{ \AA}$ .

For the calculation of the activation energy barrier, the last structure of the molecular dynamics simulation was used as a starting point, without the solvent molecules. The QM-MM border was placed on the  $\text{C}_\alpha - \text{C}_\beta$  bond. In the QM region the same amino acids were included as in model 2OM, giving a total of 133 atoms in the QM region from a total of 6836 atoms. The same functional and basis set was used for the description of the inner region, while for the outer region we used the AMBER-94 force field [CCB<sup>+</sup>95]. Due to the large number of degrees of freedom, the position of all amino acids that were more than  $12 \text{ \AA}$  from the phosphorus atom of the substrate was fixed during the optimization. All optimizations were done with the ONIOM-EE (electronic embedding) methodology [FTS<sup>+</sup>04]. In order to speed up the computations,

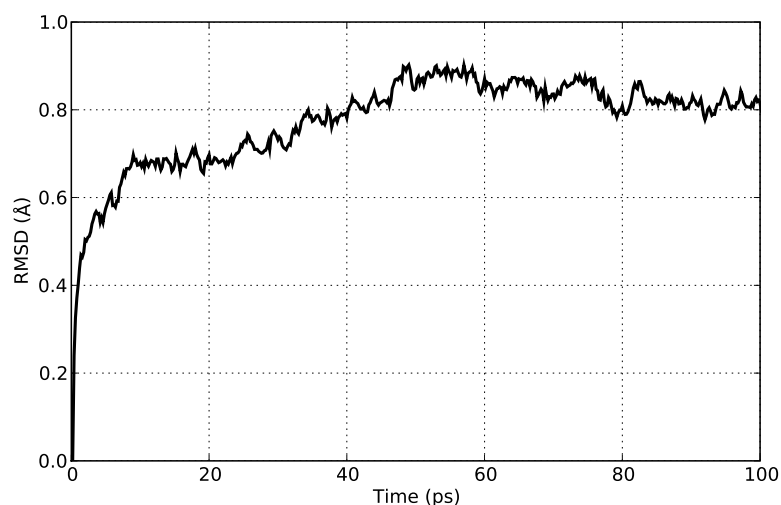


Figure 3.7: Root mean square deviation (RMSD) of the protein backbone during molecular dynamics.

the default convergence criteria from Gaussian 03 were modified by using the keyword `IOP(1/7=5000)` in the input. Consequently, the root mean square of all forces was considered optimized if lower than 0.005 a.u.

As in the case of model 2OM, the histidine His202 and His296 aid in correctly positioning the substrate for a nucleophilic attack. For the P-O( $\mu$ -OH) distance we find a value of 2.78 Å, which is in good agreement with the value obtained for the precedent ONIOM model (Table 3.3).

In order to locate the transition state we performed a relaxed PES scan. The reaction coordinate was taken as the difference between two distances R1 and R2, where R1 is the P-O(phenolate) distance and R2 is the P-O( $\mu$ -OH) distance. The results of this scan are illustrated in Figure 3.8.

We obtain a difference of 44.1 kJ/mol, close to the value obtained by using the model 2OM. We also tried to refine the transition state as was previously done for the smaller ONIOM model. Unfortunately, the large number of redundant internal coordinates made the gradient calculations computationally prohibitive due to memory limitations. Another possibility would be to refine the PES scan by increasing the number of points along the reaction coordinate. This work is currently in progress.

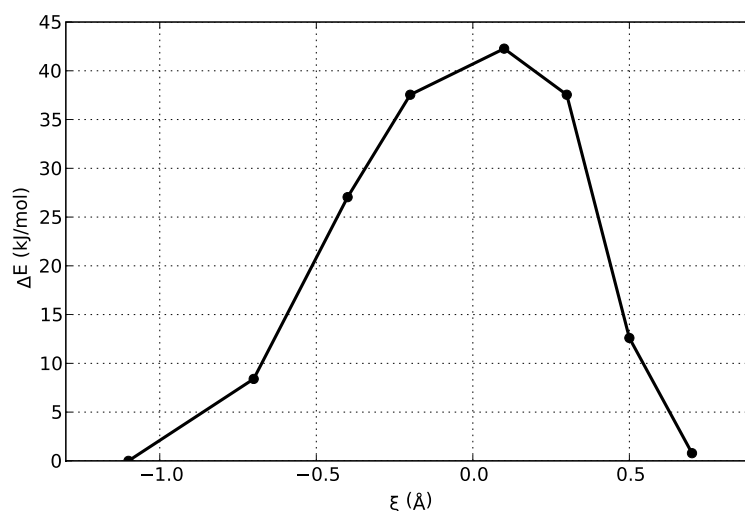


Figure 3.8: Results for the ONIOM PES scan for mechanism B. The reaction coordinate  $\xi$  is defined as  $\xi = d_{R1} - d_{R2}$ . Here  $d_{R1}$  is the distance between the phosphorous atom and the oxygen atom of the phenolate and  $d_{R2}$  is the distance between the oxygen atom of the nucleophilic hydroxide bridge and the phosphorous atom. The relative value of the QM energy for each point is plotted on the ordinate.

### 3.5 Conclusions and Perspectives

In this chapter we presented an application of the QM/MM methods for the mechanistic study of purple acid phosphatase from red kidney bean (kbPAP). Since the calculations required large amount of computational resources, we deemed necessary to realize a preliminary study on different sized models of the active site of the protein. We found that a medium model, in which the aromatic cycles of amino acids were cut by the QM-MM border, could give similar results to a more complex model. Subsequently, the activation energy barrier of the nucleophilic attack of the bridging hydroxide was evaluated by DFT and ONIOM calculations. Regarding this first ONIOM model, it is important to underline that the MM region served only as a mechanical constrain for the QM atoms, since calculations were done with mechanical embedding. Afterwards we were interested to see if the approximations made in this model had an influence on the calculated value of the activation energy barrier. A more complex model of the protein was employed, together with an ONIOM electrostatic embedding scheme and a new positioning of the QM-MM border. Surprisingly, we find a good agreement

between the two ONIOM studies. This is probably due to the cancellation of errors that takes place, since we compare a difference in energy.

From a methodological point of view, in the present study we had the opportunity of using the ONIOM method from the Gaussian 03 [FTS<sup>+</sup>04] software package. While the current implementation can be routinely used for relatively small systems, like those presented in the previous chapter, its application in the reactivity of large systems, like proteins, remains limited. This has mostly to do with the fact that the transition search algorithm does not make use of the microiteration scheme as in normal geometry optimizations. A way to circumvent this problem is to do a relaxed PES scan of the reaction coordinate, but besides this no other methodologies for locating transition state structures are implemented.

Sampling rare events in high-dimensional systems like proteins is not a trivial task. Example of proposed algorithms include: thermodynamic integration [CCHK89, SC98], umbrella sampling [PV75], weighted histogram analysis method [KRB<sup>+</sup>95], steered dynamics [Jar97], adaptive force biasing [DP01], conformational flooding [Gru95], metadynamics [LP02], to name just a few. Consequently, we turned our attention to other software packages in which some of these algorithms are implemented. Pradip Biswas has recently developed a interface [BG05] that couples CPMD [Hut] and Gromacs [SLH<sup>+</sup>05]. A large amount of time was spent on trying to run QM/MM calculations with this interface. These efforts resulted in the correction of some bugs from the code (see <http://www.tougaloo.edu/research/qmmm> for details). Unfortunately, due to the complex nature of the system (many atoms had to be included in the QM region and the high spin iron atom) the QM/MM molecular dynamics were very time consuming, and no meaningful conclusions could be formulated from the resulted short trajectories (1 ps of QM/MM simulation required approximately 20 days on 10 Itanium 2 processors).

To conclude, our attempts on modeling this system revealed the importance of His202 residue, that can act as a proton donor for the leaving group, and the fact that mechanism B is feasible, since we obtain an activation energy barrier for the nucleophilic attack of approximately 44.1 kJ/mol.



## Résumé

Ce dernier chapitre illustre les potentialités des méthodes QM/MM en drug design, application qui émerge actuellement. Après un rapide exposé sur les principales méthodes de calcul d'énergie libre, le développement d'une nouvelle méthode couplant un calcul énergétique QM/MM à un calcul de solvation PB/SA est présenté. Cette dernière est adaptée de l'approche MM/PB-SA connue pour évaluer de manière correcte les énergies d'interaction récepteur/protéine en calculant l'énergie libre du complexe en différents points d'une trajectoire d'une dynamique moléculaire. L'énergie libre est décomposée selon l'équation suivante :

$$\Delta G = \Delta E_{\text{gas}} + \Delta G_{\text{PB}} + \Delta G_{\text{SA}} - T\Delta S$$

où  $\Delta E_{\text{gas}}$  est l'énergie d'interaction dans la phase gazeuse,  $\Delta G_{\text{PB}}$  est l'énergie de solvation polaire calculée avec l'équation de Poisson Boltzmann et  $\Delta G_{\text{SA}}$  est l'énergie de solvation nonpolaire, proportionnel à la surface accessible au solvant. Enfin le terme  $T\Delta S$  représente la contribution entropique. En couplant cette méthode à une approche QM/MM en prenant, par exemple, le ligand dans la partie quantique, on s'affranchit des problèmes de paramétrisation pour chaque ligand testé et on prend en compte la polarisation exercée sur le ligand après fixation par la protéine. Cette approche appelée QM/MM-PB/SA a été développée et testée sur différents inhibiteurs de la protéine casein kinase 2 (CK2). Le choix de la protéine a été motivé par l'obtention d'une touche confirmée biologiquement lors d'un criblage de la chimiothèque de notre laboratoire. Les calculs ont été réalisés dans un premier temps sur des inhibiteurs connus afin de valider la méthode utilisée. Les positions du ligand obtenues sont en accord avec les données cristallographiques et l'analyse QM/MM-PB/SA fait ressortir l'importance de la solvation et des interactions de Van der Waals dans le mécanisme de reconnaissance. Enfin l'étude de nouveaux inhibiteurs a permis de comprendre les résultats obtenus lors du criblage.





# 4

## CK2 Protein and Its Inhibitors

### 4.1 Introduction

Nearly three decades have past since the first publications of free energy calculations on macromolecules [PBH82, TM84, JR85]. Even though the statistical theory underlying these calculations was developed long time ago [Kir35, Zwa54], scientists had to wait the advent of powerful computers, molecular sampling techniques applicable to macromolecular systems [KM83], and reliable force fields. Since the free energy dictates almost all the physical process (solvation, diffusion, binding, folding, and many others), it's not hard to understand the high interest for this subject. Within a few years, publications appeared that seemed to indicate that these methods could easily and reliably be used to determine the free energy differences between molecules and molecular states [LMW86, HW87, SBP86]. Unfortunately, when the free energy methods were taken up by laboratories to be applied on their particular systems, they could not reproduce the promised close agreement between theory and experiment. We know for a fact that these first simulations used in predicting the free energy were often very short, 10-40 ps of sampling, and thus they cannot yield the kind of predictability

they appeared to offer. Since these early days, free energy simulation techniques have been subject to intense research efforts. Only recently they have become reliable, due in part to the better sampling provided by the more powerful computers, but more importantly, to the improved theoretical approaches with better convergence properties.

In this chapter we present our contribution to the field of free energy calculation of ligand binding. Following a short theoretical review (Section 4.2), we present the system (Section 4.3) and the simulation protocol (Section 4.4) used. Section 4.5 outlines the most important results for the current study, while Section 4.6 concludes.

## 4.2 Free Energy Calculations and Ligand Binding

The following theoretical background begins with a short review of the so called rigorous approaches that are based directly on exact equations from classical statistical mechanics, the Free Energy Perturbation (FEP) and the Thermodynamic Integration (TI). This is followed by the presentation of less “exact” approaches that combine these equations with assumptions and approximations to allow simulations to be carried out more rapidly.

### 4.2.1 Free energy perturbation

Consider a system of  $N$ -particles described by the Hamiltonian  $\mathcal{H}_0(\mathbf{r}, \mathbf{p})$ , which is a function of  $3N$  Cartesian coordinates,  $\mathbf{r}$ , and their conjugated momenta  $\mathbf{p}$ . We are interested in calculating the free energy difference between this reference system and the target system characterized by the Hamiltonian  $\mathcal{H}_1(\mathbf{r}, \mathbf{p})$ .

The difference in the Gibbs free energy between the target and the reference system,  $\Delta G$ , can be written in terms of the ratio of the corresponding partition functions  $Q_1$  and  $Q_0$ :

$$\Delta G = -\frac{1}{\beta} \ln \frac{Q_1}{Q_0} \quad (4.1)$$

where  $\beta = (k_B T)^{-1}$ , and

$$Q = \frac{1}{b^{3N} N! V_0} \int dV \iint \exp\{-\beta[\mathcal{H}(\mathbf{r}, \mathbf{p}) + PV]\} d\mathbf{r} d\mathbf{p} \quad (4.2)$$

The normalization constant contains Plank's constant  $h$ , which is a measure of the elementary volume in phase space, the factor  $N!$ , which should be present only when the particles are indistinguishable, and  $V_0$ , which is the basic unity of volume and renders the partition function dimensionless.

Assuming that the  $PV$  work is negligible, Equation 4.2 can be rewritten in terms of ensemble averages [Zwa54]:

$$\Delta G = -\frac{1}{\beta} \ln \langle \exp[-\beta \Delta \mathcal{H}(\mathbf{r}, \mathbf{p})] \rangle_0 \quad (4.3)$$

where  $\Delta \mathcal{H}(\mathbf{r}, \mathbf{p}) = \mathcal{H}_1(\mathbf{r}, \mathbf{p}) - \mathcal{H}_0(\mathbf{r}, \mathbf{p})$ .

Here,  $\langle \dots \rangle_0$  denotes an ensemble average over the configurations sampled from the reference state. This is the fundamental FEP formula, and it states that  $\Delta G$  can be estimated by sampling only the equilibrium conformations of the reference state.

If the two systems of particles have the same mass, the integration implied by the statistical average is carried out over particle coordinates only. Thus the difference of kinetic energy ( $\Delta \mathcal{K}$ ) between the target and the reference systems cancels out, and  $\Delta \mathcal{H}$  can be replaced by the difference in the potential energy,  $\Delta \mathcal{V}$ :

$$\Delta G = -\frac{1}{\beta} \ln \langle \exp(-\beta \Delta \mathcal{V}) \rangle_0 \quad (4.4)$$

If we reverse the reference and the target systems, and repeat the same derivation, we obtain:

$$\Delta G = -\frac{1}{\beta} \ln \langle \exp(-\beta \Delta \mathcal{V}) \rangle_1 \quad (4.5)$$

Although Equation 4.4 and Equation 4.5 are formally equivalent, their convergence properties may be quite different. This means that there is a preferred direction to carry out the required transformation between the two states.

The FEP method will provide accurate estimation of free energy differences, only if the target systems is sufficiently similar to the reference system. If the two systems do not overlap in phase space, the target system is not expected to be sufficiently sampled during a simulation of the reference system. In such cases it is necessary to use enhanced sampling techniques such as stratification. Stratification, sometimes also called multistage sampling, works by describing the transformation from the reference system to the target system with a coupling parameter  $\lambda$ . The simplest approach

involves linear interpolation, such that  $\lambda = 0$  and  $\lambda = 1$  for the reference and the target states, respectively.

$$\mathcal{H} = \lambda \mathcal{H}_1 + (1 - \lambda) \mathcal{H}_0 \quad (4.6)$$

Here  $\mathcal{H}_0$  and  $\mathcal{H}_1$  denote, as before, the Hamiltonian of the reference and the target system, respectively, but for the sake of clarity their dependence on  $\mathbf{r}$  and  $\mathbf{p}$  was not explicitly specified. If we create  $N - 2$  intermediate states linking the reference and the target state such as  $\lambda_1 = 0$  and  $\lambda_N = 1$ , Equation 4.4 becomes:

$$\Delta G = \frac{1}{\beta} \ln \sum_{\lambda=1}^{N-1} \langle \exp(-\beta \Delta \mathcal{H}) \rangle_{\lambda} \quad (4.7)$$

Note that the transformation between the two systems by the variable  $\lambda$  may or may not correspond to a physical transformation. The free energy is a thermodynamic state function and its value is independent of the path along which the change is made, as long as it is reversible. It is also important that there is enough overlap between successive  $\lambda$  states. In general a change that involves high energy barriers will require much smaller increment of  $\lambda$  to ensure reversibility of the process with respect to a pathway with a lower barrier. Obviously, the computation time increases with the number of subintervals involved, since individual simulations are necessary for each intermediate state  $\lambda_i$ .

## 4.2.2 Thermodynamic Integration

We are interested in free energy differences,  $\Delta G$ , between two systems or states of a system linked by a coupling parameter  $\lambda$  as defined above. Assuming the free energy  $G$  is a continuous function of  $\lambda$ , we have the identity:

$$\Delta G = \int \frac{\partial G}{\partial \lambda} d\lambda \quad (4.8)$$

The key observation is that it is possible to calculate  $\partial G / \partial \lambda$  by recognizing that it is in fact equal to the following statistical average [Kir35]:

$$\frac{\partial G}{\partial \lambda} = \left\langle \frac{\partial \mathcal{H}}{\partial \lambda} \right\rangle_{\lambda} \quad (4.9)$$

where the averaging is over the ensemble belonging to  $\lambda$ . The difference between the two systems for which  $\lambda$  is zero and unity can be evaluated from the integration of Equation 4.9

$$\Delta G = \int_0^1 \left\langle \frac{\partial \mathcal{H}}{\partial \lambda} \right\rangle_{\lambda} d\lambda \quad (4.10)$$

In the early days of free energy simulations the TI approach was synonymous with what has later become called *slow growth*. This means that the value of  $\lambda$  actually was changed in each time step of an MD simulation, so that the ensemble average can be approximated by its instantaneous value (a sample over one step). This assertion cannot be rigorously proven, and the consensus today is that a “non-continuous” change in  $\lambda$  is a better choice. The primary reasons are that equilibration can be allowed at each point, extra points can be added at any time, and any pattern of spacing between the  $\lambda$ -points can be used in order to optimize the efficiency.

Both FEP and TI are carried out by systematically varying  $\lambda$  from the reference state to the target state. At each  $\lambda$  point equilibration of the system is performed, followed by a data collection to determine the value of the ensemble for the equilibrated system. It is important to note that the dividing of the pathway between the two physical endpoints is done for different reasons in FEP and TI. In FEP, we use the  $\lambda$  intermediates to reduce the difference between adjoining states when applying Equation 4.7. This improves the convergence profile for the required ensemble. In TI, the  $\lambda$  intermediates are required to approximate the continuous integral in Equation 4.10. The number of intermediates required when using TI will depend on the shape of the accumulated free energies versus  $\lambda$  profile. The greater the variation in the curvature in this profile, the more points that will be required to correctly approximate the required integral.

### 4.2.3 Thermodynamic cycle

The above equations allow us to calculate the free energy difference between any two configurations or molecules for which it is possible to postulate an interconversion pathway. It is not a straightforward exercise, in particular to determine the protein-ligand binding free energy. Important geometry modifications can occur with the complexation. In this case, free energy simulations are most frequently carried in the context of thermodynamic cycle [TM84] and we evaluate the relative binding free

energy between two ligands ( $L_1$  and  $L_2$ ) and a given receptor. Bash *et al.* [BSB<sup>+</sup>87] were the first to apply this method for the calculation of relative change in free energy of a receptor-ligand system. For this a thermodynamic cycle similar to the one depicted in Figure 4.1 is used.

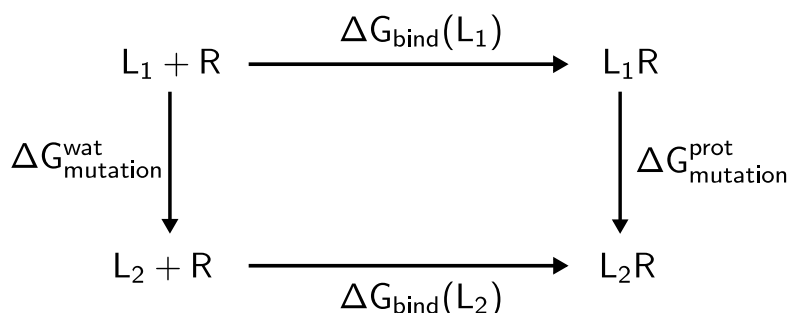


Figure 4.1: Thermodynamic cycle used in FEP/TI calculations of the relative binding free energies of two ligands,  $L_1$  and  $L_2$ , to a receptor molecule  $R$ .

$\Delta G_{\text{mutation}}^{\text{wat}}$  and  $\Delta G_{\text{mutation}}^{\text{prot}}$  in Figure 4.1 denote the difference in free energy between  $L_1$  and  $L_2$  in water and bound to the solvated receptor binding site, respectively while  $\Delta G_{\text{bind}}$  are the corresponding free energies. The relative free energy of binding hence becomes:

$$\Delta \Delta G_{\text{bind}} = \Delta G_{\text{bind}}(L_2) - \Delta G_{\text{bind}}(L_1) = \Delta G_{\text{mutation}}^{\text{prot}} - \Delta G_{\text{mutation}}^{\text{wat}} \quad (4.11)$$

With the FEP or TI approach, we calculate the free energy associated with the two unphysical paths,  $L_1 \rightarrow L_2$  and  $L_1 R \rightarrow L_2 R$  which corresponds to a mutation of  $L_1$  into  $L_2$  in the free and the bound state.

#### 4.2.4 End-Point Methods

End-point methods are computational effective alternatives to more exact methods such as FEP and TI. These methods sample only the free and bound states and compute  $\Delta G$  by taking a difference. Besides their low computational cost, the end-point methods can be applied to structural diverse compounds, since they do not need the simulation of an unphysical transformation between molecules. However, their theoretical foundations still needs to be strengthened, although efforts are being made in this direction [SHM04]. Two methods have showed promising results for the estimation of absolute and relative binding free energies, the Linear Interaction

Energy (LIE) [ÅMS94] and the Molecular Mechanics Poisson-Boltzmann Surface Area (MM-PB/SA) [SCC<sup>+</sup>98, KMR<sup>+</sup>00].

### Linear Interaction Energy

The basic idea behind LIE is to consider the absolute binding energy of a ligand, the change in free energy when transferred from solution to the solvated receptor binding site, as composed of a polar and nonpolar contribution. For the determination of the polar contribution to the binding free energy the linear response approximation (LRA) is used, while the nonpolar contribution is estimated using an empirically derived parameter that scales with the intermolecular van der Waals interaction energies from the MD simulations. This lead to an approximate equation for the binding free energy of the following type:

$$\Delta G_{\text{bind}} = \alpha \Delta \langle V_{l-s}^{\text{vdw}} \rangle + \beta \Delta \langle V_{l-s}^{\text{elec}} \rangle + \gamma \quad (4.12)$$

where  $\langle \dots \rangle$  denotes ensemble averages of the van der Waals (vdw) and electrostatic (elec) interaction of the ligand and its surrounding environment (l-s). The  $\Delta$ 's denotes the changes in these averages when the ligand is transferred from solution (free state) to the receptor binding site (bound state). The first applications of the LIE method used a  $\beta$  coefficient of 1/2 (as predicted by the LRA), along with a nonpolar coefficient  $\alpha$  of 0.16 and the additional constant  $\gamma$  set to zero. By excluding  $\gamma$  from the Equation 4.12 it was assumed that all effects of the binding process can be approximated only by van der Waals and electrostatic interactions. Although these parameters gave satisfying results, it was found using FEP calculation that  $\beta$  could be considered as a function of the ligand nature. Values of 0.5, 0.43, 0.37, and 0.33 were suggested for ionic molecules, and neutral compounds with one, two, or more hydroxyl groups, respectively [HMÅ98]. A value of 0.18 was found to be optimal for  $\alpha$ , while a non-zero value can be necessary to correctly reproduce  $\Delta G_{\text{bind}}$  for some systems [LMM<sup>+</sup>01].

Several important contributions to molecular recognition are neglected in LIE, such as the conformational rearrangement upon complexation of the ligand and the receptor, the receptor desolvation energy, and entropies. However, it has been argued that these terms are implicitly taken into account by the LRA and the adjustable parameters of the model [ZFG<sup>+</sup>01, ÅM01]. This method is generally applied to structurally related



molecules, and the cancellation of errors contributes to explain the success of the method in estimating  $\Delta G_{\text{bind}}$ .

### Molecular Mechanics Poisson-Boltzmann Surface Area

Another approach that has gained considerable attention in the last few years for estimating association free energies of molecular complexes is the MM-PB/SA method. It consists in post processing of molecular dynamics trajectories using a continuum solvent approach. The method approximates the “average” free energy of a state as:

$$\langle G \rangle = \langle E_{\text{MM}} \rangle + \langle G_{\text{PBSA}} \rangle - T \langle S_{\text{MM}} \rangle \quad (4.13)$$

where  $\langle E_{\text{MM}} \rangle$  is the average molecular mechanical energy arising from bonds, angles, torsion angle van der Waals and electrostatic terms from a regular force field. It is evaluated with no nonbonded cutoff. Solvation free energies are calculated using a numerical solution of the Poisson-Boltzmann equation [WW82, GH88] and with the solvent accessible surface area (SASA) based on an estimation of the nonpolar free energy [SSH94]. It constitutes the  $\langle G_{\text{PBSA}} \rangle$  term. Finally  $-T \langle S_{\text{MM}} \rangle$  represents the entropic term, due to the loss of degrees of freedom upon association. Recently, an approach related to MM-PB/SA, in which the electrostatic contribution to the solvation energy is determined by using a GB model [STHH90] has been introduced under the name Molecular Mechanics Generalized Born Surface Area (MM-GB/SA) [SCC<sup>+</sup>98, GKC03]. Despite its approximations, the GB model makes this variant attractive because it is much faster than PB. The nonpolar contribution to the solvation energy can be considered as the sum of a cavity term and the solute-solvent van der Waals term and can be approximated with  $\gamma \times \text{SASA} + \beta$ . This relation comes from a well known fact that the solvation free energy of saturated nonpolar hydrocarbons is linearly related to SASA [Her72]. The surface tension  $\gamma$  and the constant  $\beta$  can be set to  $0.00542 \text{ kcal mol}^{-1} \text{ \AA}^{-2}$  and  $0.92 \text{ kcal mol}^{-1}$ , respectively. Both  $\langle E_{\text{MM}} \rangle$  and  $\langle G_{\text{PBSA}} \rangle$  are obtained by averaging over a sample of representative geometries extracted from a MD trajectory of the system. For the entropic term, its contribution can be evaluated by using quasi-harmonic analysis or by normal-mode analysis. Previous studies [Hes02, BWC06] have showed that the covariance matrix of atomic fluctuations, needed to compute the entropy in a quasi-harmonic analysis, converges very slowly. In some cases even after 10 ns of simulation time, the convergence of entropy was not

achieved [GC04]. On the other hand, for example, estimation of the ligand entropies in solution from normal mode analysis around conformations from a simulation of its complex with a receptor, may be a drastic simplification. First of all, one assumes that the local minimum may be represented by an harmonic potential. Secondly, since the water molecules and counterions are striped from the snapshot, a distance dependent on a dielectric constant has to be used to mimic the solvent in the course of the minimization prior to the normal mode analysis. Finally, because the normal-mode analysis is very time consuming compared with the other energy evaluations, fewer snapshots are usually used. All these factors can lead to large errors in estimating the entropy and explains why in calculations of binding free energies for a series of similar ligands to a protein receptor, the entropic contribution calculation is often not performed [KMR<sup>+</sup>00].

The MM-PB/SA approach was initially used to study the stability of various RNA and DNA fragments [SCC<sup>+</sup>98], but in the last years has also been used to estimate ligand binding free energies [KMR<sup>+</sup>00, WMWK01, HWC<sup>+</sup>02]. In order to calculate the binding free energy between a ligand and a receptor, two alternatives exist for this methodology. The first is to evaluate the term in Equation 4.13 for the complex, receptor and ligand based on separate trajectories with a subsequent determination of  $\Delta G_{\text{bind}}$  according to:

$$\Delta G_{\text{bind}} = \langle G_{\text{complex}} \rangle - \langle G_{\text{receptor}} \rangle - \langle G_{\text{ligand}} \rangle \quad (4.14)$$

The second alternative is to determine each term in Equation 4.14 based on the snapshots taken from the trajectory of the complex only, in which case the two latter terms are estimated simply by “removing” one of the molecular partners from the trajectory. In practice, calculations with separate trajectories are often not used in receptor-ligand studies, because it is very difficult to get the  $\langle E_{\text{MM}} \rangle$  term to converge for the receptor or complex within reasonable computing time. Hence, the regular implementation of this method actually assumes that the structure of the receptor and the ligand does not change upon binding, because no intramolecular terms either in the receptor or ligand are taken into account. This is in contrast to the LIE method in which such terms are considered in terms of responses to the intermolecular interaction through the appropriate weight coefficients.

The MM-PB(GB)/SA is expected to estimate  $\Delta G_{\text{bind}}$  without adjustable parameters.

Although several studies were able to reproduce experimental  $\Delta G_{\text{bind}}$  for ligand binding free energies with an error lower than  $1 \text{ kcal mol}^{-1}$ , these results are open to discussion. Indeed, the approach contains several “hidden” parameters, like the force field used, the choice of PB or GB and that of the nonpolar solvation model, the use of one or three trajectories, and the different terms that can be included or neglected. As a consequence, it is sometimes possible to find a combination of such hidden parameters apparently allowing a fine estimation of  $\Delta G_{\text{bind}}$  for a given system. However, the transferability of such results to other systems is questionable. Nevertheless, MM-PB(GB)/SA has proven to be useful for several applications less sensitive to the choice of hidden parameters, such as the comparison of relative stabilities of macromolecular conformations, determination of relative affinities for different small ligands in drug design applications, and estimation of the effect of mutations on association processes and fold stability.

Although some studies aimed at determining absolute  $\Delta G_{\text{bind}}$  for ligand-receptor association, MM-PB(GB)/SA is usually used to estimate relative affinities for different ligands targeting the same receptor. This allows additional approximations, like as we see before, the neglect of the entropy terms for ligands of similar mass binding to the same site. Also, despite the fact that this approach is expected to tackle chemically diverse ligands, it is often applied to a series of chemically related ligands. This obviously simplifies the problem thanks to additional cancellation of errors, but it also reflects the usual drug discovery process that generally focus on families of similar ligands.

#### 4.2.5 Improved Potential Energy Models

Success in the computational study of ligand-receptor binding free energy depends critically on both sampling techniques and the energy function employed. As seen from previous sections of this chapter, sampling of relevant conformational states can be as important as providing an accurate representation of the inter- or intramolecular interactions. With that knowledge in mind, we will focus our attention on ways to improve the potential energy models, going beyond the typical empirical force fields.

Historically, force field development has focused on amino acids and nucleotides and not necessary on small organic molecules as we can find in drug design. Thus to calculate ligand-receptor binding free energy, the parameters for the ligand must

be defined. The atomic charges for the ligand were often estimated using standard quantum mechanical calculations and the van der Waals parameters were taken directly from the force field used for the rest of the system. Nowadays an increasing number of automated procedures for deriving these parameters are available. For example, the Merck molecular force field (MMFF) [Hal96] was designed to be a transferable force field for pharmaceutical compounds that accurately treats conformational energetics and nonbonded interactions. An algorithm to automatically assign molecular mechanics parameters for ligands, based on the general AMBER force field (GAFF), was also developed [WWC<sup>+</sup>04, WWKC06]. This approach derives ligand parameters that are consistent with the AMBER force field parameters used for the description of the receptor.

Perhaps one day there will be computers that will allow the computation of binding affinities from first-principle QM simulations in femtoliter solvent droplets, thus making the current approaches obsolete. Unfortunately this is not feasible yet, but steps can be taken in this direction. In the last couple of years QM/MM based methods have started to emerge as an alternative to the all classical representation of the system. By explicitly treating the ligand with a QM model, problems associated with the development of specific parameters can be overcome. It is also possible to extend this model for the treatment of the receptor and solvent, but for efficiency reasons, we have to settle for a classical representation.

The group of Iñaki Tuñón have made extensive use of the QM/MM methodologies for analyzing the interaction between ligands and receptors [AMC<sup>+</sup>07, AMCS<sup>+</sup>07, AMC<sup>+</sup>08]. In these studies the sampling of the conformational space was realized by QM/MM MD simulation, with a semiempirical representation of the ligand. They reported very good correlation between experimental binding affinity of five CDK2 inhibitors derived from O<sup>6</sup>-cyclohexylmethylpurine and the QM/MM interaction energy [AMCS<sup>+</sup>07].

With a hybrid representation of the system, polarization effects are taken into account by adding an electrostatic term to the Hamiltonian that represents the receptor and solvent by a set of point charges. Hensen *et al.* [HHN<sup>+</sup>04] offered a quantified picture of the importance of these effects in drug design. In their study of the HIV-1 protease with three high affinity inhibitors, nelfinavir, mozenavir, and triptonavir, they found that the polarization of the ligand by the enzyme environment contributes up to 39% of the total electrostatic interaction energy. Based on this they proposed

modifications to one of the inhibitors that can possibly lead to an increase in binding affinity. Using the same theoretical foundation, Garcia-Viloca *et al.* [GVTG03] have investigated the role of polarization of the substrate tetrahydrofolate, and the cofactor NADPH, at various stages of the dihydrofolate reductase-catalyzed hydride transfer reaction. The authors found that polarization contributes 4% of the total electrostatic interaction energy, and it accounts for 9 kcal/mol of the transition state stabilization relative to the reactant state.

Khandelwal *et al.* [KLC<sup>+</sup>05] used a four-tier approach that involves docking, QM/MM optimizations, constrained MD simulation, and QM/MM interaction energy calculation to accurately predict the binding affinity of a series of hydroxamate-based inhibitors of matrix metalloprotease (MMP-9). The authors used a modified version of the extended linear response theory (ELR), which is derived from LIE. The inclusion of MD sampling and QM/MM energy evaluation significantly improved both correlation and prediction.

A few years ago, Gräter *et al.* were the first to use QM/MM combined with PB/SA for the study of binding free energy of trypsin complexed with benzamidinium derivatives and FKBP inhibitors [GSD<sup>+</sup>05]. The ensemble used by the authors for the evaluation of the binding free energy was obtained via a docking protocol in the case of benzamidinium derived inhibitors while for the FKBP inhibitors was obtained including experimental data. The accuracy of the prediction was higher if the FKBP inhibitors were included in the data set (correlation coefficient = 0.56), compared to only the benzamidinium derived inhibitors (correlation coefficient = 0.20). The authors concluded that the predicted binding affinities are influenced by the quality of the structures used.

In the following, we present our results in this domain. We use a similar QM/MM-PB/SA formalism to gain a further insight into the specific binding interactions between CK2 and a number of inhibitors derived from tetrabromo-benzimidazole/imidazole. In contrast with the study of Gräter *et al.* [GSD<sup>+</sup>05], the sampling of the conformational ensemble was obtained via QM/MM molecular dynamics simulations.

### 4.3 The System

More than 500 different protein kinase are encoded in the human genome [MWM<sup>+</sup>02]. The similarity of the reaction mechanism has proven to be a major obstacle in developing specific inhibitors. CK2 (an acronym derived from the misnomer “casein kinase 2”), which catalyzes the transfer of the  $\gamma$  phosphate of ATP to serine and threonine residues of protein substrates, is the most pleiotropic protein kinase known to date. It has more than 300 substrates already identified, many of which are involved in signal transduction, gene expression and DNA synthesis and repair [MP03]. The elevated activity of CK2 in a number of tumors [TYW<sup>+</sup>01], in conjunction with the observation that many viruses relay on it to perform phosphorylation of the proteins essential to their life cycle, makes the CK2 a potential target for antineoplastic and antiinfectious drugs.

Consequently a number of ATP site directed inhibitors in the low micromolar range were described in the literature. Among these inhibitors, TBB (4,5,6,7-tetrabromobenzotriazole) has proven to be one of the most successful for in cells and in vivo studies [RPP02]. Recently the efficiency of TBB was improved by replacing the triazole ring by an imidazole, functionalized with a variety of polar functions [PAR<sup>+</sup>04]. The crystal structures of CK2 with different inhibitors and mutational analysis show a series of bulky residues which make the ATP binding site smaller in CK2, compared to other protein kinase [SdMR<sup>+</sup>03, BMS<sup>+</sup>01, BMS<sup>+</sup>05, BMC<sup>+</sup>07]. It is likely that the remarkable selectivity of TBB towards CK2 among a panel of more than 30 protein kinase [SRM<sup>+</sup>01] could be attributed to this fact.

Virtual screening of a large library of compounds is a common strategy for identifying new inhibitors. Such an approach was adopted by Prudent *et al.* [PMLR<sup>+</sup>08] in order to find innovative CK2 inhibitors. They performed high-throughput screening of multiple chemical libraries of low molecular weight compounds, including the Chemical Library of DCM. The *N*-hydroxyphthalimide derivatives, which were synthesized by Jacques Einhorn and his group (DCM-SERCO), were able to inhibit cellular CK2 activity. Especially the compound in which all the substituents of the *N*-hydroxyphthalimide skeleton are phenyl groups proved to be as active as TBB in promoting cell death. Interestingly, if in the above compound the hydroxyl group is replaced by a hydrogen, the activity of the CK2 protein is not affected.

Since X-ray structures of either of the two inhibitors complexed with the CK2

protein were not resolved, we decided to run a preliminary study on the family of inhibitors derived from TBB, for which crystallographic data together with inhibition constant determination was available. The advantage of this approach is twofold. In the first place we can evaluate and quantify the specific interactions that are responsible for the difference in binding affinities for related inhibitors. Secondly, by simulating free energies of binding for inhibitors for which extensive data is available in the literature, we can validate our simulation protocol before applying it to other inhibitors.

The rest of the chapter focuses on the application of a QM/MM-PB/SA method on determining the relative free energy of binding for the TBB derived inhibitors. With the solid data collected from these initial simulations, we tackle the inhibitors developed in our laboratory.

## 4.4 Methodology

### 4.4.1 Construction of the Initial Models

For the initial study we used 5 inhibitors that share a common core structure with TBB (Table 4.1). The inhibitors are named in conjunction to their relative inhibition power towards CK2, with L1 being the most potent one and L5 the least potent inhibitor. Their experimental inhibition constant,  $K_i$ , are all taken from the same reference [PAR<sup>+</sup>04], and it can be related to the change in free energy of binding with the equation  $\Delta G = k_B T \ln K_i$ , where  $k_B$  is the Boltzmann constant, and T is the absolute temperature. The structure of CK2 complexed with 4,5,6,7-tetrabromo-*N,N*-dimethyl-1H-benzimidazol-2-amine (L1) (PDB code: 1ZOE) [BMS<sup>+</sup>05] was used as a template to construct the initial models for the other complexes. This was accomplished with the MOLDEN package [SN00]. It should be noted that the crystal structure of CK2 complexed with L2 (PDB code: 1ZOG) [BMS<sup>+</sup>05], and L4 (PDB code: 1J91) [BMS<sup>+</sup>01] were also available. The former structure was chosen because of its higher X-ray resolution: 1.77 Å compared with 2.30 and 2.22 Å for CK2-L2 and CK2-L4, respectively. All missing residues were added using the PyMol software [DeL02]. Hydrogen atoms were placed according to the predicted  $pK_a$  of the amino acids at a pH of 7 using PDB2PQR [DNMB04], which assigns protonation states based on  $pK_a$  values calculated with PROPKA 2.0 [LRJ05].

Table 4.1: Inhibitor activity against the CK2 for the TBB derived inhibitors.

Id.	Structure	$K_i$ ( $\mu\text{M}$ )
L1		0.04
L2		0.07
L3		0.23
L4 (TBB)		0.40
L5		1.90

#### 4.4.2 Hybrid Potential Simulations

After the initial model build, the system was placed in a cubic box of water molecules with a margin of 10 Å along each dimension. All water molecules that within 2.8 Å from the complex were removed. An appropriate number of counterions were added to neutralize the system. The entire system was then minimized by per-



forming 1000 steps of CG (conjugate-gradient). Since the implementation of hybrid schemes in the fDynamo library [FAB<sup>+</sup>00] does not allow for periodic boundary conditions, the protein residues and water molecules that were more than 18 Å away from the inhibitor were fixed, and remained fixed during the whole simulation time. This means that our simulations were basically run in the NVT ensemble, since the volume of the box was kept fixed. With this setup we equilibrated the system by using Langevin-Verlet MD simulations. A successive visual inspection of the resulted trajectories, brought to our attention, that the sphere of unconstrained atoms was “missing” water molecules. This is probably due to the fact that the certain parts of the system were fixed right before the MD simulations, and a initial empirical rule of removing water molecules is not expected to reproduce the correct solvation of the system. We didn't asses the influence of this “artifact” on the resulted energies of binding, but we decided to take a more rigorous approach in resolving it.

If one wants to get the system to a correct density, MD simulations in NPT ensemble are required. This means that the volume of the box has to be variable. NPT simulations using a hybrid potential are currently not possible in the fDynamo library. So, after the initial optimization needed to remove bad contacts in the X-ray structure, the system was heated to 300 K in the course of 20 ps MD simulations in NVT ensemble, and the density of the system was subsequently brought to an approximate value of 1 by running a 100 ps MD simulation in the NPT ensemble. The inhibitor was modeled by using CHELPG charges derived from quantum mechanical calculations, while vdW parameters were adapted from the force field. In the course of these simulation the position of both the protein and the inhibitor was maintained by using harmonic constrains. After this we run 1.5 ns of QM/MM simulation, with the last 0.5 ns being used for the post processing analysis. The semiempirical AM1 model [DZHS93] was used to describe the inhibitor, while the OPLS-AA force field [JMTR96] was adopted for the protein and the three-point TIP3P model [JCM<sup>+</sup>83] for water. The use of a semiempirical QM model allows a fair sampling at a reasonable computational cost, even for large inhibitors. We used a time step of 1 fs for integrating the equations of motion, and a friction coefficient of 2 ps<sup>-1</sup> for the Langevin-Verlet algorithm. The electrostatic and Lennard-Jones interactions were calculated using a spherical cutoff distance of 12 Å with a switch function in the region 8 to 12 Å to feather the interaction energy to zero. The Lennard-Jones parameters for bromine atoms were set to 3.47 Å for  $\sigma$  and 0.47 kcal/mol for the depth of the potential well,  $\epsilon$ .

All QM/MM simulations were performed using the fDynamo library [FAB<sup>+</sup>00].

### 4.4.3 QM/MM-PB/SA and Energy Decomposition

As seen in the introduction of this chapter, the MM-PB/SA method has successfully been applied for the estimation of binding free energies of drug candidates to proteins. In this approach developed by Kollman *et al.*, all the system, inhibitor included, is represented with an empirical energy function. The current force fields have been used to successfully simulate macromolecular complexes for a long time now, thus proving a relatively high level of confidence in their parameters. This is not always the case for parameters of the organic molecules derived, for example, by the automatic procedure implemented in AMBER software package. These parameters have to be validated, and this can be a very difficult task. The situation is more severe for other force fields where no such automatic procedure exists, therefore a lot of work is required when a new inhibitor is under study.

By using a QM/MM MD simulation for the sampling required prior to the PB/SA analysis, the parameter problem is resolved, since the inhibitor is represented with a QM Hamiltonian. Additionally the electrostatic interaction between the inhibitor and its surroundings is computed at a quantum mechanical level, which includes the polarization of the inhibitor by the environment. In all other regards our approach follows closely the one developed by Kollman *et al.*

For each protein-inhibitor complex, a total of 51 equispaced snapshots were taken from the last 0.5 ns of the MD trajectory. We used a single trajectory approach which means that the energy coming from bonds, angles and torsion angles cancels out. The fDynamo library was used for the computation of the nonbonded interactions for which an infinite cutoff was employed.

The electrostatic contribution to solvation was estimated using the DELPHI software [GH88]. Grids dimensions  $169 \times 169 \times 169$  were used. The coarse density maps calculated with a scale of 0.5 grids/Å and Coulomb boundary conditions, were subsequently refined in by reducing the spacing to 2 grids/Å. Charges and van der Waals radii were taken from the OPLS-AA force field. Because the force field defines  $\sigma$ , but not van der Waals radii per se, the atomic radii were taken to be half of the values of the Lennard-Jones  $\sigma$  parameter, except for the hydrogens where we used a value of 1.15 Å. In the case of the inhibitor the charges were derived by ab initio

electrostatic surface potential (ESP) calculation (HF/6-31G\*), using the CHELPG module from Gaussian 03 [FTS<sup>+</sup>04] software package. The dielectric constant was set to unity for the solute and 80 in the case of the solvent. Since this term is computed classically, the polarization of the ligand by the solvent is neglected. Moreover, the classical estimate of the ligand solvation energy by using a charge distribution that differs from what would be obtained with the AM1 method used in the QM/MM part. A natural solution for this problem was recently proposed by Wang *et al.* [WW07] and consists in using the quantum mechanical charge density of the ligand directly in the Poisson-Boltzmann solver. However, this required modification of the DELPHI software which were beyond the scope of the present study, but which might be addressed in future implementations.

The solvent-accessible surface area was estimated using the fDynamo library with a solvent probe radius of 1.4 Å. This term was then used to compute the nonpolar component of the solvation with the equation  $\gamma \times \text{SASA} + \beta$  [SSH94], where  $\gamma$  and  $\beta$  are empirical constants, and in this work they were set to 0.00542 kcal mol<sup>-1</sup> Å<sup>-2</sup> and 0.92 kcal mol<sup>-1</sup>, respectively. As the series of compounds have similar structures, the entropic term was omitted for estimation of the binding free energy.

Additional to the QM/MM-PB/SA energy analysis, their van der Waals and electrostatic interaction energy was decomposed in a per-residue contribution. This has allowed us to identify the amino acids with the largest contribution to the binding free energy.

Finally, the polarization effects of the protein environment were quantified by decomposing the electrostatic interaction energy with a scheme that was originally developed for the study of the polarization effects for small organic compounds in solution [GX92].

In this procedure the total electrostatic interaction energy from a QM/MM simulation is separated into a permanent interaction energy ( $\Delta E_{\text{perm}}$ ), and polarization energy ( $\Delta E_{\text{pol}}$ ). The permanent interaction energy corresponds to the interaction between the protein, represented as point charges, and the unpolarized molecular wavefunction of the inhibitor in the gas phase. The polarization energy corresponds to the changes in the molecular wavefunction due to the protein environment.

$$\Delta E_{\text{elec}} = \Delta E_{\text{perm}} + \Delta E_{\text{pol}} \quad (4.15)$$

The polarization energy can be further decomposed into a polarization stabilization term and an electronic distortion term:

$$\Delta E_{\text{pol}} = \Delta E_{\text{stab}} + \Delta E_{\text{dist}} \quad (4.16)$$

where:

$$\Delta E_{\text{stab}} = \langle \psi | \mathcal{H}^{\text{QM/MM}} | \psi \rangle - \langle \psi_0 | \mathcal{H}^{\text{QM/MM}} | \psi_0 \rangle \quad (4.17)$$

and:

$$\Delta E_{\text{dist}} = \langle \psi | \mathcal{H}_0^{\text{QM}} | \psi \rangle - \langle \psi_0 | \mathcal{H}_0^{\text{QM}} | \psi_0 \rangle \quad (4.18)$$

The polarization stabilization term depicts the increase in interaction energy of the inhibitor due to its new charge distribution, while the electronic distortion term is a penalty term for reorganizing the charge distribution for the inhibitor.

## 4.5 Results and Discussion

In this section we start by presenting the structural features of all the inhibitors studied and their evolution during the QM/MM MD simulations. This is followed by an analysis of the interaction energy between the protein and the inhibitors. Both geometrical and energy terms are averaged over 51 equispaced snapshots taken from the last 0.5 ns of the simulations if not specified otherwise.

### 4.5.1 Structural Features

The average distances with residues from the hinge region (E114(O)-Br3 and V116(O)-Br2) and the distance with polar residue K68 (K68(NZ)-C1 or N2) are given in Table 4.2. These values allowed us to follow the evolution of the inhibitors during the MD simulations.

### Inhibitor L4

TBB (4,5,6,7-tetrabromo-benzotriazole), denoted L4 in this work, is one of the most widely used inhibitor of CK2. The analysis of the X-ray structure of CK2

Table 4.2: Average bond distances ( $\text{\AA}$ ) and their standard deviations between each inhibitor and selected amino acids from the active site of CK2. Values in parenthesis represent crystallographic data.

	L1	L2A	L2B	L3	L4	L5
E114(O)-Br3	$3.61 \pm 0.39$ (3.43)	$5.37 \pm 0.32$ (3.11)	$5.50 \pm 0.40$ (5.42)	$3.39 \pm 0.21$	$6.10 \pm 0.43$ (6.56)	$3.44 \pm 0.24$
V116(O)-Br2	$3.39 \pm 0.33$ (3.24)	$4.44 \pm 0.34$ (3.27)	$5.77 \pm 0.47$ (5.30)	$3.39 \pm 0.24$	$7.62 \pm 0.51$ (6.99)	$3.40 \pm 0.20$
K68(NZ)-C1 or N2	$6.77 \pm 0.46$ (6.35)	$4.22 \pm 0.19$ (5.32)	$9.51 \pm 0.41$ (9.14)	$4.64 \pm 0.27$	$3.18 \pm 0.21$ (3.29)	$5.04 \pm 0.33$

complexed with different inhibitors, leads Battistutta *et al.* [BMC<sup>+</sup>07] to notice a shift in the position of the inhibitor in the active site in comparison with the other tetrabromo-benzimidazole derivatives K25 and K37 denoted in this study L1 and L2, respectively. Due to this shift, no halogen bonds are established between L4 and the carbonyl backbone of residues E114 and V116 in the hinge region. The authors hypothesize that this different orientation is due to the distinct chemical properties of the five-membered ring, which would make L4 an anion at neutral pH. To test this assumption, we have performed two distinct simulations with the neutral and the anionic form of L4. The final structure for the two simulations together with the X-ray structure are showed in Figure 4.2. The evolution of the anionic and neutral form during the MD is indeed different. In the case of the neutral form, the triazole ring establishes two hydrogen-bond with the water molecules close to the active site. This leads to an in-plane clockwise rotation of the inhibitor relative to its initial position. Additionally, the initial halogen bonds of the inhibitor with residues E114 and V116 are shifted from Br3 and Br2 to Br2 and Br1.

On the other hand, for the anionic form of L4, we observed three hydrogen-bond between the five-membered ring and water molecules. One of these interactions is established with the water molecule that lies in-plane with the inhibitor and mediates the electrostatic interaction with the positively charged K68. This observation coupled with the fact that the anionic form of L4 has a  $C_{2v}$  symmetry explains its orientation for the first 0.9 ns of MD simulation. For this simulation interval the average distances with the residues from the hinge region are  $3.71 \pm 0.39$  Å for E114(O)-Br3 and  $4.09 \pm 0.67$  Å for V116(O)-Br2. For the last 0.5 ns of the MD simulation, we observed a different orientation in which the anionic form of L4 is shifted away from the hinge region towards the positively charged K68. The distance from the hinge region residues increases, while the distance with K68 decreases (Figure 4.2d). This suggests that the electrostatic interaction is able to orient the inhibitor in the active site. The average distances for the production run are in good agreement with those measured from X-ray structure (Table 4.2).

It is worth mentioning the presence of two water molecules W1 and W2 in the active site of the anionic form. These water molecules occupy positions similar to the water molecules in the X-ray structure. No resident water molecules were observed in the case of protonated form. These observations suggest that the anionic form is most likely to be present at the neutral pH and that a relative long equilibration time, more

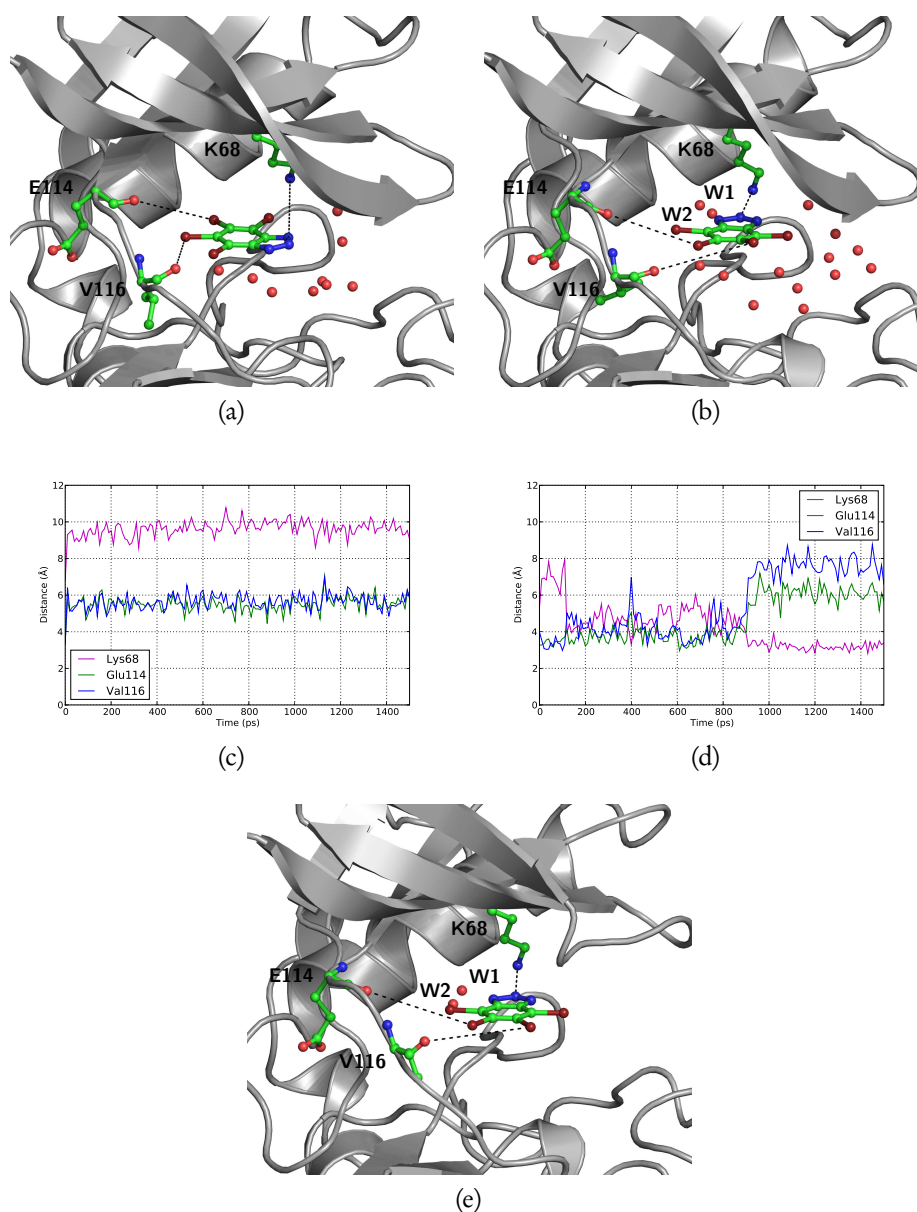


Figure 4.2: Final snapshot from the MD trajectory of CK2 complexed with L4 in its neutral form (a), anionic form (b), and crystal structure of the complex (e). The inhibitor, residues K68, E114, V116 and all water molecules within 5 Å of the inhibitor are represented in ball-and-stick model. The protein is represented with grey cartoons. Hydrogens are removed for clarity. Measured distances are depicted as black dashed lines. Evolution of the selected distances between the neutral form and CK2, and the anionic form and CK2, are depicted in figures (c) and (d), respectively.

than 1 ns, is required for these particular systems.

### Inhibitors L3 and L5

These inhibitors have a fifth bromine atom which is bound to the C1 atom of imidazole ring. They differentiate by the fact that the hydrogen atom bound to N1 in L3 is replaced by an acetamido group in the case of L5. This acetamido group makes more interactions with the water molecules around the active site, shifting the position of the inhibitor during the course of the molecular dynamics, allowing water molecules to enter in the region between the inhibitor and the protein. In the case of L3 these interactions are reduced, consequently the occasional shift towards the exterior is not observed and we have only one water molecule present between the inhibitor and the protein. In the first 0.6 ns of the simulation, this water molecule flips between the positions W1 and W2. After this equilibration period the water molecule settles in the W2 position.

The final orientations of the two inhibitors are similar. This is reflected by the fact that the average distance with the residues from the hinge region are close (Table 4.2). Also it should be noted that no direct interaction is established with residue K68 as in the case of the anionic form of L4. For L3 and L5 this interaction is mediated by the water molecule W2, leading to an increase in the distance K68(NZ)-C1.

### Inhibitors L1 and L2

These inhibitors have the highest inhibition constants in the series under study with L1 being the most potent one [PAR<sup>+</sup>04]. The position of L1 doesn't change dramatically during the course of the molecular dynamic simulation and the average distances for the whole 1.5 ns MD simulation are in good agreement with those from Table 4.2,  $3.68 \pm 0.40$  Å for E114(O)-Br3 distance,  $3.45 \pm 0.34$  Å for V116(O)-Br2 distance and  $6.74 \pm 0.43$  for the K68(NZ)-C1 distance. This can be explained by the fact that the starting structure for this simulation was the X-ray structure of CK2 complexed with this inhibitor. Secondly, we have seen that in the case of anionic or polar inhibitors, the orientation of the inhibitor in the active site is strongly influenced by the electrostatic interaction with the positively charged K68 (in the case of L4 anion) or by the hydrogen bonds formed with water molecules close to the active site (in



the case of L4 neutral). For L1 the two CH<sub>3</sub> groups on the nitrogen atom hinder these interactions and even the presence of a hydrogen bond (N1-H···OH<sub>2</sub>) cannot “overcome” the halogen bond interaction of the inhibitor with the residues from the hinge region.

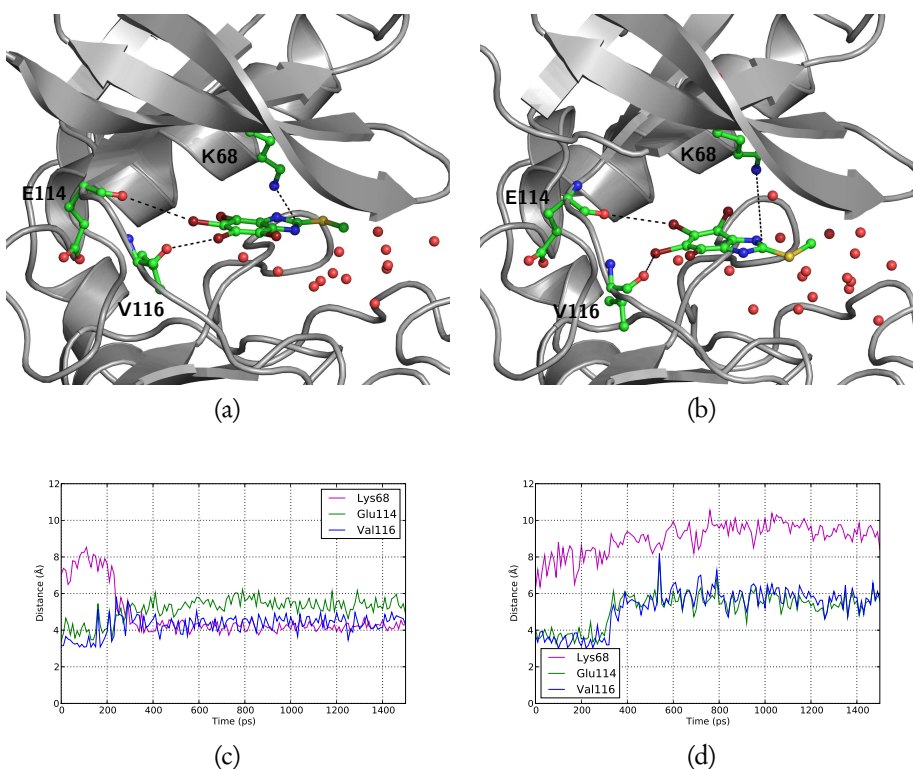


Figure 4.3: Final snapshot from the MD trajectory of CK2 complexed with L2 in simulation L2A (a), and simulation L2B (b). Representation is made as in Figure 4.2. Evolution of the selected distances between the inhibitor and CK2 in simulation L2A, and simulation L2B, are depicted in figure (c) and (d), respectively.

In the case of L2, two distinct simulations were run (L2A and L2B), while changing only the random velocities generated at the beginning of the simulation. This was motivated by the fact that the crystal structure of CK2 complexed with L2 showed two possible orientations for this inhibitor in the active site of CK2.

In simulation L2A, even though at the beginning the CH<sub>3</sub> group on the sulfur atom was orientated towards residue K68, we noticed a flip after 0.2 ns (Figure 4.3c). This leads to a strong electrostatic interaction with residue K68 which reduces the distance K68(NZ)-C1 by more than 3 Å, when compared to the initial value. In the crystal

structure this interaction is mediated by a water molecule, which leads to a placement of the inhibitor closer to the hinge region (Table 4.2).

For simulation L2B, the flip of the CH<sub>3</sub> group is not observed. As a result the inhibitor executes a clockwise rotation similar to that of the neutral form of L4 (Figure 4.3b), and exposes the imidazol ring to the solvent.

For both simulations, L2A and L2B, we find a very good agreement between the averaged distances of the 0.5 ns production run and the X-ray values (Table 4.2).

## 4.5.2 Relative Binding Free Energies

The present QM/MM-PB/SA study allows us to analyze the individual energy components contributing to the total binding energy. Calculated values together with experimentally measured relative difference of free energy of binding are presented in Table 4.3.

In a previous study of CDK2 inhibitors, Tuñón *et al.* obtained a high coefficient of determination ( $R^2 = 0.92$ ), when they expressed the activity as a function of QM/MM interaction energy [AMCS<sup>+</sup>07]. Based on their results, they concluded that for inhibitors belonging to the same family the QM/MM interaction energy can be used as a predictor of biological activity.

We performed the same linear regression analysis, using the experimental binding energy ( $\Delta\Delta G_{\text{exp}}$ ) and QM/MM interaction energy ( $\Delta\Delta E_{\text{QM/MM}}$ ) as the input data set (for L2 only the calculated value from the B simulation was used to perform this analysis). In our case we obtain a value of 0.14 for  $R^2$ . The poor correlation can be explained by taking a closer look at the individual electrostatic and van der Waals energies that are summed up to give  $\Delta E_{\text{QM/MM}}$ . Especially the calculated values for  $\Delta E_{\text{elec}}$  show large variation within this set of inhibitors. For example the presence of the acetamido group on L5 makes the electrostatic interaction approximately 9 kcal/mol larger when compared to L3. The van der Waals interaction energy also favors L5 by approximately 6 kcal/mol. On the contrary, experimental data disfavors L5 by -1.3 kcal/mol. The most striking finding of Table 4.3 is the value obtained for electrostatic interaction energy in the case of L4, which is one order of magnitude larger when compared with the other inhibitors. Due to the different polarities of these inhibitors, especially L4 and L5, it's very difficult to obtain good correlation between

Table 4.3: Binding energies computed with the QM/MM-PB/SA method.<sup>a</sup>

	L1	L2A	L2B	L3	L4	L5
$\Delta E_{\text{elec}}$	$2.84 \pm 1.44$	$-4.52 \pm 2.25$	$-6.81 \pm 4.05$	$-4.01 \pm 1.57$	$-104.15 \pm 7.29$	$-12.99 \pm 4.39$
$\Delta E_{\text{vdw}}$	$-41.00 \pm 1.71$	$-40.68 \pm 1.89$	$-39.46 \pm 1.64$	$-38.43 \pm 1.53$	$-31.55 \pm 1.48$	$-44.67 \pm 1.55$
$\Delta G_{\text{PB}}$	$4.04 \pm 3.14$	$13.76 \pm 3.19$	$14.88 \pm 4.47$	$16.22 \pm 1.74$	$108.88 \pm 5.56$	$31.40 \pm 4.76$
$\Delta G_{\text{SA}}$	$-2.72 \pm 0.06$	$-2.35 \pm 0.06$	$-2.41 \pm 0.06$	$-2.17 \pm 0.05$	$-1.89 \pm 0.05$	$-2.62 \pm 0.05$
$\Delta G_{\text{polar}}^{\text{b}}$	$7.88 \pm 3.08$	$9.24 \pm 3.11$	$8.07 \pm 2.64$	$12.21 \pm 2.17$	$4.73 \pm 3.06$	$18.41 \pm 2.98$
$\Delta G_{\text{nonpolar}}^{\text{c}}$	$-42.72 \pm 1.71$	$-43.03 \pm 1.90$	$-41.87 \pm 1.64$	$-40.60 \pm 1.54$	$-33.44 \pm 1.49$	$-47.29 \pm 1.55$
$\Delta\Delta E_{\text{QM/MM}}^{\text{d}}$	0.00	8.04	9.11	5.28	98.54	20.50
$\Delta\Delta G_{\text{cal}}^{\text{e}}$	0.00	-1.05	-1.04	-6.45	-6.13	-5.96
$\Delta\Delta G_{\text{exp}}$	0.00	-0.34	-0.34	-1.05	-1.38	-2.31

<sup>a</sup> All energies are in kcal/mol;

<sup>b</sup>  $\Delta G_{\text{polar}} = \Delta E_{\text{elec}} + \Delta G_{\text{PB}}$ ;

<sup>c</sup>  $\Delta G_{\text{nonpolar}} = \Delta E_{\text{vdW}} + \Delta G_{\text{SA}}$ ;

<sup>d</sup>  $\Delta\Delta E_{\text{QM/MM}}$  refers to the relative QM/MM interaction energy; for each inhibitor  $\Delta E_{\text{QM/MM}} = \Delta E_{\text{elec}} + \Delta E_{\text{vdW}}$ ;

<sup>e</sup>  $\Delta\Delta G_{\text{cal}}$  refers to the relative binding energy calculated with QM/MM-PB/SA method;

for each inhibitor  $\Delta G_{\text{cal}} = \Delta E_{\text{elec}} + \Delta E_{\text{vdW}} + \Delta G_{\text{PB}} + \Delta G_{\text{SA}}$ .

experimental data and calculated  $\Delta\Delta E_{\text{QM/MM}}$  without taking into account the effects of both ligand and protein desolvation.

Within the traditional MM-PB/SA framework, it has been demonstrated that the solvation energy is a key element in correctly ranking inhibitors belonging to the same family. As can be seen in Table 4.3, the favorable electrostatic protein-inhibitor interaction energy is canceled by an unfavorable polar desolvation contribution ( $\Delta G_{\text{PB}}$ ), which has the largest value for L4 and L5. Because the absolute value of  $\Delta E_{\text{elec}}$  is smaller, net positive polar components ( $\Delta G_{\text{pol}}$ ) are obtained for all the inhibitors. Similarly, the nonpolar component of the protein-inhibitor binding free energy ( $\Delta G_{\text{nonpolar}}$ ) can be regarded as the sum of the van der Waals energy term ( $\Delta E_{\text{vdW}}$ ) and the nonpolar term in the solvation energy ( $\Delta G_{\text{SA}}$ ). This quantity is a large negative number for all the inhibitors studied. As a result we can conclude that the latter is the driving force for the binding of these CK2 inhibitors. With the inclusion of solvation energy, the coefficient of determination resulted from the linear regression analysis is improved from  $R^2 = 0.14$  to  $R^2 = 0.69$ . This clearly indicates the importance of solvation effects.

### 4.5.3 Electrostatic Energy Decomposition

The previously described electrostatic energy components are presented in Table 4.4. As previously underlined, the inhibitors have different orientations in the active site. In the case of L2A, L3, L4, and L5, the electrostatic interaction with residue K68 favors an orientation in which the imidazole (triazole) ring interacts with this residue. For L2B this interaction is shielded by the  $\text{CH}_3$  group orientated towards K68, and the inhibitor turns in the active site of CK2.

The interaction energy between the unpolarized wavefunction in gas phase and the protein environment ( $\Delta E_{\text{perm}}$ ) is not always favorable to binding. This is closely related to the position of the inhibitor and its polarity. In the case of L2 the average value obtained from simulation A is unfavorable to binding, while in the case of simulation B this value becomes favorable (Table 4.4). An unfavorable contribution is also obtained for L1 and L3. Even though L5 is positioned similarly in the active site, we obtain a negative value, suggesting that the polarity of the inhibitor can also influence  $\Delta E_{\text{perm}}$ . This becomes evident in the case of L4 for which we obtain a value of  $-97.10 \pm 6.90$  kcal/mol, one order of magnitude larger when compared with the other inhibitors.

The polarization energies associated to passing from gas phase to the protein

Table 4.4: Components of the electrostatic interaction energy<sup>a</sup>

	L1	L2A	L2B	L3	L4	L5
$\Delta E_{\text{perm}}$	$8.25 \pm 1.83$	$4.25 \pm 2.13$	$-2.57 \pm 4.30$	$0.49 \pm 1.46$	$-97.10 \pm 6.90$	$-7.43 \pm 4.02$
$\Delta E_{\text{pol}}$	$-5.41 \pm 0.75$	$-8.76 \pm 1.07$	$-4.24 \pm 0.67$	$-4.50 \pm 0.74$	$-7.05 \pm 0.79$	$-5.57 \pm 0.77$
$\Delta E_{\text{stab}}$	$-10.81 \pm 1.49$	$-17.58 \pm 2.17$	$-8.47 \pm 1.33$	$-9.02 \pm 1.49$	$-14.07 \pm 1.57$	$-11.15 \pm 1.54$
$\Delta E_{\text{dist}}$	$5.39 \pm 0.75$	$8.82 \pm 1.09$	$4.23 \pm 0.67$	$4.52 \pm 0.75$	$7.03 \pm 0.78$	$5.58 \pm 0.77$

<sup>a</sup> All energies are in kcal/mol.

have similar values for all the inhibitors, and are always favorable. It is important to underline that, with the exception of L4,  $\Delta E_{\text{pol}}$  has the same order of magnitude as  $\Delta E_{\text{perm}}$ , indicating the importance of electronic polarization effects. Again, due to its particular electronic structure, L4 makes a discordant note in this family of inhibitors, with  $\Delta E_{\text{pol}}$  constituting about 6% of  $\Delta E_{\text{elec}}$ . However, this does not imply that polarization effects are not significant. The reason for the small percentage is the large interaction energy of charged species with the rest of the system [GVTG03].

Another interesting finding is that the absolute value of the polarization stabilization energy ( $\Delta E_{\text{stab}}$ ) is two times larger than the absolute value of electronic distortion energy ( $\Delta E_{\text{dist}}$ ) for each protein-inhibitor complex. This is consistent with linear response theory, and provides further validation for the energy decomposition scheme [GX92].

#### 4.5.4 Energetic Contribution of Individual Amino Acids

Because one of the goals for the present study was to gain insight into the specific interactions that are established between the CK2 and this family of inhibitors, we performed a further decomposition of the van der Waals and electrostatic energy into terms involving each amino acid. In the case of electrostatic energy we used both the gas-phase and the polarized wavefunctions.

The ATP-binding site of CK2 can be divided in three separate regions: the adenine pocket, the ribose region and the triphosphate region (Figure 4.4).

As can be seen from Table 4.5 the values resulted from the per-residue van der Waals interaction energy decomposition are influenced by the position of the inhibitor in the active site. L2A and L4 have the largest van der Waals interaction energy with residues from the triphosphate site. If we consider only the van der Waals interaction energy coming from residues located in the adenine pocket, the situation is inverted. For the interaction with residues from the ribose region we find similar values for all inhibitors.

The high specificity of this family of inhibitors towards CK2 was attributed to the particularity of its ATP-binding site. Three of the residues that found in the active site of CK2 (I66, M163 and I174) are specific to this protein, and are replaced by less bulkier residues in other kinase. The importance of I66 and I174 was assessed by mutational studies in which they were individually or collectively mutated to one or more alanines

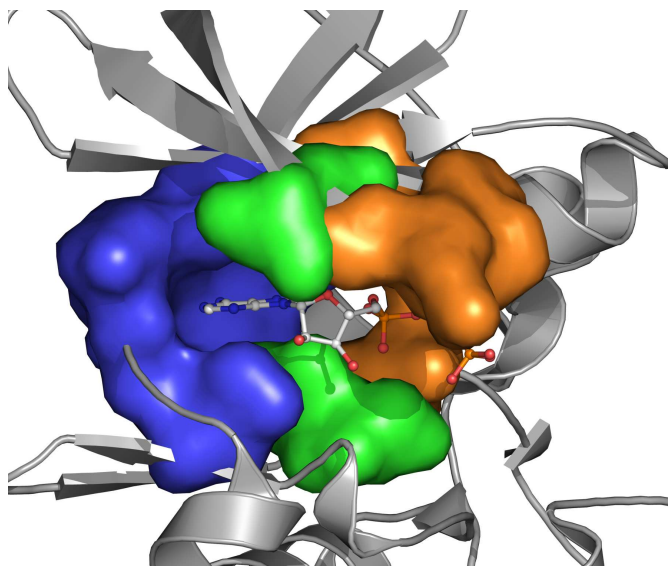


Figure 4.4: The X-ray structure of CK2 with AMPPNP (PDB code: 1DAW). The ANP is represented in ball-and-stick model, while the CK2 protein is represented with grey cartoons. The adenine pocket (blue) comprise the residues I66, V95, F113, E114, Y115, V116 and M163. The ribose region (green) comprise the residues V45, V53, N161 and I174. The triphosphate binding region (orange) comprise the residues G46, R47, G48, S51, K68 and D175.

[MPM<sup>+</sup>04]. Interestingly, we find that these residues have the largest contribution to the final van der Waals interaction energy. Additionally residues V45, V53, V95, F113, D175 were found to stabilize all the inhibitors, but to a lesser extent. These results confirm the recent study of Nie *et al.* [NPE<sup>+</sup>07] that targeted the hydrophobic pocket outlined by V95 and F113. With the addition of a nitrile group on a pyrazolo[1,5-a][1,3,5]triazine skeleton which interacts with these hydrophobic residues a  $\sim 10$ -fold increase of the potency was observed.

While comparing different X-ray crystal structure of CK2 complexed with different inhibitors, Battistutta *et al.* [BMC<sup>+</sup>07] noticed a tendency of clustering around a unique zone of the ATP-binding pocket locate near K68, for inhibitors that bear a negatively charge moiety. We have seen previously that this interaction can play an important role in orienting the inhibitors in the active site of the CK2. As can be seen from Table 4.6, K68 has a large contribution to the total electrostatic interaction energy. This interaction is more stabilizing in the case of L2A and L4. It should be noted that this favorable contribution is partially canceled by two unfavorable contributions

Table 4.5: Individual amino acid residues contribution to the van der Waals interaction energy.<sup>a</sup>

	Residue	L1	L2A	L2B	L3	L4	L5
adenine pocket	I66	-4.06	-2.71	-4.12	-4.25	-2.44	-4.22
	V95	-1.33	-1.18	-1.46	-1.39	-0.80	-1.26
	F113	-2.76	-2.76	-2.76	-2.46	-1.91	-2.43
	E114	-1.00	-0.31	-1.09	-1.12	-0.25	-1.16
	Y115	-0.86	-0.22	-1.08	-1.26	-0.19	-1.16
	V116	-1.84	-2.28	-2.27	-2.59	-0.67	-2.53
	M163	-3.90	-2.34	-3.93	-3.62	-2.43	-4.41
	Total	-15.75	-11.8	-16.71	-16.69	-8.69	-17.17
ribose region	V45	-1.46	-1.76	-2.98	-1.62	-1.73	-2.89
	V53	-3.14	-3.56	-3.14	-2.83	-3.74	-3.34
	N161	-0.66	-0.60	-0.35	-0.43	-0.46	-0.77
	I174	-4.85	-4.96	-4.58	-4.36	-4.28	-5.82
	Total	-10.11	-10.88	-11.05	-9.24	-10.21	-12.82
triphosphate region	G46	-0.66	-1.17	-1.02	-0.55	-1.38	-1.30
	R47	-1.01	-1.28	-0.74	-0.38	-1.30	-0.73
	G48	-0.65	-0.66	-0.12	-0.16	-0.29	-0.14
	S51	-0.58	-1.12	-0.20	-0.59	-0.59	-0.42
	K68	-1.70	-3.47	-1.41	-2.17	-1.96	-1.82
	D175	-2.26	-3.43	-1.69	-2.11	-2.50	-1.90
	Total	-6.86	-11.13	-5.18	-5.96	-8.02	-6.31

<sup>a</sup> All energies are in kcal/mol.

coming from residues D81 and E175. The values from Table 4.6 demonstrate once again the importance of polarization. For example in the case of electrostatic interaction between L2A and K68 the inhibitor is stabilized by more than 6 kcal/mol when the polarized wavefunction is used. Nie *et al.* [NPE<sup>+</sup>07] found that inhibitors that take advantage of these electrostatic interactions have a  $\sim 20$ -fold increase in their inhibition power.

The inhibitors belonging to this family of compounds rely mainly on hydrophobic contacts to inhibit the CK2 protein. The strongest inhibitor L1 fits perfectly in the adenine pocket of the protein, thus maximizing the van der Waals interaction, but has a weak electrostatic interaction with the charged residues clustered around K68. On the



Table 4.6: Individual amino acid contribution to the electrostatic interaction energy calculated using the gas-phase ( $\psi_0$ ) and the polarized ( $\psi$ ) semiempirical wavefunction for the description of the inhibitor.<sup>a</sup>

Residue	L1		L2A		L2B		L3		L4		L5	
	$\psi_0$	$\psi$	$\psi_0$	$\psi$	$\psi_0$	$\psi$	$\psi_0$	$\psi$	$\psi_0$	$\psi$	$\psi_0$	$\psi$
K68	0.01	-2.73	-15.00	-21.23	-2.08	-4.33	-5.21	-9.78	-76.13	-84.08	-4.35	-9.11
D81	0.48	1.55	4.59	7.12	1.46	2.43	1.76	3.72	43.43	46.22	1.62	3.64
E175	-3.00	0.01	3.80	8.15	1.29	3.72	-1.61	2.69	55.38	60.45	-0.06	4.04

<sup>a</sup> All energies are in kcal/mol.

other hand L2A, has a strong electrostatic interaction with these residues. Because of this strong interaction, the inhibitor shifts away from the hinge region, thus reducing the van der Waals interaction energy. Based on these observations we can state that the addition of sufficiently large polar group on the tetrabromo-benzimidazole skeleton, could increase the electrostatic interaction with the triphosphate binding region, while keeping the inhibitor close to the hinge region, and thus maximizing the van der Waals interaction with the adenine pocket.

#### 4.5.5 Application of the QM/MM-PB/SA Method to New Inhibitors

The simulation protocol for the inhibitors derived from *N*-hydroxyphthalimide follows closely the one developed above for the simulation of the tetrabromo-benzimidazole/benzotriazole family of inhibitors. As discussed in Section 4.3 no crystallographic data was available for these inhibitors. This renders our discussion regarding the structural features of these inhibitors in the course of the MD simulations without an experimental basis to compare with. But, since the previous simulations have showed to reproduce experimental data quite well, we can pursue our goal of understanding the energetics of the binding with confidence.

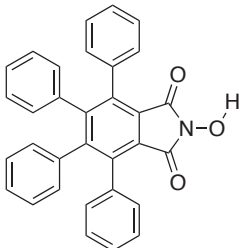
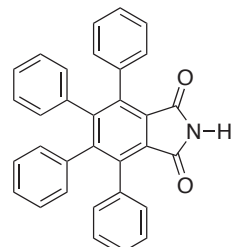
As can be seen from Table 4.7, the structures of the two inhibitors are very similar, the only difference is the presence of a hydroxyl group in the case of L6, in place of the hydrogen atom in L7.

One can observe that core structure of these inhibitors, is somewhat similar to the benzotriazole/benzimidazole skeleton of the previously studied inhibitors, with the notable difference that the bromine atoms are substituted with phenyl rings.

The results of the QM/MM-PB/SA analysis together with the electrostatic interaction energy decomposition are presented in Table 4.8.

Not surprisingly the values obtained for the binding free energy for the two inhibitors are close, with a difference of 3 kcal/mol. This difference is mainly due to the polar components of the calculated binding free energy. Apparently,  $\Delta E_{\text{elec}}$  is more favorable in the case of L6 by approximately 2.5 kcal/mol. As in the case of the first inhibitors studied, the protein-inhibitor electrostatic interaction energy is canceled by an unfavorable polar desolvation contribution. For  $\Delta G_{\text{PB}}$  we obtained values that favor the binding of L7 by approximately 4.8 kcal/mol. This is somewhat expected,

Table 4.7: Inhibitor activity against the CK2 for the *N*-hydroxyphthalimide derived inhibitors.

Id.	Structure	IC <sub>50</sub> ( $\mu$ M)
L6		1.75
L7		N/A

since the penalty of solvating should be greater for L6 due to the presence of a hydroxyl group, that in L7 is substituted by a hydrogen atom. This difference is also reflected in the van der Waals interaction energy, but to a far lesser extent.

The absolute values for the terms resulting from the decomposition of electrostatic interaction energy are larger for L6. Again we notice that for both inhibitors the polarization energy in passing to the gas phase to the protein represents almost half of the total electrostatic interaction energy, suggesting that this energy component can be very important in problems related to ligand binding.

Our results are pertinent, but unfortunately they could not explain the experimental observations. Results from *in vitro* studies suggest that L6 should be an inhibitor of the CK2 protein in the micromolar range, while L7 should not present any inhibition activity. Our theoretical results indicate that they have comparable inhibition constants, with L7 being more potent.

An explanation would be that, as in the case of L4 (TBB), L6 can be an anion at a neutral pH. Consequently, we decided to run another simulation in which the hydrogen from the hydroxyl group of L6 was removed.

The different energy terms for the anionic form of L6 are presented in Table 4.8.

Table 4.8: Binding energies and electrostatic interaction energy decomposition.<sup>a</sup>

	L6	L7	L6 anionic
$\Delta E_{\text{perm}}$	$-10.02 \pm 3.44$	$-8.88 \pm 2.40$	$-125.21 \pm 8.98$
$\Delta E_{\text{pol}}$	$-9.23 \pm 1.28$	$-8.29 \pm 1.29$	$-11.32 \pm 1.72$
$\Delta E_{\text{stab}}$	$-18.52 \pm 2.57$	$-16.66 \pm 2.59$	$-22.48 \pm 3.38$
$\Delta E_{\text{dist}}$	$9.29 \pm 1.29$	$8.37 \pm 1.30$	$11.15 \pm 1.66$
$\Delta E_{\text{elec}}$	$-19.52 \pm 3.46$	$-17.18 \pm 2.64$	$-136.53 \pm 10.32$
$\Delta E_{\text{vdw}}$	$-52.18 \pm 2.24$	$-52.43 \pm 2.41$	$-55.01 \pm 2.47$
$\Delta G_{\text{PB}}$	$44.25 \pm 5.49$	$39.50 \pm 4.31$	$154.62 \pm 8.29$
$\Delta G_{\text{SA}}$	$-4.44 \pm 0.08$	$-4.44 \pm 0.10$	$-4.62 \pm 0.06$
$\Delta G_{\text{polar}}^{\text{b}}$	$25.01 \pm 3.98$	$22.32 \pm 3.90$	$18.09 \pm 4.53$
$\Delta G_{\text{nonpolar}}^{\text{c}}$	$-56.62 \pm 2.25$	$-56.88 \pm 2.42$	$-59.27 \pm 2.47$
$\Delta G_{\text{calc}}^{\text{d}}$	$-31.61 \pm 4.67$	$-34.56 \pm 4.45$	$-41.18 \pm 4.77$

<sup>a</sup> All energies are in kcal/mol;

<sup>b</sup>  $\Delta G_{\text{polar}} = \Delta E_{\text{elec}} + \Delta G_{\text{PB}}$ ;

<sup>c</sup>  $\Delta G_{\text{nonpolar}} = \Delta E_{\text{vdW}} + \Delta G_{\text{SA}}$ ;

<sup>d</sup>  $\Delta G_{\text{calc}} = \Delta G_{\text{polar}} + \Delta G_{\text{nonpolar}}$ ;

Because the inhibitor has a negative charge present, we notice a  $\sim 7$ -fold increase in the electrostatic interaction energy when compared with the neutral form. This large favorable energy contribution is canceled by the  $\Delta G_{\text{PB}}$  term which is larger in absolute value, which results in an unfavorable polar contribution to binding. The van der Waals interaction energy is also influenced by this change, but the differences are smaller in magnitude, and are due to the fact that the negative charge of the inhibitor interacts with the charged residues around the K68, slightly shifting the position of the inhibitor in the active site. Overall we find a difference of approximately 7 kcal/mol between the calculated free energy of binding of L7 and the anionic form of L6. Having in mind that L6 was a micromolar inhibitor of the CK2 protein, this difference in calculated binding affinity would make L7 a molecule with a very low affinity towards the CK2 protein. For example if the relative binding affinity of two inhibitors differs by 8.2 kcal/mol, the rapport of their inhibition constant would be one million. Thus the basic function present in L6 and not in L7 should explain why the two inhibitors behave differently towards the CK2.

## 4.6 Conclusions

We propose an extension of the classical end-point approach MM-PB/SA by using a hybrid QM/MM MD simulation to sample the conformational space prior to the energy analysis. In this new approach, the inhibitor is described with a semiempirical Hamiltonian which avoids the development of specific empirical parameters and allows an explicit treatment of the polarization that the protein environment exerts on the inhibitor. The resulted trajectories were analyzed by combining QM/MM interaction energy evaluation with the Poisson-Boltzmann/surface area model for the evaluation solvation terms.

This new methodology was applied for two families of inhibitors derived from tetrabromo-benzimidazole/benzotriazole and from *N*-hydroxyphthalimide.

For the first inhibitors studied, our analysis indicates that the van der Waals interactions have a significant favorable contribution to binding. The inclusion of solvation effects proved to be essential to the proper damping of electrostatic interaction. Their importance is clearly demonstrated by the coefficient of determination resulted from the linear regression analysis between the experimental binding energy and the calculated one. The value of  $R^2$  is greatly improved from 0.14, for the case when only the QM/MM interaction energy is used, to 0.69 when solvation effects are also taken into account.

The van der Waals and electrostatic interaction energies were decomposed in terms pertaining to each amino acid. Based on the information provided by this analysis we propose a way for improving the current inhibitors, by taking advantage of the electrostatic interaction with the residues located around K68. The addition of a sufficiently large polar group on the tetrabromo-benzimidazole/benzotriazole skeleton could increase this electrostatic interaction without losing the favorable van der Waals interactions with the bulky residues I66, M163 and I174 specific to the CK2.

The knowledge gathered in this first study was then used to give an explanation to some questions regarding the inhibitors derived from *N*-hydroxyphthalimide. For this study we applied the QM/MM-PB/SA methodology to two inhibitors with very similar structures. Even though the experimental studies indicated that the two inhibitors should behave very differently towards the CK2, our initial simulations suggested the opposite. A new simulation was ran considering the anionic form of one inhibitor. These new results managed to explain experimental data.

Altogether, this approach provides a detailed and reliable enough picture of the interactions in the CK2-inhibitor complex, and may be used for studying new inhibitors without the need of developing force field parameters for each inhibitors. By using a semiempirical representation for the inhibitor, the computational time increases when compared with classical simulations, but still remains within reasonable limits.



# Conclusions Générales

Les calculs de chimie quantique restent encore de nos jours limités aux molécules contenant une centaine d'atomes. Au delà, on doit se tourner vers d'autres méthodes. Une solution est l'utilisation des méthodes mixtes QM/MM qui permettent de combiner les avantages de la mécanique quantique et de la mécanique moléculaire, en limitant la description quantique à une sous partie du système et en traitant tout le reste à l'aide de champs de force classiques.

Dans cette thèse ces méthodes ont été utilisés pour résoudre des problèmes concrets allant de la structure de composés organométalliques, à la réactivité de métalloprotéine, au drug design. Sur chaque système, une analyse rigoureuse des modèles utilisés a été effectuée et des réponses quantitatives ont été apportées. Ainsi dans une première étude, différentes coupures QM/MM et leur influence dans la modélisation de la structure électronique d'une série de ligands phosphines sont décrites. Pour cela le calcul du potentiel électrostatique apparaît comme un mode d'évaluation rapide et fiable. Le deuxième exemple nous a permis de mettre en évidence l'importance de prendre en compte l'environnement des protéines acid phosphatase pourpre pour modéliser leur réactivité, les approches quantiques pures sur de petits clusters modèles donnant des énergies d'activation beaucoup trop grandes. Là encore différents calculs ONIOM ont été réalisés en jouant sur la taille de la partie quantique. Enfin au cours de la dernière étude, nous avons été amené à adapter une méthode théorique originale appelée QM/MM-PB/SA permettant de décrire de façon réaliste les énergies d'interaction entre une protéine et un substrat. Cette dernière a été testée et validée sur différents inhibiteurs de la protéine CK2 et a permis de proposer une explication aux résultats obtenus lors d'un criblage de la chimiothèque de notre laboratoire.

Différents schémas de méthodes QM/MM ou QM/MM/MD ont été utilisés. Au delà des problèmes résolus, nous montrons leurs domaines de validité. Ainsi si l'approche ONIOM a donné de bons résultats pour l'étude des composés organométal-



liques, pour la modélisation statique de la réactivité des systèmes biologiques, elle ne permet pas de prendre en compte les aspects dynamiques essentiels pour décrire les systèmes biologiques. C'est pourquoi nous nous sommes tournés vers d'autres implémentations des méthodes QM/MM comme le code dynamo utilisé dans la dernière étude ou encore le code couplant CPMD et Gromacs. Les connaissances acquises sur ces codes nous ont permis d'écrire un nouveau module QM/MM-PB/SA adaptable à la librairie dynamo, prometteur pour la modélisation des interactions ligand-protéine. Les programmes mis en place permettent d'envisager une utilisation aisée sur d'autres systèmes originaux.

# Concluding Remarks and Perspectives

Quantum chemistry calculations are computationally demanding and therefore restricted to small isolated molecules. At the present time, larger systems are not amenable for a full quantum mechanical description. One approach to the simulation of such systems is to use a combination of quantum mechanics and molecular mechanics (QM/MM). The reacting parts of the system are described quantum mechanically, with the remainder is modeled using a force field. Simulations based on such mixed quantum-classical description have contributed to a better understanding of many catalytic and enzymatic processes. In the first part of this thesis, we review the QM/MM methods, with a special emphasis on the ONIOM method developed by Morokuma *et al.*

One area where hybrid QM/MM methods appear to be particularly well suited is homogeneous catalysis. This is due do the fact that catalysts are often bulky transition metal complexes, and a full quantum mechanical representation can be computationally very expensive. With a hybrid description of the system, the steric interaction of the bulky substituents can be accurately described by the MM method, while the transition metal center and its immediate environment can be described with an accurate QM method. In the first study presented we investigated how the placement of the QM-MM border influences the electronic effects of a series of phosphine ligands. For this we proposed a computationally fast approach based on molecular electrostatic potential (MESP).

In the subsequent two chapters we were interested in the application of the QM/MM methods to biological problems. For the study of red kidney bean purple acid phosphatase (kbPAP) we explored different theoretical approaches, ranging from simple cluster models of the active site to complex models that comprise the whole protein. A mechanism proposed for the hydrolysis of phosphorylated substrates by kbPAP involves a bridging hydroxide as nucleophile. Part of the experimental studies argued that the nucleophilicity of this bridging hydroxide would be to low to assure efficient

reactivity. However, we find an activation energy barrier of 44.1 kJ/mol for this mechanism. To our knowledge this is the first theoretical attempt on simulating this protein with a hybrid methodology. The last study deals with the application of QM/MM methods to drug design. We proposed a new method (QM/MM-PB/SA) to estimate free energies of ligand binding from molecular dynamic simulations with a hybrid Hamiltonian. This methodology was applied to the inhibitors of casein-kinase 2 (CK2). Our calculations allowed us to propose a way to improve the current tetrabromobenzimidazole/benzotriazole inhibitors. Additionally, we managed to explain the difference in affinity towards the CK2 protein of two inhibitors synthesized by the SERCO group from DCM.

Throughout this thesis, we have shown different applications of the QM/MM approach. From a methodological point of view this has allowed a direct contact with problems related to their various implementations. While the current implementation of the ONIOM method in the Gaussian 03 software package was found to be adapted for the study of organometallic systems, it lacks a certain flexibility when describing biological systems. As a result, we turned our attention to more adapted implementations of the QM/MM methods, like those from the fDynamo library or the CPMD/Gromacs interface. The knowledge acquired from using them was employed in writing a small module for the fDynamo library that was used for the calculation of the free energy of ligand binding. This module is easily adaptable for other systems, and it is currently used for the evaluation of binding interactions in sugar/*Pseudomonas aeruginosa* lectin complexes.

# Bibliography

- [AAL83] B. C. Antanaitis, P. Aisen, and H. R. Lilienthal. Physical characterization of two-iron uteroferrin. Evidence for a spin-coupled binuclear iron cluster. *J. Biol. Chem.*, 258(5):3166–3172, 1983. 62
- [AEH<sup>+</sup>03] G. Andersson, B. Ek-Rylander, K. Hollberg, Y. Wang, and S. J. Zhang. TRACP as an Osteopontin Phosphatase. *J. Bone Miner. Res.*, 18(10):1912–1915, 2003. 62
- [AFAG00] P. Amara, M. J. Field, C. Alhambra, and J. Gao. The Generalized Hybrid Orbital Method for Combined Quantum Mechanical/Molecular Mechanical Calculations: Formulation and Tests of The Analytical Derivatives. *Theor. Chim. Acta.*, 104:336–343, 2000. 39
- [ALS94] M. A. S. Aquino, J. S. Lim, and G. Sykes. Mechanism of the reaction of different phosphates with the iron (II) iron (III) form of purple acid phosphatase from porcine uteri (uteroferrin). *J. Chem. Soc., Dalton Trans.*, 1994(4):429–436, 1994. 63, 65
- [ÅM01] J. Åqvist and J. Marelus. The Linear Interaction Energy Method for Predicting Ligand Binding Free Energies. *Comb. Chem. High Throughput Screen.*, 4:613–626, 2001. 87
- [AMC<sup>+</sup>07] C. N. Alves, S. Martí, R. Castillo, J. Andrés, V. Moliner, I. Tuñón, and E. Silla. Calculation of Binding Energy Using BLYP/MM for the HIV-1 Integrase Complexed with the S-1360 and Two Analogues. *Bioorg. Med. Chem.*, 15(11):3818–3824, 2007. 91
- [AMC<sup>+</sup>08] C. N. Alves, S. Martí, R. Castillo, J. Andrés, V. Moliner, I. Tuñón, and E. Silla. A Quantum Mechanic/Molecular Mechanic Study of the

Wild-Type and N155S Mutant HIV-1 Integrase Complexed with Diketo Acid. *Biophys. J.*, 94:2443–2451, 2008. 91

- [AMCS<sup>+</sup>07] J. H. Alzate-Morales, R. Contreras, A. Soriano, I. Tuñón, and E. Silla. A Computational Study of the Protein-Ligand Interactions in CDK2 Inhibitors: Using Quantum Mechanics/Molecular Mechanics Interaction Energy as a Predictor of the Biological Activity. *Biophys. J.*, 92:430–439, 2007. 91, 105
- [ÅMS94] J. Åqvist, C. Medina, and J. E. Samuelsson. A new method for predicting binding affinity in computer-aided drug design. *Protein Eng.*, 7(3):385–391, 1994. 87
- [AR96] X. Assfeld and J.-L. Rivail. Quantum Chemical Computations on Parts of Large Molecules: The Ab Initio Local Self Consistent Field Method. *Chem. Phys. Lett.*, 263:100–106, 1996. 39
- [AT98] I. Antes and W. Thiel. On the Treatment of Link Atoms in Hybrid Methods. In J. Gao and M. A. Thompson, editors, *Hybrid Quantum Mechanical and Molecular Mechanical Methods*, ACS Symposium Series, pages 50–65. American Chemical Society, 1998. 38
- [AWF<sup>+</sup>00] N. Z. Angel, N. Walsh, M. R. Forwood, M. C. Ostrowski, A. I. Cassady, and D. A. Hume. Transgenic mice overexpressing tartrate-resistant acid phosphatase exhibit an increased rate of bone turnover. *J. Bone Miner. Res.*, 15(1):103–110, 2000. 62
- [BCCK93] C. I. Bayly, P. Cieplak, W. Cornell, and P. A. Kollman. A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges: the RESP model. *J. Phys. Chem.*, 97(40):10269–10280, 1993. 74
- [BdJZ<sup>+</sup>88] J. Beck, J. de Jersey, B. Zerner, M. Hendrich, and P. Debrunner. Properties of the Fe(II)-Fe(III) Derivative of Red Kidney Bean Purple Phosphatase - Evidence for a Binuclear Zn-Fe Center in the Native Enzyme. *J. Am. Chem. Soc.*, 110:3317–3318, 1988. 62

- [BDLW97] G. Battistuzzi, M. Dietrich, R. Löcke, and H. Witzel. Evidence for a conserved binding motif of the dinuclear metal site in mammalian and plant purple acid phosphatases:  $^1\text{H}$  NMR studies of the di-iron derivative of the Fe(III)Zn(II) enzyme from kidney bean. *Biochem. J.*, 323(Pt 3):593, 1997. 62
- [Bec93] A. D. Becke. Density-Functional Thermochemistry. III. The Role of Exact Exchange. *J. Chem. Phys.*, 98:5648, 1993. 50, 65
- [BG05] P. K. Biswas and V. Gogonea. A regularized and renormalized electrostatic coupling Hamiltonian for hybrid quantum-mechanical-molecular-mechanical calculations. *J. Chem. Phys.*, 123:164114, 2005. 77
- [BKJZ84] J. L. Beck, D. T. Keough, J. D. Jersey, and B. Zerner. Enzymatically active zinc, copper and mercury derivatives of the one-iron form of pig allantoic fluid acid phosphatase. *Biochim. Biophys. Acta*, 791(3):357–363, 1984. 62
- [BMC<sup>+</sup>07] R. Battistutta, M. Mazzorana, L. Cendron, A. Bortolato, S. Sarno, Z. Kazimierczuk, G. Zanotti, S. Moro, and L. A. Pinna. The ATP-binding site of protein kinase CK2 holds a positive electrostatic area and conserved water molecules. *Chembiochem*, 8(15):1804–1809, 2007. 93, 99, 111
- [BMJZ88] J. L. Beck, M. J. McArthur, J. D. Jersey, and B. Zerner. Derivatives of the purple phosphatase from red kidney bean: Replacement of zinc with other divalent metal ions. *Inorganica Chim. Acta*, 153(1):39 – 44, 1988. 62
- [BMS<sup>+</sup>86] J. Beck, L. McConachie, A. Summors, W. Arnold, J. de Jersey, and B. Zerner. Properties of a Purple Phosphatase from Red Kidney Bean - a Zinc-Iron Metalloenzyme. *Biochim. Biophys. Acta*, 869:61–68, 1986. 62
- [BMS<sup>+</sup>01] R. Battistutta, E. D. Moliner, S. Sarno, G. Zanotti, and L. A. Pinna. Structural Features Underlying Selective Inhibition of Protein Kinase CK2 by ATP-Binding Site-Directed Tetrabromo-2-Benzotriazole. *Protein. Sci.*, 10:2200–2206, 2001. 93, 94

- [BMS<sup>+</sup>05] R. Battistutta, M. Mazzorana, S. Sarno, Z. Kazimierczuk, G. Zanotti, and L. A. Pinna. Inspecting the Structure-Activity Relationship of Protein Kinase CK2 Inhibitors Derived from Tetrabromo-Benzimidazole. *Chem. Biol.*, 12:1211–1219, 2005. 93, 94
- [BSB<sup>+</sup>87] P. A. Bash, U. C. Singh, F. K. Brown, R. Langridge, and P. A. Kollman. Calculation of the Relative Change in Binding Free Energy of a Protein-Inhibitor Complex. *Science*, 235:574–576, 1987. 86
- [BSW04] P. V. Bernhardt, G. Schenk, and G. J. Wilson. Direct electrochemistry of porcine purple acid phosphatase (uteroferrin). *Biochemistry*, 43(32):10387–10392, 2004. 62
- [BT96] D. Bakowies and W. Thiel. Hybrid Models for Combined Quantum Mechanical and Molecular Mechanical Approaches. *J. Phys. Chem.*, 100:10580–10594, 1996. 31, 33, 38
- [BWC06] N. Basdevant, H. Weinstein, and M. Ceruso. Thermodynamic Basis for Promiscuity and Selectivity in Protein-Protein Interactions: PDZ Domains, a Case Study. *J. Am. Chem. Soc.*, 128:12766–12777, 2006. 88
- [CCB<sup>+</sup>95] W. D. Cornell, P. Cieplak, C. I. Bayly, I. R. Gould, K. M. Merz, D. M. Ferguson, D. C. Spellmeyer, T. Fox, J. W. Caldwell, and P. A. Kollman. A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules. *J. Am. Chem. Soc.*, 117(19):5179–5197, 1995. 74
- [CCHK89] E. A. Carter, G. Ciccotti, J. T. Hynes, and R. Kapral. Constrained reaction coordinate dynamics for the simulation of rare events. *Chem. Phys. Lett.*, 156(5):472–477, 1989. 77
- [CCR92] C. J. Casewit, K. S. Colwell, and A. K. Rappe. Application of a universal force field to organic molecules. *J. Am. Chem. Soc.*, 114(25):10035–10046, 1992. 50, 65
- [CFB<sup>+</sup>01] C. E. Check, T. O. Faust, J. M. Bailey, B. J. Wright, T. M. Gilbert, and L. S. Sunderlin. Addition of polarization and diffuse functions

- to the LANL2DZ basis set for p-block elements. *J. Phys. Chem. A*, 105(34):8111–8116, 2001. 65
- [DA82] J. C. Davis and B. A. Averill. Evidence for a spin-coupled binuclear iron unit at the active site of the purple acid phosphatase from beef spleen. *Proc. Natl. Acad. Sci. U.S.A.*, 79(15):4623, 1982. 65
- [dBMGG04] T. J. M. de Bruin, A. Milet, A. E. Greene, and Y. Gimbert. Insight into the Reactivity of Olefins in the Pauson-Khand Reaction. *J. Org. Chem.*, 69(4):1075–1080, 2004. 49
- [dBMR+01] T. J. de Bruin, A. Milet, F. Robert, Y. Gimbert, and A. E. Greene. Theoretical Study of the Regiochemistry-Determining Step of the Pauson-Khand Reaction. *J. Am. Chem. Soc.*, 123(29):7184–5, 2001. 49
- [dBMV+06] T. J. M. de Bruin, C. Michel, K. Vekey, A. E. Greene, Y. Gimbert, and A. Milet. First C-C bond formation in the Pauson-Khand reaction: Influence of carbon-carbon triple bond polarization on regiochemistry A density functional theory study. *J. Organomet. Chem.*, 691(20):4281–4288, 2006. 49
- [DeL02] W. DeLano. The PyMOL Molecular Graphics System, 2002. DeLano Scientific, Palo Alto, CA, USA, 2002. 51, 94
- [DES+99] A. Durmus, C. Eicken, B. Sift, A. Kratel, R. Kappl, J. Huttermann, and B. Krebs. The active site of purple acid phosphatase from sweet potatoes (*Ipomoea batatas*) - Metal content and spectroscopic characterization. *Eur. J. Biochem.*, 260:709–716, 1999. 62
- [DFAC02] A. Dikiy, E. G. Funhoff, B. A. Averill, and S. Ciurli. New insights into the mechanism of purple acid phosphatase through (1) H NMR spectroscopy of the recombinant human enzyme. *J. Am. Chem. Soc.*, 124(47):13974–13975, 2002. 63
- [DKB+99] S. Dapprich, I. Komáromi, K. S. Byun, K. Morokuma, and M. J. Frisch. A New ONIOM Implementation in Gaussian98. Part I. The Calculation of Energies, Gradients, Vibrational Frequencies and Electric Field Derivatives. *J. Mol. Struct. (THEOCHEM)*, 462:1–21, 1999. 36



- [DMSW91] M. Dietrich, D. Munstermann, H. Suerbaum, and H. Witzel. Purple acid phosphatase from bovine spleen. Interactions at the active site in relation to the reaction mechanism. *Eur. J. Biochem.*, 199(1):105–113, 1991. 63
- [DNMB04] T. J. Dolinsky, J. E. Nielsen, J. A. McCammon, and N. A. Baker. PDB2PQR: An Automated Pipeline for the Setup of Poisson-Boltzmann Electrostatics Calculations. *Nucleic Acids Res.*, 32:665–667, 2004. 94
- [DP01] E. Darve and A. Pohorille. Calculating free energies using average force. *J. Chem. Phys.*, 115:9169, 2001. 77
- [DQ90] S. S. David and L. J. Que. Anion binding to uteroferrin. Evidence for phosphate coordination to the iron(III) of the dinuclear active site and interaction with the hydroxo bridge. *J. Am. Chem. Soc.*, 112:6455–6463, 1990. 66
- [dVSC+99] A. H. de Vries, P. Sherwood, S. J. Collins, A. M. Rigby, M. Rigutto, and G. J. Kramer. Zeolite Structure and Reactivity by Combined Quantum-Chemical-Classical Calculations. *J. Phys. Chem. B*, 103:6133–6141, 1999. 39
- [DYP93] T. Darden, D. York, and L. Pedersen. Particle mesh Ewald: An  $N \cdot \log(N)$  method for Ewald sums in large systems. *J. Chem. Phys.*, 98:10089, 1993. 74
- [DZHS93] M. J. S. Dewar, E. G. Zoebisch, E. F. Healy, and J. J. P. Stewart. Development and Use of Quantum Mechanical Molecular Models. 76. AM1: A New General Purpose Quantum Mechanical Molecular Model. *J. Am. Chem. Soc.*, 115(12):5348–5348, 1993. 96
- [EBD+93] A. W. Ehlers, M. Böhme, S. Dapprich, A. Gobbi, A. Höllwarth, V. Jonas, K. F. Köhler, R. Stegmann, A. Veldkamp, and G. Frenking. A Set of f-Polarization Functions for Pseudo-Potential Basis-Sets of the Transition-Metals Sc-Cu, Y-Ag and La-Au. *Chem. Phys. Lett.*, 208:111–114, 1993. 50

- [ECBH96] K. P. Eurenium, D. C. Chatfield, B. R. Brooks, and M. Hodoscek. Enzyme Mechanisms with Hybrid Quantum and Molecular Mechanical Potentials. I. Theoretical Considerations. *Int. J. Quantum Chem.*, 60:1189–1200, 1996. 32, 36
- [ES73] S. W. F. Edwards, J. D. and A. G. Sykes. Reactions of  $\mu$ -hydroxodicobalt(III) Complexes. Part X. Kinetic studies of the Reaction of the  $\mu$ -Amido- $\mu$ -hydroxo-bis[tetra-ammine-cobalt(III)] Complex with Phosphate Ions in Aqueous Perchloric Acid Solutions and the Characterisation of the  $\mu$ -Amido- $\mu$ -phosphato-bis[tetra-ammine-cobalt(III)]. *J. Chem. Soc., Dalton Trans.*, pages 2757–2781, 1973. 66
- [ETHP99] M. Eichinger, P. Tavan, J. Hutter, and M. Parrinello. A Hybrid Method for Solutes In Complex Solvents: Density Functional Theory Combined with Empirical Force Fields. *J. Chem. Phys.*, 110:10452–10467, 1999. 37, 39
- [FAB<sup>+</sup>00] M. J. Field, M. Albe, C. Bret, F. P.-D. Martin, and A. Thomas. The Dynamo Library for Molecular Simulations Using Hybrid Quantum Mechanical and Molecular Mechanical Potentials. *J. Comput. Chem.*, 21:1088–1100, 2000. 37, 94, 96
- [FAR02] N. Ferré, X. Assfeld, and J.-L. Rivail. Specific Force Field Parameters Determination for the Hybrid Ab Initio QM/MM LSCF Method. *J. Comput. Chem.*, 23:610–624, 2002. 39
- [FBA05] E. G. Funhoff, M. Bollen, and B. A. Averill. The Fe(III)Zn(II) form of recombinant human purple acid phosphatase is not activated by proteolysis. *J. Inorg. Biochem.*, 99(2):521–529, 2005. 62
- [FBK90] M. J. Field, P. A. Bash, and M. Karplus. A Combined Quantum Mechanical and Molecular Mechanical Potential for Molecular Dynamics Simulations. *J. Comput. Chem.*, 11:700–733, 1990. 19, 22, 32, 36
- [FKS<sup>+</sup>01] E. G. Funhoff, C. H. W. Klaassen, B. Samyn, J. V. Beeumen, and B. A. Averill. The Highly Exposed Loop Region in Mammalian Purple Acid

Phosphatase Controls the Catalytic Activity. *Chembiochem*, 2(5):355–363, 2001. 62

- [FLW<sup>+</sup>01] E. G. Funhoff, J. Ljusberg, Y. Wang, G. Andersson, and B. A. Averill. Mutational analysis of the interaction between active site residues and the loop region in mammalian purple acid phosphatases. *Biochemistry*, 40(38):11614–11622, 2001. 62
- [FTS<sup>+</sup>04] M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, J. A. Montgomery, Jr., T. Vreven, K. N. Kudin, J. C. Burant, J. M. Millam, S. S. Iyengar, J. Tomasi, V. Barone, B. Mennucci, M. Cossi, G. Scalmani, N. Rega, G. A. Petersson, H. Nakatsuji, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, M. Klene, X. Li, J. E. Knox, H. P. Hratchian, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, P. Y. Ayala, K. Morokuma, G. A. Voth, P. Salvador, J. J. Dannenberg, V. G. Zakrzewski, S. Dapprich, A. D. Daniels, M. C. Strain, O. Farkas, D. K. Malick, A. D. Rabuck, K. Raghavachari, J. B. Foresman, J. V. Ortiz, Q. Cui, A. G. Baboul, S. Clifford, J. Cioslowski, B. B. Stefanov, G. Liu, A. Liashenko, P. Piskorz, I. Komaromi, R. L. Martin, D. J. Fox, T. Keith, M. A. Al-Laham, C. Y. Peng, A. Nanayakkara, M. Challacombe, P. M. W. Gill, B. Johnson, W. Chen, M. W. Wong, C. Gonzalez, and J. A. Pople. Gaussian 03, Revision C.02, 2004. Gaussian, Inc., Wallingford, CT, 2004. 50, 74, 76, 97
- [GAAF98] J. Gao, P. Amara, C. Alhambra, and M. J. Field. A Generalized Hybrid Orbital (GHO) Method for the Treatment of Boundary Atoms in Combined QM/MM Calculations. *J. Phys. Chem. A*, 102:4714–4721, 1998. 36, 39
- [GC04] H. Gohlke and D. A. Case. Converging Free Energy Estimates: MM-PB(GB)SA Studies on the Protein-Protein Complex Ras-Raf. *J. Comput. Chem.*, 25:238–250, 2004. 88

- [GH88] M. K. Gilson and B. Honig. Calculation of the Total Electrostatic Energy of a Macromolecular System: Solvation Energies, Binding Energies, and Conformational Analysis. *Proteins*, 4:7–18, 1988. 88, 97
- [GKC03] H. Gohlke, C. Kiel, and D. A. Case. Insights into Protein-Protein Binding by Binding Free Energy Calculation and Free Energy Decomposition for the Ras–Raf and Ras–RalGDS Complexes. *J. Mol. Biol.*, 330:891–913, 2003. 88
- [GLM<sup>+</sup>03] Y. Gimbert, D. Lesage, A. Milet, F. Fournier, A. E. Greene, and J. C. Tabet. On Early Events in the Pauson-Khand Reaction. *Org. Lett.*, 5:4073–4076, 2003. 49
- [GMH<sup>+</sup>99] L. Guddat, A. McAlpine, D. Hume, S. Hamilton, J. de Jersey, and J. Martin. Crystal structure of mammalian purple acid phosphatase. *Structure*, 7:757–767, 1999. 62
- [Gru95] H. Grubmüller. Predicting slow structural transitions in macromolecular systems: Conformational flooding. *Phys. Rev. E*, 52(3):2893–2906, 1995. 77
- [GSD<sup>+</sup>05] F. Gräter, S. M. Schwarzl, A. Dejaegere, S. Fischer, and J. C. Smith. Protein/Ligand Binding Free Energies Calculated with Quantum Mechanics/Molecular Mechanics. *J. Phys. Chem. B*, 109:10474–10483, 2005. 92
- [GVTG03] M. Garcia-Viloca, D. G. Truhlar, and J. Gao. Importance of Substrate and Cofactor Polarization in the Active Site of Dihydrofolate Reductase. *J. Mol. Biol.*, 327:549–560, 2003. 91, 107
- [GX92] J. Gao and X. Xia. A Priori Evaluation of Aqueous Polarization Effects Through Monte Carlo QM-MM Simulations. *Science*, 258:631–635, 1992. 33, 98, 107
- [Hal96] T. A. Halgren. Merck Molecular Force Field. I. Basis, Form, Scope, Parameterization, and Performance of MMFF94. *J. Comput. Chem.*, 17:490–519, 1996. 91

- [HB96] F. A. Hicks and S. L. Buchwald. Highly Enantioselective Catalytic Pauson-Khand Type Formation of Bicyclic Cyclopentenones. *J. Am. Chem. Soc.*, 118:11688–11689, 1996. 48
- [HBBF97] B. Hess, H. Bekker, H. J. C. Berendsen, and J. Fraaije. LINCS: A linear constraint solver for molecular simulations. *J. Comput. Chem.*, 18(12):1463–1472, 1997. 74
- [Her72] R. B. Hermann. Theory of Hydrophobic Bonding. II. Correlation of Hydrocarbon Solubility in Water with Solvent Cavity Surface Area. *J. Phys. Chem.*, 76:2754–2759, 1972. 88
- [Hes02] B. Hess. Convergence of Sampling in Protein Simulations. *Phys. Rev. E*, 65:031910, 2002. 88
- [HHN<sup>+</sup>04] C. Hensen, J. C. Hermann, K. Nam, S. Ma, J. Gao, and H.-D. Höltje. A combined QM/MM approach to protein-ligand interactions: polarization effects of the HIV-1 protease on selected high affinity inhibitors. *J. Med. Chem.*, 47:6673–6680, 2004. 91
- [HJB<sup>+</sup>96] A. R. Hayman, S. J. Jones, A. Boyde, D. Foster, W. H. Colledge, M. B. Carlton, M. J. Evans, and T. M. Cox. Mice lacking tartrate-resistant acid phosphatase (Acp 5) have disrupted endochondral ossification and mild osteopetrosis. *Development*, 122(10):3151–3162, 1996. 62
- [HLY<sup>+</sup>97] J. D. Hoyer, C. Y. Li, L. T. Yam, C. A. Hanson, and P. J. Kurtin. Immunohistochemical demonstration of acid phosphatase isoenzyme 5 (tartrate-resistant) in paraffin sections of hairy cell leukemia and other hematologic disorders. *Am. J. Clin. Pathol.*, 108(3):308–315, 1997. 62
- [HMÅ98] T. Hansson, J. Marelius, and J. Åqvist. Ligand binding affinity prediction by linear interaction energy methods. *J. Comput. Aided Mol. Des.*, 12(1):27–35, 1998. 87
- [HP73] P. C. Hariharan and J. A. Pople. The Influence of Polarization Functions on Molecular Orbital Hydrogenation Energies. *Theor. Chim. Acta.*, 28(3):213–222, 1973. 50

- [HPR<sup>+</sup>08] H. P. Hratchian, P. V. Parandekar, K. Raghavachari, M. J. Frisch, and T. Vreven. QM:QM Electronic Embedding Using Mulliken Atomic Charges: Energies and Analytic Gradients in an ONIOM Framework. *J. Chem. Phys.*, 128:034107–11, 2008. 41
- [Hut] J. Hutter. CPMD Software Package. *Max-Planck-Institut für Festkörperforschung and IBM Zürich (1995–1999)*. 77
- [HW85] P. J. Hay and W. R. Wadt. Ab Initio Effective Core Potentials for Molecular Calculations. Potentials for K to Au Including the Outermost Core Orbitals. *J. Chem. Phys.*, 82(1):299–310, 1985. 50, 65
- [HW87] J. K. Hwang and A. Warshel. Semiquantitative Calculations of Catalytic Free Energies in Genetically Modified Enzymes. *Biochemistry*, 26(10):2669–2673, 1987. 81
- [HWC<sup>+</sup>02] S. Huo, J. Wang, P. Cieplak, P. A. Kollman, and I. D. Kuntz. Molecular Dynamics and Free Energy Analyses of Cathepsin D-Inhibitor Interactions: Insight into Structure-Based Ligand Design. *J. Med. Chem.*, 45:1412–1419, 2002. 89
- [HWKA00a] K. Hiroi, T. Watanabe, R. Kawagishi, and I. Abe. Asymmetric catalytic Pauson-Khand Reactions With Chiral Phosphine Ligands: Dramatic Effects of Substituents in 1,6-Enyne Systems. *Tetrahedron Lett.*, 41:891–895, 2000. 48
- [HWKA00b] K. Hiroi, T. Watanabe, R. Kawagishi, and I. Abe. Catalytic use of chiral phosphine ligands in asymmetric Pauson-Khand Reactions. *Tetrahedron: Asymmetry*, 11:797–808, 2000. 48
- [Jar97] C. Jarzynski. Nonequilibrium Equality for Free Energy Differences. *Phys. Rev. Lett.*, 78(14):2690–2693, 1997. 77
- [JCM<sup>+</sup>83] W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, and M. L. Klein. Comparison of Simple Potential Functions for Simulating Liquid Water. *J. Chem. Phys.*, 79:926–935, 1983. 74, 96

- [JMTR96] W. Jorgensen, D. Maxwell, and J. Tirado-Rives. Development and Testing of the OPLS All-Atom Force Field on Conformational Energetics and Properties of Organic Liquids. *J. Am. Chem. Soc.*, 118:11225–11236, 1996. 96
- [JR85] W. L. Jorgensen and C. Ravimohan. Monte Carlo simulation of Differences in Free Energies of Hydration. *J. Chem. Phys.*, 83:3050–3054, 1985. 81
- [JSC00] N. Jeong, B. K. Sung, and Y. K. Choi. Rhodium(I)-Catalyzed Asymmetric Intramolecular Pauson-Khand-Type Reaction. *J. Am. Chem. Soc.*, 122(28):6, 2000. 48
- [Kim00] E. Kimura. Dimetallic hydrolases and their models. *Curr. Opin. Chem. Biol.*, 4(2):207–213, 2000. 63
- [Kir35] J. G. Kirkwood. Statistical Mechanics of Fluid Mixtures. *J. Chem. Phys.*, 3(5):300–313, 1935. 81, 84
- [KKP<sup>+</sup>73] I. U. Khand, G. R. Knox, P. L. Pauson, W. E. Watts, and M. I. Foreman. Organocobalt Complexes. Part II. Reaction of Acetylenehexacarbonyldicobalt Complexes,  $(R^1C_2R^2)Co_2(CO)_6$ , With Norbornene and Its Derivatives. *J. Chem. Soc. Perkin Trans. I*, 1973:977–981, 1973. 48
- [KKPW71] I. U. Khand, G. R. Knox, P. L. Pauson, and W. E. Watts. A Cobalt Induced Cleavage Reaction and a New series of Arenecobalt Carbonyl Complexes. *J. Chem. Soc. Chem. Commun.*, 1971(1):36–36, 1971. 48
- [KKPW73] I. U. Khand, G. R. Knox, P. L. Pauson, and W. E. Watts. Organocobalt Complexes. Part I. Arene Complexes Derived From Dodecacarbonyltetracobalt. *J. Chem. Soc. Perkin Trans. I*, 1973:975–977, 1973. 48
- [KLC<sup>+</sup>05] A. Khandelwal, V. Lukacova, D. Comez, D. M. Kroll, S. Raha, and S. Balaz. A Combination of Docking, QM/MM Methods, and MD Simulation for Binding Affinity Estimation of Metalloprotein Ligands. *J. Med. Chem.*, 48:5437–47, 2005. 92

- [KM83] M. Karplus and J. A. McCammon. Dynamics of Proteins: Elements and Function. *Annu. Rev. Biochem.*, 52:263–300, 1983. 81
- [KMR<sup>+</sup>00] P. A. Kollman, I. Massova, C. Reyes, B. Kuhn, S. Huo, L. Chong, M. Lee, T. Lee, Y. Duan, and W. Wang. Calculating Structures and Free Energies of Complex Molecules: Combining Molecular Mechanics and Continuum Models. *Acc. Chem. Res*, 33(12):889–897, 2000. 87, 89
- [KRB<sup>+</sup>95] S. Kumar, J. M. Rosenberg, D. Bouzida, R. H. Swendsen, and P. A. Kollman. Multidimensional Free-Energy Calculations Using the Weighted Histogram Analysis Method. *J. Comput. Chem.*, 16(11):1339–1350, 1995. 77
- [KSF<sup>+</sup>96] T. Klabunde, N. Sträter, R. Frohlich, H. Witzel, and B. Krebs. Mechanism of Fe(III)-Zn(II) purple acid phosphatase based on crystal structures. *J. Mol. Biol.*, 259:737–748, 1996. 62, 63, 64, 65, 69, 73
- [LJK<sup>+</sup>99] Y. Lindqvist, E. Johansson, H. Kaija, P. Vihko, and G. Schneider. Three-dimensional structure of a mammalian purple acid phosphatase at 2.2 Å resolution with a  $\mu$ -(hydr)oxo bridged di-iron center. *J. Mol. Biol.*, 291:135–147, 1999. 62, 64
- [LMM<sup>+</sup>01] K. B. Ljungberg, J. Marelius, D. Musil, P. Svensson, B. Norden, and J. Åqvist. Computational Modelling of Inhibitor Binding to Human Thrombin. *Eur. J. Pharm. Sci.*, 12(4):441–446, 2001. 87
- [LMW86] T. P. Lybrand, J. A. McCammon, and G. Wipff. Theoretical Calculation of Relative Binding Affinity in Host-Guest Systems. *Proc. Natl. Acad. Sci. U.S.A.*, 83(4):833–835, 1986. 81
- [LP02] A. Laio and M. Parrinello. Escaping free-energy minima. *Proc. Natl. Acad. Sci. U.S.A.*, 99(20):12562–12566, 2002. 77
- [LR03] G. Lamoureux and B. Roux. Modeling Induced Polarization with Classical Drude Oscillators: Theory and Molecular Dynamics Simulation Algorithm. *J. Chem. Phys.*, 119:3025–3039, 2003. 35



- [LRJ05] H. Li, A. D. Robertson, and J. H. Jensen. Very Fast Empirical Prediction and Rationalization of Protein  $pK_a$  Values. *Proteins*, 61:704–721, 2005. 74, 94
- [LT05] H. Lin and D. G. Truhlar. Redistributed Charge and Dipole Schemes for Combined Quantum Mechanical and Molecular Mechanical Calculations. *J. Phys. Chem. A*, 109:3991–4004, 2005. 38
- [LYP88] C. Lee, W. Yang, and R. G. Parr. Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density. *Phys. Rev., B*, 37(2):785–789, 1988. 65
- [MA98] M. Merckx and B. A. Averill. The activity of oxidized bovine spleen purple acid phosphatase is due to an Fe (III) Zn (II) 'impurity'. *Biochemistry*, 37(32):11223–11231, 1998. 62
- [MCAK93] E. Mueller, M. Crowder, B. Averill, and J. Knowles. Purple Acid-Phosphatase - a Diiron Enzyme That Catalyzes a Direct Phospho Group Transfer to Water. *J. Am. Chem. Soc.*, 115:2974–2975, 1993. 63
- [MLT<sup>+</sup>96] G. Monard, M. Loos, V. Théry, K. Baka, and J.-L. Rivail. Hybrid Classical Quantum Force Field for Modeling Very Large Molecules. *Int. J. Quantum Chem.*, 58:153–159, 1996. 39
- [MM95] F. Maseras and K. Morokuma. IMOMM: A New Integrated Ab Initio + Molecular Mechanics Geometry Optimization Scheme of Equilibrium Structures and Transition States. *J. Comput. Chem.*, 16:1170–1179, 1995. 36, 38, 40
- [MP03] F. Meggio and L. A. Pinna. One-thousand-and-one Substrates of Protein Kinase CK2? *FASEB J.*, 17:349–368, 2003. 93
- [MPA99] M. Merckx, M. W. H. Pinkse, and B. A. Averill. Evidence for nonbridged coordination of p-nitrophenyl phosphate to the dinuclear Fe (III)-M (II) center in bovine spleen purple acid phosphatase during enzymatic turnover. *Biochemistry*, 38(31):9914–9925, 1999. 63, 65

- [MPF00] R. B. Murphy, D. M. Philipp, and R. A. Friesner. Frozen Orbital QM/MM Methods for Density Functional Theory. *Chem. Phys. Lett.*, 321:113–120, 2000. 39
- [MPM<sup>+</sup>04] F. Meggio, M. Pagano, S. Moro, G. Zagotto, M. Ruzzene, S. Sarno, G. Cozza, J. Bain, M. Elliott, A. Deana, A. Brunati, and L. Pinna. Inhibition of Protein Kinase CK2 by Condensed Polyphenolic Derivatives. An in Vitro and in Vivo Study. *Biochemistry*, 43:12931–12936, 2004. 110
- [MTS07] J. Mathew, T. Thomas, and C. H. Suresh. Quantitative Assessment of the Stereoelectronic Profile of Phosphine Ligands. *Inorg. Chem.*, 46(25):10800–10809, 2007. 50
- [MWGG08] P. Maji, W. Wang, A. E. Greene, and Y. Gimbert. Studies directed toward the synthesis of chiral tungsten and molybdenum carbonyl complexes. *J. Organomet. Chem.*, 693(10):1841–1849, 2008. 49
- [MWM<sup>+</sup>02] G. Manning, D. B. Whyte, R. Martinez, T. Hunter, and S. Sudarsanam. The Protein Kinase Complement of the Human Genome. *Science*, 298:1912–1934, 2002. 92
- [NPE<sup>+</sup>07] Z. Nie, C. Perretta, P. Erickson, S. Margosiak, R. Almassy, J. Lu, A. Averill, K. M. Yager, and S. Chu. Structure-Based Design, Synthesis, and Study of Pyrazolo[1,5-a][1,3,5]triazine Derivatives as Potent Inhibitors of Protein Kinase CK2. *Bioorg. Med. Chem. Lett.*, 17:4191–4195, 2007. 110, 111
- [PAR<sup>+</sup>04] M. A. Pagano, M. Andrzejewska, M. Ruzzene, S. Sarno, L. Cesaro, J. Bain, M. Elliott, F. Meggio, Z. Kazimierczuk, and L. A. Pinna. Optimization of Protein Kinase CK2 Inhibitors Derived From 4,5,6,7-Tetrabromobenzimidazole. *J. Med. Chem.*, 47:6239–6247, 2004. 93, 94, 103
- [PBH82] J. P. M. Postma, H. J. C. Berendsen, and J. R. Haak. Thermodynamics of Cavity Formation in Water. A Molecular Dynamics Study. *Faraday Symp. Chem. Soc.*, 17:55–67, 1982. 81

- [PCV<sup>+</sup>92] J. P. Perdew, J. A. Chevary, S. H. Vosko, K. A. Jackson, M. R. Pederson, D. J. Singh, and C. Fiolhais. Atoms, Molecules, Solids, and Surfaces: Applications of the Generalized Gradient Approximation for Exchange and Correlation. *Physical Review B*, 46(11):6671–6687, 1992. 50
- [PF99] D. M. Philipp and R. A. Friesner. Mixed Ab Initio QM/MM Modeling Using Frozen Orbitals and Tests with Alanine Dipeptide and Tetrapeptide. *J. Comput. Chem.*, 20:1468–1494, 1999. 39
- [PMA99] M. W. H. Pinkse, M. Merckx, and B. A. Averill. Fluoride inhibition of bovine spleen purple acid phosphatase: characterization of a ternary enzyme-phosphate-fluoride complex as a model for the active enzyme-substrate-hydroxide complex. *Biochemistry*, 38(31):9926–9936, 1999. 62
- [PMLR<sup>+</sup>08] R. Prudent, V. Moucadel, M. López-Ramos, S. Aci, B. Laudet, L. Mouawad, C. Barette, J. Einhorn, C. Einhorn, and J. N. Denis. Expanding the Chemical Diversity of CK2 Inhibitors. *Mol. Cell. Biochem.*, 316:71–85, 2008. 93
- [PV75] G. N. Patey and J. P. Valleau. A Monte Carlo method for obtaining the interionic potential of mean force in ionic solution. *J. Chem. Phys.*, 63:2334, 1975. 77
- [RCC<sup>+</sup>92] A. K. Rappe, C. J. Casewit, K. S. Colwell, W. A. G. III, and W. M. Skiff. UFF, a full periodic table force field for molecular mechanics and molecular dynamics simulations. *J. Am. Chem. Soc.*, 114(25):10024–10035, 1992. 50, 65
- [RCC93] A. K. Rappe, K. S. Colwell, and C. J. Casewit. Application of a universal force field to metal complexes. *Inorg. Chem.*, 32(16):3438–3450, 1993. 50, 65
- [RDMK00] N. Reuter, A. Dejaegere, B. Maigret, and M. Karplus. Frontier Bonds in QM/MM Methods: A Comparison of Different Approaches. *J. Phys. Chem. A*, 104:1720–1735, 2000. 38

- [RMG<sup>+</sup>01] F. Robert, A. Milet, Y. Gimbert, D. Konya, and A. E. Greene. Regiochemistry in the Pauson-Khand Reaction: Has a Trans Effect Been Overlooked? *J. Am. Chem. Soc.*, 123(23):5396–5400, 2001. 49
- [RPP02] M. Ruzzene, D. Penzo, and L. A. Pinna. Protein kinase CK2 inhibitor 4,5,6,7-tetrabromobenzotriazole (TBB) induces apoptosis and caspase-dependent degradation of haematopoietic lineage cell-specific protein 1 (HS1) in Jurkat cells. *Biochem. J.*, 364:41–47, 2002. 93
- [SBC<sup>+</sup>01] G. Schenk, C. Boutchard, L. Carrington, C. Noble, B. Moubaraki, K. Murray, J. de Jersey, G. Hanson, and S. Hamilton. A purple acid phosphatase from sweet potato contains an antiferromagnetically coupled binuclear Fe-Mn center. *J. Biol. Chem.*, 276:19084–19088, 2001. 62
- [SBP86] T. P. Straatsma, H. J. C. Berendsen, and J. P. M. Postma. Free Energy of Hydrophobic Hydration: A Molecular Dynamics Study of Noble Gases in Water. *J. Chem. Phys.*, 85:6720–6727, 1986. 81
- [SC98] M. Sprik and G. Ciccotti. Free energy from constrained molecular dynamics. *J. Chem. Phys.*, 109:7737, 1998. 77
- [SCC<sup>+</sup>98] J. Srinivasan, T. Cheatham, P. Cieplak, P. A. Kollman, and D. A. Case. Continuum Solvent Studies of the Stability of DNA, RNA, and Phosphoramidate-DNA Helices. *J. Am. Chem. Soc.*, 120:9401–9409, 1998. 87, 88, 89
- [SdMR<sup>+</sup>03] S. Sarno, E. de Moliner, M. Ruzzene, M. A. Pagano, R. Battistutta, J. Bain, D. Fabbro, J. Schoepfer, M. Elliott, P. Furet, F. Meggio, G. Zanotti, and L. A. Pinna. Biochemical and three-dimensional-structural study of the specific inhibition of protein kinase CK2 by [5-oxo-5,6-dihydroindolo-(1,2-a)quinazolin-7-yl]acetic acid (IQA). *Biochem. J.*, 374:639–646, 2003. 93
- [SEL<sup>+</sup>08] G. Schenk, T. W. Elliott, E. Leung, L. E. Carrington, N. Mitic, L. R. Gahan, L. W. Guddat, P. M. Gorman, S. Kim, and M. Guo. Crystal structures of a purple acid phosphatase, representing different steps of this enzyme’s catalytic cycle. *BMC Struct. Biol.*, 8(1):6, 2008. 62, 63, 64

- [SGC<sup>+</sup>99] G. Schenk, Y. Ge, L. Carrington, C. Wynne, I. Searle, B. Carroll, S. Hamilton, and J. de Jersey. Binuclear metal centers in plant purple acid phosphatases: Fe-Mn in sweet potato and Fe-Zn in soybean. *Arch. Biochem. Biophys.*, 370:183–189, 1999. 62
- [SGC<sup>+</sup>05] G. Schenk, L. Gahan, L. Carrington, N. Mitic, M. Valizadeh, S. Hamilton, J. de Jersey, and L. Guddat. Phosphate forms an unusual tripodal complex with the Fe-Mn center of sweet potato purple acid phosphatase. *Proc. Natl. Acad. Sci. U.S.A.*, 102:273–278, 2005. 62
- [SGM89] J. Schindelmeiser, F. Gullotta, and D. Munstermann. Purple acid phosphatase of human brain macrophages in AIDS encephalopathy. *Pathol. Res. Pract.*, 185(2):184–186, 1989. 62
- [SHF<sup>+</sup>96] M. Svensson, S. Humbel, R. D. J. Froese, T. Matsubara, S. Sieber, and K. Morokuma. ONIOM: A Multilayered Integrated MO + MM Method for Geometry Optimizations and Single Point Energy Predictions. A Test for Diels-Alder Reactions and Pt(P(t-Bu)<sub>3</sub>)<sub>2</sub> + H<sub>2</sub> Oxidative Addition. *J. Phys. Chem.*, 100:19357–19363, 1996. 40
- [SHM04] J. M. J. Swanson, R. H. Henchman, and J. A. McCammon. Revisiting Free Energy Calculations: A Theoretical Connection to MM/PBSA and Direct Calculation of the Association Free Energy. *Biophys. J.*, 86:67–74, 2004. 86
- [SK86] U. C. Singh and P. A. Kollman. Applications to the CH<sub>3</sub>Cl + Cl-Exchange Reaction and Gas Phase Protonation of Polyethers. *J. Comput. Chem.*, 7:718–730, 1986. 19, 22, 32, 35, 36, 38
- [SKT<sup>+</sup>95] N. Sträter, T. Klabunde, P. Tucker, H. Witzel, and B. Krebs. Crystal-Structure of a Purple Acid-Phosphatase Containing a Dinuclear Fe(III)-Zn(II) Active-Site. *Science*, 268:1489–1492, 1995. 62, 64
- [SLH<sup>+</sup>05] D. V. D. Spoel, E. Lindahl, B. Hess, G. Groenhof, A. E. Mark, and H. J. C. Berendsen. GROMACS: fast, flexible, and free. *J. Comput. Chem.*, 26:1701–1718, 2005. 74, 77

- [SN00] G. Schaftenaar and J. Noordik. Molden: a pre- and post-processing program for molecular and electronic structures. *J. Comput. Aided Mol. Des.*, 14:123–134, 2000. 94
- [SQW+02] S. K. Smoukov, L. Quaroni, X. Wang, P. E. Doan, B. M. Hoffman, and L. Q. Jr. Electro-nuclear double resonance spectroscopic evidence for a hydroxo-bridge nucleophile involved in catalysis by a dinuclear hydrolase. *J. Am. Chem. Soc.*, 124(11):2595–2603, 2002. 63, 64, 65
- [SRM91] J. Schindelmeiser, H. J. Radzun, and D. Munstermann. Tartrate-resistant, purple acid phosphatase in Gaucher cells of the spleen. Immuno- and cytochemical analysis. *Pathol. Res. Pract.*, 187(2-3):209–213, 1991. 62
- [SRM+01] S. Sarno, H. Reddy, F. Meggio, M. Ruzzene, S. P. Davies, A. Donella-Deana, D. Shugar, and L. A. Pinna. Selectivity of 4,5,6,7-tetrabromobenzotriazole, an ATP Site-Directed Inhibitor of Protein Kinase CK2 (‘Casein Kinase-2’). *FEBS Lett.*, 496:44–48, 2001. 93
- [SSH94] D. Sitkoff, K. A. Sharp, and B. Honig. Accurate Calculation of Hydration Free Energies Using Macroscopic Solvent Models. *J. Phys. Chem.*, 98(7):1978–1988, 1994. 88, 98
- [STHH90] W. C. Still, A. Tempczyk, R. C. Hawley, and T. Hendrickson. Semianalytical Treatment of Solvation for Molecular Mechanics and Dynamics. *J. Am. Chem. Soc.*, 112:6127–6129, 1990. 88
- [Sur06] C. H. Suresh. Molecular electrostatic potential approach to determining the steric effect of phosphine ligands in organometallic chemistry. *Inorg. Chem.*, 45(13):4982–4986, 2006. 50
- [Swa03] M. Swart. AddRemove: A New Link Model for Use in QM/MM Studies. *Int. J. Quantum Chem.*, 91:177–183, 2003. 37
- [TL99] Y. Tu and A. Laaksonen. On the Effect of Lennard-Jones Parameters on the Quantum Mechanical and Molecular Mechanical Coupling in a Hybrid Molecular Dynamics Simulation of Liquid Water. *J. Chem. Phys.*, 111:7519–7525, 1999. 33

- [TM84] B. L. Tembre and J. A. McCammon. Ligand-Receptor Interactions. *Comput. Chem.*, 8:281–283, 1984. 81, 85
- [TRR<sup>+</sup>94] V. Théry, D. Rinaldi, J.-L. Rivail, B. Maigret, and G. G. Ferenczy. Quantum Mechanical Computations on Very Large Molecular Systems: The Local Self-Consistent Field Method. *J. Comput. Chem.*, 15:269–282, 1994. 36, 39
- [TS99] M. B. Twitchett and A. G. Sykes. Structure, Properties and Reactivity of the FeIIFeIII and ZnIIFeIII Purple Acid Phosphatases. *European Journal of Inorganic Chemistry*, 1999(12):2105–2115, 1999. 63, 65
- [TYW<sup>+</sup>01] S. Tawfic, S. Yu, H. Wang, R. Faust, A. Davis, and K. Ahmed. Protein kinase CK2 signal in neoplasia. *Histol. Histopathol.*, 16:573–582, 2001. 93
- [ULS<sup>+</sup>99] J. Uppenberg, F. Lindqvist, C. Svensson, B. Ek-Rylander, and G. Andersson. Crystal structure of a mammalian purple acid phosphatase. *J. Mol. Biol.*, 290:201–211, 1999. 62, 64
- [VBK<sup>+</sup>06] T. Vreven, K. S. Byun, I. Komáromi, S. Dapprich, J. A. Montgomery, K. Morokuma, and M. J. Frisch. Combining Quantum Mechanics Methods with Molecular Mechanics Methods in ONIOM. *J. Chem. Theory Comput.*, 2:815–826, 2006. 41
- [VCA91] J. B. Vincent, M. W. Crowder, and B. A. Averill. Evidence for a phosphoryl-enzyme intermediate in phosphate ester hydrolysis by purple acid phosphatase from bovine spleen. *J. Biol. Chem.*, 266(27):17737–40, 1991. 63
- [WHD<sup>+</sup>91] D. L. Wang, R. C. Holz, S. S. David, L. Q. Jr, and M. T. Stankovich. Electrochemical properties of the diiron core of uteroferrin and its anion complexes. *Biochemistry*, 30(33):8187–8194, 1991. 62
- [WHGP90] B. Waszkowycz, I. H. Hillier, N. Gensmantel, and D. W. Payling. A Theoretical Study of Hydrolysis by Phospholipase A2: The Catalytic Role of the Active Site and Substrate Specificity. *J. Chem. Soc. Perkin Trans. II*, N/A:1259–1268, 1990. 38

- [WHWQ00] X. D. Wang, R. Y. N. Ho, A. K. Whiting, and L. Que. Spectroscopic characterization of a ternary phosphatase-substrate-fluoride complex. Mechanistic implications for dinuclear hydrolases. *J. Am. Chem. Soc.*, 122(33):8103–8103, 2000. 63
- [WKSV06] T. Waratrujiwong, B. Krebs, F. Spener, and P. Visoottiviseth. Recombinant purple acid phosphatase isoform 3 from sweet potato is an enzyme with a diiron metal center. *FEBS Journal*, 273(8):1649–1659, 2006. 62
- [WL76] A. Warshel and M. Levitt. Theoretical Studies of Enzymic Reactions: Dielectric, Electrostatic and Steric Stabilization of the Carbonium Ion in the Reaction of Lysozyme. *J. Mol. Biol.*, 103:227–249, 1976. 19, 21, 35, 39
- [WMWK01] J. Wang, P. Morin, W. Wang, and P. A. Kollman. Use of MM-PBSA in Reproducing the Binding Free Energies to HIV-1 RT of TIBO Derivatives and Predicting the Binding Mode to HIV-1 RT of Efavirenz by Docking and MM-PBSA. *J. Am. Chem. Soc.*, 123:5221–5230, 2001. 89
- [WQ98] X. Wang and L. J. Que. Extended X-ray absorption fine structure studies of the anion complexes of FeZn uteroferrin. *Biochemistry*, 37(21):7813–7821, 1998. 64
- [WW82] J. Warwicker and H. C. Watson. Calculation of the Electric Potential in the Active Site Cleft Due to Alpha-Helix Dipoles. *J. Mol. Biol.*, 157:671–9, 1982. 88
- [WW07] M. Wang and C. F. Wong. Rank-ordering protein-ligand binding affinity by a quantum mechanics/molecular mechanics/Poisson-Boltzmann-surface area model. *J. Chem. Phys.*, 126:026101, 2007. 97
- [WWC<sup>+</sup>04] J. Wang, R. M. Wolf, J. W. Caldwell, P. A. Kollman, and D. A. Case. Development and Testing of a General Amber Force Field. *J. Comput. Chem.*, 25:1157–1174, 2004. 91
- [WWKC06] J. Wang, W. Wang, P. A. Kollman, and D. A. Case. Automatic Atom Type and Bond Type Perception in Molecular Mechanical Calculations. *J. Mol. Graph. Model.*, 25:247–260, 2006. 91



- [YMS97] Y. S. Yang, J. M. McCormick, and E. I. Solomon. Circular dichroism and magnetic circular dichroism studies of the mixed-valence binuclear non-heme iron active site in uteroferrin and its anion complexes. *J. Am. Chem. Soc.*, 119(49):11832–11842, 1997. 62, 65
- [ZFG<sup>+</sup>01] R. Zhou, R. A. Friesner, A. Ghosh, R. C. Rizzo, W. L. Jorgensen, and R. M. Levy. New Linear Interaction Method for Binding Affinity Calculations Using a Continuum Solvent Model. *J. Phys. Chem. B*, 105:10388–10397, 2001. 87
- [Zwa54] R. W. Zwanzig. High-Temperature Equation of State by a Perturbation Method. I. Nonpolar Gases. *J. Chem. Phys.*, 22(8):1420–1426, 1954. 81, 83