Universitat Autònoma de Barcelona

Prediction of biometallic interactions: challenges and applications

Ph.D.Thesis

Elisabeth Ortega Carrasco

Supervisors: Jean-Didier Maréchal, Agustí Lledós i Falcó

Theoretical and Computational Chemistry Ph.D. Program

Departament de Química

Facultat de Ciències

2015

Memòria presentada per aspirar al Grau de Doctor per Elisabeth Ortega Carrasco.

Elisabeth Ortega Carrasco

Vist i plau:

Dr. Jean-Didier Maréchal

Prof. Agustí Lledós Falcó

Bellaterra, 4 de juny del 2015.

The best way to make your dreams come true is to wake up.

Paul Valery

Abbreviations

ATP	Adenosine triphosphate
CD	Circular dichroism
*cd***HO**	*Corynebacterium Diphteriae* Haemoxygenase
DNA	Deoxyribonucleic acid
DFT	Density Functional Theory
EDA	Energy Decomposition Analysis
MM	Molecular Mechanics
NMA	Normal Mode Analysis
PES	Potential Energy Surface
QM	Quantum Mechanics
QM/MM	Quantum Mechanics/Molecular Mechanics
TS	Transition State

Abstract

The presence of transition metals in nature is not a man-made fact. Indeed, the structural and electronic properties provided by that particular pieces of the periodic table are key in the regulation of biological processes such as the oxygen transport in animals or the photosynthesis in plants.
Pharmacology, chemistry, medicine and biotechnological fields learned the lessons of organometallic chemistry given by Nature, evolving their research to more specific processes.

In this thesis, titled "Prediction of biometallic interactions: challenges and applications", all the efforts have been focused in the description of the structural and electronic effects of the binding of an organometallic compound to a biological host (DNA, proteins and peptides), by the application of molecular modelling techniques.
The molecular modeller's toolbox includes a huge variety of methods to choose. The election of one of other not only depends on the system where it has to be applied, but also the chemical interaction is an important variable to take into account. Molecular modelling techniques can be divided in two fields: those based in the principles of quantum mechanics, and the other which are described using the fundamentals of molecular mechanics. Quantum mechanics methods accurately describes the electronic effect of the metal, but are more time consuming that molecular mechanics ones, which are the best option to study large systems.

This thesis is composed by four chapters presenting the results obtained during the last years of research. In all four, different chemical and biological processes involving the interaction of transition metals and biological hosts have been studied using different computational tools. In these chapters we investigated on 1) the use of protein-ligand docking with organometallic ligands, 2) the description of the resting state and activation process of an artificial metalloenzyme using the hybrid Quantum Mechanics/Molecular Mechanics approach, 3) the study of the short and long range effects of the coordination of cisplatin in DNA and proteins mixing Quantum Mechanics and Normal Mode Analysis, and 4) on the clues on the enantioselective formation of artificial metallopeptides combining theoretical and experimental knowledge. The election of the best technique at each stage of research was key to describe the chemical system and succeed on the description of the involved biological process.

Contents

1 ... Introduction

The secret of getting ahead is getting started.
The secret of getting started is breaking your complex overwhelming tasks into small manageable tasks, and starting on the first one.

Mark Twain

1.1 Transition metals in living systems: general concepts, potential uses and applications

Transition metal ions are fundamental in Nature. Despite the general opinion, their presence in biological processes is not a man-made fact but a process that prompted early during the evolution. Transition metals have provided life with efficient catalysts and other optimal processes that would not be reached by pure organic systems. In fact, key mechanisms of primitive and evolved organisms are based on transition metal compounds. Haemoglobin, Haemocyanin, Vitamin B12, Cytochrome C or Cytochromes P450 are respectively involved in respiration, DNA regulation, electron transfers and metabolism of endo- and exobiotics, and they are only few examples of natural biomolecules containing metals.

The general idea of society regarding these particular pieces of the periodic table is still somehow negative. Indeed, metals are also of the worst pollutants on Earth and the balance between their harmful and positive effects on living organisms is biochemically complex. Such phenomenon is highly dependent on their concentration inside and outside the cells and exacerbated for the so-called heavy metals. One only has to write on his/her favourite internet browser the sentence "Toxicity of heavy metals" to see the huge amount of information that are published about how dangerous can be the presence of heavy metals in living organisms and its environment (2.360.000 hits in the google motor on the 11th may 2015).[1,2]

Biotechnological applications of the interactions between metal and biological systems have been inspired by Nature. Transition metals can interact with biological species such as proteins, DNA or peptides and form a wide variety of chemical complexes, such complementarity is now used in numerous fields: from medicine to the syntesis of new biosensors. In this section we will discuss on the presence of transition metals in biological processes, starting from the basics of bioinorganic chemistry and the properties of transition metals and finishing focusing on how metals interact with biological targets.

1.1.1 A drop of bioinorganic chemistry

Bioinorganic chemistry is the field of the science involved on the study of inorganic compounds (metals, transition metals and metalloids) interacting with biological systems such as proteins or DNA. These biological systems are basically composed by carbon, hydrogen, nitrogen, oxygen, phosphorous and sulphur, which are defined as **primary bioelements**.[a] Other elements, which are called essential **secondary bioelements**, are present in living organisms but in lower concentration than the primary ones. Examples of them are calcium, sodium, potassium, magnesium, chlorine, iron and iodine. **Trace elements** are those that can be found in some living organisms in very low concentrations (boron, bromine, copper, fluorine, manganese and silicon). In accordance with this classification, the abundance of elements present in living organisms is not the same than in

[a] Indeed, some books refer to those as CHNOPS (Carbon, Hydrogen, Nitrogen, Oxygen, Phosphorous and Sulphur).

non-living inorganic environments. In Figure 1.1, the amount of the most representative elements in the Earth's crust is compared with the presence of the same elements in living organisms. The different percentages of primary bioelements in both fields gives an idea of the distinctive type of chemical behaviour of living and non-living worlds. [3]

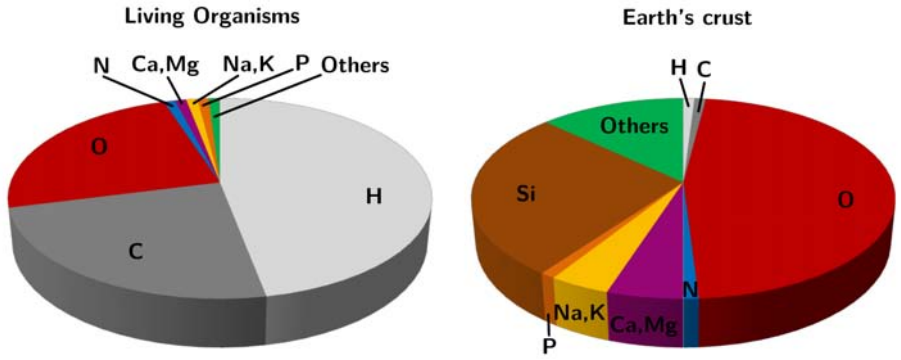


Figure 1.1: Abundance of the most representative elements in the Earth's crust compared with their abundance in living organisms.

The basic idea behind the concept of bioinorganic chemistry is to take profit of the best properties of both biological and chemical fields. On the one side, biological compounds have specific functions in nature and are fully recognized by their environment. On the other side, transition metals have special capabilities given by the presence of electrons in the *d* subshell which confers them catalytic and/or structural properties not present in the main group elements. Combining these features with the recognition of biological systems, a wide range of new regio and/or stereoselective compounds with specific functions and features have appeared in nature or can be constructed artificially.

The scope of bioinorganic chemistry is broad. Despite of its multi-disciplinarity, bioinorganic chemistry has been recognised as a relevant field of research for some important institutions. One example can be found in the Royal Society of Chemistry (RSC), who, in 2008, established a biennially award to the best researchers in the bioinorganic chemistry field, called The RSC Bioinorganic Chemistry Award.[b] The first winner of this prize was Professor Chris Orvig from the University of British Columbia, in 2009. Prof. Orvig was awarded by his contributions in understanding the behaviour of metal complexes in biological systems and for the development of new therapeutic agents. In 2011, Professor James A. Cowan, from the Ohio State University received the RSC Bioinorganic Chemistry Award for his creative research in the areas of cellular iron chemistry, novel advances in the understanding of the biological chemistry of magnesium and metalloproteases, and for the development of a new field on the research in catalytic metallodrugs. The last researcher being awarded by the RSC was Professor Tom O'Halloran, from the Northwestern University, in 2013.

[b] Further information can be found at http://www.rsc.org/ScienceAndTechnology/Awards/BioinorganicChemistry/

His studies in the role of intracellular fluctuation of transition metal availability in the regulation of cellular physiology, and the discovery of proteins involved in this process, let to Professor O'Halloran being rewarded for his research (Figure 1.2). In 2015 other talented researcher will be awarded by the Dalton Division Awards Committee but in the date which this dissertation was written the name of the winner is unknown.

Prof. Cris Orvig
Awarded in 2009

Prof. James A. Cowan
Awarded in 2011

Prof. Tom O'Halloran
Awarded in 2013

Figure 1.2: The three winners of the RSC Bioinorganic Chemistry Award from its fundation in 2008 to 2013.

1.1.2 Transition metal chemistry

Transition metals are defined by the IUPAC as elements with an incomplete *d* sub-shell, or which can give rise to cations with an incomplete *d* sub-shell.[c] Taking a look on the periodic table, these elements are comprised between the groups 3 and 11 (including the *f* block), between the ones with a complete *s* sub-shell (group 2) and the complete *d* sub-shell (group 12).

Having an incomplete *d* sub-shell causes transition metals having a rich chemistry due to their electronic configuration, specially if electrons present in the *d* sub-shell participate in the valence shell.

One of these characteristics corresponds to the wide range of oxidation states that transition metals can reach. Almost all transition metals are able of being oxidized to more than two different states, depending on the redox conditions. Also the oxidation number can vary from I (Sc) to VII (Mn) but only in rare cases (V, Mo, Fe and W) the oxidation state of metals involved in biological processes is higher than IV.[5] The oxidation or reduction of metal centres can have implications in key processes in the bioinorganic chemistry area. An example of a biological mediated redox process can be found in one of the steps of the catalytic activity of the haemoxygenase, which is the degradation of haem group to biliverdin and Fe(II) (Equation 1.1):

$$Haem + 3O_2 + 3\frac{1}{2}NADPH + 3\frac{1}{2}H^+ + 7e^- \rightarrow Biliverdin + Fe^{2+} + CO + 3\frac{1}{2}NADP^+ + 3H_2O \quad (1.1)$$

Other particularity of transition metals is how the behaviour of the metal can change depending on the position of the electrons in the different *d* orbitals: the spin configuration. In a brief, two

[c] In some textbooks, elements from group 3, as well as lanthanides and actinides, are also often considered not transition metals because their *d* electron is lost to form ionic species.[4]

major configurations can be considered: the low spin configuration,[6] where the electrons are as paired as possible, and the high spin one, where the electrons occupy all the orbitals they can. An other feature related to the arrangement of electrons in the *d* orbitals is the behaviour of the transition metal under a magnetic field. If all the electrons are paired, the metal has a diamagnetic behaviour and is repelled by an external magnetic field. In contrast, if the metal is attracted by an external applied magnetic field, that compound present paramagnetic characteristics.

The energy differences between the *d* orbitals of the metal, in conjunction with their occupancy in number of electrons and its position, can determine the final geometry of the coordination compound. In Figure 1.3 one can see the crystal field diagram of the geometries most frequently found at the active sites of metal-containing biomolecules.

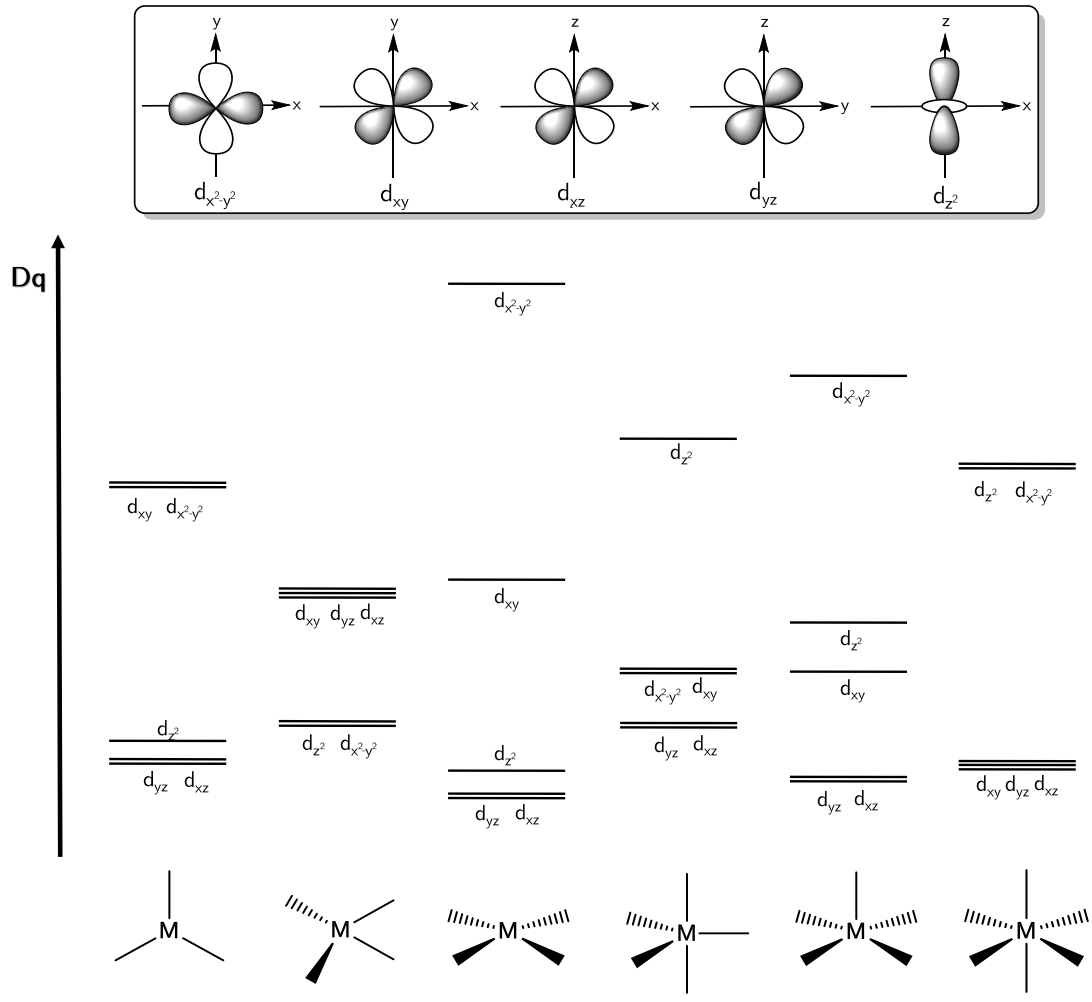


Figure 1.3: Crystal field diagrams of the most frequently geometries found in bioinorganic systems.[7,8]

The most important feature of transition metals described here is the possibility to form coordination compounds. Described by Alfred Werner in 1893,[9] these compounds are made by a central atom, normally a cation, which is interacting with other chemical entities by means of

a coordination bond, which is formed by the interaction of the orbitals of the metal with a pair of electrons of the coordinated ligand. Metal-ligand interactions have their principles in the HSAB theory (hard and soft (Lewis) acids and bases). Hard metals and ligands have a small radius, a high oxidation state and a low polarizability. Co^{+3}, Cr^{+3}, H_2O and NH_3 are examples of hard metals and ligands. In contrast, soft metals and ligands have a large radius, a low coordination state and a high polarizability. Examples of soft metals and ligands are Pt^{+2}, Ag^{+2}, CO and I^-.[4]

Combining the oxidation state of the metal, its spin configuration and the availability of ligands to coordinate the metal, researchers can have a total control of the behaviour of the system.[10,11] The most exploded application of transition metal is the organometallic catalysis field, but emerging research areas, such as biomaterials[12,13] are also taking into account all the possibilities that transition metal chemistry have for them.

1.1.3 Transition metals involved in biological processes

Metallobiomolecules are formed by the interaction of a biological host with a transition metal. In Figure 1.4, a classification of the main functions and applications is resumed.

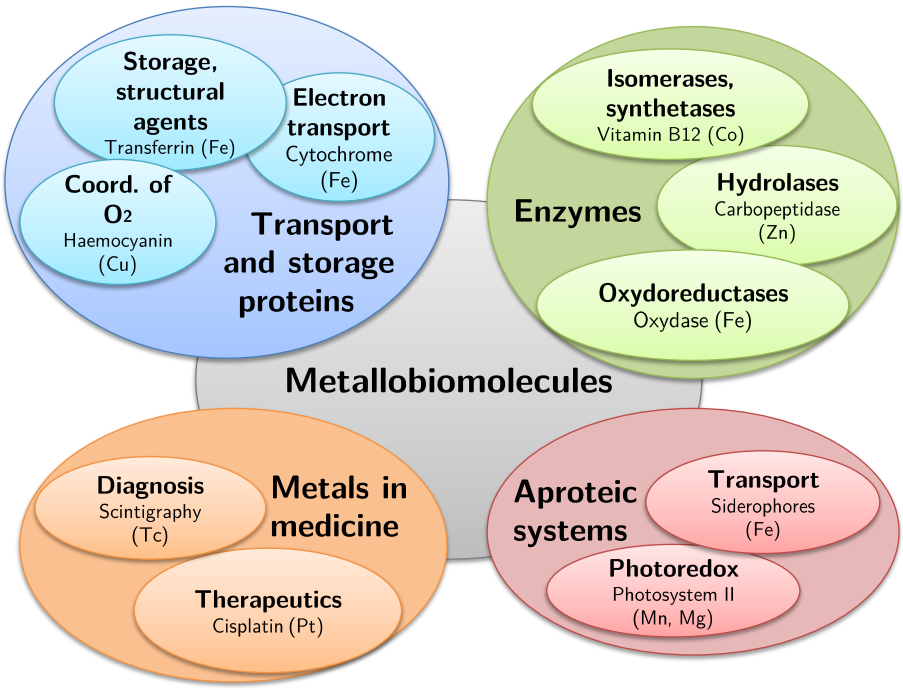


Figure 1.4: Classification of some metallobiomolecules according to its function.

Metal mediated biological processes take profit of the recognition of the biological host from the media and the structural and catalytic properties of the transition metal. Metals are able to coordinate labile external ligands, such as O_2, and carry them to the place where they are required

or storing them for a later utilization using the biological partner as a transport media. In metalloenzymes involved in biological processes, such as photosynthesis and water oxidation, the presence of the metal is mandatory to perform their activity. Additionally, the use of inorganic drugs in health has appeared as an increasing alternative due to the special structural and chemical properties conferred by the metal centre. As a difference with the limited binding geometries available for carbon atoms (linear, trigonal planar and tetrahedral), the coordination modes of transition metals includes a long range of new possibilities. For example, hexacoordinated organimetallic systems are capable to form 30 possible isomers, while tetrahedral compounds are only able to form two (Figure 1.5).

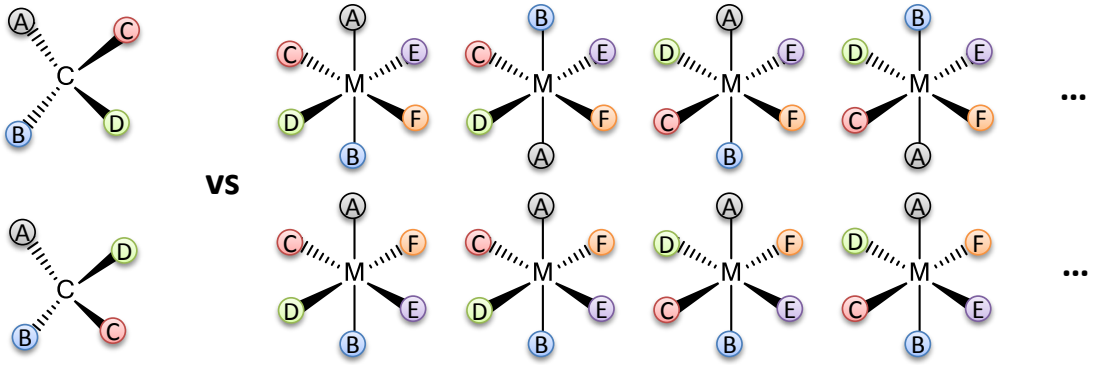


Figure 1.5: Comparative of the number of possible stereoisomers for tetrahedral carbon compounds and octahedral organometallic complexes.

In the following lines, the presence of metals in medicine and aspects of the bioaccumulation of transition metals will be discussed, as well as the chemical interaction between metals and proteins.

Metallodrugs or "metallopoisons"? Potential uses of metals in medicine

Metals, including transition metals, are required for living organisms to complete some vital activities. However, in high concentrations, all metals become hazardous to the health. Some examples of the effects of metals on human health can be shown in Table 1.1. [14]

The design of novel therapeutic and diagnostic agents is crucial for understand and treat diseases which nowadays are intractable. In this field, due to the special features of transition metals in molecular recognition processes, they are great candidates to be used with therapeutic purposes.

One of the first metallodrugs was an arsenic-based antimicrobial agent used in the treatment against syphilis. It was reported by Paul Ehrlich in 1912 under the name of Salvarsan. [15] Since then, the use of metallodrugs in medicine has been dramatically increased, being key in both diagnostic and therapeutic fields. However, for many researchers the first compound that appears joining the words "metal" and "medicine" is cisplatin (*cis*-diammine-dichloroplatinum(II)) which is still one of the most successful therapeutic drugs nowadays. [16] Moreover, metallodrugs have major potential

Table 1.1: An overview of some health injuries in humans provoked for metals.

Cadmium	Renal disfunction, Lung disease, Osteoporosis, Gastrointestinal disorder
Chromium	Damage to the nervous system, Fatigue
Copper	Anemia, Kidney and liver damage, Stomach and intestinal irritation
Lead	Development delay, Fatal infant encephalopathy, Congenital paralysis, Damage to the nervous system, Gastrointestinal damage
Manganese	Damage to central nervous system
Mercury	Tremors, Spontaneous abortion, Damage to the nervous system
Zinc	Corrosive

mainly in anticancer therapies, but also in other fields:[17]

- Antiarthritic metallodrugs: myochrysine (Au),[18], solganol (Au).[19]
- Antidiabetes metallodrugs: BMOV (V), BEOV (V).[20,21]
- Antiviral metallodrugs: CTC-96 (Co),[22] HPA-23 (W).[23]

Designing new drugs is not a procedural work. Some of them, i.e. cisplatin, are discovered by serendipity. The complete process in drug design can take more than ten years and includes numerous steps such as the drug discovery, the compound optimization, activity tests, preclinical and clinical validation and the medicament listing. The term "drug design" mainly enters on the three first tasks. Here, there are two different manners to proceed: a "trial-and-error" or having a starting hypothesis about how the new drug works, being the last one defined as "rational drug design". To succeed in the first stages of the rational drug design process, one have to take into account all the different interactions which can appear between a candidate and its host, as well as their effects in the coordination, in terms of metallodrugs, where things become a little bit complicated. The presence of the metal implies an extra complexity in the structural possibilities of the new compound.[24]

The interaction of biological targets with metals

In drug design and other related fields, the number of variables to take into account to have an accurate description of the chemical process under study is substantial. When dealing with metallodrugs, one of these variables is related to the effect of the binding strength between the metal and the biological partner and its relationship with the conformational flexibility of the ligand. In other words, the effect of transition metals in the structure of its biological partner.[25] The special features of transition metals provides them several ways to interact with biological systems. As biological targets, amino acids present in peptides and proteins, and nitrogenous bases forming DNA and RNA are susceptible to coordinate metallic cations (Figure 1.6).

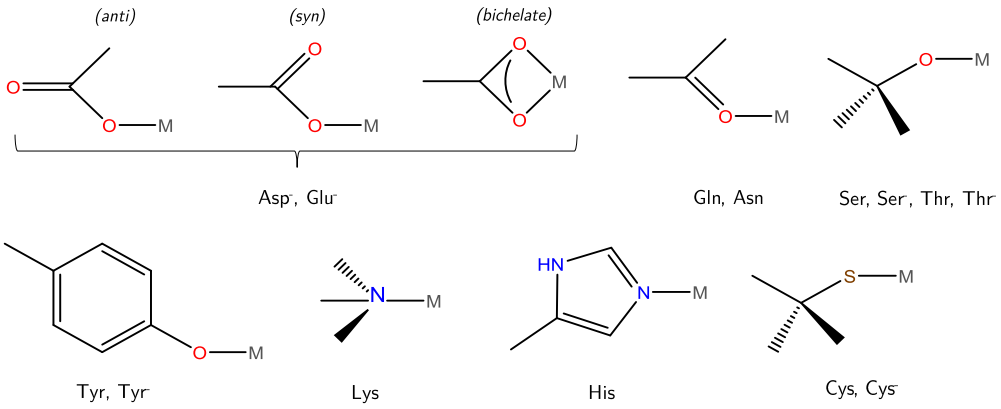


Figure 1.6: Schematic representation of the binding sites of metals in selected amino acids

Metals can interact primarily with the biological counterparts in two ways (Figure 1.7). The first one is a direct coordination of the metal to amino acids side chains or nucleobases from the DNA. The metal geometry is constrained by the pre-fixed orientation of the residues placed in the binding site. In most cases the first coordination sphere of the metal is completed by solvent molecules. In the second interaction scheme, the metal interacts with the protein accompanied with an organometallic cofactor, being the metal able to coordinate to the biological host by its vacant sites.

Pre-fixed residue orientation

Pre-fixed metal geometry

Figure 1.7: Examples of pre-organized binding site (left) and pre-fixed geometry of the metal (right) in X-ray structres (Superoxyde Dismutase: 2SOD, and single-domain globin: 2WY4), respectivelly.

Furthermore, the electronic and structural properties of transition metals can affect the binding to the biological partner. One example is the oxidation-reduction processes performed by many metalloenzymes. Changing the population of the *d* subshell of the metal, the most stable geometry of the system can change. This effect can provoke the release of a ligand and the creation of a free vacant site accessible for a new substituent. [26]

Transition metals have important effects in biology, being key in a huge variety of biological processes due to its structural and electronic properties which main group elements are not able to afford. However, in metallodrug design, artificial metalloenzyme synthesis and related fields, the

prediction of the effect of the metal on the biological host is a complex multivariable problem and is not an easy task to resolve. Even though the efforts performed in the synthetic area, experimentalist are not capable to figure out all the possibilities that transition metals confers to biological partners. Molecular modelling techniques have a great advantage at this stage and could represent an interesting alternative.[27]

1.2 Molecular Modelling in Bioinorganic Chemistry

By definition, molecular modelling includes all the computational techniques and theoretical methods used to mimic the behaviour of molecules on its media at atomistic scale. Its application encompasses a wide range of chemical and biological systems: from the smallest ones, which only accounts a couple of atoms, to the larger biological or material compounds. Computational chemistry, drug design, computational biology and materials science are some examples of the fields where molecular modelling can be applied. In this dissertation we focus at the field of computational bioinorganics.

Almost a century have elapsed since the first mathematical approximation to chemistry were performed. D. H. Andrews[28] and Hartree[29] put the foundation stone in the fields of molecular mechanics and quantum mechanics areas, respectively. In different works, they described the same natural process using two different techniques, or what it is the same, two different models.

1.2.1 Does a model represent the reality?

What is a model? What it represents? "Model" is not an easy word to describe. The vast definition of what is a model is: "a model is the representation of the reality". But now the question is, how far one has to go in order to get a good representation of the reality? To answer this question, a visual example is given.

Three different representations of the same reality (the author of the manuscript) are shown in Figure 1.8. In model A there is a stickman which can represent that this manuscript has been written by a human.[d] In model B it is more clear that the manuscript is written by a woman, with certain physical attributes as her height and weight. Finally, in model C the face of the author can be seen being fully recognizable. All three models are correct, but choosing one or other only depends on the accuracy needed (or affordable).

1.2.2 An overview on bioinorganic modelling

The concept of model previously described also applies in molecular modelling. Depending on the employed level of theory, different information can be obtained as a result. In Figure 1.9, a representation of the different levels of theory which can be applied depending on the accuracy, size of the

[d] Indeed, in software engineering the stickman represents the user of the program. There, the stickman has the name of "actor"

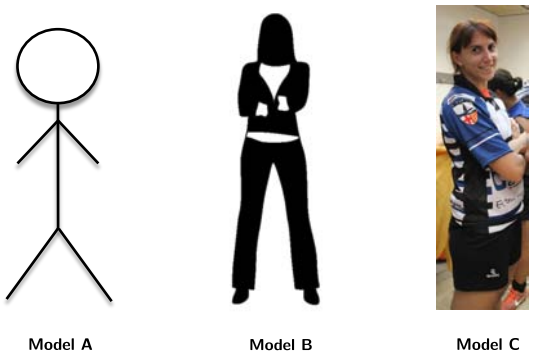
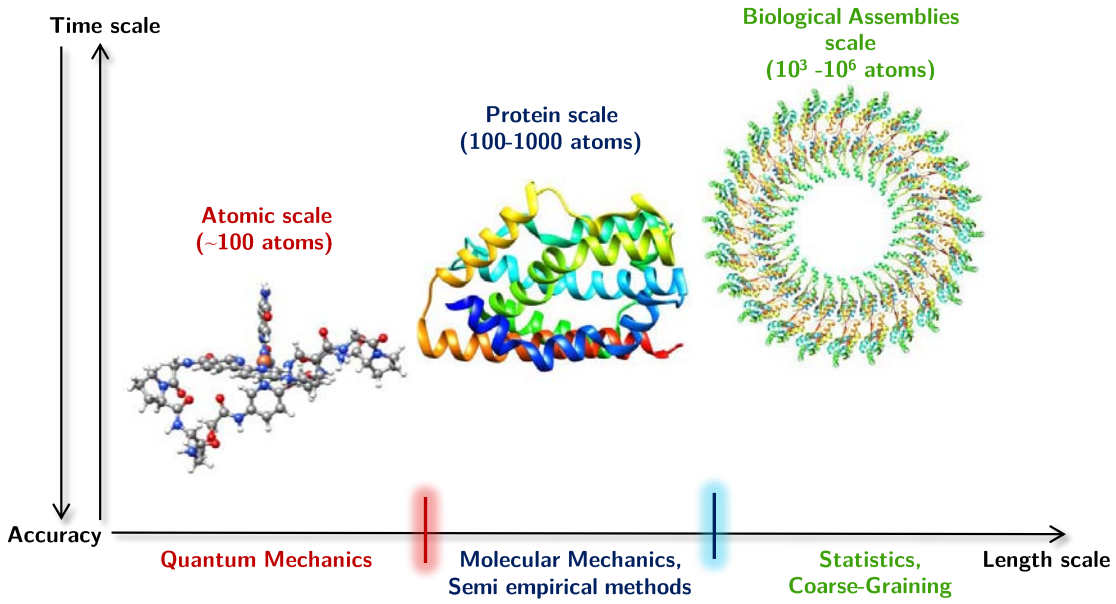


Figure 1.8: Three different models of the author of this manuscript.

system and time is shown. Simulations performed at low level of theory (i.e. molecular mechanics) will give global results for big system, or quick exploration for small compounds, even for vast ensembles. On the other hand, simulations performed at higher level (i.e. quantum mechanics) yields to a more detailed outcome at higher computational cost.



the

Figure 1.9: Schematic representation of the different levels of theory and their correspondence with the size of the system under study.

In a vast number of cases, one can choose easily the appropriate model only taking into account the scale where their studied systems have to be fitted, and then, designate the correct tool to achieve the desired results. In bioinorganic chemistry this "simple" task can be more complicated than expected merely by the presence of a metal.

In precedent sections, the peculiarities of transition metals have been described (oxidation state, change of coordination, etc.). But in fact, these features are one of the bottlenecks of modelling bioinorganic systems. For example, the flexibility of the coordination bonds of the metal, which, on the one hand complicates the parametrization of the system (if one is at molecular mechanics level of theory) and, on the other, increases the computational cost of the calculation (at quantum mechanics level). Normally, these pitfalls are overcome in the "quick-way", adding the needed parameters (in Molecular Mechanics) or spending a little bit more of time in the calculation (if Quantum Mechanics is applied).

A problem of size

In computational bioinorganic chemistry we deal with big systems. Combining the size of the complex and the presence of metals which can participate or not in the activity of its biological hosts, the application of multiscale approaches is in a vast variety of cases cases a good choice. Combining different levels of theory in a single system can help the researcher to have an overall point of view of the complex, without missing any important information about the behaviour of the metal. One example of the application of multiscale approaches is in the study of metalloenzymes. These systems have to be modelled combining quantum mechanics, to describe the catalytic process, and molecular mechanics, to account with the rest of the system. Multiscale approaches are well accepted by the scientific community, being M. Karplus, M. Levitt and A. Warshel awarded with the Nobel prize in Chemistry in 2013 for the development of multiscale methods for complex chemical systems.

The impact of the water

From the cellular media to the blood stream, metallodrugs are feasible to interact with water molecules.[30] Having an accurate description of the solvent is crucial, particularly in the study of the binding process or in the metalloenzyme field. Due to the abundant presence of water in the biological field, it is important not to neglect its structural and electrostatic effects. On the one hand, structural effects can be seen in two major ways: by a direct coordination to the metal and bridging protein and ligand in the binding site (Figure 1.10).

The coordination of a water molecule directly to the metal stabilizes the system fulfilling its first coordination sphere. In fact, water molecules can also displace a coordinated ligand, if it is labile enough. Otherwise, when the solvent acts as a bridge between the biological host and the drug it is important to take into account the placement of solvent molecules during the simulation. However, it is not easy to detect the presence of solvent in the binding pocket by means of X-ray spectroscopy, and sometimes the addition of water molecules in the correct place is an assigned task to the computational researcher. This assignment can be done using the previous knowledge of the investigator or by computational means either by using molecular dynamics approaches[31] or using a water model, such as TIP3P,[32] normally followed by a molecular or quantum mechanics

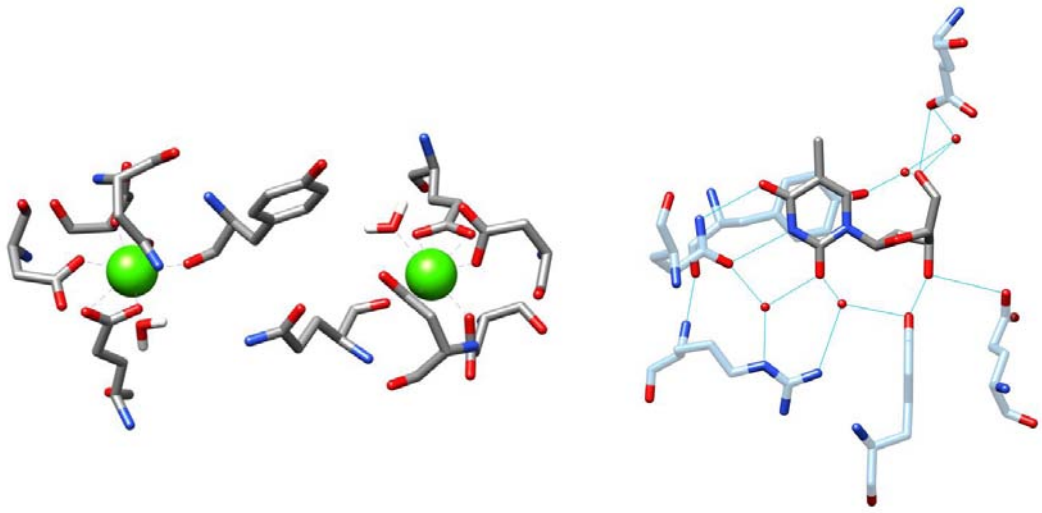


Figure 1.10: Bound (left, PDB code: 1CLL) and brigde (right, PDB code: 1KIM) representation of the effects of water in protein-ligand interactions.

geometry optimization.

Other procedure to mimic the effects of the solvent is using an implicit representation of it, were the solvent is simulated by a continuum polarizable environment with its corresponding value of the dielectric constant (ε). This method is faster and simplest than adding explicit solvent molecules and is good enough to reproduce the bulk effects of the system. From all the available continuum solvent model algorithms, Cosmo was the elected in this manuscript. [33–35] This continuum approach generates a conducting polygonal surface around the system at the van der Waals distance. Other application of the continuum solvent model used in this manuscript is to simulate a heterogeneous environment, such as the blood stream, the cell plasma and the inner cavity of a protein. [36]

In some cases it can be useful to mix both explicit and implicit models, being some solvent molecules placed in the region of the system where the researcher expects solute-solvent interactions and treating the rest of the system with a continuum model.

Active and inert first coodination sphere

The prediction of the binding of organometallic compounds to biological entities is one of the recipes of molecular modelling cookbooks. Indeed, to tackle the study of a new bioinorganic systems, molecular modellers have to first identify the complementarity between the chemical and the biological partner. Sometimes this preliminary work is not straightforward due to the complexity to explore intensively the conformational space.

To encompass the analysis of the binding processes, it can be differentiated in two major fashions,

depending on the implication of the first and second coordination sphere of the inorganic part. When binding occurs by means of a direct coordination of the metal to the biological host, changes in the electronic structure of the metal or in the coordinating amino acid can affect the entire system. In other words, the metal has an active first coordination sphere. The second kind of binding processes refers to those where the metal is not directly involved in the coordination to the host and only has a structural role holding the coordinating ligands, which are the ones involved in the binding. In this case one can say that the metal has an inert first coordination sphere because its properties do not change during the binding process. In Figure 1.11 there is a visual representation of both kinds of interactions.

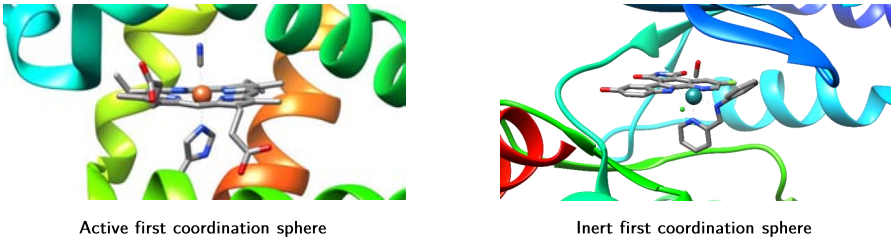


Figure 1.11: Representation of active first coordination sphere (left, PDB code: 1WZD) and inert first coordination sphere processes (right, PDB code: 3FXZ).

To account long range and short range effects of the interaction of metals with its biological partners, multiscale procedures have to be applied to cover the areas where each computational level of theory fails. In molecular modelling, a wide representation of the metal is given by a single simulation. The pitfall here is the description of the metal and its interaction with the coordinated ligands. However, it is possible to find the correct parameters for a given molecule by literature search or using quantum mechanical methods,[37] but it is still not able to accurately describe the electronic effects of the metal.

In contrast, quantum mechanical methods are capable to perform a precise prediction of the electronic effects of the metal. The inconvenient with this level of theory is the impossibility to take into account the entire system due to the high computational cost that it implies. Multiscale methods combine the accuracy of quantum mechanical simulations with the cheap cost of molecular modelling techniques. In this dissertation the most used multiscale method is the hybrid quantum mechanics/molecular mechanics (QM/MM) technique, which describes the first coordination sphere of the metal at quantum mechanics level of theory and includes also a molecular mechanics description of the rest of the system, in order to consider the complete system in the simulation process. A better description of all the computational tools mentioned here and others which are stored the computational chemist toolbox is given in the Materials and methods chapter.

1.3 References

[1] Fu, J.; Zhou, Q.; Liu, J.; Liu, W.; Wang, T.; Zhang, Q.; Jiang, G. *Chemosphere* **2008**, *71*, 1269–1275.

[2] Wongsasuluk, P.; Chotpantarat, S.; Siriwong, W.; Robson, M. *Environ. Geochem. Hlth.* **2014**, *36*, 169–182.

[3] Alberts B, L. J., Johnson A *Molecular Biology of the Cell. 4th edition*; Garland Science, New York, 2002.

[4] Rayner-Canham, G.; Overton, T. *Descriptive inorganic chemistry*; Macmillan, 2003.

[5] Holm, R. H.; Kennepohl, P.; Solomon, E. I. *Chem. Rev.* **1996**, *96*, 2239–2314.

[6] Cirera, J.; Ruiz, E.; Alvarez, S. *Inorg. Chem.* **2008**, *47*, 2871–2889.

[7] Companion, A. L.; Komarynsky, M. *J. Chem. Educ.* **1964**, *41*, 257.

[8] Zuckerman, J. *J. Chem. Educ.* **1965**, *42*, 315.

[9] Werner, A. *Z. Anorg. Chem.* **1893**, *3*, 267–330.

[10] Swart, M. *Inorg. Chim. Acta* **2007**, *360*, 179–189.

[11] Sun, Y.; Kim, Y.-H.; Zhang, S. *J. Am. Chem. Soc.* **2007**, *129*, 12606–12607.

[12] Dudev, T.; Lim, C. *Chem. Rev.* **2014**, *114*, 538–556.

[13] Chen, Q.; Thouas, G. A. *Mat. Sci. Eng. R.* **2015**, *87*, 1–57.

[14] Singh, R.; Gautam, N.; Mishra, A.; Gupta, R. *Indian J. Pharmacol.* **2011**, *43*, 246.

[15] Ehrlich, P.; Bertheim, A. *Ber. Dtsch. Chem. Ges.* **1912**, *45*, 756–766.

[16] Alderden, R. A.; Hall, M. D.; Hambley, T. W. *J. Chem. Educ.* **2006**, *83*, 728–734.

[17] Mjos, K. D.; Orvig, C. *Chem. Rev.* **2014**, *114*, 4540–4563.

[18] Bau, R. *J. Am. Chem. Soc.* **1998**, *120*, 9380–9381.

[19] Mazid, M. A.; Razi, M. T.; Sadler, P. J.; Greaves, G. N.; Gurman, S. J.; Koch, M. H.; Phillips, J. C. *J. Chem. Soc. Chem. Comm.* **1980**, 1261–1263.

[20] Sun, Y.; James, B. R.; Rettig, S. J.; Orvig, C. *Inorg. Chem.* **1996**, *35*, 1667–1673.

[21] Thompson, K. H.; Lichter, J.; LeBel, C.; Scaife, M. C.; McNeill, J. H.; Orvig, C. *J. Inorg. Biochem.* **2009**, *103*, 554–558.

[22] Schwartz, J. A.; Lium, E. K.; Silverstein, S. J. *J. Virol.* **2001**, *75*, 4117–4128.

[23] Dormont, D.; Spire, B.; Barre-Sinoussi, F.; Montagnier, L.; Chermann, J. *Ann. Inst. Pasteur Vir.* **1985**, *136*, 75–83.

[24] Hocharoen, L.; Cowan, J. A. *Chem.-Eur. J.* **2009**, *15*, 8670–8676.

[25] Cisnetti, F.; Maréchal, J.-D.; Nicaise, M.; Guillot, R.; Desmadril, M.; Lambert, F.; Policar, C. *Eur. J. Inorg. Chem.* **2012**, *2012*, 3308–3319.

[26] Ortega-Carrasco, E.; Lledós, A.; Maréchal, J.-D. *J. R. Soc. Interface* **2014**, *11*, 20140090.

[27] Comba, P.; Hambley, T. W.; Martin, B. *Molecular modeling of inorganic compounds*; John Wiley & Sons, 2009.

[28] Andrews, D. H. *Phys. Rev.* **1930**, *36*, 544.

[29] Hartree, D. R. *Math. Proc. Cambridge*; 1928; Vol. 24, pp 89–110.

[30] Trachtman, M.; Markham, G. D.; Glusker, J. P.; George, P.; Bock, C. W. *Inorg. Chem.* **1998**, *37*, 4421–4431.

[31] Bizzarri, A. R.; Cannistraro, S. *J. Phys. Chem. B* **2002**, *106*, 6617–6633.

[32] Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926–935.

[33] Pascual-Ahuir, J.; Silla, E.; Tomasi, J.; Bonaccorsi, R. *J. Comput. Chem.* **1987**, *8*, 778–787.

[34] Klamt, A.; Schüürmann, G. *J. Chem. Soc.* **1993**, 799–805.

[35] Klamt, A. *J. Phys. Chem.-US* **1995**, *99*, 2224–2235.

[36] Ortega-Carrasco, E.; Cossío, F. P.; Lledós, A.; Maréchal, J.-D. *J. Inorg. Biochem.* **2012**, *117*, 230–236.

[37] Seminario, J. M. *Int. J. Quantum Chem.* **1996**, *60*, 1271–1277.

2 ... Objectives

Insanity is doing the same thing again over again and expecting different results.

Albert Einstein

The present thesis is centred in the description of the different interactions existing between transition metals and biological partners, using molecular modelling techniques. To do so, several aspects have to be taken into account: from the characteristics of the system under study to the affordable technical level one could reach. To provide the necessary molecular information needed, an inspection of the structural characteristics of the binding between the organometallic ligand and the host has to be performed, taking into account:

- Inert biometallic binding.
- Active biometallic binding.
- Long range effects of the binding of an organometallic compound to a biological host.
- The balance between the first and second coordination sphere processes.

To do so, a set of known molecular modelling methods have been applied: ones following the procedures shown in the bibliography, being others evaluated on their limits. The choice of one or other technique at each stage not only depends on the system where it has to be applied, but also the chemical interaction is an important variable to take into account.

In detail, each chapter aims to accomplish the following challenges:

- **Chapter 4**. Protein-ligand docking with inert first coordination sphere: the case study of Ruthenium based kinase inhibitors.
  - Assessing protein-ligand docking methods for the prediction of the binding of organometallic ligands to proteins, with an inert first coordination sphere. Calibration of docking methods to reproduce reported X-ray structures of ruthenium based kinase inhibitors.
- **Chapter 5**. Active first coordination sphere: resting state and activation process of the *Corynebacterium diphteriae* Haem Oxygenase.
  - Describe the electronic state of the transition metal in the resting state of an artificial metalloenzyme.
  - Identify the local and global changes needed to activate the metalloenzyme.
- **Chapter 6**. Novel insights on the molecular interactions of cisplatin on DNA and proteins.
  - Identification of the local effects of the binding of cisplatin to DNA.
  - Translate these local effects to the overall structure in order to identify their global implications.
  - Study the possibility of tri-coordinated cisplatin-protein adducts in solvent.

- **Chapter 7**. Enantioseletive formation of artificial metallopeptides.
  - Describe the effect of the configuration of the stereogenic centre of chiral residues present in the peptide sequence of a metallopeptide and its consequences in the organization of the first coordination sphere of the metal.

3 ··· Materials and Methods

$\uparrow\uparrow\downarrow\downarrow\leftarrow\rightarrow\leftarrow\rightarrow B\ A$

The Konami Code

3.1 Computational methods

Since the late 50's,[a] scientists and engineers did not start to use computers to perform their day-by-day tasks, evolving their research by means of computer programming, teaching to a "box" to solve tedious equations which can not be easily done by hand. In chemistry, researchers did not hesitate in applying their theories to computer science, being one researcher of the MIT, in 1956, the first one to perform *ab-initio* calculations.[1] Since then, computational chemistry has evolved to an unpredicted level, hand by hand with the advances in technology.

Computational chemistry can be split in two major areas: quantum chemistry and molecular mechanics. The main differences between both are the representation of the main components of the energetic form, and the scale of the systems under study. For large atomic scale, classical physics based methods (molecular mechanics) are enough to describe the position of the atoms and the interplay between them by means of non bonded interactions. Going deeper to the electron scale, taking into account the nuclei and the electrons, implies upgrade the level of theory to quantum mechanics.

In the following section, a general idea of quantum mechanics, molecular mechanics and hybrid methodologies will be given, as well as a brief explanation of the computational techniques applied in the results chapter of this manuscript.[b]

3.1.1 Quantum Chemistry

The objective of quantum chemistry is to interpret and predict the electronic structure and the reactivity of chemical systems by means of quantum mechanics principles. Nowadays, there are two major quantum chemistry schemes: the wave function-based methods and the density functional theory.

Wave function-based methods

Wave function-based methods use the Schrödinger Equation without any experimental parameter.[c] In Equation 3.1, the simplest representation of the time-independent Schrödinger equation is written. $\hat{H}$ represents the Hamiltonian operator, Ψ the wave function which describes the system under study and E the energy value associated to the current state of the system.

$$\hat{H}\Psi = E\Psi \tag{3.1}$$

[a] The first digital computer fully functional, ENIAC, was constructed in 1946 by J.P. Eckert and J. Mauchly

[b] Some of the techniques can be enclosed in both quantum and molecular mechanics fields (i.e. Potential Energy Surfaces). The placement of the technique depends exclusively on how it was applied in this manuscript.

[c] That is way they are also called *ab-initio* (from the beginning).

Wave function-based methods are widely used, and contains two main computational methodologies.

In a brief, these methods are:

- **The Hartree-Fock approximation:** provides a simple description of electrons occupying orbitals. This method was initially described by Hartree in 1948 and latter improved by Fock. To solve the Schödinger equation, the coordinates of the electrons have to be separable. The Coulomb repulsion term does not fulfil this requirement because of its $\frac{1}{r_{1,2}}$ term. For multi-electron systems, this part is approximated to only depend on the coordinates of a single electron. The resulting one-electron wavefunctions (φ), after being applied to a multi-electron Hamiltonian (or Fock operator, $\hat{F}$) will give the lowest energy of a multi-electron system. These wavefunctions are called Hartree-Fock wavefunctions. Hartree-Fock equations (Equation 3.2) are Schödinger-like equations able to operate with the mentioned Hartree-Fock wavefunctions.

$$\hat{F}\varphi_i = \epsilon_i \varphi_i \tag{3.2}$$

  The complete form of the Hartree-Fock equation can be seen in Equation 3.3.

$$\hat{F} = \hat{H}^0 + \sum_{j=1}^{N}(2\hat{J}_j - \hat{K}_j) = -\frac{\hbar^2}{2m}\nabla^2 - \frac{Ze^2}{4\pi\epsilon_0 r} + \sum_{j=1}^{N}(2\hat{J}_j - \hat{K}_j) \tag{3.3}$$

  $\hat{H}^0$ includes the kinetic and potential energy of a single electron. The second term, $\sum_{j=1}^{N}(2\hat{J}_j - \hat{K}_j)$ accounts for the potential energy of one electron in an average field created by all the other electrons in the system. Separately, $\hat{J}$ is called Coulomb operator, and operates on a function of one electron in an orbital. The other term, $\hat{K}$, is the exchange operator. The electron is exchanged between the two orbitals i and j. This overlap contribution to the charge density and potential energy is a quantum mechanical effect. It is a consequence of the wave-like properties of electrons.

- **Post Hartree-Fock methods:** adds electron correlation to the original Hartree-Fock approximation by means of the calculation of the almost "exact" energy of the system. It includes three categories which differs in how this "exact" energy is calculated: Configuration Interaction, Coupled Cluster and Many Body Perturbation Theory.

The use of one or the other approach depends on the system,[d] the needed accuracy and the available computational resources, being Post Hartree-Fock methodologies clearly far more computer demanding than Hartree-Fock.

[d] The Hartree-Fock approximation completely fails in the calculation of dissociation process.

Density Functional Theory

Density Functional Theory (DFT) represents an alternative to Post Hartree-Fock methods, including the effects of the electron correlation into the solution of the electronic Schrödinger equation. DFT relies on the premise that the electronic properties of a molecular system can be determined by the electronic density. Kohn and Hohenberg demonstrate the veracity of this hypothesis in 1964, showing that the electron density determines the Hamiltonian. This fact makes possible to calculate any observable of a non-degraded ground state as a functional of the electronic density.[2] Equation 3.4 shows the effect of the electronic density of a system (ρ) associated to a certain external potential (υ) field on the total energy of the system. The $F[\rho_0]$ parameter (also known as universal functional) includes the kinetic energy and the electron-electron repulsion of a given system with an electron density defined by ρ_0 (Equation 3.5). This value is independent of the external potential field.

$$E = E_\upsilon[\rho_0] = \int \rho_0(r)\upsilon(r)dr + F[\rho_0] \tag{3.4}$$

where,

$$F[\rho_0] =< T[\rho_0] > + < V_{ee}[\rho_0] > \tag{3.5}$$

However, $F[\rho_0]$ can not be solved exactly. The dissociated form of Equation 3.5 is:

$$F\rho_0 = T[\rho_0] + V_{ee}[\rho_0] = T_s[\rho_0] + V_{ee}[\rho_0] + J[\rho_0] + E_{xc}[\rho_0] \tag{3.6}$$

where, T_s is the kinetic energy of a non-interacting reference system (T is unknown), J is the classical Coulomb interaction (can be exactly calculated), and E_{xc} describes the electron exchange and correlation. This last term can be split in E_x and E_c, and is the unique approximated term on the density functional. Exchange term, E_x, includes the effect between identical particles, in other words, the interaction of two electrons with the same spin. Correlation term, E_c represents the repulsion between electrons.

The evolution of the Density Functional Theory relies on the research of the most appropriate description of the E_{xc}, being from the beginning until nowadays one of the most active topics in computational chemistry. Depending on the applied methodology to describe the E_{xc} term, functionals can be classified in several fields. In complexity order (starting from the less complex), these are:

- **Local Density Approximation (LDA) functionals:** uses the approximation of the uniform electron gas for the local density at each point of the space. The term E_x is analytically know, but E_c has to be estimated. Some examples of LDA functionals are: VWN,[3] PZ[4] and CP.[5]
- **General Gradient Approximation (GGA) functionals:** uses the gradient of the density

to take into account the heterogeneity of the true local electron density. These functionals have the E_x and E_c parts fully combinable, and can be divided in exchange and correlation functionals. Some examples these functionals are: B,[6] PBE[7] (exchange); LYP,[8] P86[9] (correlation); and PW91[10] (both).

- **Meta-Generalized Gradient Approximation (meta-GGA) functionals:** incorporates an explicit dependence on kinetic energy density in the exchange term. Examples: M06-L[11] and TPSS.[12]

- **Hybrid functionals (hyper-GGA):** include some contribution of the exact exchange energy calculated by the Hartree-Fock method mixed with the DFT exchange-correlation. Examples of hybrid functionals are: B3LYP,[13] PBE0[14] and M06.[15]

Quantum Mechanics based computational tools

In this part, an introduction on the quantum-based computational tools used in the research described in this dissertation is given.

I) Potential Energy Surface (PES)

Based on the fact that for a given structure and electronic state of a molecule, it has a specific energy. The representation of the energy in front of the coordinates of the molecules in a particular electronic state receives the name of potential energy surface.

The relationship between structure and energy relies in the Born-Oppenheimer approximation, which discriminates the motion of the nuclei in front of the motion of the electrons and allows us calculate the energy of the molecule taking into account only the electronic effects.

In theory, the energy of every geometry that could be adopted by the system can be analysed in a potential energy surface. Nonetheless, the amount of possible conformations that can exist makes it an infeasible task, being almost impossible calculate the potential energy surface for the 3N-6 dimensions[e] of the system. The general use of the potential energy surfaces is to study reactivity processes, and its explanation will be focused on the description of the reaction PES.

The PES is formed by a tridimensional representation of the system, being two coordinates the ones that describes the interatomic distance between the involved atoms, and the last for the energy (Figure 3.1). To make the PES analysis feasible, only the key stationary points are taken into account: the minima and the saddle points. From a chemical point of view, minima correspond to reactants, products and intermediates, whereas saddle points represents the transition state structures and the 2^{nd} order saddle points.

From this plot one can obtain the energy values to describe the reaction process. The energy difference between the product and the reactant valley corresponds to the energy of the reaction. Additionally, the energy difference between the transition state and the reactant is the energy barrier of the reaction. The thermal corrections and the zero-point energies needed to calculate the

[e] N is the number of atoms

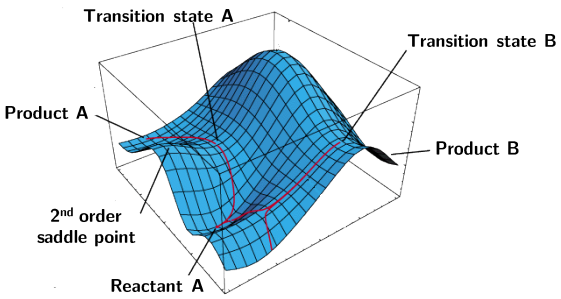


Figure 3.1: Schematic representation of a potential energy surface.

free energy differences can be found by the vibrational motion of the molecule about the reactant and product minima.

II) Geometry optimization procedure

Geometry optimization is the process which the structure of the molecule is adjusted until a stationary point on the potential energy surface is found. The relationship between energy and coordinates is described by the energy derivatives (Equation 3.7).

$$E(\mathbf{x}) = E(\mathbf{x}_0) + \mathbf{g}_0^T \Delta\mathbf{x} + 1/2\Delta\mathbf{x}^T \mathbf{H}_0 \Delta\mathbf{x} \tag{3.7}$$

$\mathbf{g}_0$ is the gradient (dE/dx) and $\mathbf{H}_0$ is the Hessian matrix (d^2/dx^2). In the stationary points of the potential energy surface, the value of the gradient is zero. To discriminate between a minima or a saddle point, one has to take a look into the eigenvalues of the Hessian matrix: if all of them are positive the system is placed on a minima, if only one is negative the system corresponds to a transition state, in other cases, *n* negative eigenvalues, the system is placed in a *n*-th order saddle point. [16]

A general geometry optimization process is described in Figure 3.2. For a given system with a given geometry and a guessed Hessian, the energy and the gradient of the molecule is calculated, and then, compared to a threshold value. If the obtained values at the gradient and/or displacement are lower than the threshold, the minimization finishes. If the threshold is overcame, depending on the applied minimization algorithm, an update of the Hessian or a minimization along the gradient direction is performed, followed by the calculation of the next minimization step and an update of the geometry. At this point, the procedure is repeated until convergence or other criteria, such as reaching the maximum number of minimization cycles.

There are dozens of geometry optimization algorithms available in computational chemistry software packages. Only to enumerate it, and without entering in a deep description, two major minimization algorithms can be found in the most known quantum chemistry suites: the Broyden-Fletcher-Goldfarb-Shanno (BFGS) [17] one, implemented in ADF [18] and ORCA, [19] and the Berny optimization procedure [20] which can be used by default in Gaussian. [21]

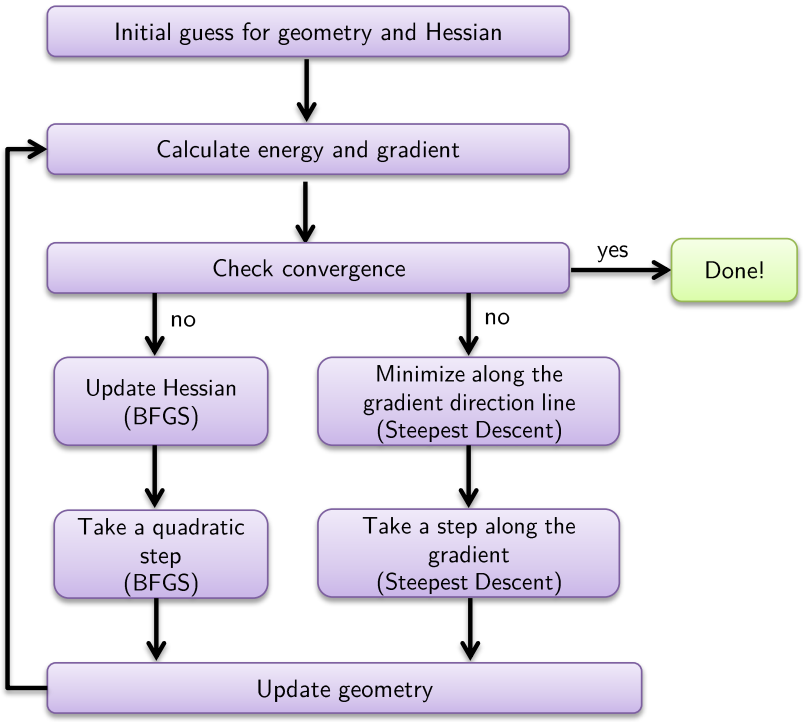


Figure 3.2: Geometry optimization flowchart.

III) Energy Decomposition Analysis (EDA)

Performing an Energy Decomposition Analysis is useful to describe the chemical bonding in molecules. This technique allows the researcher splitting the system in fragments (or partitions) and calculate the energy interaction between them. EDA was initially described by Morokuma [22] and Bickelhaupt [23] and implemented in different computational chemistry suites, such as ADF. [18]

The total bond energy term, ΔE is divided in two components (Equation 3.8). The first one, ΔE_{prep}, corresponds to the energy required to deform both fragments from their isolated equilibrium structure to the geometry they acquire in the final molecule, in both geometrical ($\Delta E_{prep,geo}$) and electronic ($\Delta E_{prep,el}$) manners. The second term, ΔE_{int}, takes into account the interaction energy between both fragments. This last term can be decomposed in three parts: ΔV_{elst}, ΔE_{pauli} and ΔE_{io}. The first one, ΔV_{elst} describes the electrostatic interaction between the unperturbed charge distributions of both fragments, as a sum of their electron densities ($\rho_A + \rho_B$). This term is normally positive for uncharged systems. The second term, ΔE_{pauli}, is also known as "steric" interaction, comprises the destabilizing interactions between occupied orbitals. The last component, ΔE_{io}, refers to the orbital-interaction energy, taking into account the electron pair bonding, charge transfer and polarization.

$$\Delta E = \Delta E_{prep} + \Delta E_{int} = \Delta V_{elst} + \Delta E_{pauli} + \Delta E_{io} \tag{3.8}$$

It is important to take into account that the ΔE_{prep} is generally not accounted in the Energy Decomposition Analysis output file. The researcher has to calculate it by a simple subtraction of the energy of each fragment to the total energy of the system ($E_{complex} - E_A - E_B$), something that can be sometimes a tedious exercise if one of the fragments is an unstable specie.

3.1.2 Molecular Mechanics

Molecular Mechanics encompasses the representation of the chemical systems using classical physics principles. Here, atoms are described as a single particle entity with a given radius and charge, bonds approximated to springs, and molecules are defined as balls on springs following the fundamentals of the Hooke's Law. In molecular mechanics, the potential energy is calculated using force fields, which contains all the parameters needed for the correct description of the system. The potential energy equation is composed by the bonded terms (see Figure 3.3) and the non-bonded ones.

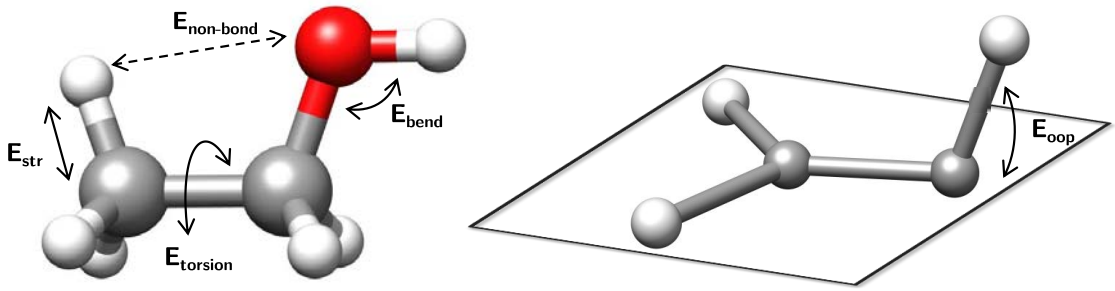


Figure 3.3: Geometrical energy terms of a force field.

Bonding terms encompasses all the interactions between the atoms that are directly linked: bond-stretching (stretch), bond-bending (bend), torsion motions (torsion) and out-of-plane potentials (out-of-plane). In contrary, the non-boded interactions refers on the effect of the environment: van der Waals potential (vdW) (usually taken to be Lennard-Jones 12-6 type), the Coulomb (or electrostatic) interactions, hydrogen bonds and cross terms, being the last two only taken into account in some force fields. A standard mathematical representation of a force field can be seen in Equations 3.9 and 3.10.

$$U = \sum_{stretch} U_{AB} + \sum_{bend} U_{ABC} + \sum_{torsion} U_{ABCD} + \sum_{out-of-plane} U_{ABCD} + \sum_{vdW} U_{AB} + \sum_{Coulomb} U_{AB} \quad (3.9)$$

To have a correct description of the system, these contributions are composed by the mathematical relationship between the different parameters which describes the system under study (Equation

3.10).

$$U = \sum_{stretch} \frac{1}{2} k_{AB}(R_{AB} - R_{e,AB})^2 + \sum_{bend} \frac{1}{2} k_{ABC}(\theta_{ABC} - \theta_{e,ABC})^2 + \sum_{torsion} \frac{U_0}{2}(1 - cos(n(\chi - \chi_0))) + \sum_{out-of-plane} \frac{k}{2sin^2\psi_e}(cos\psi - cos\psi_e)^2 + \sum_{vdW}(\frac{C_{AB}^{12}}{R_{AB}^{12}} - \frac{C_{AB}^{6}}{R_{AB}^{6}}) + \frac{1}{4\pi\epsilon_0} \sum_{charges} \frac{Q_A Q_B}{R_{AB}} \quad (3.10)$$

There is a huge number of different force fields available nowadays, being ones widely used, and others only employed in special cases. In the specific case of this dissertation, only the AMBER (Assisted Model Building with Energy Refinement) force field [24] was employed one due to its inclusion on the software used in this dissertation.

Whatever force field, the major limitation is that the change of electronic character along the simulation is not taken into account. This is a huge pitfall for studying transition metal systems involved in biological processes and particularly the changes in coordination, oxidation and spin state that can occur even during a single breathing motion or a dehydration process. However, some approaches have been proposed to feed the force field with acceptable metal-ligand parameters. One example is the Ligand Field Molecular Mechanics (LFMM). [25] That method describes the metal-ligand interactions using the crystal field theory. A second example is the use of quantum mechanics calculations to create the parameter of the metal. This can be done using the FUERZA code implemented by Seminario and coworkers. [26]

Molecular mechanics based computational tools

Molecular mechanics tools are computationally inexpensive, compared to the Quantum Mechanics based ones, and are able to deal with systems composed by a vast number of atoms. This feature makes molecular mechanics an indispensable tool for virtual screening or drug design, being the first choice in industry. [27] Some molecular mechanics based methodologies will be described in the following lines.

I) Protein-ligand Docking

Protein-ligand docking technique aims to predict the final structure of a intermolecular complex formed by the binding of a ligand to a protein scaffold. This final structure is determined by the conformation and orientation of the ligand into the binding site of the protein. The Gibbs energy associated to the binding process can be easily calculated by the law of mass action (Equation 3.11).

$$\Delta G_{binding} = -RTlnK_{eq} \quad (3.11)$$

where K_{eq} is the equilibrium constant associated to the binding process:

$$K_{eq} = \frac{[Complex]}{[Protein][Ligand]} \quad (3.12)$$

The protein-ligand docking procedure is almost the same in all the known protein-ligand docking software suites (Figure 3.4). At the beginning, both protein and ligand can be prepared by a third party software, then the binding site of the protein is defined by the proper software. The central part is the most complex, being composed by the combination of two algorithms, one that performs the conformational search and a latter one called scoring function which calculates the energy of the binding. Finally, protein-ligand fitting results are ranked accordingly to their binding energy.

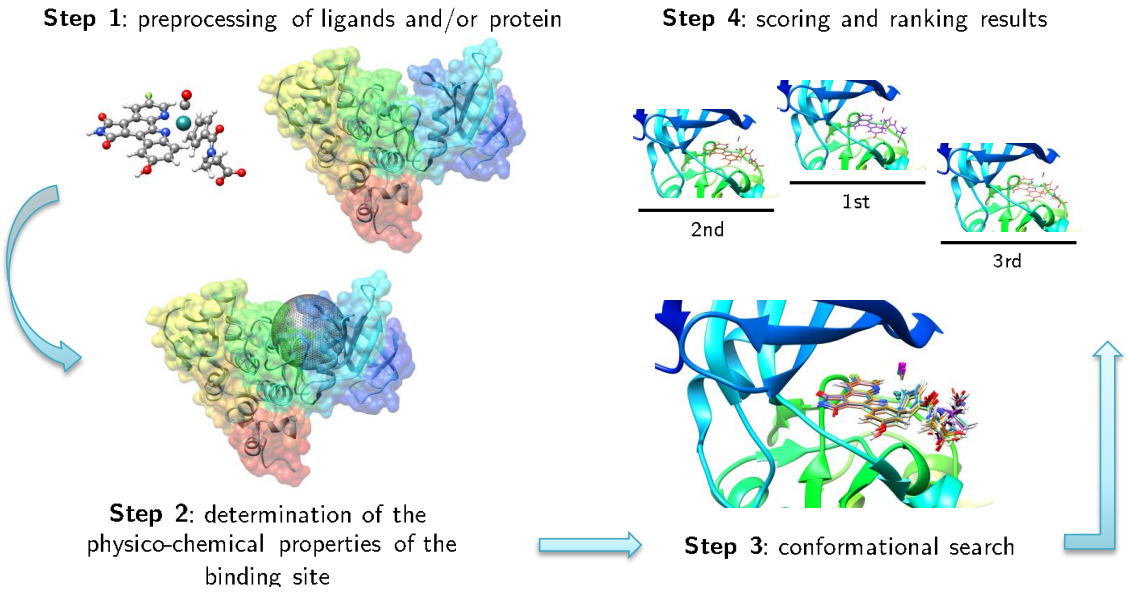


Figure 3.4: General procedure of a protein-ligand docking simulation.

The search algorithm explores the conformational space of the complex, generating all the possible orientations of the protein paired with the ligand. Many algorithms can be used for this purpose. Depending on the degrees of freedom associated to the ligand and the flexibility of the protein, search algorithms can be divided in rigid-body, flexible-ligand and flexible-protein docking. Rigid-body search algorithms only considers the rotational and translational degrees of freedom of the ligand. Flexible-ligand docking considers the protein as a rigid body and a full or partial flexibility of the ligand. Depending on how is performed the exploration of the conformational space, flexible-ligand search algorithms can be classified in three categories: systematic search (explores most conformational degrees of freedom of the ligand), stochastic (or random) methods (performs random changes in the conformation of the system) and molecular simulation algorithms (performing molecular dynamics and energy minimization calculations). Furthermore, flexible-protein algorithms explore the conformation of the residues near to the binding site. Some examples of search algorithms are Conformational search (DOCK [28]), Monte Carlo (Glide [29,30]), Genetic algorithm (GOLD [31–33]) and Simulated annealing (AutoDock [34,35]).

Due to the enormous number of candidates generated during the conformational search, accurate energy calculations of the complexes are practically impossible to perform, even if at Molecular

Mechanics level of theory. Scoring functions are used to attempt to approximate the binding Gibbs energy for the ligand binding to the receptor. Currently, the number of scoring functions available for predicting protein-ligand interactions is large and increases day-by-day. Depending on the origin of the parameters which composes the scoring functions, they can be divided in four categories:

- **Force Field Based Scoring functions:** based on the non-bonded terms of the molecular mechanics force field, estimates the interaction between the ligand and the host instead of calculating the energy of the entire system. One example is GoldScore.[31]
- **Empirical Based Scoring functions:** decomposes the binding Gibbs energy into several energetic terms such as van der Waals and electrostatic interactions, hydrogen bond terms, desolvation energy, entropy effects. The different energy terms are obtained by a regression analysis on a set of protein-ligand complexes with known binding affinities. ChemScore[36,37] and Piecewise Lineal Potential (PLP)[38] are examples of empirical based scoring functions.
- **Knowledge-Based Scoring functions:** purely statistical being designed to reproduce experimental structures instead of binding affinities. The parameters are obtained by simple statistical potentials which estimate the frequency of occurrence or non-occurrence of different atom-atom pair contacts. AutoDock Vina[39] is one example of this kind of scoring functions.
- **Consensus Scoring functions:** this kind of scoring functions are not related to a proper description of the energy or structural effects of the binding process. Consensus scoring process are employed as a post-processing step after docking runs. By definition, the application of any combination of scoring functions can be defined as a consensus scoring function, but there are some examples of standalone methods, such as GFscore.[40]

In the last years, with the evolution of both search and scoring algorithms, docking approaches have achieved several improvements. However, the current state of protein-ligand docking approaches are far from the excellence. The validation of the docking protocol, the effectiveness of covalent docking, dealing with crystallographic water molecules and the treatment of the entropy are still a weak point. Nowadays, the emergence of new computational tools, such as machine learning, offers to protein-ligand docking a reawakening.[41]

II) Normal Mode Analysis (NMA)

Molecules move. Independently of the size and function, from the smallest compound to the larger protein, molecules are in constant movement only stopping at 0 Kelvin degrees. For a certain energy state, molecules show natural movements, also called normal modes of vibration. Normal Mode Analysis (NMA) accounts for this motions by means of the analytical study of the harmonic potential well where the system is placed. Despite of the applicability of this technique in the characterization of the lower energy movements of a macromolecular system, is not suited to the study of conformational translations involving jumping from one potential well to other.

Each potential energy well represents a local minimum on the potential energy surface of a macromolecular system being this minimum energy state the starting point of normal mode analysis. The harmonic approximation of the potential well in the actual conformation of the system is constructed following the Equation 3.13.

$$U(\mathbf{r}) = \frac{1}{2}(\mathbf{r} - \mathbf{R})\mathbf{K}(\mathbf{R})(\mathbf{r} - \mathbf{R}) \tag{3.13}$$

R represents a 3N-dimensional vector which describes the stable conformation at the center of the well, **r** is a 3N-dimensional vector with the current conformation of the system and **K** is the mathematical approximation of the shape of the well. Formally, only small displacements from the minima (**r** - **R**) are correctly described with the defined harmonic approximation. From the parameters present in Equation 3.13, **R** (also called $\mathbf{R}_{min}$) is obtained by an initial minimization of the system or its X-ray structure, and **K** is computed by the resolution of the second derivative of the potential (Equation 3.14).

$$\mathbf{K}_{ij} = \left[\frac{\partial_2 U}{\partial \mathbf{r}_i \partial \mathbf{r}_j}\right]_{\mathbf{r}=\mathbf{R}_{min}} \tag{3.14}$$

The **K** matrix contains the frequencies related to the normal modes associated to the studied system. Each frequency is correlated with the energetic cost of the movement, being low energy movements associated to collective or delocalized deformations, and increasing the energy value with the localization of the displacement. Therefore, the low-frequency modes are the most relevant ones as they dictate the large-scale collective movement of the system.[f]

The use of normal mode analysis is limited due to the particular localization of the harmonic potential, which avoids larger explorations of the conformational space. Taking into account the advantages and pitfalls of normal mode analysis, this technique can be applied in two major ways. The first one is using a large set of modes to have a complete representation of the harmonic potential well. Here, the only limitation is the choice of the subset. The second use is a individual inspection of the modes, where only few of the lowest models are sufficiently separated in energy to be certain that those are not degenerate. Here, normal mode analysis results can be interpreted towards the type of motion they describe (i.e. rotation of a helix).

Despite of the limitations of the method, the use of normal mode analysis usually yields results that are in agreement with experimental knowledge. One example of it can be found in the works of J. Ma and coworkers which applied normal mode analysis to describe the conformational variants of a fatty acid synthase.[42] The predicted conformational variants using normal mode analysis of the vibration models show a good agreement with the experimental X-ray structure of the fatty acids previously synthesized.

[f] The first six modes are associated to the translation and rotation movements or a rigid-body, being meaningless in the analysis

3.1.3 Hybrid QM/MM Methods

The application of molecular modelling in the bioinorganic chemistry field have restrictive limitations in terms of size and accuracy. Simulations in a reasonable scale of time accounts for big systems modelled at lower computational level or very centred on a specific region of the complex. In addition, force field approaches are unable to describe the details of bond breaking or electron transfer processes. These limitations can be overcome by combining some of they into new integrative approaches. One example of them is the hybrid Quantum Mechanics / Molecular Mechanics (QM/MM) method, which allows to take into consideration the entire complex without losing accuracy in the binding site. This approach has been introduced in 1976 by Warshel and Levitt,[43] and in 2013, along with Karplus, they awarded the Nobel Prize in Chemistry for "the development of multiscale models for complex chemical systems". An explanation of the technicalities of QM/MM methods is given below, with the description of ONIOM, which is the chosen framework to apply the hybrid methodology in the research projects described in this manuscript.

The ONIOM framework

The ONIOM (Our own N-layered Integrated Molecular Orbital Molecular Mechanics) method allows to split the system in two (or more) layers, describing each of them by a different level of theory. The usual configuration of hybrid QM/MM includes only two layers. The total energy of the system E_{tot} accounts for the energy of both layers (E_{QM} and E_{MM}) and the interaction between them ($E_{QM/MM}$). There are two alternative schemes to represent the total energy, known as additive and substractive approaches. Both schemes are equivalent, only differing in some aspects of their implementation.[44] In the additive scheme (Equation 3.15) the total sum of energy includes the QM energy of the system included in the QM layer, the MM energy of the entire system except the QM part, and the interaction between both regions:

$$E_{tot} = E_{QM} + E_{MM} + E_{QM/MM} \tag{3.15}$$

For the subtractive scheme, or in other words, the ONIOM scheme[45,46], the total energy of the system is divided into the real and the model layers and substracts the double counted terms: E_{real}^{MM} accounts the energy of the entire complex calculated with MM, E_{model}^{QM} is the energy of the QM part, and E_{model}^{MM} is the energy of the QM part calculated at MM level (Equation 3.16 and Figure 3.5).

$$E_{total} = E_{real}^{MM} + E_{model}^{QM} - E_{model}^{MM} \tag{3.16}$$

In the cases where the QM and MM regions belong to the same molecule, a key variable in QM/MM methods is how to treat the frontier between both regions. Generally, the frontier is located at the covalent bonds that link the MM and QM regions, and the division is performed cutting the involved bonds. It yields a final structure with unpaired electrons in the QM part, which have to be paired again. To solve this problem, three different boundary treatments can be

Figure 3.5: Schematic view of the decomposition of the total energy terms in substractive QM/MM methods

found: a hydrogen link-atom, a pseudo-halogenated link atom,[47] and by the localized molecular orbital.[48,49] The most common way is pairing the unpaired electrons of the metal with a hydrogen link atom. The charge of the link atom in the MM part, as well as the van der Waals interactions, are set to zero to avoid double counting. If a semi empirical method is employed, the hydrogen link atom can be substituted by a pseudo-halogen one which is previously parametrized in order to be as similar as possible to the original bonded atom. The last alternative is a description of the termination of the QM part using localized molecular orbitals. It is not clear of which is the best option to represent the boundary between the QM and the MM part, but the most frequently used is the hydrogen link atom procedure, due to the simplicity on its implementation. However it is important to take into account the limitations of the hydrogen link atom method. This new atom is seen only by the QM part, which means that some bending and torsional terms involving boundary atoms can be neglected by the software.

Both regions can interact between them in three different embedding levels. The lower level of interaction is the mechanical embedding. Here only the the bonded and steric energies of both regions are included in the interaction term. The following level is called electronic embedding, being the atoms in the MM part allowed to polarize the QM region. In the electronic embedding expression, the mechanical embedding term is also included. The final level of improvement is known as polarizable embedding, being QM atoms allowed to polarize the MM region. However, this last embedding technique is only available if the lowest level is described by a semi empirical method.

QM/MM methods are not restricted to the modelling of big systems, which are impossible to study using pure QM approaches. An other application is the inclusion of explicit solvent molecules in the surroundings of small chemical systems, in order to take into account its structural effect. The chemical compound is placed in the QM region and the solvent in the MM part. Usually, only mechanical embedding is used to describe the interaction of both parts. However, to have an accurate description of the solvent, the effective fragment method can be used to include the quadrupoles and polarizabilities in the MM region.[50]

Despite of the technicalities present in general QM/MM methods, the most difficult decision that researchers have is the election of the correct QM/MM partition. Taking into account that the ONIOM software implemented in Gaussian09 uses the hydrogen link atom boundary description,

an accurate cut is key. For example, aromatic rings have to remain unbroken, as well as highly polarized bonds such as C-O or C-N. An other consideration to take into account is the number of QM atoms connected to the same link. If there are more than one QM atom connected to a link, the ONIOM approach will fail in the description of the boundary. Additionally, as lower is the number of link atoms, easier will be the simulation process. To avoid the lose of quality in the model, it is important also to consider where the reaction is taking place. In example, to model an enzymatic reaction, one can not set the boundary close to the reaction center. On the other hand, making the boundary too far away will increase unnecessarily the computational time.

3.1.4 What method should I use? and the need to combine them

In the molecular modelling field, the choice of the correct tool is not always an easy task. Every computational chemist knows that all the results he/she obtain will be always approximate, independently on the method and the system. The three variables that modellers have to fit are: size of the system, computational time and expected level of accuracy, weighting the three according the requirements. For example, a big system cannot be accurately modelled if the computational time is scarce. A solution is use the "divide and conquer" strategy, applying separate techniques at different steps of the molecular modelling process. This approaches are known as integrative procedures.

Integrative procedures

The use of integrative (multilevel or multiscale) approaches can help in the study of complex molecular processes. Here, the entire procedure can be divided in different steps, accordingly to the molecular event which is taking place at this point. One example of integrative procedure, proposed by our research group, whih is described in Figure 3.6.

This approach is divided in three different steps: an initial conformational search of a ligand (or cofactor) outside the proteic receptor is performed by Quantum Mechanics calculations, a flexible docking simulation accounting different conformations of the protein and a final structure refinement using QM/MM.

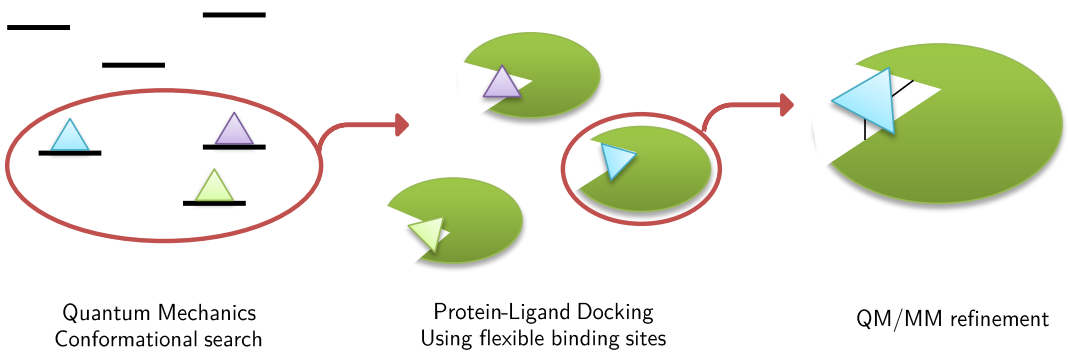


Figure 3.6: Schematic representation of our integrative protocol.

In most of the cases, a single computational chemistry software is not able to account for all the stages of the integrative procedure, being mandatory the combination of two or more to describe the system. The communication between them is crucial to avoid loss of information, but normally this step is not an easy task. By means of in house scripts written in Python code, in our group we develop an interface involving some commercial available software packages (like Gaussian and GOLD) which we implemented in the UCSF Chimera code. These scripts, facilitates the exchange of information between different models allowing an easy application of integrative procedures.

3.2 References

[1] Bolcer, J. D.; Hermann, R. B. *Rev. Comp. Ch.* **1994**, *5*, 1–63.

[2] Hohenberg, P.; Kohn, W. *Phys. Rev.* **1964**, *136*, B864.

[3] Vosko, S.; Wilk, L.; Nusair, M. *Can. J. Phys.* **1980**, *58*, 1200–1211.

[4] Perdew, J. P.; Zunger, A. *Phys. Rev. B* **1981**, *23*, 5048–5079.

[5] Cole, L. A.; Perdew, J. P. *Phys. Rev. A* **1982**, *25*, 1265–1271.

[6] Becke, A. D. *Phys. Rev. A* **1988**, *38*, 3098.

[7] Perdew, J. P.; Burke, K.; Ernzerhof, M. *Phys. Rev. Lett.* **1996**, *77*, 3865.

[8] Lee, C.; Yang, W.; Parr, R. G. *Phys. Rev. B* **1988**, *37*, 785–789.

[9] Perdew, J. P. *Phys. Rev. B* **1986**, *33*, 8822–8824.

[10] Perdew, J. P.; Chevary, J.; Vosko, S.; Jackson, K. A.; Pederson, M. R.; Singh, D.; Fiolhais, C. *Phys. Rev. B* **1992**, *46*, 6671.

[11] Zhao, Y.; Schultz, N. E.; Truhlar, D. G. *J. Chem. Theory Comput.* **2006**, *2*, 364–382.

[12] Tao, J.; Perdew, J. P.; Staroverov, V. N.; Scuseria, G. E. *Phys. Rev. Lett.* **2003**, *91*, 146401.

[13] Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 5648–5652.

[14] Adamo, C.; Barone, V. *J. Chem. Phys.* **1999**, *110*, 6158–6170.

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

[16] Schlegel, H. B. *WIREs Comput. Mol. Sci.* **2011**, *1*, 790–809.

[17] Eckert, F.; Pulay, P.; Werner, H.-J. *J. Comput. Chem.* **1997**, *18*, 1473–1483.

[18] Te Velde, G.; Bickelhaupt, F. M.; Baerends, E. J.; Fonseca Guerra, C.; van Gisbergen, S. J.; Snijders, J. G.; Ziegler, T. *J. Comput. Chem.* **2001**, *22*, 931–967.

[19] Neese, F. *WIREs Comput. Mol. Sci.* **2012**, *2*, 73–78.

[20] Li, X.; Frisch, M. J. *J. Chem. Theory Comput.* **2006**, *2*, 835–839.

[21] Frisch, M. J. et al. *Gaussian09 Revision D.01*, Gaussian Inc. Wallingford CT 2009.

[22] Kitaura, K.; Morokuma, K. *Int. J. Quantum Chem.* **1976**, *10*, 325–340.

[23] Bickelhaupt, F. M.; Baerends, E. J. *Rev. Comp. Ch.* **2007**, *15*, 1–86.

[24] Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. *J. Am. Chem. Soc.* **1995**, *117*, 5179–5197.

[25] Burton, V. J.; Deeth, R. J.; Kemp, C. M.; Gilbert, P. J. *J. Am. Chem. Soc.* **1995**, *117*, 8407–8415.

[26] Seminario, J. M. *Int. J. Quantum Chem.* **1996**, *60*, 1271–1277.

[27] Andricopulo, A. D.; Guido, R. V.; Oliva, G. *Curr. Med. Chem.* **2008**, *15*, 37–46.

[28] Kuntz, I. D.; Blaney, J. M.; Oatley, S. J.; Langridge, R.; Ferrin, T. E. *J. Mol. Biol.* **1982**, *161*, 269–288.

[29] Friesner, R. A.; Banks, J. L.; Murphy, R. B.; Halgren, T. A.; Klicic, J. J.; Mainz, D. T.; Repasky, M. P.; Knoll, E. H.; Shelley, M.; Perry, J. K. et al. *J. Med. Chem.* **2004**, *47*, 1739–1749.

[30] Halgren, T. A.; Murphy, R. B.; Friesner, R. A.; Beard, H. S.; Frye, L. L.; Pollard, W. T.; Banks, J. L. *J. Med. Chem.* **2004**, *47*, 1750–1759.

[31] Jones, G.; Willett, P.; Glen, R. C.; Leach, A. R.; Taylor, R. *J. Mol. Biol.* **1997**, *267*, 727–748.

[32] Verdonk, M. L.; Cole, J. C.; Hartshorn, M. J.; Murray, C. W.; Taylor, R. D. *Proteins-Struct Funct Genet* **2003**, *52*, 609–623.

[33] Jones, G.; Willett, P.; Glen, R. C. *J. Mol. Biol.* **1995**, *245*, 43–53.

[34] Morris, G. M.; Goodsell, D. S.; Halliday, R. S.; Huey, R.; Hart, W. E.; Belew, R. K.; Olson, A. J. et al. *J. Comput. Chem.* **1998**, *19*, 1639–1662.

[35] Goodsell, D. S.; Morris, G. M.; Olson, A. J. *J. Mol. Recognit.* **1996**, *9*, 1–5.

[36] Eldridge, M. D.; Murray, C. W.; Auton, T. R.; Paolini, G. V.; Mee, R. P. *J. Comput. Aid. Mol. Des.* **1997**, *11*, 425–445.

[37] Murray, C. W.; Auton, T. R.; Eldridge, M. D. *J. Comput. Aid. Mol. Des.* **1998**, *12*, 503–519.

[38] Gehlhaar, D. K.; Verkhivker, G. M.; Rejto, P. A.; Sherman, C. J.; Fogel, D. R.; Fogel, L. J.; Freer, S. T. *Chem. Biol.* **1995**, *2*, 317–324.

[39] Trott, O.; Olson, A. J. *J. Comput. Chem.* **2010**, *31*, 455–461.

[40] Betzi, S.; Suhre, K.; Chétrit, B.; Guerlesquin, F.; Morelli, X. *J. Chem. Inf. Model.* **2006**, *46*, 1704–1712.

[41] Durrant, J. D.; Amaro, R. E. *Chem. Biol. Drug. Des.* **2015**, *85*, 14–21.

[42] Brink, J.; Ludtke, S. J.; Kong, Y.; Wakil, S. J.; Ma, J.; Chiu, W. *Structure* **2004**, *12*, 185–191.

[43] Warshel, A.; Levitt, M. *J. Mol. Biol.* **1976**, *103*, 227–249.

[44] Banáš, P.; Jurečka, P.; Walter, N. G.; Šponer, J.; Otyepka, M. *Methods* **2009**, *49*, 202–216.

[45] Maseras, F.; Morokuma, K. *J. Comp. Chem.* **1995**, *16*, 1170–1179.

[46] Svensson, M.; Humbel, S.; Froese, R. D.; Matsubara, T.; Sieber, S.; Morokuma, K. *J. Chem. Phys.* **1996**, *100*, 19357–19363.

[47] Antes, I.; Thiel, W. *J. Phys. Chem. A* **1999**, *103*, 9290–9295.

[48] Théry, V.; Rinaldi, D.; Rivail, J.-L.; Maigret, B.; Ferenczy, G. G. *J. Comp. Chem.* **1994**, *15*, 269–282.

[49] Gao, J.; Amara, P.; Alhambra, C.; Field, M. J. *J. Phys. Chem. A* **1998**, *102*, 4714–4721.

[50] Arora, P.; Slipchenko, L. V.; Webb, S. P.; DeFusco, A.; Gordon, M. S. *J. Phys. Chem. A* **2010**, *114*, 6742–6750.

4 ... Protein-Ligand docking with inert first coordination sphere: the case study of Ruthenium based kinase inhibitors

Related paper:
Appendix A.3: Assessing protein-ligand docking for the binding of organometallic compounds to proteins

There are sadistic scientists who hurry to hunt down errors instead of establishing the truth.

Marie Salomea Skłodowska Curie

4.1 Computational study of inert first coordination sphere processes

For years now, molecular modelling has become a great partner for the field of drug design. In this chapter, we would like to assess how these methods could work on systems where interaction of proteins with organometallic compounds is driven by second coordination sphere processes, where the metal has only a structural role and does not participate directly in the binding process.

Current protein-ligand docking algorithms can deal with metals but only if it is part of the receptor.[1] In this case, some docking programs have parameters for non-transition metals or the option to parametrize newer ones. Nevertheless, the main question is, are the parameters of the metal necessary for the calculation of the score in inert first coordination sphere processes? the short answer is no, for two main reasons: the first one is because in protein-ligand docking the first coordination sphere of the ligand remains rigid during all the simulation, and the second one is due to the metal is not included in the energy calculation because it does not participate in the binding.

But, how the ligand is computed in protein-ligand docking if the metal is not included? The answer is so simply: if the metal does not participate in the binding process having only a structural role, the metal can be swapped by a parametrized entity forcing the substituents to be coordinated there. Controlling only this structural feature, the protein-ligand docking method could be able to reproduce inert first coordination sphere processes. In order to test the effectiveness of protein-ligand docking in the prediction of inert first coordination sphere complexes to biological hosts, ruthenium based kinase inhibitors, which act as anticancer drugs, were used as benchmark system.

4.2 Ruthenium complexes as anticancer metallodrugs

Some ruthenium compounds are used nowadays as anticancer metallodrugs. This is the case of ruthenium based kinase inhibitors. Before embark in the specific interaction of ruthenium with kinases, a brief explanation of the carcinogenic effect of kinases is given.

4.2.1 The role of kinases in cancer

Kinases are enzymes with one specific function: transfer phosphate groups from a donor system to an specific substrate.[2] Kinases are classified depending on the target of the phosphorilation, being the tyrosine kinase the one which is considered in this study. This specific kinase is placed inside the cell, controlling its replication by means of a "traffic-light" process: if the extracellular media has enough nutrients to feed a new cell, the kinase is stimulated. Then a cascade of phosphorilation events within the cell starts. This is message which informs that the cell is ready to divide. But, how it works?

Tyrosine kinases are part of a structure called epidermal growth factor (EGF) which is the biological entity that connects the cell with the extracellular environment. The structure of the

EGF is constituted by the extracellular domain (the EGF binding site), the tyrosine kinase and the cytosolic tail. Normally, the EFG are in pairs and close to each other. The activation of the tyrosine kinase starts with the binding of an external ligand to the EGF binding site. If both kinases are activated, the clusterization of the receptor starts, allowing a cross-phosphorilation of the tyrosine residues present in the cytosolic tail. At this moment, the cell can replicate itself. In Figure 4.1, the structure and activation of a kinase (in this case, a tyrosine kinase) is illustred.

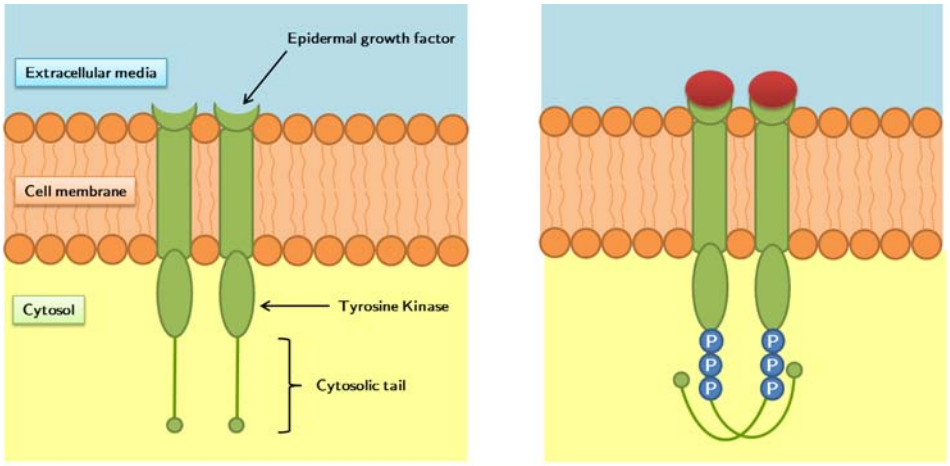


Figure 4.1: Schematic representation of the EGF (left) and cross-phosphorilation of the cytosolic tail (right).

However, in some cases tyrosine kinases are free of the EGF (those are called *nonreceptor tyrosine kinases*).[3] The main particularity of the nonreceptor tyrosine kinases is their permanent active state. Once a nonreceptor tyrosine kinase is close to an activated EGF, the clusterization and the subsequent phosphorilation is going on. At this point the cell receives the replication signal and an uncontrolled division can take place. This is the starting point of many diseases, including cancer.[4] For that reason, the use of kinase inhibitors are pursued as new anticancer therapeutics.

4.2.2 Ruthenium based kinase inhibitors

In 2004, Prof. Meggers and coworkers published the first study in ruthenium kinase inhibitors.[5] In this work, they presented a metallocomplex derived of staurosporine, which was christened octasporine due to the octahedral geometry of the metal centre. This metallocomplex is classified as a Type 1 kinase inhibitor, which recognizes the active conformation of the kinase and binds to the active site in competition with ATP. Avoiding the binding of ATP to the active site of the kinase, the phosphorilation process will be truncated. Both staurosporine and ruthenium inhibitors have a structure similar to ATP. This structural match allows the inhibitors to bind exactly in the same way than ATP (see Figure 4.2).

Despite of the high similarity between staurosporine and the ruthenium inhibitor, the second has a great advantage in front of the first: the metal centre in combination with the coordinating ligands controls the shape of the organometallic compound conferring high specificity to a specific

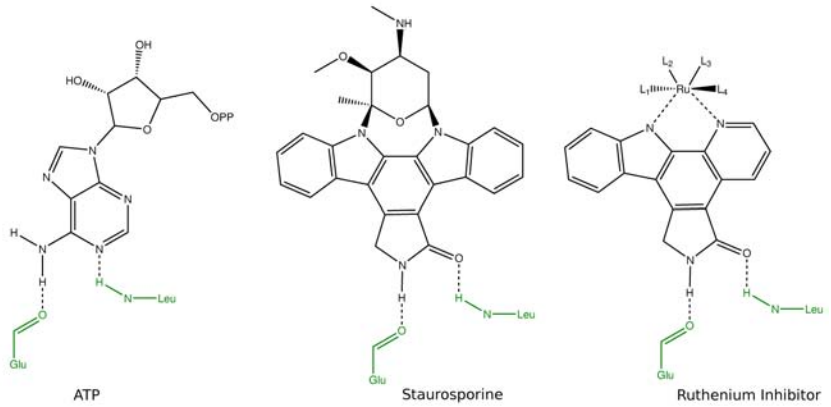


Figure 4.2: Chemical structure of ATP, staurosporine and ruthenium inhibitor compound in the ATP binding pocket of CDK2 kinase.

host.[6]

4.3 Material and Methods

4.3.1 Setting the benchmark systems

Docking calculations have been performed on 16 different crystallized kinases containing a docked organometallic kinase inhibitor. These structures were sintetized and reported by Meggers and coworkers.[7] From the starting sixteen kinase-inhibitor complexes, fifteen of the total contain a ligand which belongs to the family of the pyrocarbazoles: eleven of them are ruthenium pyridocarbazole half-sandwich complexes (2JLD **(1)**,[8] 3CST **(4)**,[9] 3Q4C **(5)**,[10] 2BZI **(6)**,[11] 2BZH **(8)**,[12] 2IWI **(9)**,[12] 3CSF **(10)**,[9] 2BZJ **(11)**,[13] 3M1S **(12)**,[13] 3FY0 **(13)**[11] and 2OI4 **(15)**[14]), one corresponds to an osmium pyridocarbazole half-sandwich (3BWF **(7)**[15]) and the last three belongs to the family of ruthenium pyridocarbazole octasporine (2YAK **(2)**,[16] 3FXZ **(3)**[11] and 3PUP **(14)**[16]). The compound number sixteen corresponds to a Schiff base Fe(III) chelate (1G3C **(16)**[17]). In Figure 4.3 the X-ray structure of two example compounds is drawn, as well as all the molecular structures of the isolated organometallic ligands in Figure 4.4.

Mostly all the inorganic compounds are docked to different kinases: GSK-3 (structures 1, 12 and 14), DAPK1 (2), PAK1 (3, 13), PI3K (4), BRAF (5), PIM1 (6, 7, 8, 11, 15), PIM2 (9) and BetaTrypsin (13).

4.3.2 Computational procedure

In protein-ligand docking there are lots of possible variables which one can adjust: number of hydrogen bonds between the ligand and the host, specific interactions with certain residues, minimization of the Van der Waals contacts, etc.

In this work, we performed a real discovery exercise in which the binding of inert oganometallic

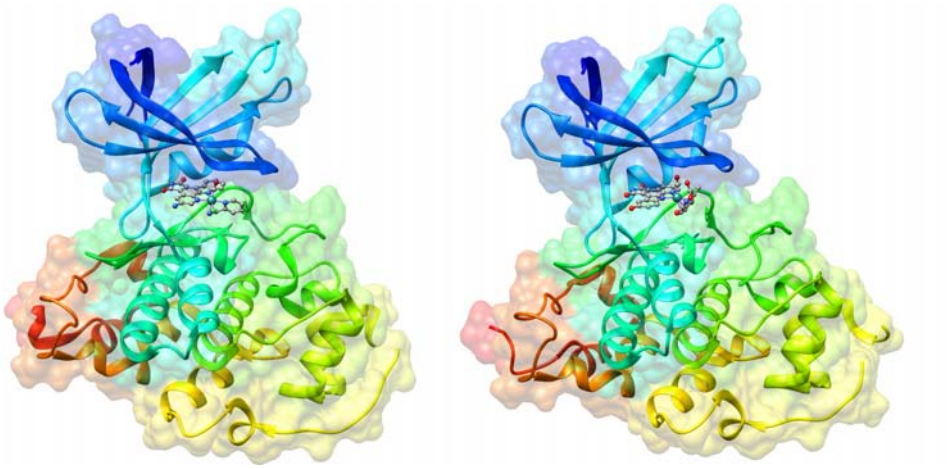


Figure 4.3: X-ray structure of 3PUP[16] (left) and 2JLD[8] (right).

compounds to proteins has been tested by means of the application of the protein-ligand docking technique. To do so, three separate variables have been explored: scoring function, flexibility of the receptor and geometry of the ligand (if experimental or optimized). In this case all the simulations were performed with the GOLD5.1 suite.[18,19] GOLD includes four built-in scoring functions: ChemPLP,[20] ChemScore,[21,22] GoldScore,[23] and ASP.[24] All four have been tested.

To proceed, four docking schemes have been defined, depending on the starting structure of the ligand and the allowed flexibility of the host:

- **I**: experimental structure of the ligand and fixed binding site.
- **II**: optimized structure of the ligand and fixed binding site.
- **III**: experimental structure of the ligand and flexible binding site.
- **IV**: optimized structure of the ligand and flexible binding site.

A total of sixteen Protein-Ligand docking series have been carried out looking for the best combination of scoring function, structure of the ligand and flexibility of the receptor. Accounting for the number of X-ray structures used in this benchmarking study, a total of 256 protein-ligand docking simulations were already studied and evaluated.

4.4 Towards a protein-organometallic ligand protocol

Results are presented in three different sections. In the first one, the importance of pre-processing the ligand is evaluated. Experimental and optimized geometries were compared to get a initial idea of how the organometallic ligand is affected during the binding process. Then, protein-ligand docking results were evaluated both structural and energetically, paying special attention on the reproducibility of results independently of the chosen scoring function, structure of the ligand and

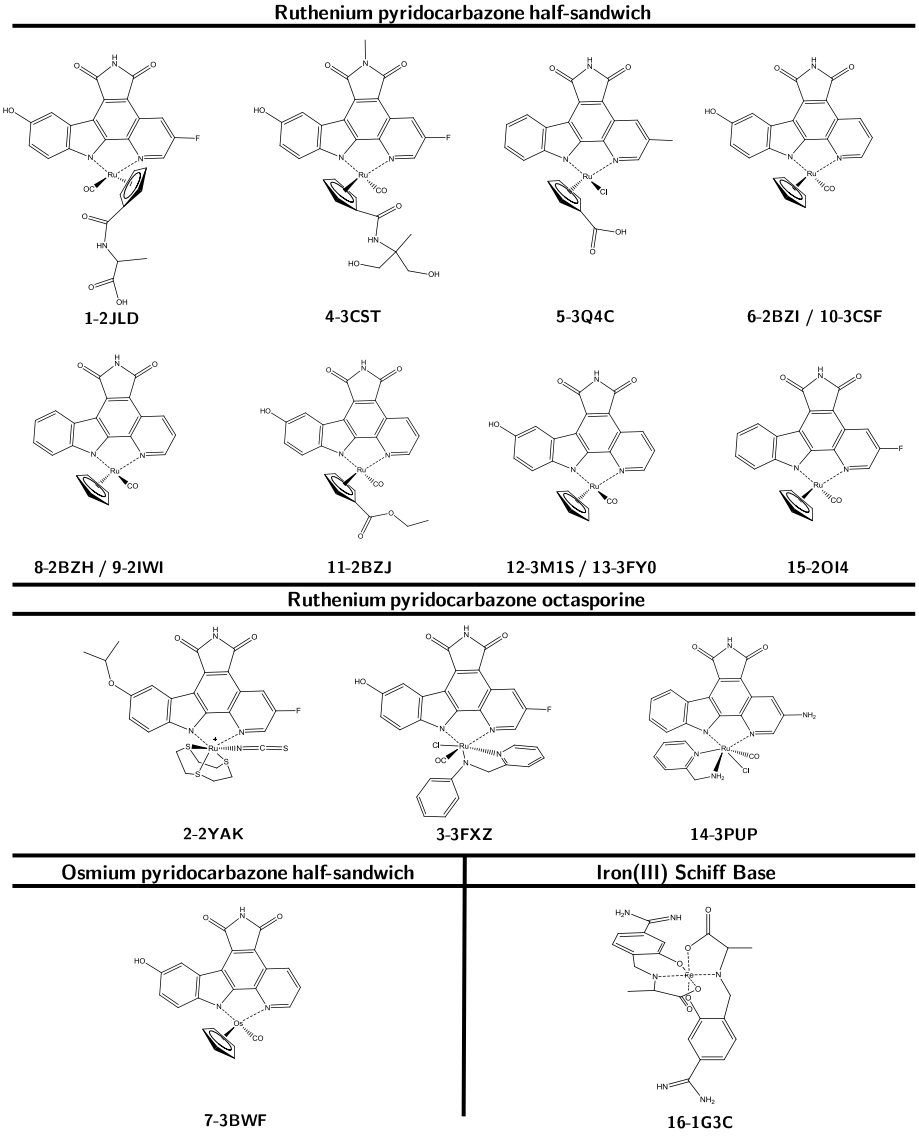


Figure 4.4: Structure of the different ligands used in this work.

flexibility of the receptor. Finally, the obtained energy results were compared with the experimental value of IC_{50}[a], the measure of the effectiveness of the inhibition process.

4.4.1 Pre-docking considerations

In protein-ligand docking it is common to perform a pre-processing work of the structure of the ligands and the receptor. In the case of the ligand one have to take under consideration the binding and the crystallization processes. At this point the ligand can suffer some deformations which can have some impact on the final docked structure. For organic ligands, performing a low cost

[a] IC_{50} is the half maximal inhibitory concentration. This quantitative measure indicates how much of a particular drug or other substance (inhibitor) is needed to inhibit a given biological process by half

geometry optimization is more than enough. For organometallic ligands, because of the electronic and structural flexibility of transition metals, molecular modelling methods are not accurate enough and the use of quantum level tools is mandatory. In this work, the DFT:B3LYP [25,26] approach has been used in combination with a mixed basis set (6-31G* [27] for the main group elements and LANL2DZ [28] for the transition metal). Calculations were performed with the Gaussian09 suite. [29]

No notable differences are observed between the experimental geometry of the ligand and the minimum energy conformation calculated at the quantum level (Table 4.1). Only in the case of half-sandwich compounds in which the cyclopentadienyl ligand has a flexible and polar substituent (structures **1**, **4** and **5**), differences are remarkable. Nevertheless, this rearrangement is expected to have a limited impact on the docking simulation because the polar substituent location is quite far from the binding site.

Table 4.1: RMSD values (in Å) between the experimental structures of the organometallic ligands (from the Protein Data Bank) and the minimizes structures of the same systems. Values in parenthesis corresponds to those obtained removing the substituent of the cyclopentadienyl group.

1-2JLD	**2- 2YAK**	**3- 2FXZ**	**4- 3CST**
0.88 (0.17)	0.43	0.28	3.17 (0.23)
5- 3Q4C	**6- 2BZI**	**7- 3BWF**	**8- 2BZH**
0.91 (0.57)	0.28	0.28	0.13
9- 2IWI	**10- 3CSF**	**11- 2BZJ**	**12- 3M1S**
0.13	0.36	0.47	0.26
13- 3FY0	**14- 3PUP**	**15- 2IO4**	**16- 1G3C**
0.29	0.29	0.22	0.79

Figure 4.5 shows a superposition of the calculated (green) and experimental (shadow) structures for **6** (left) and **4** (right) as an example of the geometrical impact of the substituent in the cyclopentadienyl due to the geometry minimization.

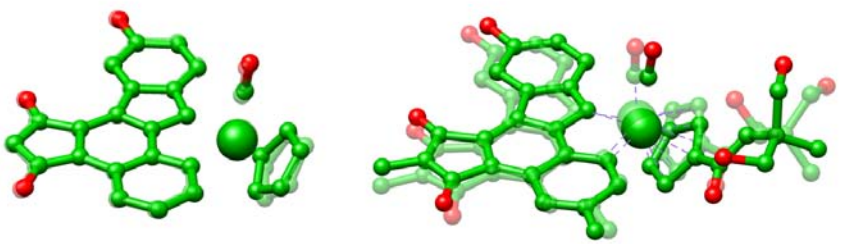


Figure 4.5: Superposition of theoretical (green) and experimental (shadow) structures of the ligands of complex **6** (left) and **4** (right).

These results clearly sustain that the first coordination sphere of the metal is not affected by the protein environment. Moreover, the rest of the ligand can be reorganized to get better anchoring with the protein host. In conclusion, performing a DFT geometry optimization over the structure of the ligands is apparently convenient for their predocking process.

Not only the ligand can be pre-processed, the receptor could also be prepared before starting the docking simulation. This step was skipped in this work because the binding site has already the optimal organization to host the organometallic ligand (both protein and ligand were extracted from a single X-ray structure) and the docked ligand is unmodified from the experimental PDB file. However, Fisher and coworkers have been recently published about the importance of taking into consideration different "snapshots" of the crystal structure, considering the lower energy conformations of the protein and weighting the docking results depending on their population.[30] In contrast, Barril in a comment of the work of Fisher, specifies the importance of having a great accordance with the experimental data. However, because the accurate exploration of the binding site of the host is not the scope of this study, the use of only the experimental structure of the host is enough to achieve our goal.

4.4.2 Protein-ligand docking Results

In order to perform a complete benchmark of the protein-ligand docking tools, three factors have been studied: the geometry of the ligand, the flexibility of the receptor and the fitness function. Altogether, a set of 16 different protein-ligand docking simulations have been performed for each of the 16 kinase-inhibitor pairs. Every docking simulation generates a total of 20 solutions, which corresponds to the 20 best fitting structures. A possible way to organize the complete benchmark study is dividing the simulations in four groups, I to IV as mentioned earlier.

With the purpose of evaluating the main capabilities of protein-ligand docking methods, these results have been analysed in two different ways: the structural one, where the pose of the resulting protein-ligand docked structure is analysed in order to find the best matching set of docking scheme and scoring function; and an energetic one, where a correlation of the score of the different structures with its experimental IC_{50} value have been carried out to have a analytical evidence of the predictiveness of protein-ligand docking techniques.

4.4.3 The structural point of view

To illustrate the accuracy of GOLD in terms of structural matching, geometrical differences between the experimental model and the docked structures were calculated. At this point, root mean square deviation (RMSD) values were calculated with a cutoff of 2.5 Å to discern between the accurate docking poses and the approximated ones. In Figure 4.6 the RMSD values between the experimental structures and the best ranked solutions are reported.

For most of the series, about a 75% of the results are under the cutoff of 2.5 Å, being in agreement with the average quality of docking methods.[31] In Figure 4.7 structural differences

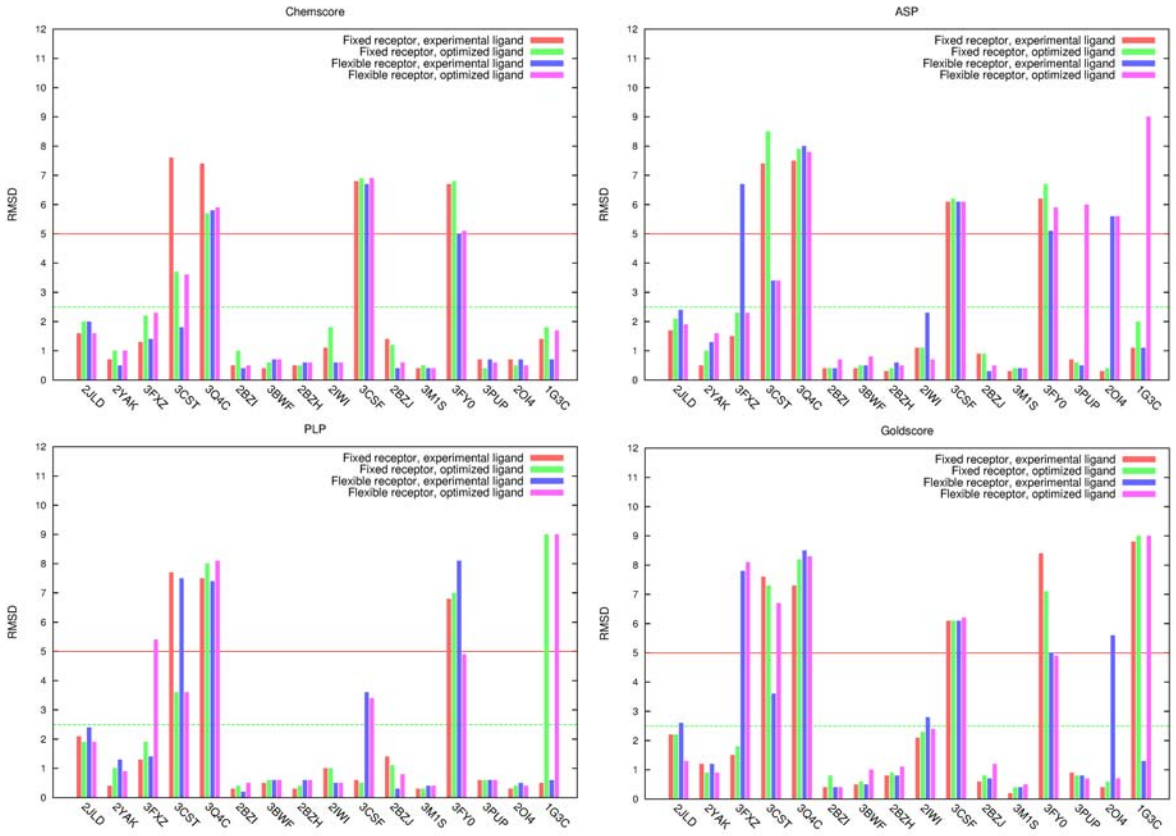


Figure 4.6: RMSD values (in Å) for the lower energy structure for the four different configurations of the protein-ligand docking Simulation for the available Scoring Functions in GOLD 5.1.

between experimental structure (in yellow) and the best matched result (in blue) are shown for both success and failure cases. Depending on the scoring function, 3 to 8 structures are not well predicted, being GoldScore the fitting function with the lowest rate of success. Additionally, a high exposure to the solvent, the presence of Van der Waals contacts between the ligand and the host and a low number of hydrogen bonds between both entities can penalise the quality of the prediction.

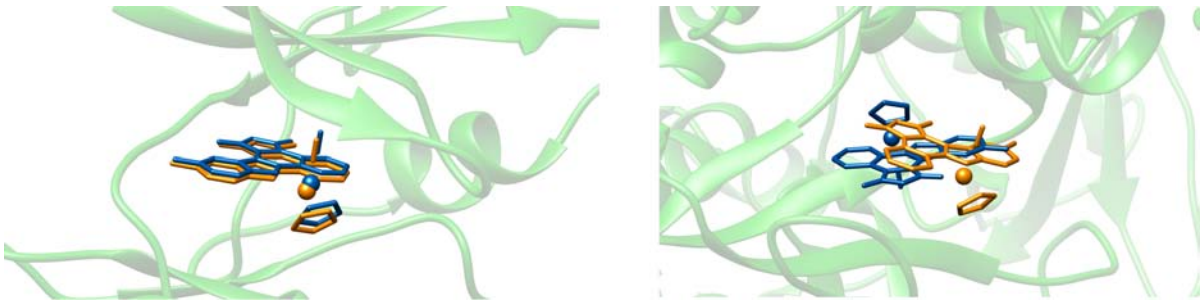


Figure 4.7: Structural differences between the experimental (yellow) and theoritically predicted (blue) ligand poses for both successfull (3M1S, left) and a failed (3CSF, right) cases.

Specifically, systems **4**, **5**, **10** and **13** (Figure 4.4) are the ones which can not be accurately

predicted due to the reasons commented before. In Figure 4.8 an example of a high exposed ligand can be found.

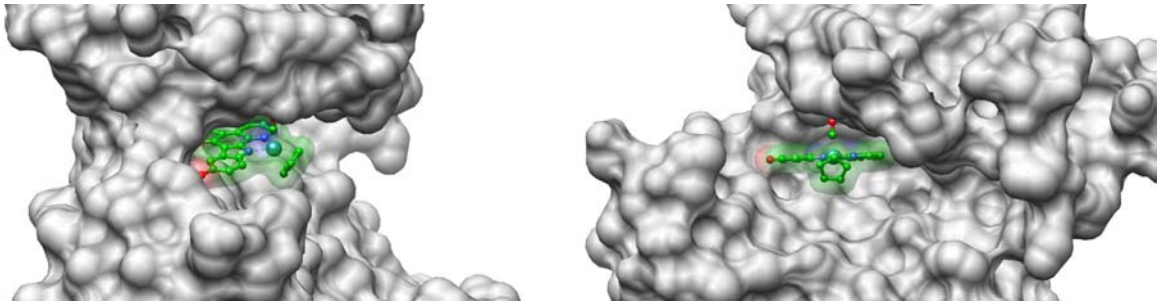


Figure 4.8: The high exposition to the solvent can be seen by different sides on the X-ray structure of the 3FY0 complex.

Detailed analysis of the success rate of the different docking schemes is also revealing. Experiments carried out with a rigid conformation of the receptor (docking schemes I and II), do not show a noticeable impact of the optimization of the metallodrug in the results. Only one (in the case of Chemscore and ASP fitting functions) or two (for ChemPLP) simulations show improved results. Only Goldscore shows a little worsening on the prediction.

In the case of simulations carried out without pre-processing the ligand and considering a fixed receptor (docking scheme I), only two structures can not be predicted (systems **5** and **13**) for the four scoring functions used in this study. The success rate decreases for the GoldScore fitting function, failing a total of five times.

Pre-processing the ligand does not affect, in overall, the results. The reason lays on the fact that the ligand mainly maintains the geometry obtained in the crystal after the DFT minimization in gas phase is performed. Results only worsens in the cases where the geometry of the experimental ligand and the optimized one are notably different due to the presence of a large substituent attached to the cyclopentadienyl group (see Table 4.1)

Increasing the flexibility of the receptor has different consequences on the quality of the docking simulation. For each system, 5 to 8 residues allocated in the binding site (at 5.0 Å from the ligand or less) (see Figure 4.9) were allowed to rotate following the Dunbrack rotamer library.[32] Non-interacting residues were skipped from this selection in order to minimize the conformational space.

When dealing with the ligand in its experimental conformation (docking scheme III), these simulations have almost the same quality than the ones using a rigid receptor (docking scheme I), improving in one case for ChemScore, and worsens in one, two and three cases for ChemPLP, ASP and GoldScore respectively. When docking simulations are performed with the optimized geometries of the ligand (docking scheme IV), the number of correctly predicted structures get worse by one point for ChemScore and ASP, by two points for ChemPLP and gets better by one point in the case of GoldScore.

To sum up, ASP, ChemScore and ChemPLP fitting functions have a good rate of success in reproducing the geometry of the experimental docked system. ChemScore appears to be the most

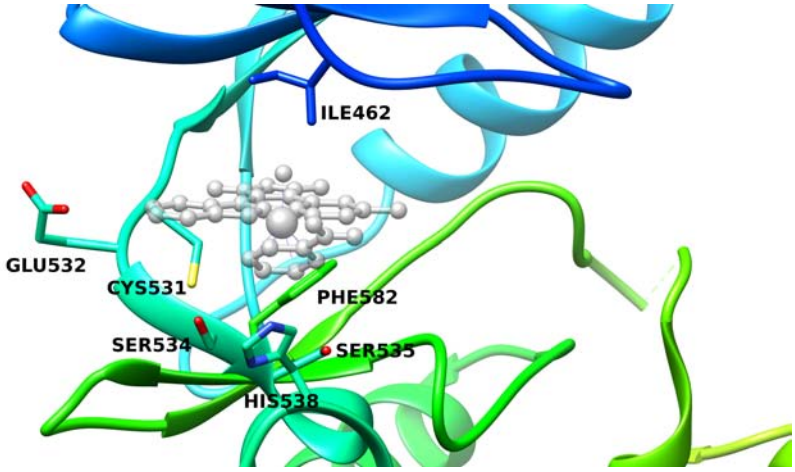


Figure 4.9: Example of aminoacids allowed to rotate (complex 3Q4C).

robust fitting function, followed by ASP and ChemPLP, which are slightly more sensitive. GoldScore, having the worst success rate, is the worst scoring function used in this part of the work.

4.4.4 The energetic point of view

As we conclude in the past section, ASP, Chemscore and ChemPLP scoring functions perform accurate solutions in terms of geometrical prediction. The next step is the analysis of the energy of the lowest energy binding poses. To do so, the energy of the best matched solution of each simulation is graphically represented in Figure 4.10 for each one of the four scoring functions available in GOLD.

The different columns represent the different docking schemes: fixed receptor and experimental ligand in red, fixed receptor and geometry optimized ligand in green, flexible receptor and experimental ligand in blue, and flexible receptor and geometry optimized ligand in pink. Because of both ligand and protein were crystallized in the same structure, one can hypothesize that the docking scheme I (fixed receptor and experimental ligand structure) will get the most favourable scoring values. However, only few simulations from the docking scheme I (in red) have the most favourable score. When the flexibility of the host is increased (green and pink columns) an improvement of the score values can be seen. This effect is modulated depending on the geometry of the ligand, because the flexibility of the protein can also accommodate the geometrical changes in the pre-processed ligand. In the specific case where the binding site is too exposed to the solvent (this is the particular case of 3CST, 3Q4C, 3CSF and 3FY0 X-ray solutions), the most reliable group of simulations are the ones which include flexibility of the host and the pre-processed ligand (docking scheme IV), followed close by the docking scheme III, where the ligand is not optimized. For protein-ligand docking techniques, predict the pose of a ligand in a exposed binding site is not an easy task. So, the capacity of improving the score of such kind of simulations is an example of a success story of the protocol presented here.

Comparing now the best result of each simulation with the four different scoring functions used

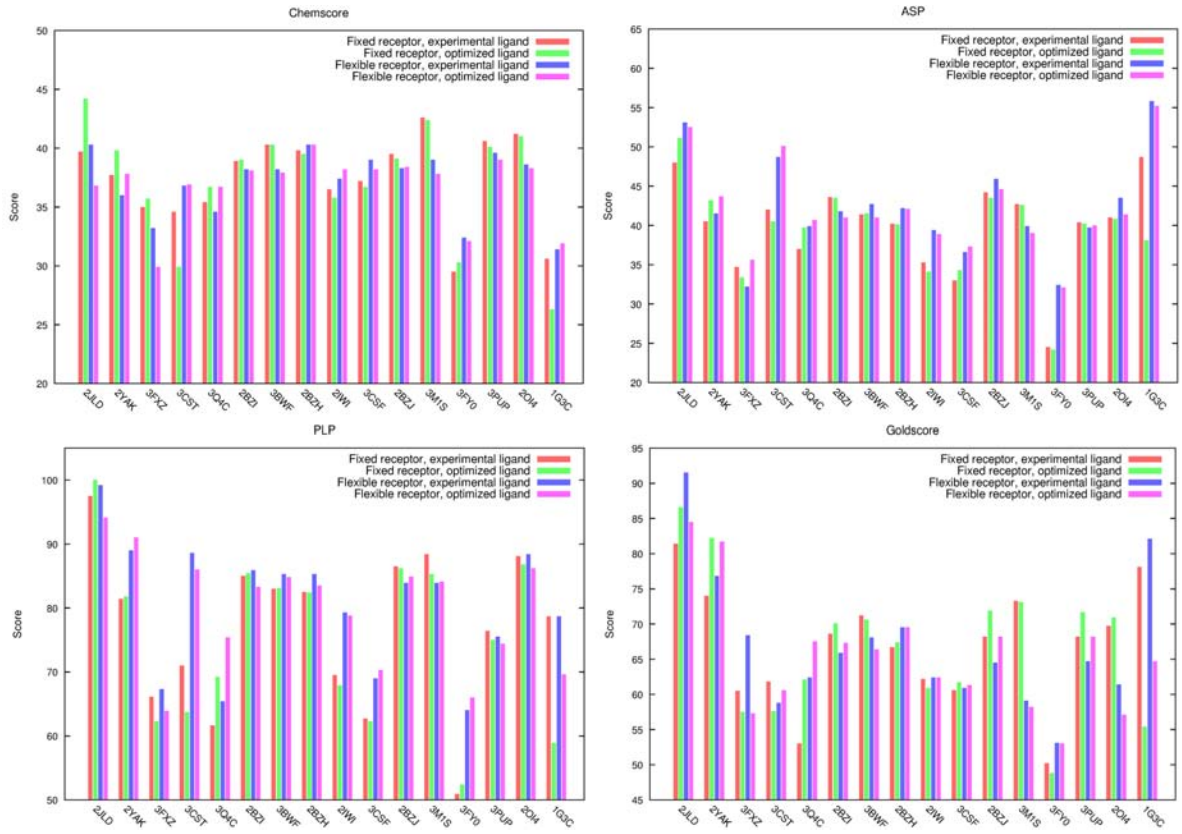


Figure 4.10: Score energies for the lower energy structure for the four different configurations of the protein-ligand docking Simulation for the available Scoring Functions in GOLD 5.1.

here, one can see an agreement between them in 9 of the total 16 cases, being 3 of them the most challenging ones (where the binding site is exposed to the solvent). The coincidence of results show that doing only an energetic analysis is not possible to determine which is the best scoring function. Then, the next step is correlating the scoring results with an experimental values in order to get evidences of which of the four available scoring function is the most favourable in the prediction of the interaction of proteins with organometallic ligands.

4.4.5 Fitting theory and experiment upon binding

After the observation of the structural and energetic effects of each scoring function in all four docking schemes, the next step is performing a comparative study between experimental evidences and theoretical results. In this case, the IC_{50} value is the chosen experimental variable to compare with the score values calculated in the precedent section. This coefficient represents the half maximal inhibitory concentration and indicates the concentration of the inhibitor to needed to inhibit the half of the biological activity. As lower is the number, best inhibition power will have the compound.

Only 7 of the total 16 inhibitors used in this study have the experimental value of IC_{50} reported in the bibliography (see Table 4.2), which corresponds to the structures 1 to 7. In this section,

these seven complexes were used as a training set to test the energetic predictiveness of GOLD. Despite of the low number of members of the training set, the correlation of the IC_{50} value and the score will shed some light upon the capacity of docking methods in the prediction of the binding of organometallic compounds inside proteic hosts.

Table 4.2: Description of the organometallic ligands of this study.

ID	PDB code	Receptor	IC_{50} (nM)
1	2JLD	GSK-3	0.04
2	2YAK	DAPK1	2
3	3FXZ	PAK1	130
4	3CST	PI3K	0.04
5	3Q4C	BRAF	370
6	2BZI	PIM1	0.2
7	3BWF	PIM1	0.2

The correlation of the score with the IC_{50} value has been fitted to a linear regression, which follows the next equation:

$$Score = ln(IC_{50})m + b \tag{4.1}$$

Where m and b are constants.

As an example, in Figure 4.11 one can see two plots which correspond to simulations performed with the ASP scoring function. The correlation coefficients of both (and the ones obtained with the other scoring functions) can be seen in Table 4.3.

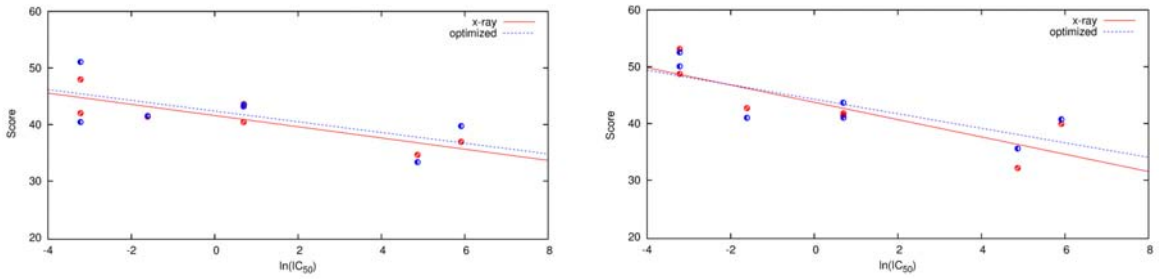


Figure 4.11: ASP Score versus ln(IC_{50}) for x-ray (red) and optimized (blue) ligand geometries, and fixed (left) and flexible (right) binding site of the receptor.

The regression coefficient is highly variable and goes from from R^2=0.03 to R^2=0.84. One of the reasons of this huge variation is the reduced set used to perform the correlations. Indeed, an unique outlier is able to impact on the regression trend.

Table 4.3: Regression coeficients for each docking scheme and fitting functions for all four docking schemes.

	ChemScore	ASP	ChemPLP	GoldScore
I	0.28	0.77	0.51	0.09
II	0.03	0.45	0.23	0.10
III	0.75	0.64	0.84	0.49
IV	0.32	0.54	0.54	0.19

First, all the simulations carried out with the experimental geometry of the ligand (docking schemes I and III) present better performance than those with the pre-processed ligand (docking schemes II and IV). As a clear example, calculations performed with the ASP scoring function have a good agreement with the experiment when the experimental structure of the ligand is used (R^2 = 0.77) which drops drastically if the preprocessed structure is the chosen one (R^2 = 0.45). This is a clear example of how the structural rearrangements promoted by the DFT minimization can affect the predictiveness of protein ligand docking techniques.

Additionally, in the precedent section is demonstrated how the results are improved if the flexibility of the binding site of the host is considered. Comparison of results obtained in docking schemes I and II with those obtained in the docking schemes III and IV also validate the hypothesis. In the specific case of ChemPLP, adding flexibility to the host improves the results from the non correlation (R^2 = 0.51) to the highest desirable fitting between theory and experiment (R^2 = 0.84).

Fitting values between experiment and theory is possible to check if the obtained results in the previous tests, where the structure of the docked complex and the score of the best matched solutions were evaluated, are correct. On the one hand, ChemScore and ChemPLP fitting functions present a better correlation with experimental data. However, ASP fitting function is more robust than the others, having reasonably good correlation values in two of the four docking schemes. On the other hand, taking into account the flexibility of the binding sites also improves the result. This point is also seen comparing the correlation of the four different docking schemes.

To sum up, the accuracy of the four scoring functions in terms of agreement between score and the experimental IC_{50} value follows the next order: ASP, being the most robust because of its reproducibility independently from the experiment, is better than ChemPLP, which is reasonably better than ChemScore. GoldScore presents the worst correlations. In fact, this scoring function is not recommended.[33] About the flexibility of the host, is important to take it into account, especially if the ligand is preprocessed and its geometry is quite far from the experimental one.

4.5 Conclusions

It is well known that the conjunction between theory and experiment is fundamental to improve the expansion of the modern fields of research. The development of metallodrugs is not an exception but sometimes technical limitations in modelling are one step back from the current state of experimental research. Molecular modelling tools are a good complement on the study of the interaction of

organometallic ligands interacting with biological hosts but sometimes dealing with the metal is not an easy job. In this chapter, a new protocol of how describe the binding between inert first coordination sphere ligands and a proteic host using protein-ligand docking methods has been tested. To do so, an evaluation the effectiveness of the scoring functions implemented in GOLD in predicting the structural features of the binding of organometallic compounds to biological partners has been performed, as well as other considerations typical of docking experiments. The set of systems developed by Prof. Meggers and co-workers was computationally reproduced using the protein-ligand docking approach. From a structural point of view, protein-ligand docking is able to reproduce the correct orientation of the organometallic species in the binding site of the receptor with a success rate up than 75%. The scoring function performing best matching solutions with the experimental structures is ChemScore, followed by ASP and ChemPLP, being GoldScore the poorer one. When the calculated binding constant (score) is compared with the experimental IC_{50} value by means a linear regression, the R^2 value reaches 0.8 for the best energetic correlations. Energetically, ASP fitting functions is the most robust one, followed close by ChemPLP and ChemScore. Again, GoldScore underperforms the correlation presenting results of poor quality. This study is, to our knowledge, the first work on how the interaction of organometallic compounds with biological entities can be predicted by means of protein-ligand docking approaches.

4.6 References

[1] Seebeck, B.; Reulecke, I.; Kämper, A.; Rarey, M. *Proteins* **2008**, *71*, 1237–1254.

[2] Zhang, J.; Yang, P. L.; Gray, N. S. *Nat. Rev. Cancer* **2009**, *9*, 28–39.

[3] Becker, W. M.; Kleinsmith, L. J.; Hardin, J. *World of the Cell*; Benjamin-Cummings Publishing Company, 2012.

[4] Nair, P. *Curr. Sci.* **2005**, *88*, 890–898.

[5] Zhang, L.; Carroll, P.; Meggers, E. *Org. Lett.* **2004**, *6*, 521–523.

[6] Feng, L.; Geisselbrecht, Y.; Blanck, S.; Wilbuer, A.; Atilla-Gokcumen, G. E.; Filippakopoulos, P.; Kraling, K.; Celik, M. A.; Harms, K.; Maksimoska, J. et al. *J. Am. Chem. Soc.* **2011**, *133*, 5976–5986.

[7] Meggers, E. *Chem. Commun.* **2009**, 1001–1010.

[8] Atilla-Gokcumen, G. E.; Pagano, N.; Streu, C.; Maksimoska, J.; Filippakopoulos, P.; Knapp, S.; Meggers, E. *ChemBioChem* **2008**, *9*, 2933–2936.

[9] Xie, P.; Williams, D. S.; Atilla-Gokcumen, G. E.; Milk, L.; Xiao, M.; Smalley, K. S. M.; Herlyn, M.; Meggers, E.; Marmorstein, R. *ACS Chem. Biol.* **2008**, *3*, 305–316.

[10] Xie, P.; Streu, C.; Qin, J.; Bregman, H.; Pagano, N.; Meggers, E.; Marmorstein, R. *Biochemistry* **2009**, *48*, 5187–5198.

[11] Maksimoska, J.; Feng, L.; Harms, K.; Yi, C.; Kissil, J.; Marmorstein, R.; Meggers, E. *J. Am. Chem. Soc.* **2008**, *130*, 15764–15765.

[12] Bullock, A. N.; Russo, S.; Amos, A.; Pagano, N.; Bregman, H.; Debreczeni, J. E.; Lee, W. H.; von Delft, F.; Meggers, E.; Knapp, S. *PLoS one* **2009**, *4*, e7112.

[13] Atilla-Gokcumen, G. E.; Di Costanzo, L.; Meggers, E. *J. Biol. Inorg. Chem.* **2011**, *16*, 45–50.

[14] Pagano, N.; Maksimoska, J.; Bregman, H.; Williams, D. S.; Webster, R. D.; Xue, F.; Meggers, E. *Org. Biomol. Chem.* **2007**, *5*, 1218–1227.

[15] Maksimoska, J.; Williams, D. S.; Atilla-Gokcumen, G. E.; Smalley, K. S. M.; Carroll, P. J.; Webster, R. D.; Filippakopoulos, P.; Knapp, S.; Herlyn, M.; Meggers, E. *Chem.-Eur. J.* **2008**, *14*, 4816–22.

[16] Feng, L.; Geisselbrecht, Y.; Blanck, S.; Wilbuer, A.; Atilla-Gokcumen, G. E.; Filippakopoulos, P.; Kräling, K.; Celik, M. A.; Harms, K.; Maksimoska, J.; Marmorstein, R.; Frenking, G.; Knapp, S.; Essen, L.-O.; Meggers, E. *J. Am. Chem. Soc.* **2011**, *133*, 5976–5986.

[17] Toyota, E.; Ng, K. K.; Sekizaki, H.; Itoh, K.; Tanizawa, K.; James, M. N. *J. Mol. Biol.* **2001**, *305*, 471–479.

[18] Jones, G.; Willett, P.; Glen, R. C. *J. Mol. Biol.* **1995**, *245*, 43–53.

[19] Jones, G.; Willett, P.; Glen, R. C.; Leach, A. R.; Taylor, R. *J. Mol. Biol.* **1997**, *267*, 727–748.

[20] Korb, O.; Stützle, T.; Exner, T. E. *J. Chem. Inf. Comp. Sci.* **2009**, *49*, 84–96.

[21] Eldridge, M. D.; Murray, C. W.; Auton, T. R.; Paolini, G. V.; Mee, R. P. *J. Comput. Aid. Mol. Des.* **1997**, *11*, 425–445.

[22] Baxter, C. A.; Murray, C. W.; Clark, D. E.; Westhead, D. R.; Eldridge, M. D. *Proteins* **1998**, *33*, 367–382.

[23] Verdonk, M. L.; Cole, J. C.; Hartshorn, M. J.; Murray, C. W.; Taylor, R. D. *Proteins* **2003**, *52*, 609–623.

[24] Mooij, W. T. M.; Verdonk, M. L. *Proteins* **2005**, *61*, 272–287.

[25] Lee, C.; Yang, W.; Parr, R. G. *Phys. Rev. B* **1988**, *37*, 785–789.

[26] Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 5648–5653.

[27] Fogarasi, G.; Zhou, X.; Taylor, P. W.; Pulay, P. *J. Am. Chem Soc.* **1992**, *114*, 8191–8201.

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

[29] Frisch, M. J. et al. *Gaussian 09 Revision D.01*, Gaussian Inc. Wallingford CT 2009.

[30] Fischer, M.; Coleman, R. G.; Fraser, J. S.; Shoichet, B. K. *Nature Chem.* **2014**, *6*, 560–561.

[31] Sousa, S. F.; Fernandes, P. A.; Ramos, M. J. *Proteins* **2006**, *65*, 15–26.

[32] Dunbrack, R. L. *Curr. Opin. Struc. Biol.* **2002**, *12*, 431–440.

[33] Liebeschuetz, J. W.; Cole, J. C.; Korb, O. *J. Comput. Aid. Mol. Des.* **2012**, *26*, 737–748.

5 ... Active first coordination sphere: resting state and activation process of the Corynebacterium diphteriae Haem Oxygenase

Related paper:
Appendix A.1: What can molecular modelling bring
to the design of artificial inorganic cofactors?
Appendix A.4: Unravelling novel synergies between organometallic
and biological partners: a quan- tum mechanics/molecular mechanics
study of an artificial metalloenzyme

I'm not going to limit myself just because people won't accept the fact that I can do something else.

Dolly Parton

5.1 Tuning proteins with metals

Proteins are involved in a vast array of biochemical processes in living organisms. This includes catalysis, transport and recognition. A special case are enzymes, which are proteins involved in catalysis. Usually, the catalytic process comprises the binding of chemical species (small or large molecules) in a specific part of the host, the active site. In the specific case in which an inorganic cofactor is present in the binding site, the system receives the name of metalloenzyme. Metalloenzymes are not an isolated group of biological entities; they represent nearly half of all proteins in nature. Metalloenzymes have been converted in a gold mine in the field of bioinorganic chemistry being nowadays a vivid area of multidisciplinary research.

5.1.1 Artificial Metalloenzymes: mixing the best from inorganic and biological fields

Organometallic chemistry can provide versatile catalysts although most have low specificity and are generally not eco-friendly. On the other side, natural enzymatic processes are selective and specific in nature. How can both fields be mixed to achieve the best catalyst? Using artificial metalloenzymes, the versatility of organometallic compounds is combined with the selectivity of enzymes in the quest for the perfect catalysts.[1]

Artificial metalloenzymes can be constructed in many different ways: two of them are the *de novo*-designed scaffols or the use of native proteins. In the case of *de novo*-designed proteins, a polypeptide sequence is constructed without any pattern. This step is not obvious because the number of possible sequences is enormous: for a sequence of "n" residues, the complete number of different systems that can be constructed is 20^n (i.e. for a sequence of 100 residues the total number of possible sequences is $5.4{\cdot}10^{44}$).[2] Native scaffolds skip the tedious work of finding the correct sequence. However, they have a pre-organized binding pocket configuration, which can difficult the recognition of the inorganic cofactor.

To the design of artifical metalloenzymes, one have to ensure the localization of the cofactor inside the host protein. Two main strategies are used to solve efficiently this task, which are the covalent and non-covalent inclusion of a metal-containing cofactor. On the one hand, covalent incorporation entails a chemical modification of specific amino acids. This is the case of the copper(II)-based antibody sintetized by Janda and coworkers, in which the sidechain at the active-site lysine of the antibody 38C2 has been changed by an imide-linked bisimidazole one.[3] On the other, no chemical transformation is done, neither in the host nor the cofactor. One example of this kind of host-cofactor interaction is the insertion of Cr(III) complexes into the active site of apomyoglobin, reported by Watanabe.[4]

In metalloenzyme design, the structural information that experimentalists can obtain is insufficient to go ahead in the research. Molecular modelling represents an interesting alternative to

describe the structural features of possible candidates. One example of an alliance between experiment and theory can be seen in this work and its precedents,[5,6] where a combination of protein-ligand docking, statistics, quantum mechanics and quantum mechanics/molecular mechanics techniques were applied in order to get an accurate prediction of the binding pose, the resting state and the activation process of a novel metalloenzyme syntetized by Ueno and coworkers (PDB code: 1WZD).[7]

5.1.2 A case study: Artificial Haem Oxygenases

The modification of naturally existing proteins has a strong repercussion in biochemical, biomedicine and pharmaceutical areas. Well known enzymes can be modified to provide them new catalytic features. This is the case of Haem Oxygenases, which have been used as template for metalloenzyme design. These enzymes, were reported by first time in 1968 by Schmid and coworkers[8] and are still studied nowadays in several projects including drug design processes.[9] The main biological activity of Haem Oxygenase is to catalyse the oxidative degradation of iron-protophorphyrin to biliberdin (a bilirubin precursor), Fe(II) and carbon monoxide. [a] Haem molecule can also be degraded by using a reducing agent, needed to activate O_2 and reduce the oxidation state of Fe(III) to Fe(II).[10]

In 2006 Ueno and coworkers reported a new metalloenzyme incorporating a Fe(Schiff base) into the heme cavity of *Corynebacterium diphteriae* heme oxygenase (*cd***HO**).[7] Its X-ray structure shows a non-expected conformation, notably different from the one adopted by the coordination of a Haem group in the natural enzyme.

The first difference between heme group and Fe(Schiff base) is the chelation to the enzyme: haem complexes binds to the host by a coordination bond between His20 and the metal centre. The Fe(Schiff base)·*cd***HO** is also coordinated to His20 but, additionally to the residue Glu24 (see Figure Figure 5.1). The second difference comes from its distorted octahedral structure: the equatorial environment of the metal corresponds to a N2O of the Schiff Base plus the O_γ for Glu24, and the axial positions are occupied by the remaining oxygen from the Schiff base plus the N_ε from His20. This last difference has implications on the chemical activity of Fe(Schiff base) which is unable to coordinate an external molecule in the axial position of the inorganic cofactor because of the lack of any vacancy of the metal and the distorted shape of the cofactor.

Both natural and artificial systems have common critical processes in their catalytic mechanisms. The Fe(III) atom has to be reduced by means of the action of the cytocrome P450 reductase, being the Fe(II) atom able to coordinate a molecule of dioxygen. At this point, differences between the natural and the artificial compound start. While the oxidative process of the natural hemeoxygenase ends with the cleavage of the heme group, the artificial enzyme stops its activity once the oxygen is activated to form superoxide. The activity of the artificial enzyme acting as superoxidase differs from the natural function of the *cd***HO**, something surprising taking into account the high chemical similarity between porphyrinic cofactors and salophen Schiff bases.

[a] In humans, this reaction is the only one which produces carbon monoxide.

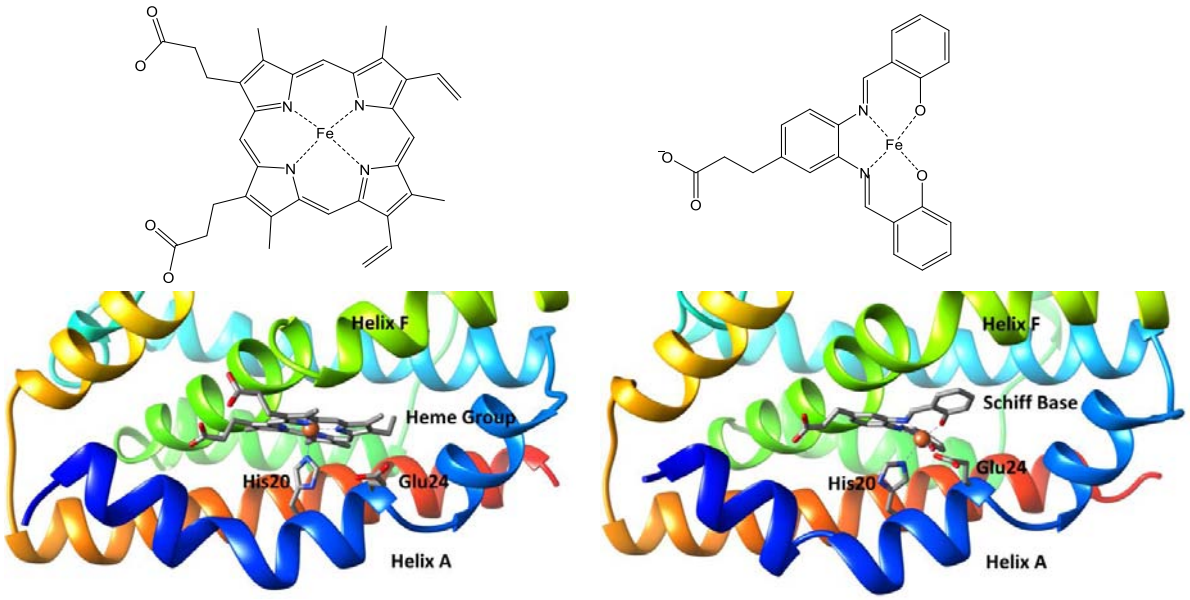


Figure 5.1: X-ray structure of Haem (left) and Fe(Schiff base) (right) bound heme oxygenase

5.2 Previous work in the structural determination of cdHO

The computational prediction from scratch of the most stable Fe(Schiff base) cd**HO** represents a complex exercise in modelling. An extensive exploration of the conformational space has to be carried out in order to find any evidence of how the organometallic moiety interacts with the protein receptor. Here, protein-ligand docking approaches seems to be, apparently, the optimum method to apply. However, this methodology is unable to simulate the changes that could occur in the first coordination sphere of the metal. Increasing the level of theory to quantum mechanics, is possible to accurately predict the first coordination sphere phenomena but the computational cost of a complete conformational exploration is not affordable.

In 2010, in our group we designed an integrative computational approach able to predict organometallic compound-protein interactions, being tested in the *Corynebacterium diphteriae* Haem Oxygenase system.[5] Because of the huge exploration of the conformational space that this work demands and the accuracy needed to fully describe the behaviour of the organometallic ligand, no computational tool available at this time was suitable to resolve the system. Our idea was based in the "divide and conquer" methodology, dividing all the work in four steps using, each of them, a different molecular modelling tool:

1. Preliminary treatment of the organometallic cofactor by means of quantum mechanical geometry optimization processes. The starting geometry of the ligand was taken from the original PDB X-ray file of the Fe(II)(Schiff base)*cd***HO**.

2. Exploration of the conformational space using protein-ligand docking techniques. The struc-

ture of the protein was extracted from the PDB X-ray Fe(II)(Schiff base)*cd***HO**, and the geometry of the ligand is the one obtained in the precedent step. To choose the binding site not only the one observed in the PDB file was explored. Other possibilities accounting the total volume of the ligand were also evaluated. Also, due to the incapability of protein-ligand docking techniques to carry on with metals, the Fe atom is exchanged by a pseudometal atom type able to reproduce qualitatively the metal-ligand electrostatic interactions.

3. Identification of the residues that could alter the first coordination sphere of the metal upon binding. Here, the residues that are able to coordinate to the metal are evaluated and weighted depending on two spatial features: the closeness of its C_α to the metal, and the facilities of the side chain to adopt an orientation were the coordination of the metal is possible.

4. Refinement of the docked structures by means of hybrid Quantum Mechanics/Molecular Mechanics (QM/MM) energy minimization procedures. The two best binding poses predicted in the Protein-Ligand Docking step were optimized in order to find the most stable one, which is the final result of all the protocol.

Following this methodology, we succeed in the reproduction of the pose of the ligand inside the host, being the root mean square deviation (RMSD) value between the experimental structure and the predicted one only 0.5Å.

5.3 The quest for the resting state and the activation process of Fe(Schiff base)cdHO

As a consequence of the previously described work, we kept interested in the same artificial metalloenzyme due to the experimental observations given by Ueno and coworkers.[7] As we introduce previously, they expected similar behaviour between the natural haemoxygenase and the new artificial one, because of the high structural complementaries between both complexes. However, in the case of natural haemoxygenases, the entire oxidative process is performed ending with the cleavage of the haem, but the Fe(Schiff base) bound system stops its activity in the activation of oxygen step. The crystal structure of the complex presents a distorted octahedral geometry of the cofactor which do not corresponds to an active form, having all the valences of the Fe center fulfilled and no room on the axial position to allow the binding of an external ligand.

In this section we will embark on the description of the electronic state of the cofactor in the reported X-ray structure, which will correspond to the resting state of the system, followed by the investigation on a possible activation process. All these studies have been performed by means of the application of Quantum Mechanics/Molecular Mechanics techniques in the description of chemobiological processes.

5.3.1 Computational methodology

This work have been entirely performed with the Quantum Mechanics/Molecular Mechanics approach using the ONIOM (DFT:AMBER) method[11] implemented in Gaussian09.[12] This technique allows to differentiate the system in two regions: the part of the system where the reaction is taking part (i.e. binding site), which is described by quantum mechanics, and the rest of the system, simulated with molecular mechanics. To a correct description of the system at different stages of the work, two different QM/MM layouts have been used (see Figure 5.2). The bigger one (P1 in Figure 5.2) is the default for all work long and it is constituted by all the Fe(Schiff base) except the propionate tail, meanwhile the small one (P2) is formed only by the first coordination of the metal (the N2O2 moiety, the linking carbons and the proper hydrogen atoms) being useful to study the effects of the first coordination sphere in the latest part of the work. In both subsystems His20 and Glu24 residues are included in the QM part in their 4-methylimidazole and propionate form, respectively. Also, the flexibility of the system was defined, being the cofactor and part of the Helix A of the protein free to move. The rest of the bioinorganic entity remain fixed (see Table 5.1 for more details).

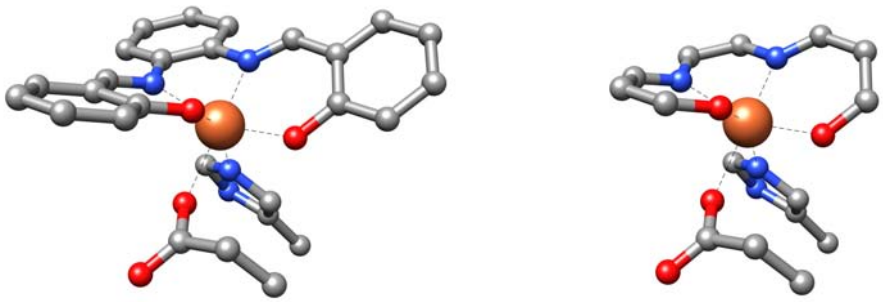


Figure 5.2: Representation of the QM part of the two QM/MM layouts used in this study: the big one (referred as P1) on the left and the small one (P2) on the right.

Table 5.1: Breakdown of the total number of atoms that constitutes the entire system.

Total	Flexible	QM partition (P1)	QM partition (P2)
3347	192	55	37

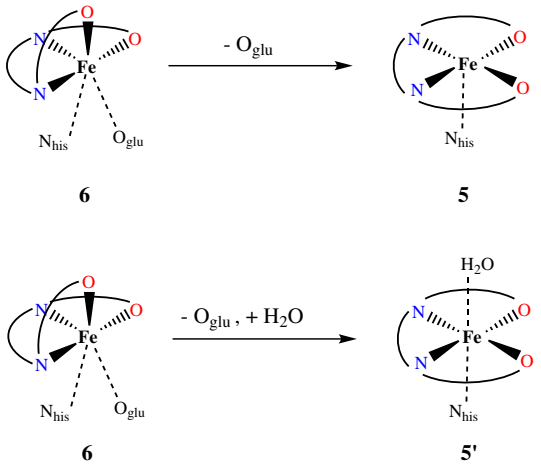
QM/MM minimizations were undertaken at B3LYP[13,14] DFT level of theory using a triple-ζ basis set for all the atoms: 6-311+g*[15] for the main group elements and aug-cc-pVTZ[16,17] for the iron atom. In addition, the reliability of the B3LYP results was tested by means of single point calculations on the optimized structures using other DFT functionals as M06L[18], PBE[19] and B97D[20], and also testing the effect of using electronic embedding in the B3LYP functional. Due to the high similarity of relative energies with other functionals and accounting the embedding effect, B3LYP with no electronic embedding applied will be the level of theory used in all the study. The MM calculation was performed with AMBER 99.[21]

All the calculations have been performed for Fe(II) and Fe(III) in their low, intermediate and high spin configurations (multiplicities 1, 3, 5 and 2, 4, 6 for Fe(II) and Fe(III), respectivelly).

The starting geometry of the hexacoordinated complex is the original one from its X-ray structure. From now, this complex will be defined as **6**. To create the pentacoordinate structure, the rotamer of the Glu24 have been changed for other out of the first coordination sphere using the Dunbrack rotamer library,[22] also included in the UCSF Chimera package. From now, this structure is referred as **5**.

Charges and protonation states were automatically determined using the UCSF Chimera package[23] with the exception of the His20 chelation to Fe atom, which was manually defined. The entire complex has a total charge of -7 or -8, depending on the oxidation state of the metal. In all the cases, the total charge of the Molecular Mechanics part is divided in -10 for the isolated protein, and +2 or +3 for the organometallic ligand. The charge of the QM partition is for both P1 and P2 cases +1 in the case of Fe(II) and +2 for Fe(III).

Additionally, to test the capability of the complex **5** to coordinate an external ligand in the axial position, calculations of the **5** complex with a water molecule linked in the axial vacancy were also performed. This water molecule was included in the Quantum Mechanics partition of the ONIOM scheme. From now, this complex is referred as complex **5'**. See Scheme 5.1 and Figure 5.3 for a visual description.



Scheme 5.1: Schematic representation of 6, 5 and 5'.

5.3.2 Determination of the resting state

To determine the resting state of *Corynebacterium diphteriae* Haem Oxygenase (Fe(Schiff base)*cd***HO**), the X-ray structural geometry has been used as the starting point. Here, the inorganic cofactor presents an unusual highly distorted octahedral configuration due to the coordination of two residues of the protein (His20 and Glu24) to the metal centre. With this geometry, the enzyme is not able to coordinate molecular oxygen, which is the next step in its catalytic activity. This

Figure 5.3: Structural differences between the hexacoordinated (**6**) (left) and pentacoordinated (**5**) (right) structures of the *Corynebacterium diphteriae* Haem Oxygenase system.

fat suggests that the resting state of the complex under study is different from other Fe(III) haem containing enzymes, where only one residue from the receptor is bound to the metal leading to a resulting square pyramidal geometry which allows the coordination of an external ligand because of the presence of a free vacancy on the axial position of the cofactor.[24–27]

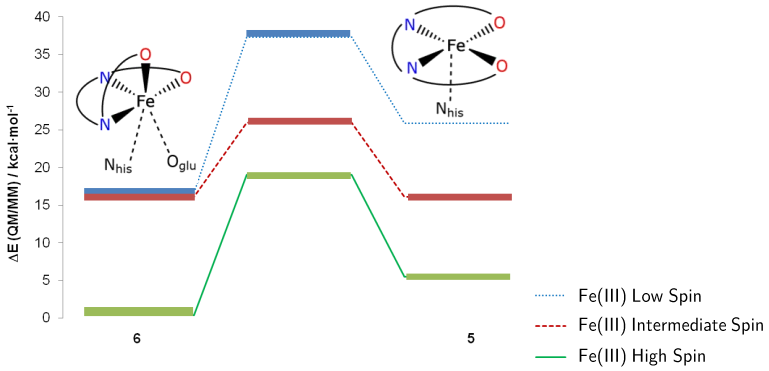
Calculations have been carried out on the crystallographic system (**6**), as well as the pentacoordinated one (**5**) and that with a water molecule on the axial position (**5'**). At this point, calculations have been performed with Fe(III) taking into account different spin states of the Fe metal ion: doublet, quadruplet and sextet.

Optimized geometries of the system **6** do not show any particular difference respect the X-ray structure for each spin state. However, the structure of high spin configuration complex is the closest one to the experimental geometry. The removal of Glu24 away from the first coordination sphere (structure **5**) produced a reorganization of the octahedral geometry to a square pyramidal one, where the salen moiety is almost planar. The host is also modified after the geometry optimization process: Helix A is reorganized pointing its C_{ter} end of the helix to the solvent instead of remaining pointing inside the protein, as can be observed in the experimental structure. **5'** structure suffer almost the same deformations that the **5** one, being here the cofactor slightly more planar. Water molecule only remains coordinated to Fe(III) in its low spin configuration.

Energetically, in the case of the hexacoordinated geometry, the high spin system is the most stable one being 16 kcal/mol lower in energy respect the other spin configurations. Additionally, for the same spin state, the energy differences between **6** and both pentacoordinated structuresto **6** by 5.1 and 8.0 kcal/mol for **5** and **5'** respectively.

The transition path was therefore investigated. As the water chelating the iron in 5 0 is dissociated in the high- and intermediate-spin function, further calculations were carried out only on the transition between **6** and **5**.Transition states have been characterized at Fe-O_{glu} distances of 3.01, 2.55 and 3.37 Å for low, intermediate and high spin complexes respectively. An unique nega-

tive vibration mode was observed for all the electronic configurations and corresponds to the same movement: Fe-O_{glu} bond is cleaved by a movement of the Glu24 residue which now points to the solvent, the Fe(Schiff base) becomes planar and the Helix A is reorganized. Transition state are located 20.7, 9.8 and 18.6 kcal/mol above the reactant structure for low, intermediate and high spin species respectively. All of them are sufficient large to discard a transition from **6** to **5** at room temperature (Scheme 5.2).



Scheme 5.2: Simplified representation of the potential energy surface for the transition from configuration 6 to configuration 5 Fe(III)(Schiff base) . *cd*HO in doublet, quadrupet and sextuplet states.

In this part of the work the structure of the experimental resolved X-ray Fe(Schiff base)*cd***HO** has been tested as resting state. The results point out that its electronic state corresponds to a Fe(III) configuration in a sextuplet state. Differences in energy between other electronic states of the metal show that an equilibrium between **6** and **5** is not reachable in solution. Then the activation process would be triggered by a reduction of Fe(III) to Fe(II) which involves the participation of the first coordination sphere. This rearrangement prepares the system for the next step: the activation process.

5.3.3 Study of the activation process

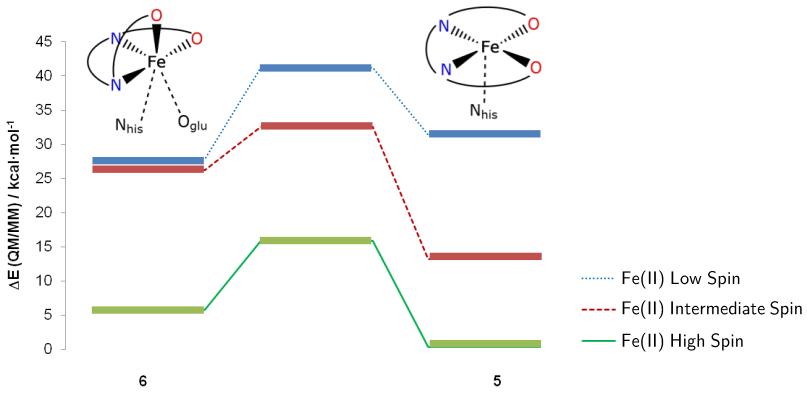
In the previous section, the energetic path from **6** to **5** shows that the pentacoordinated structure is not energetically favourable for Fe(III) systems. Also the geometry of the experimental Fe(Schiff base) (**6**) is not optimal for the activation process because the oxygen molecule can not reach the Fe atom due to the organization in the first coordination sphere. Geometry minimizations performed on the pentacoordinated structures (**5**) show how the Fe(Schiff base) becomes almost planar in its N2O2 moiety allowing the formation of a room in the axial position of the inorganic cofactor able to enclose a small ligand.

To proceed with the study of the activation state, a reduction of the metal have been supposed as first stage. Quantum Mechanics / Molecular Mechanics geometry optimization of Fe(II) **6**, **5** and **5'** species have been performed to study a possible activation process. Low (singlet), intermediate (triplet) and high spin (quintuplet) states of the metal are also taken into account in this part of

the study.

The general structural profiles of these species are similar than those observed in Fe(III) structures. Metal-ligand distances of the first coordination sphere are slightly longer (less than 0.1 Å for salen atoms and 0.5 Å as much for the His20 coordination) due to the lost of a positive charge which decreases the electrostatic energy of the mentioned bond. The removal of the Glu24 in **5** provokes a reorganization of the Helix A, due to its accommodation in a stable rotameric pose away from the first coordination sphere of the metal. Fe(Schiff base) is also rearranged from a distorted octahedral structure to a square pyramidal one: the salen ring of the Schiff base adopts a planar distribution forming a vacancy in the axial position which can host a small ligand. **5'** structure show the same behaviour than in the Fe(III) case. As in the case of Fe(III) only the low spin system keeps the water molecule bonded to the metallic centre.

Optimizations show that the transition between **6** to **5** is exothermic for all low, intermediate and high spin configurations. The high spin complex is the most energetically favourable, being 12.8 and 13.9 kcal/mol lower in energy than intermediate an low spin systems respectively (see Scheme 5.3). For the transition between **6** and **5'**, high spin wave-function also remains the most probable spin state. The transition path were therefore investigated. As equal than the precedent section, only the transition from **6** to **5** is calculated.



Scheme 5.3: Simplified representation of the potential energy surface for the transition from configuration 6 to configuration 5 Fe(II)(Schiff base) . *cd*HO in doublet, quadrupet and sextuplet states.

The transition state has been characterized at Fe-O_{glu} distances of 3.44, 2.73 and 3.02 Å for low, intermediate and high spin complexes respectively. The general features of the lowest vibration mode are similar to those of the ferric system as reported in the previous section. In addition, the transition state of these different species are located at 13.4, 2.5 and 10.1 kcal/mol above the hexacoordinated structure for low, intermediate and high spin systems, in the same order.

Both, resting state and activation process were yet described. In a brief, once the metal ion is reduced from Fe(III) to Fe(II), the removal of the Glu24 becomes exothermic. In all the process, not only the organometallic centre is affected, also a rearrangement of the helix A has to be performed in order to accommodate the released Glu24 and the new coordination state of the metal. However,

which of these features dominates in the process is yet to be determined.

5.4 Who drives the process?

In this section, a deeper study of the system have been performed to get some information about the effect of the organometallic moiety in the entire binding process process. The contribution of the first coordination sphere and also the protein-cofactor complementaries have been evaluated. At this point two different analysis were performed: the first one shows the effect of only the first coordination sphere in front of the rest of the system, and the last corresponds to the study of both inorganic and biological entities individually.

5.4.1 First coordination sphere vs. entire system

In order to computationally differentiate the first coordination sphere from the rest, the QM/MM partition used at this stage is P2 (previously described in Figure 5.2). The complete path from **6** to **5** have been calculated for Fe(II) and Fe(III), both in its low, medium and high spin configuration. The resulting QM/MM energy have been decomposed in its Quantum Mechanical (QM) and Molecular Mechanical (MM) terms. Figure 7.5 shows the relative contribution of both parts, as well as the total QM/MM energy.

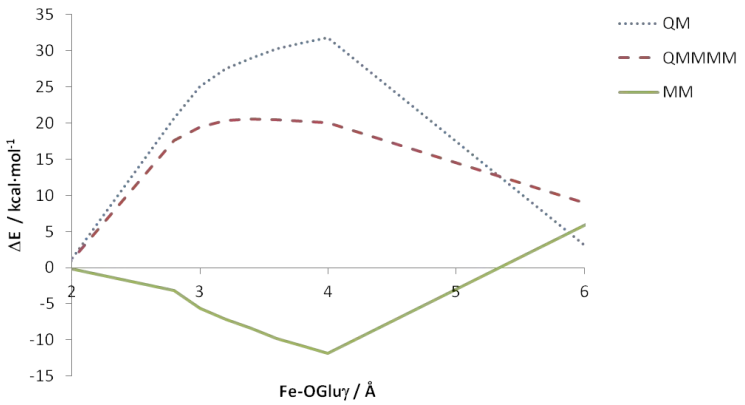


Figure 5.4: QM/MM Energy breakdownfor Fe(III) High Spin system in the **6** to **5** transformation.

Results show that the first coordination sphere (QM part) dictates the shape of the overall profile, being slightly modulated by the MM term. However, it is still interesting to note that generally the MM part remains almost unaffected or even stabilized in the first part of the Fe-O_{glu} bond cleavage. In the specific case of Fe(III) high spin system (the resting state) once reached the transition state, a destabilization of the system can be observed, reaching the 15 kcal/mol.

An structural analysis, displays the following evidences: at the first part of the transition the Helix A is relaxed while the glutamate is removed; then, the glutamate is forced to point to the

solvent and the Helix A is reorganized. Clashes between the flexible and fixed parts of the protein also destabilize the system, disfavouring the MM contribution on the transition path.

5.4.2 Protein-Ligand complementarities

The previous discussion was focused on the effect of the first coordination sphere in the reactivity of the system. Here, both ligand and protein were detached to investigate their energetic changes during the transition from **6** to **5**. In general, for all the studied electronic configurations of the metal, the energy of both parts decreases when the Fe-O_{glu} distance grows (see Figure 5.5).

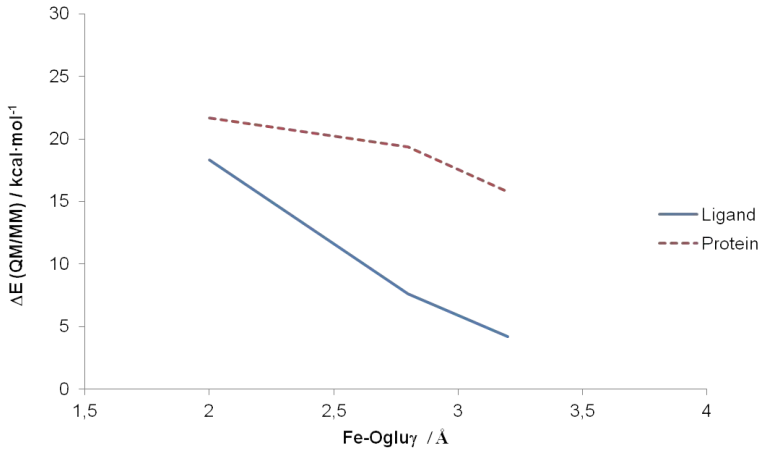


Figure 5.5: Example of the energetic path for both protein and ligand (Fe(III) High Spin) systems along the transition from **6** to **5**.

The removal of Glu24 has a stabilizing effect in both entities, even in the case of Fe(III) complexes, were the same reaction coordinate is endothermic for the attached complex, as we seen in previous sections (Scheme 5.2). The Fe(Schiff base) adopts a planar configuration when the Glu24 is removed, releasing a free vacancy at the bottom side of the cofactor. Now the cofactor has not any sterical effect which could block the relaxation of the organometallic compound. For the protein, the change of the rotameric state of Glu24 induces a reorganization of the entire Helix A. This rearrangement also makes the C_{ter} of the Helix A pointing towards the solvent, adopting a conformation which is consistent with current knowledge on porphyrinic, salen and salophen moieties. [28]

5.5 Conclusions

In this study, both resting state and the activation process of the Fe(Schiff base) docked to *Corynebacterium diphteriae* Haem Oxygenase were characterized. Calculations present a good agreement with the experimental results of Ueno *et al*,[7] being the metal in oxidation state III in the resting state, and II in the activation process, in both cases with a high spin configuration. These electronic configurations have structural effects. The initial geometry of the inorganic cofactor is characterized by the coordination of two amino acids (His20 and Glu24) from the host, something

unexpected for Haem Oxygenases. Here, the ligand presents a distorted octahedral structure which remains stable for Fe(III) complexes but does not allow the coordination of any external ligand. After a reduction of the metal centre, the Glu24 residue is removed from the first coordination sphere releasing a free vacancy in the Fe(Schiff base), which now can be reached by an external ligand. This process is entirely controlled by the first coordination sphere but implies an overall structural change of the proximal helix and the conformation of the cofactor. This process, despite of being exothermic for both isolated cofactor and protein is not naturally occurring due to the structural restrictions imposed by the binding.
This work sheds light on the global effects that a organometallic ligand can provoke on the structure of its host. In addition, the capacities of Quantum Mechanics/Molecular Mechanics methods in the field of active coordination sphere processes have been successfully evaluated.

5.6 References

[1] Deuss, P. J.; den Heeten, R.; Laan, W.; Kamer, P. C. *Chem.-Eur. J.* **2011**, *17*, 4680–4698.

[2] Beasley, J. R.; Hecht, M. H. *J. Biol. Chem.* **1997**, *272*, 2031–2034.

[3] Nicholas, K. M.; Wentworth, P.; Harwig, C. W.; Wentworth, A. D.; Shafton, A.; Janda, K. D. *P. Nat. Acad. Sci. USA* **2002**, *99*, 2648–2653.

[4] Ohashi, M.; Koshiyama, T.; Ueno, T.; Yanase, M.; Fujii, H.; Watanabe, Y. *Angew. Chem. Int. Edit.* **2003**, *115*, 1035–1038.

[5] Robles, V. M.; Ortega-Carrasco, E.; Fuentes, E. G.; Lledós, A.; Maréchal, J.-D. *Faraday Discuss.* **2011**, *148*, 137–159.

[6] Ortega-Carrasco, E.; Lledós, A.; Maréchal, J.-D. *J. R. Soc. Interface* **2014**, *11*, 20140090.

[7] Ueno, T.; Yokoi, N.; Unno, M.; Matsui, T.; Tokita, Y.; Yamada, M.; Ikeda-Saito, M.; Nakajima, H.; Watanabe, Y. *P. Nat. Acad. Sci. USA* **2006**, *103*, 9416–9421.

[8] Tenhunen, R.; Marver, H. S.; Schmid, R. *P. Nat. Acad. Sci. USA* **1968**, *61*, 748.

[9] Motterlini, R.; Foresti, R. *Antiox. Redox. Sign.* **2014**, *20*, 1810–1826.

[10] Maines, M. D. *Annu. Rev. Pharmacolog.* **1997**, *37*, 517–554.

[11] Dapprich, S.; Komáromi, I.; Byun, K. S.; Morokuma, K.; Frisch, M. J. *J. Mol. Struct-.THEOCHEM* **1999**, *461*, 1–21.

[12] Frisch, M. J. et al. *Gaussian 09 Revision D.01*, Gaussian Inc. Wallingford CT 2009.

[13] Lee, C.; Yang, W.; Parr, R. G. *Phys. Rev. B* **1988**, *37*, 785–789.

[14] Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 5648–5653.

[15] Krishnan, R.; Binkley, J. S.; Seeger, R.; Pople, J. A. *J. Chem. Phys.* **1980**, *72*, 650–654.

[16] Dunning Jr, T. H. *J. Chem. Phys.* **1989**, *90*, 1007–1024.

[17] Kendall, R. A.; Dunning Jr, T. H.; Harrison, R. J. *J. Chem. Phys.* **1992**, *96*, 6796–6807.

[18] Y. Zhao,; D. G. Truhlar, *J. Chem. Phys.* **2006**, *125*, 1–18.

[19] J.P. Perdew,; K. Burke,; M. Ernzerhof, *Phys. Rev. Lett.* **1996**, *77*, 3865–3868.

[20] S. Grimme, *J. Comput. Chem.* **2006**, *27*, 1787–1799.

[21] Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. *J. Am. Chem. Soc.* **1995**, *117*, 5179–5197.

[22] Dunbrack, R. L. *Curr. Opin. Struc. Biol.* **2002**, *12*, 431–440.

[23] Pettersen, E. F.; Goddard, T. D.; Huang, C. C.; Couch, G. S.; Greenblatt, D. M.; Meng, E. C.; Ferrin, T. E. *J. Comput. Chem.* **2004**, *25*, 1605–1612.

[24] Sicking, W.; Korth, H.-G.; Jansen, G.; de Groot, H.; Sustmann, R. *Chem.-Eur. J.* **2007**, *13*, 4230–4245.

[25] Kuramochi, H.; Noodleman, L.; Case, D. A. *J. Am. Chem. Soc.* **1997**, *119*, 11442–11451.

[26] Aissaoui, H.; Bachmann, R.; Schweiger, A.; Woggon, W.-D. *Angew. Chem. Int. Edit.* **1998**, *37*, 2998–3002.

[27] Sono, M.; Roach, M. P.; Coulter, E. D.; Dawson, J. H. *Chem. Rev.* **1996**, *96*, 2841–2888.

[28] Meunier, B.; De Visser, S. P.; Shaik, S. *Chem. Rev.* **2004**, *104*, 3947–3980.

6 ... Novel insights on the molecular interactions of cisplatin on DNA and proteins

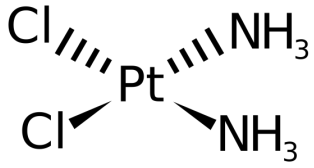
Related paper:
Appendix A.2: Computational insights on the possibility of tri-coordinated cisplatinated adducts with protein models

Science and everyday life cannot and should not be separated.

Rosalind Franklin

6.1 Cisplatin as anticancer drug and its side effects

Cisplatin is one of the most successful drugs in anticancer therapy. Its IUPAC name is cis-diamminedichloroplatinum(II) which corresponds to the formula *cis*-$Pt(NH_3)_2Cl_2$ (Scheme 6.1). Cisplatin was first reported in 1845 by Michele Peyrone in its work *"Ueber die einwirkung des ammoniaks auf platinchlorür"*,[1] but its chemotherapic effects were not discovered until 1965 by Barnet Rosenberg and coworkers, when using a platinum electrode in their experiments in *Escherichia Coli* bacteria they discovered a stop in the proliferation of cells.[2]



Scheme 6.1: Chemical structure of cisplatin.

This unexpected effect encouraged Rosenberg and coworkers to investigate the antitumor activity of platinum compounds in mice with an initial report in 1969.[3] Some years later, in 1978, the use of cisplatin in humans was approved by the Food and Drugs Administration (FDA) for the treatment of metastasic testicular and ovarian cancer.[4,5]

In chemotherapy, cisplatin is administered by an intravenous injection for chemotherapy treatment and has effect in a reasonable number of cancer affected areas including testicles, ovaries, bladder, head, neck, cervix, lung, brain and pleura. One of the commercial names of cisplatin is Platinol®[6] (see label in Figure 6.1) but this branded product is no longer available on the market. Nevertheless, some generic alternatives are available under the name of cis-DDP, cis-Platinum II, DDP or cis-Diamminedichloroplatinum.[7]

Figure 6.1: Platinol drug label.

After its administration, cisplatin circulates in the blood in its dichloride form due to the high concentration of chloride ions in plasma (∼100mM). Then cisplatin enters the cell via passive diffusion or facilitated transport and once there, one or both chloride ions can be exchanged by water molecules, generating positively charged species, which can cross the nuclear membrane via diffusion

or transport and form cisplatin-DNA adducts (Figure 6.2). Additionally, both monoaquo and diaquo compounds can react with nucleophilic sites from RNA, proteins and membrane phospholipids.[8]

Understanding the molecular grounds for cisplatin interaction with DNA and other biomolecules is fundamental to improve the efficiency of the drug, decrease its cytotoxicity and limit its side effects. However, this is still one of the grails in metal based therapy.

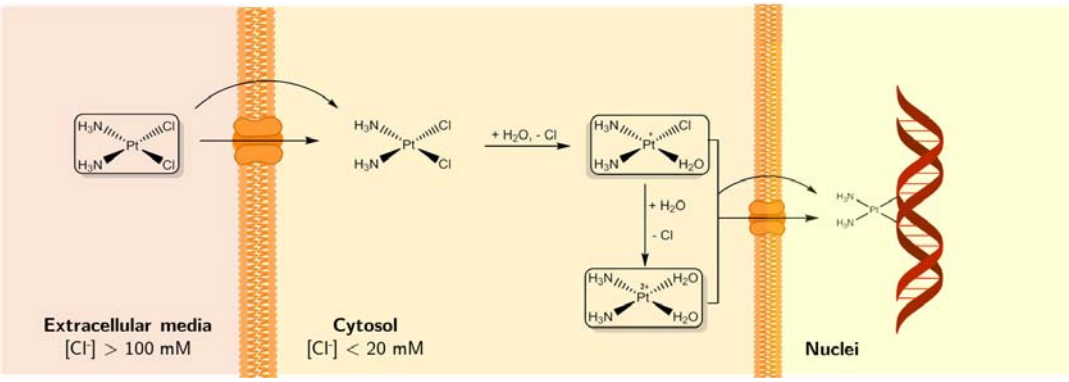


Figure 6.2: Schematic representation of the mode of action of Cisplatin.

In the specific case of cisplatin-DNA adducts, platinum is coordinated to DNA by the N7 atom from the purine bases, preferentially to guanine. Cisplatin can form five different kinds of DNA adducts. In abundance order, the binding modes are[a]: 1,2-intra(GpG) (65%), 1,2-intra(ApG) (25%), 1,3-intra(GG) (6%), 1,2-inter(GG) (3%) and *cis*-Pt$(NH_3)_2$ (GMP) (1%).[9,10]

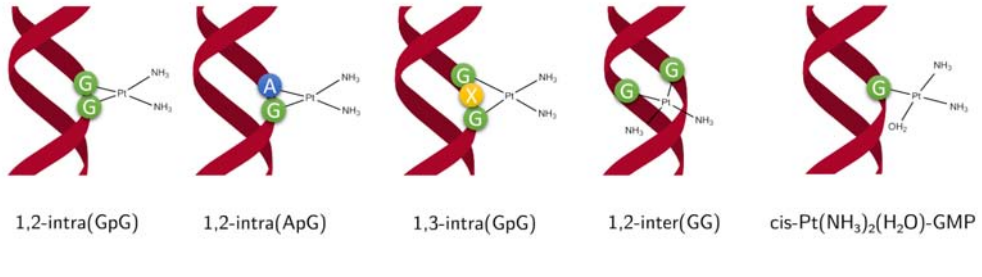


Figure 6.3: Cisplatin-DNA adducts, in order of abundance.

Despite the high similarity between guanine and adenine nucleobases, Baik and coworkers demonstrated by computational means that the main reason for the preference between both in their binding with cisplatin is the presence of the electron withdrawing oxo group in guanine, a phenomenon that decreases the energy of the lone-pair orbital at N7 of guanine with respect to adenine by 4.6kcal/mol. As a result, better orbital overlap between the coordinated N7 atom from the guanine and the Pt atom is expected.[11] A latter work of Baik also demonstrated

[a]The nomenclature of these adducts is: "1,2" (or 1,3) for the position of the adjacent purines, "G" and "A" for guanine and adenine respectivelly, "N" for any base, "p" denotes the phospate group linker, "intra" and "inter" differentiates from the intrastrand and the interstrand cross-links (see Figure 6.3). Guanine monophosphate group (GMP) does not follow these conventions and its molecular formula is used as identification.

the reason of the impossibility of formation of the 1,2-intrastrand(GpA) cisplatin-DNA adduct. Thermodynamically, (GpA) and (ApG) are closer in energy, but the reaction barrier at the formation of the bifunctional (GpA) compound from the monofunctional (Pt-G) one is too high, being the reaction prohibitively slow. For the (ApG) cross-link, a hydrogen bond is formed between the axial amine ligand and the phosphate backbone in the transition state structure, which decreases the energy barrier making the reaction affordable in normal conditions. The same hydrogen bond can not be formed in the (GpA) mode due to steric restrictions. [12]

In addition to its interaction with DNA, cisplatin also interacts with a broad range of proteins, a phenomenon on which underly most of its side effects. In fact, it has been shown that shortly after administration, more than the 65% of the drug is already coordinated to plasma proteins. Cisplatin and other derivates (oxoplatin and carboplatin) can be coordinated to Haemoglobin which in turn induces the release of the haem group from its apoprotein. [13]

Long range of side effects derived by the coordination of cisplatin to the protein where reported, being the most serious ones the hair loss (which can be permanent in some cases), allergic reactions and kidney injuries. Nevertheless, some of them can be treated with chemoprotective agents, drugs that protect healthy tissues from the toxic effects of chemotherapic pharmaceuticals. [14,15]

A better description of the mechanism of action of cisplatin, as well as deeper studies on chemotherapy side effects are nowadays the mayor fields of research in cisplatin based anticancer therapy. In addition to the great efforts of experimental research groups, the application of molecular modelling techniques can be key on the description of biological processes at molecular level.

6.2 Computational studies on cisplatin-DNA interactions

In-silico studies on cisplatin-DNA interactions started in 1980 by the contribution of Carsey in the electronic structure of platinum-guanine complexes using the Self Consistent Modified Extended Hückel molecular orbital method (SC-MEH). [16] In 1988 and 1989, Lipinski also reported his investigations on the electronic structure of platinum(II) antitumor complexes and their interaction with nucleic acid bases, performed with the semiempirical all-valence method (GRINDOL). [17,18] But not all the efforts were focused in the study of the cisplatin-DNA binding. Basch, using Hartree-Fock self-consistent-field calculations [19] and Carloni with the Density Functional Theory (DFT) [20] reported their results on the study of the hydrolysis of cisplatin to form aquo complexes, which is the previous step before the chelation to DNA. Post Hartree-Fock methodologies were also used in the description of small molecules of cisplatin-GG complexes. [21] In 2000, Carloni and coworkers reported a study of the interaction of cisplatin and DNA performed with *ab-initio* Molecular Dynamics (AIMD) taking under consideration explicit solvation with water molecules. [22] In this work they found, after a relatively short MD simulation, how the orientation of guanine bases was distorted. Spiegel in 2004 repeated the simulation using hybrid Quantum Mechanics/Molecular

Mechanics (QM/MM) Car-Parinello Molecular Dynamics simulations. This computational approach allows the simulation of bigger systems, splitting the system in two regions, depending on the level of accuracy needed. Spiegel found that not only the coordinated guanines suffer the effect of cisplatin binding, but also the surrounding bases were affected.[23] Other calculations using the awarded hybrid Quantum Mechanics / Molecular Mechanics (QM/MM) technique[b] were also performed in the field of cisplatin-DNA interactions. Examples of the application of QM/MM methods in this field are the works of Platts and coworkers.[24–27] In these articles, the author prove the efficiency of QM/MM technique as well as the importance of well defined explicit solvation effects in order to reproduce accurately the impact of cisplatin in the DNA double strand. However, such techniques are not able to study large conformational changes on the biomolecule that could result from the binding of the drug.

In this section, we present the results of our study of the structural changes induced by cisplatin on the DNA double strand structure using molecular modelling techniques. The particular objectives of this section of the thesis are the description of:

1. The energetic and structural effects at the local binding of cisplatin on DNA (1^{st} coordination sphere processes).
2. The collective distortion of the DNA strand by cisplatin binding (2^{nd} coordination sphere and onwards).
3. The structural impact of the decoordination of the cisplatin binding to proteins.

Combining a local and a collective study of the effect of cisplatin binding to DNA is not feasible using a single tool. In addition, without a previous exploration of the first coordination sphere processes is not possible to understand the reason of the structural changes on the entire system and its implications on the chemical behaviour of the adduct. Using different methodologies is a mandatory step to transfer the knowledge from the beginning of the binding process to its final structural consequences.

6.2.1 Local effects of cisplatin on DNA

At it is has been previously introduced, the major binding mode of cisplatin to DNA is through a direct coordination with the N7 atom of the two coordinating guanines (see labels in Figure 6.4), that are disposed in a conformation which needs to be distorted by the cisplatin, in order to achieve the final square planar geometry proper to a d^8 metal.

To explore the structural effect of the coordination of cisplatin to DNA, some crystallized cisplatin-DNA adducts published in the Protein Data Bank (PDB) were compared with the native geometry of DNA. From now, only the Guanine-Guanine adducts of cisplatin-DNA will be

[b] In 2013 the Nobel Prize in 2013 was for Prof. Karplus, Levit and Warshel for the development of multiscale models for complex chemical systems

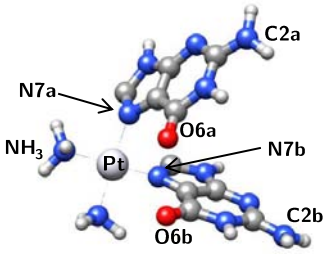


Figure 6.4: Atomic labels of the cisplatin-GG compound.

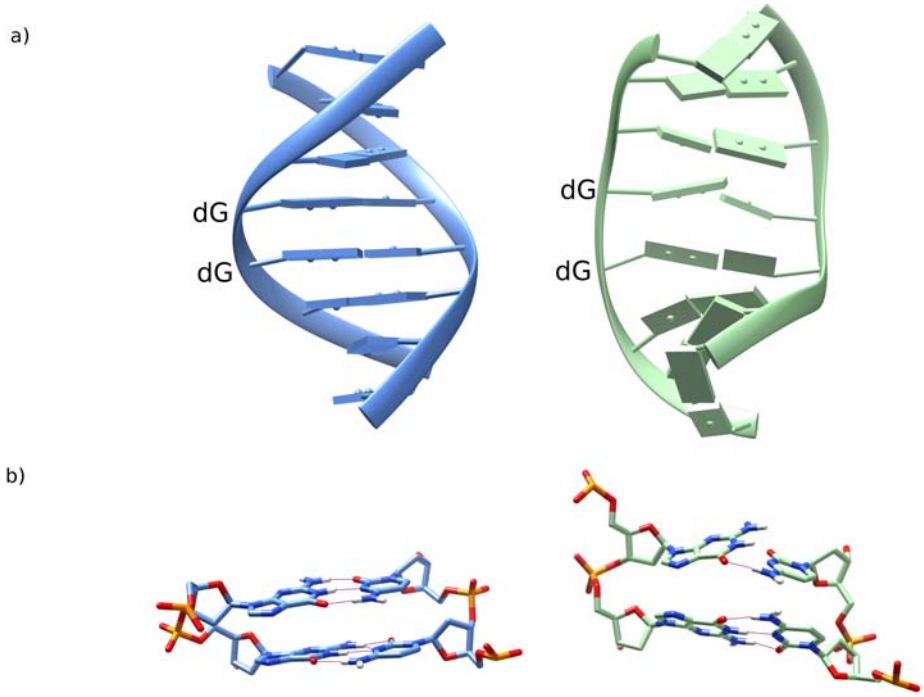


Figure 6.5: Structural effects of cisplatin. In blue, the native structure (PDB code: 2K0W), in green a cisplatin-DNA adduct (PDB code: 1A84). a) Difference in the position of the base pairs, b) Effect on the hydrogen bonds between G-C pairs. Cisplatin drug in 1A84 model was removed to get a clear view of the position of guanine bases.

described and further analysed. The binding of cisplatin to DNA not only affects the position of both coordinating guanines, but also a global effect can be seen. (Figure 6.5.a). More in detail, the parallel position of the coordinating guanines present in the native structure is lost, being deformed in a "funnel" fashion in the cisplatin-DNA adduct. The effect of the distortion is not local, being also the position of cytosine partners affected due to the impossibility to form the broken H-bonds (Figure 6.5.b).

Taking on board the geometries of the cisplatin-DNA adducts present in the Protein Data Base (PDB), the major changes in the position of the basis bound to the metal correspond to a distortion on the parallel position of both coordinating guanines to achieve the coordinating geometry. The

relative position of the two aromatic moieties of the GG pair can be described with two particular angles, which correspond to the tilt (τ) and twist (Ω) ones.[c] τ corresponds to the angle involving the atoms C2a, N7b and C2b, and represents the vertical separation between the coordinating bases. In the same way Ω is defined by the C2a, N7a, N7b and C2b atoms and represent the horizontal separation between guanines (see Figure 6.6 for a visual explanation of both tilt and twist angles).

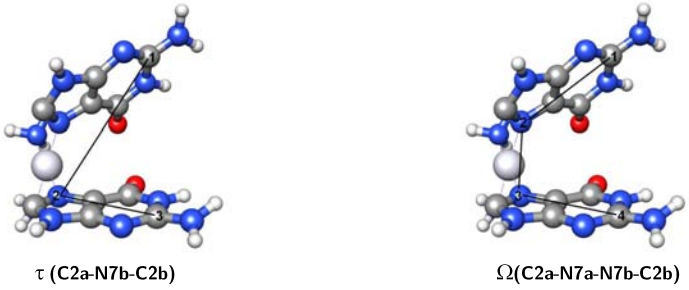


Figure 6.6: Tilt, τ (left) - Twist, Ω (right) schema.

In Table 6.1 the τ and Ω values of crystallographic structures of cisplatin-DNA adducts are reported.

Table 6.1: τ and Ω values (in $^\circ$) for experimental cisplatin-DNA structures.

Interaction	Structure	τ	Ω	Interaction	Structure	τ	Ω
Native DNA	2K0V[28]	50.9	36.7	1,2-inter(GG)	1A2E[29]	67.9	47.0
	2K0W[28]	47.8	17.9		1DDP[30]	54.2	32.0
	2NQ1[31]	43.1	26.0		1I1P[32]	69.6	49.1
1,2-intra(GpG)	1AU5[33]	72.7	20.7	1,3-intra(GpG)	1DA4[34]	79.8	14.2
	1CKT[35]	87.6	14.7		1DA5[34]	84.2	16.2
	1A84[36]	73.9	7.4		3O62[37]	96.5	7.0
	3LPV[38]	69.5	27.1				
	1AIO[39]	68.7	10.8				
	2NPW[31]	72.6	26.8				
	2NQ0[31]	72.9	28.9				

For native DNA unbound structures its τ value, which goes from 43 to 51 degrees, is the smallest one when comparing with cisplatin-DNA adducts. In ascending order, the τ values for the platinated systems are between 54 to 70 degrees for the 1,2-intra(GpG) structure, 69 to 88 for the 1,2-inter(GG), and 80 to 97 for 1,2-intra(GpG) adduct. In the case of Ω angle the ordering changes: 1,3-intra(GpG) and 1,2-intra(GpG) structures have similar angle values (7 to 16 and 7 to 29 degrees, respectively), which are slightly lower than those from the native DNA (18 to 37) and far from the higher values (32 to 49) provided by the 1,3-intra(GpG) adduct. Having in mind the structure of these different adducts (see Figure 6.3), these numbers get meaning: τ represents

[c] The name and the labelling of these angles are the same as those defined for the DNA macrostructure.

the distortion of the parallel position of the two affected guanines once the cisplatin is coordinated to the N7 atom. As farther are the guanines, bigger distortion have, and for this reason the 1,3-intra(GpG) systems are the most τ-distorted ones. On the other hand, Ω describes the loss of π-stacking interaction between two guanines due to the displacement of one base over the other. Ω values have to be similar in the intra(GpG) motifs and the native structure, and noticeably bigger for the 1,2-inter(GG) adduct because of the two guanines are in different chains.

To have an idea of the energetic effects of the two torsional angles and the sensibility of the first coordination sphere of the metal when interacts with the DNA, full Quantum Mechanics calculations were performed with a small model of cisplatin-GG adducts, where only the cisplatin compound and the purine ring of both coordinated guanines are included. Calculations were carried out using the Gaussian09 package[40] at DFT level using and BHandH.[41] Tests with B3LYP[42,43] were also performed but this functional fails for dispersion interactions, which is the physical effect of π-stacking. 6-31+G**,[44,45] was the chosen basis set for the main group elements, and LANL2DZ[46] for Pt and its pseudopotential.

We first considered as a starting point geometry for the scan the absolute minimum of the isolated cisplatin-GG model. In this optimized geometry, two hydrogen bonds (one per base) were present, both between the O6 atom of the purine and one hydrogen from the ammonia group (see Figure 6.7). We notice that this structure can not be formed in DNA bound complexes because the large separation between bases is not consistent with the double strand interactions (τ=150° and Ω=121°).

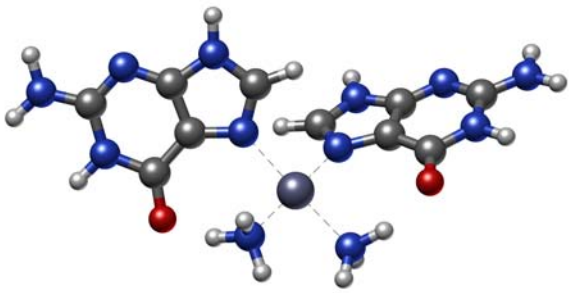


Figure 6.7: Structure of the minimum energy geometry of the cispatin-GG compound in gas phase.

To maintain as much as possible the realism of our model, the selected starting geometry is the isolated cisplatin-GG model with a spatial restraint to correspond to the experimental structure of an 1,2-intra(GpG) (PDB code: 1AU5). In order to get a complete exploration of the relative position of both guanines, two coordinates were chosen. The first one corresponds to the τ angle, and the later to the Ω one, both with a step size of 5°. For τ, the range goes from 35 to 100°. In Figure 6.8 one can see a comparative between structures with different τ value (35, 60 and 100°). For lower

values than 35° the angle is so close that Van der Waals contacts appear between both guanines. On the other hand, over 100° guanines can contact with the bases placed above and below the model.

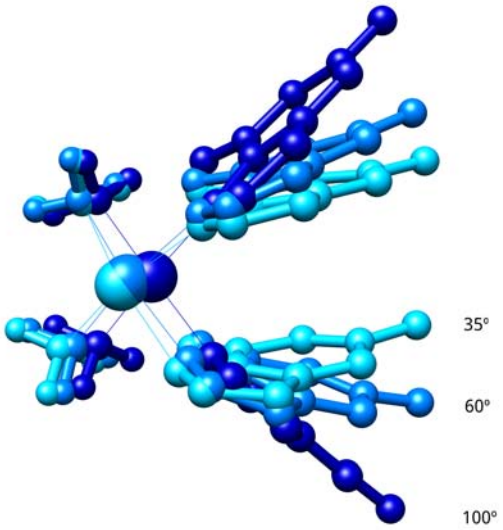


Figure 6.8: Comparison between structures with different τ value.

For Ω, the exploration range goes from 0° (where both bases are completely parallel to each other) to 50°, more than the higher value observed in experimental structures for Ω. A comparison of different Ω values can be seen in Figure 6.9.

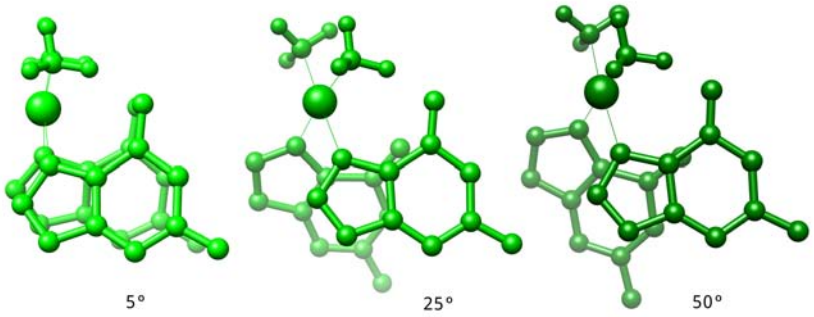


Figure 6.9: Comparison between structures with different Ω value.

To have a complete exploration on the structural effects of the binding of cisplatin to DNA, the following step is to calculate the energy surface given by the exploration of both τ and Ω angles simultaneously (Figure 6.10). The lowest energy combination of both angles includes the region from 70 to 100 degrees for τ, and 0 to 50 for Ω. This region almost cover all the different values of each cisplatin-DNA adduct, but it is out of range for the native DNA structures.

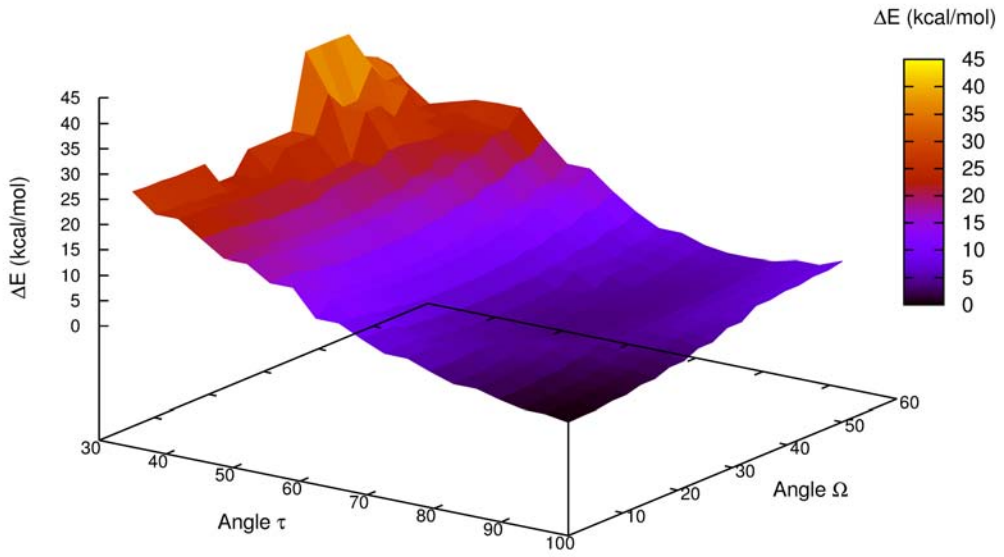


Figure 6.10: Bidimensional exploration of τ and Ω angles

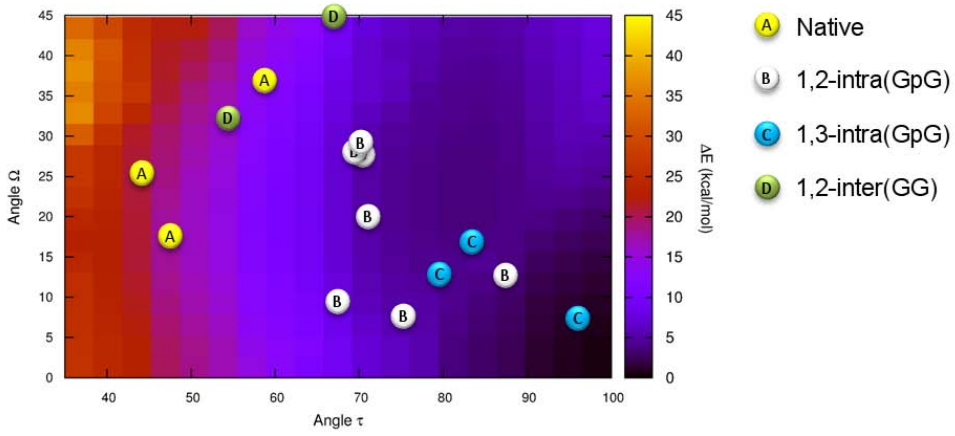


Figure 6.11: 2-D representation of the bidimensional exploration of τ and Ω angles

In Figure 6.11 there is a two-dimensional representation of the energy profile shown in the Figure 6.10. In different colours, the dots represent the point of the energy profile were experimental structures appear, depending only on its τ and Ω values. Dots placed in the highest energy region (in yellow) corresponds to the native geometries, and are up to 20 kcal/mol over the minimum. The following are the 1,2-inter(GG) structures (in green) which are 15-20 kcal/mol higher than the

lowest one. Last, the 1,2-intra(GpG) and 1,3-inter(GpG) adducts (white and cyan respectively) are both about 10 kcal/mol less stable than the minimum. These relative energy differences are in agreement with the experimental abundance of the different adducts for the 1,2-inter(GG) and intra cases.

The high energy values in conformations with the basis as they stand in the native DNA show that a deformation of the coordinating guanine bases stabilizes the complex generated by the coordination of the cisplatin drug. Regarding the energy of the first coordination sphere of the three different cisplatin-GG adducts considered in this study, one can see how the geometrical restrictions imposed by the 1,2-inter(GG) compounds are stronger than any intra adduct. However, there is no clear difference in energy between those adducts resulting from 1,2-intra(GpG) and 1,3-intra(GpG) interactions. Based on this part of the work, it is clear that the specificity of the interaction does not result from the pure energetic profile of the first coordination sphere of the metal. We therefore decided to explore geometrical variables further than the first coordination sphere and particularly the problem of the DNA reorganization upon the metal binding.

6.2.2 Long-range effects of cisplatin on DNA

In the previous section, local effects of cisplatin binding in DNA were studied. Using two geometrical variables describing the relative positions of the coordinated guanine bases is possible to discern between inter and the intra adducts. Here we intent to find evidences of additional geometrical and energetic variables that could explain the different of abundance of 1,2 and 1,3 intrastrand binding modes.

6.2.2.1 QM/MM calculations of cisplatin-DNA adducts

The first step is to find the initial evidences on the deformation of the DNA double strand once cisplatin is coordinated, expanding the size of the system to a more real one which includes also the surrounding bases of the two coordinated guanines. To do so, we performed geometry optimizations in some cisplatin-DNA adducts fragments, in particular 1AU5, 1A84, 1CKT and 1AIO 1,2-intra(GpG) adducts, 1A2E and 1DDP 1,2-inter(GG) ones, and 1DA4 1,3-intra(GpG) structures, applying the hybrid QM/MM technique using the ONIOM[47] utility implemented in Gaussian09. This method allows us to accurately describe the region of the system where the interaction is taking place accounting for the rest of the system using a cheaper method. In the present case study, the quantum mechanical part was described at DFT level using BHandH as functional. The basis set is the same than in the precedent section, 6-31+G** for main group atoms and LANL2DZ for Pt and its pseudopotential. For the molecular mechanics part, the AMBER99[48] force field is the chosen one to simulate the environment and ANTECHAMBER[49] code is used to add the corresponding charges to each atom of the system.

Crystallographic cisplatin-DNA adducts available in the Protein Data Bank present insufficient

solvent molecules. The experimental systems used in this study were resolved by X-ray or NMR analysis. In the first case, structures have to be solved with very high resolution to clearly identify all the well organized water molecules but even so could miss the relevant ones.[50] In the second one, this method does not provide any single water molecule. The lack of solvent molecules could explain the reason of the absence of a hydrogen bonding network in some of the experimental adducts. Moreover, this could also lead to erroneous predictions of the structure of cisplatin-DNA adducts by computational means if the solvation shell is not extended to a more realistic one.

For that reason, all the adducts were explicitly solvated with explicit water molecules using the TIP3BOX algorithm[51] included in the AMBER package and implemented in UCSF Chimera.[52] A 10Å solvent shell wraps the fragment with up to 1000 water molecules (Figure 6.12). To improve the localization of explicit water molecules, a previous minimization of the solvent around the adduct has to be performed. This step was also performed with ONIOM using the same DFT functional and bases, keeping fixed the cisplatin-DNA fragment and allowing free movement to the solvent shell. Once the solvent is correctly placed, the minimization of the entire system has been carried out.

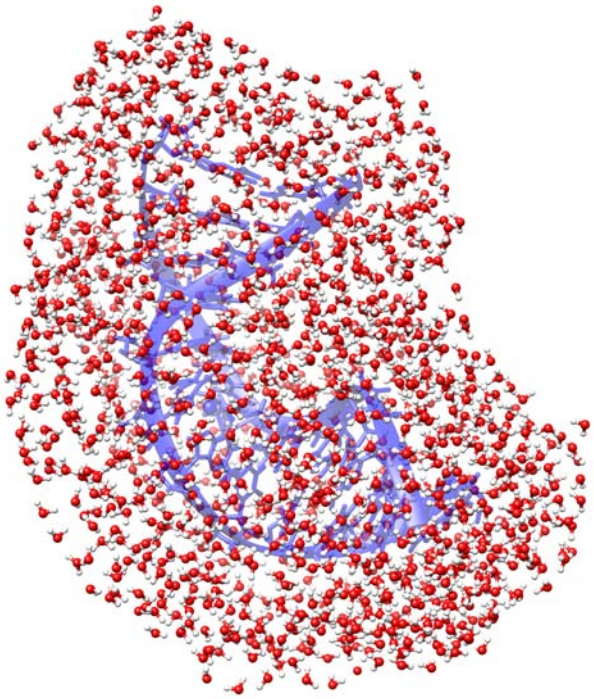


Figure 6.12: A 10 Å thick layer wraps the cisplatin-DNA adduct. This example corresponds to the 1A84 model, solvated with 1401 water molecules.

To proceed with the study of the long range effects of the coordination of cisplatin to DNA, three different QM/MM partitions have been used. In these three systems, the Quantum Mechanical part corresponds to:

1. Cisplatin-GG adduct (from now, 2 bases system). The aim of this calculation is to study the

effect of cisplatin in the relative positions of both guanines in DNA. The QM layer includes 39 atoms.

2. Cisplatin-GG adduct and the complementary bases (4 bases system). To have an accurate description of the H-bond network between the coordinated guanines and its complementary bases. Here, the QM layer includes 63 atoms.

3. Cisplatin-GG adduct, the bases above and below them and the corresponding complementary bases (8 bases system). Here, to the effect of the H-bond network is added the π-stacking interaction of the surrounding bases. The total size of the QM layer is 118 atoms.

2 bases system

This is the most similar case to the previously studied with cluster models constraining the geometry of the adduct. Only the cisplatin-GG adduct is included in the quantum mechanical part. The rest of the system and the solvent were computed at molecular mechanics level. In Table 6.2, the most important geometrical values are reported for the calculated and experimental structure of each type of adduct. In Figure 6.13 there is also shown the superposition between experimental and calculated geometries for each kind of coordination.

Table 6.2: Geometrical values of optimized and experimental (in parenthesis) cisplatin-GG adducts with the 2 bases model.

Type	PDB	τ	Ω	G-Pt-G	G-G-NH_3-NH_3
1,2-intra(GpG)	1AU5	67.9 (72.7)	22.8 (20.7)	85.4 (87.4)	0.5 (0.3)
	1CKT	71.7 (87.6)	5.1 (14.7)	85.0 (90.3)	1.6 (3.7)
	1A84	101.2 (73.9)	2.8 (7.4)	89.2 (90.0)	1.3 (0.0)
	1AIO	62.6 (68.7)	18.8 (10.8)	88.1 (99.9)	1.0 (4.0)
1,2-inter(GG)	1A2E	60.9 (67.9)	41.2 (47.0)	84.7 (86.8)	5.5 (0.1)
	1DDP	62.4 (54.2)	23.7 (32.0)	87.6 (92.2)	5.2 (0.0)
1,3-intra(GpG)	1DA4	71.9 (79.8)	24.8 (14.2)	90.0 (85.2)	9.5 (0.0)

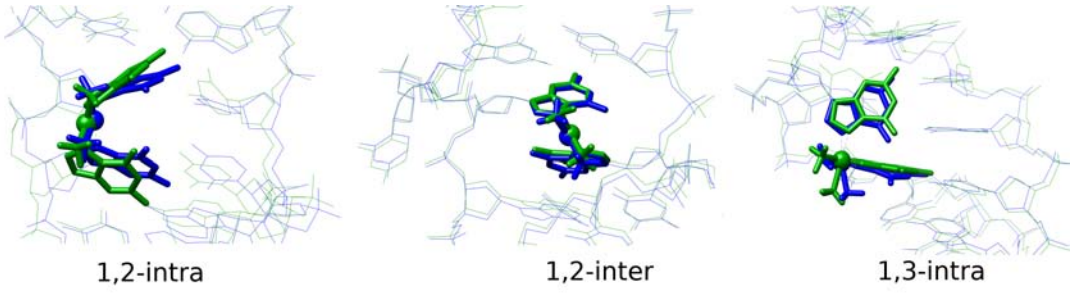


Figure 6.13: Structural differences between calculated (blue) and experimental (green) cisplatin-GG adducts. The 1,2-intra(GpG) adduct corresponds to 1A84, the 1,3-intra(GpG) to 1DA4 and the last one to structure 1A2E.

In general, we do not observe major changes in τ and Ω values with respect to the experimental structures after QM/MM minimization. This is also the case for any variables related to the first coordination sphere of the metal, such as the distortion of the square planar platinum(II) and the Guanine-Platinum-Guanine angle.

In the specific case of 1,2-intra(GpG) adducts, τ values are lower in the calculated structures than in the experimental ones, except for the case of the 1A84 which arises the 100°. On the other hand, Ω values do not show any clear tendency, and its variation its not higher than 9° respect to the experimental structure. Guanine-Platinum-Guanine angle was slightly decreased (5° as much) from the expected 90°. Finally, the planarity of the square planar complex is being modestly corrected to the ideal 0° value. In overall, the calculated systems are not so far in geometry from the experimental ones, which demonstrates the weak structural restrictions imposed by the surrounding bases at this level of theory.

For the 1,2-inter(GG) complexes, the τ values are mainly disperse which suggest little preference of the system for this motion. However, the system tends to relax along a clear decreasing of Ω coordinate angle to get a most accommodated structure according to the potential energy surface (Figure 6.10). The Guanine-Platinum-Guanine angle is also slightly decreased being lower than 90°. The only important geometrical variation is in the planarity of the complex, which is 5° away from the experimental structures.

Finally, the unique 1,3-intra(GpG) adduct studied in this section presents a lower value of the τ variable and a higher value for the Ω one than the experimental structure, both distancing the system from the experimental minimum (Figure 6.10). The angle G-Pt-G is optimized to the expected 90° and the planarity of the square planar complex is negatively affected.

Despite of the different DNA sequence of all the adducts, the quantum mechanical partition includes the same number of atoms in all of them. Because of that, a comparison of the QM energy of all the calculated systems can be done, as it can be seen in Table 6.3.

Table 6.3: Energy difference (in kcal/mol) of each cisplatin-GG adduct respect the minimum (1A84).

Type	PDB	ΔE(QM)
1,2-intra(GpG)	1AU5	7.7
	1CKT	16.2
	1A84	0.0
	1AIO	4.7
1,2-inter(GG)	1A2E	5.5
	1DDP	13.7
1,3-intra(GpG)	1DA4	3.9

The minimum energy adduct corresponds to the 1A84 structure, which agrees with the relation between τ and Ω values and energy seen in Figure 6.10. For the rest of the systems any tendency

can be inferred using only the studied variables. For example, taking into account the energy profile shown in Figure 6.10, the second structure lower in energy expected due its angle values is the 1CKT, which is modestly higher in energy than the minimum. Surprisingly, distances between the oxygen (O6) of both guanines and the O6-NH_3 one also play a role, being the first unfavourable due to the electron repulsion, and the second stabilizing because it can facilitate the formation of H-bonds between both atoms. Only few structures are affected by these geometrical variables: 1A84, 1AIO and 1DA4 are all stabilized by one hydrogen bond between O6 and NH_3, and 1AU5 and 1CKT are unfavoured by electron repulsion of lone pairs in O6.

4 bases system

The following part of the study is to consider the complementary cytosine bases of the interacting guanines with cisplatin. Here, not only the position of the guanines is optimized, the corresponding H-bonds between purines and pyrimidines are also taken into account in the high level ONIOM layer. This study is only performed in the case of intrastrand adducts. Interstrand models does not present any kind of interaction between the coordinated guanines to cisplatin and its complementary bases.

In Figure 6.14 are overlapped the structure of the experimental and calculated cisplatin-GG adducts for one example of each 1,2-intra(GpG) and 1,3-intra(GpG) adducts. There, one can see that the main structural deviations of the experimental and geometry optimized structures come from the high ONIOM layer. 1,2-intra(GpG) model shows a stabilization of the upper guanine to allow the formation of three hydrogen bonds which are not present in the experimental system, and also an additional hydrogen bond between the NH_3 groups and the O6 from both guanines is formed. In the case of the 1,3-intra(GpG) model, no important structural difference between the experimental and calculated geometries can be appreciated. However, it is important to take into account the presence of the complementary partner of the base placed in the middle of both guanines, which corresponding base stabilizes the upper guanine by to the formation of a pair of H-bonds. Additionally, the middle base restricts the reorganization of the coordinating bases, being one of the causes of the slight differences between the theoretical and the experimental models.

Differences in energy between the 1A84 model (taken in Table 6.3 as reference) and the other intrastrand systems show particular variations with respect to the 2 bases system (see Table 6.4). The energy of the QM/MM system is no longer comparable to those values of the particular τ and Ω movements because distortions caused in the complementary chain also affect the stability of the system. This is the reason why the energy of the 1,3-intra(GpG) model has been chosen as a reference point, once discarding the 1AIO structure which presents a lower energy value because DNA chain is shorter than in the rest of the cases and provokes less constraints in the movement of both bases, yielding to a more relaxed optimized geometry.

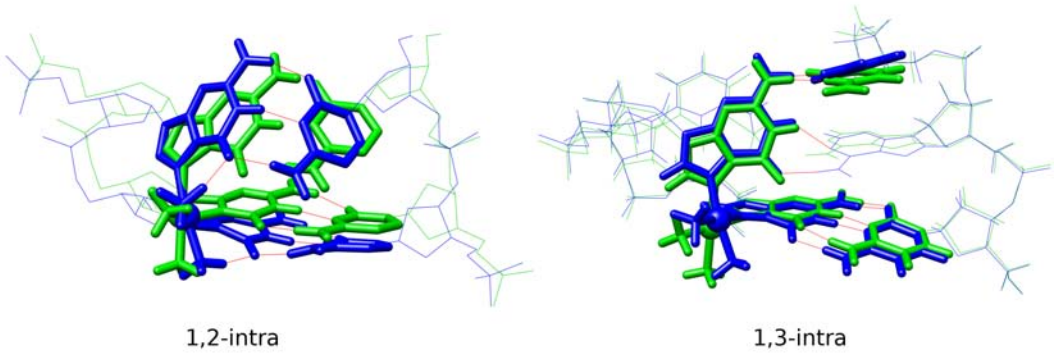


Figure 6.14: Structural differences between calculated (blue) and experimental (green) cisplatin-GG adducts. 1,2-intra(GpG) adduct corresponds to 1A84 model and the 1,3-intra(GpG) to the 1DA4 one. H-bonds are shown in red.

Table 6.4: Energy differences (in kcal/mol) of each cisplatin-GG adduct respect the 1A84 adduct using the 4 bases model.

Type	PDB	ΔE(QM)
1,2-intra(GpG)	1AU5	-6.1
	1CKT	2.6
	1A84	0.0
	1AIO	-10.9
1,3-intra(GpG)	1DA4	0.1

8 bases system

In the precedent section, the cisplatin-GG adduct plus the complementary cytosines where included in the Quantum Mechanical layer. The hydrogen bonding effects were studied in the past section, but this is not enough to have a realistic idea of the structural impact of the coordination of cisplatin to DNA. The next step is to include the surrounding bases in the QM layer, having 8 nitrogenated bases, plus the cisplatin, included in the QM partition. Geometry minimized structures are not so far from the experimental ones, as in the precedent sections (see Figure 6.15). A small reorganization takes place in the position of the bases, in order to maximize the stabilizing effect of hydrogen bonds. Also, the interaction between the NH_3 group and the O6 disappeared from the 1A84 model, which presents two hydrogen bonds in the 4 bases system.

Energetically, the 8 bases models of 1A84 and 1CKT can not be comared. Despite of having the same number of atoms, the DNA sequence of the binding strand is different (TGGT for 1CKT and TGGA for 1A84).

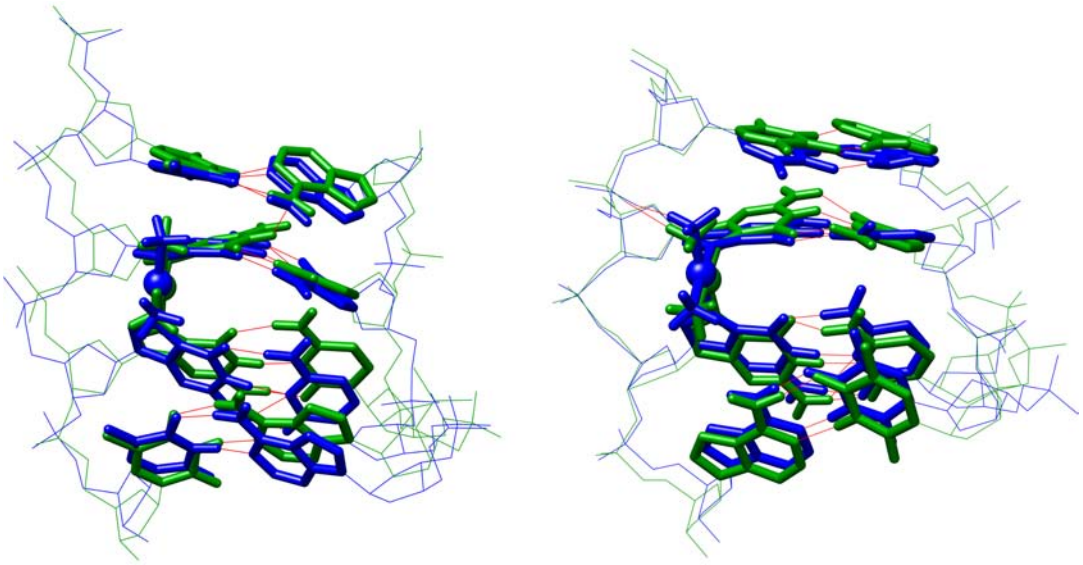


Figure 6.15: Structural differences between calculated (blue) and experimental (green) cisplatin-GG adducts for 1A84 (left) and 1CKT (right).

Is it necessary increase infinitely the size of the QM layer?

In the past sections, the size of the Quantum Mechanical part was increased in a power of two in order to describe at each step one determined phenomena. By definition, this layer has to include the region of the system were the chemical interaction is taking place. In our study, the QM part has to include at least the cisplatin-GG adduct. Gkionis and coworkers demonstrate the efficiency of including this small region in the QM layer in the comparative of different anticancer drugs. [26,27] Increasing the size including the complementary bases is useful to take into account the G-C pair interactions. Adding the bases above and below the cisplatin-GG adduct is a good idea to consider the packing of the bases from the same chain due to the π-stacking effect. [24] However, the difficulty and computational cost of the calculation increases exponentially with the size of the QM layer, being practically impossible full Quantum Mechanical calculations in cisplatin-DNA adducts.

Other ONIOM partitions could be used in this study. For example, including the sugar and phosphate groups linked to guanines in the QM layer can be useful to account for the effect of the drug in the ribbon. However, due to the negative charge of the PO_4^{-3} group, an undesired interaction between the platinum centre and the phosphate can appear, leading to a final structure where the platinum presents a pentacoordinated environment. This unexpected interaction could be avoided if the system is correctly solvated with explicit water molecules, as it can be seen in the following section.

The effect of the solvent

In precedent sections where the QM/MM partition was evaluated, all the calculations were carried out including explicit solvent molecules, because the use of implicit solvent effects accounted by the

Polarizable Conductor Calculation Model (CPCM)[53,54] is inadvisable. Tests performed with this technique show an unusual unwinding of the double helix of the DNA.

As a first stage, calculations were performed without taking into account solvent molecules in the first coordination sphere, but a direct coordination of the terminal O atom of the closest PO_4^{-3} group to the platinum centre have been seen as an artifact of the calculation. Then, the inclusion of the crystallographic water molecules was studied but the number of water molecules placed in the first coordination sphere is insufficient (Figure 6.16).

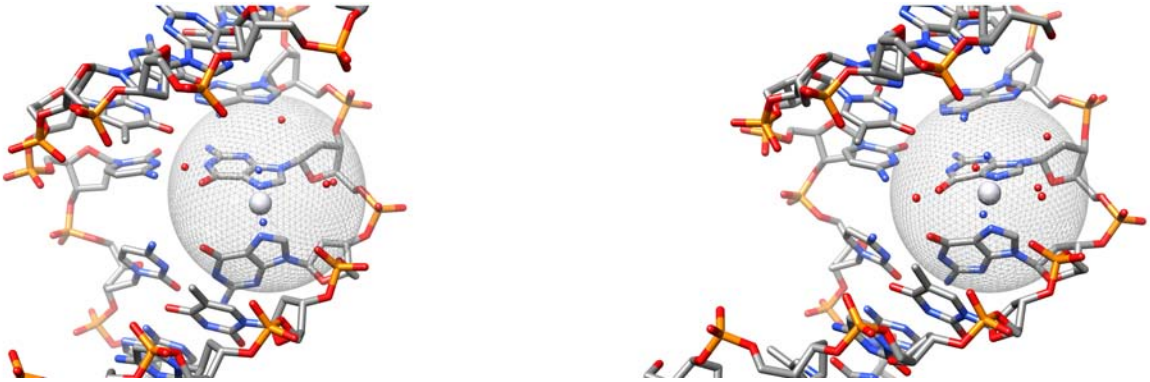


Figure 6.16: Representation of the water molecules present in a region of 5 Å from the platinum centre for the X-ray structure of 1CKT (left) and the one after minimize the systems with the 2 bases partition (right).

At structural level, calculations show that the presence of water molecules affect the conformation of the first and second coordination sphere of the cisplatin drug. Interestingly, some water molecules lay between the platinum centre and the backbone, which represents a weak coordination of the solvent to the metal leading to a distorted octahedron. This fact is not surprisingly since electrostatic interactions at the two axial sites of the metal are reasonable. However, the solvent was modelled at Molecular mechanics level, which means that electronics effect between the platinum centre and the solvent are not taken into consideration. Performing QM/MM minimizations of DNA (with or without cisplatin) in a unsolvated media produces severe distortions at the ends of the strings. Wrapping the system with a thick layer of solvent prevent the unwinding of the DNA double strand system.

6.2.2.2 Normal Mode Analysis of intrastrand models

Using the ONIOM framework we explored the effects of cisplatin on DNA in a small area (from 2 to 8 bases). However, it is not possible to see the consequences of the binding on the entire double strand using only this method and asses if there is a long range involvement for specificity. An available computational solution to go ahead with the exploration is Molecular Dynamics but the enormous number of degrees of freedom make the study very time consuming. To proceed with the exploration, we carried out a series of Normal Mode Analysis (NMA) of hexameric cisplatin-GG adducts. NMA is a technique frequently used to describe the motions of a structure at its equilibrium

geometry. For large molecular architectures like biomolecules, the vibrations of lowest energy are associated with collective movements that are related to their physiological mechanisms. They are generally associated with breathing and aggregation processes and are intrinsically related to the shape and size of the systems[55]. To date, NMA studies are far less reported in the literature for double strand DNA than for proteins or RNA, the relatively reduced conformational variability of the former offering a narrower range of dynamical behaviours to investigate. However, the lowest energy modes of DNA have long been characterized and correspond to open-close and elongation motions. In a way, these motions could be related to the wrapping mechanism of one strand against the other and the bending properties of the entire strand. In this part of the study, we intent to identify by means of NMA calculations the possible mechanical alterations of the DNA double strand suffered from the binding of cisplatin in forming 1,3-intra(GpG) and 1,2-intra(GpG) complexes.

Normal Mode Analysis was implemented in the Molecular Mechanics Toolkit (MMTK) by Prof. Hinsen[56] and later exported to UCSF Chimera[52] by our research group. In our case, calculations were performed using the Elastic Network Model (ENM), also developed by Prof. Hinsen. The atomic force field used is AMBER99 and the previously published set of parameters for the cisplatin moiety were added.[57]

In this study, the normal modes from the PDB experimental structures were compared to those generated in the cisplatin unbound systems. This is not the first case of NMA applied on DNA. Chou[58] proved in 1984 the accuracy of this method reproducing the "breathing" motions of DNA.

In order to get only a general idea of the long range effects of cisplatin, simulations have been performed for one example of each intra cisplatin-GG adduct: 1AU5 PDB model for 1,2-intra(GpG) system and 1DA4 for the 1,3-intra(GpG) one. To avoid inconsistencies in the comparative of the results, all the systems were shortened to 8 bases per strand. Modes 1 to 6 correspond to rotational and translational movements and were discarded from the analysis. In both adducts, the normal modes from 7 to 20 present similar ranges of frequencies from about 9cm^{-1} to 26 cm^{-1}.

A first analysis consisted in classifying individual modes in open-close and elongation categories. For the 1,2-intra(GpG) adduct, the mode 7 (9.2 cm^{-1}) corresponds to an open-close motion, and the mode 8 (11.6 cm^{-1}) to an elongation mode. All the remaining modes are motions mixing both elongation and open-close motions. For the 1,3-intra(GpG) adduct, the relative order of elongation and open-close is inverted. The mode 7 (8.6 cm^{-1}) corresponds to an elongation mode and mode 9 (12.1 cm^{-1}) to the open-close. All the remaining modes are a combination of both displacements. Those results were compared to native hexameric structure of double strand DNA (PDB code 1F6E). In this case, the two lowest modes correspond, in order, to the open-close motion (9.7 cm^{-1}) and the elongation mode (11.4 cm^{-1}). This result first indicates that the 1,2-intra(GpG) adduct presents a dynamical behaviour substantially closer to a native strand and that alterations of the intrinsic dynamics of the DNA are more pronounced in the 1,3-intra(GpG) system.

A latter analysis of the individual contributions of the 100 lowest vibrational modes also supports the similarity between the 1,2-intra(GpG) and the unbound system. For 1,2-intra(GpG) models,

the most important contribution (16%) is observed in the 7^{th} mode alone. However, for the 1,3-intra(GpG), the lowest modes (7 and 8) are not involved in the transition to native DNA. The contribution of the mode which describes the open-close motion (mode 9) is reduced to 11%. That fact indicates that the formation of the 1,2-intra(GpG) model takes place along the lowest energy motion of the native DNA while the 1,3-intra(GpG) fashion requires a more complicated denaturation of the macromolecule.

To understand how the open-close motions of each kind of intrastrand adduct are affected by the binding of cisplatin, their displacement vectors were compared. Two differentiated dynamical patterns were observed. In the 1,2-intra(GpG) adduct, the displacement of the 3' and 5' ends of both strands is highly symmetric with both of them participating to almost equal magnitudes to the wrapping of one strand against the other (Figure 6.17-left). This behaviour is similar to the one observed in native DNA double strands. However, in the 1,3-intra(GpG) adduct, this motion is highly asymmetric with 3'-OH end of both strands displaying a far lower amplitude of motion (Figure 6.17-right). In this case, the open-close motion is substantially altered meaning that reformation of a structure close to the native DNA is unlikely. This last result sustains that the binding of the metallodrug does not affect the double strand of DNA in the formation of an 1,2-intra(GpG) adduct but that the complexation in an 1,3- fashion blocks part of the DNA and the denaturation of the polymer pushes the system towards a different dynamical space from the native and 1,2-intra(GpG) systems.

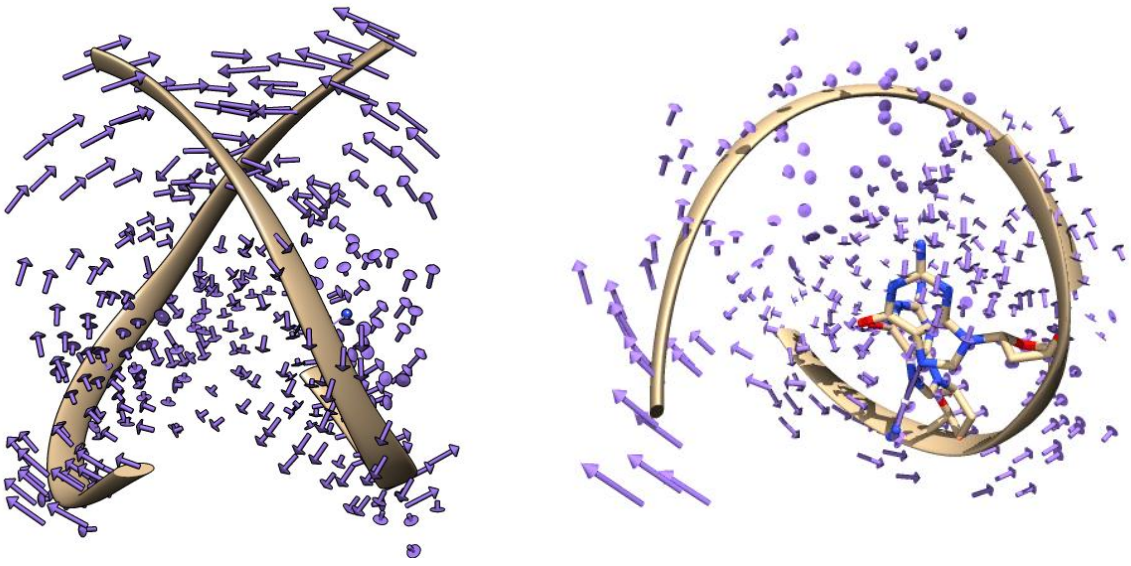


Figure 6.17: Displacement vectors of the lowest energy open-close modes of the 1,2-intra(GpG) (left) and 1,3-intra(GpG) (right) cisplatin-GG adducts. The backbone structures of each strands are represented in golden ribbons, cisplatin and its coordinating bases are depicted in stick. Arrows are reported to indicate the magnitude and direction of the atom displacement associated to the collective vector. Those associated to the hydrogen atoms are omitted.

6.3 Interaction of cisplatin with proteins as a side effect of anti-cancer drug

In 1998, Neault and coworkers reported the first spectroscopic evidence on the binding of cisplatin to human serum albumin (HSA), which is the most abundant protein in the human blood plasma.[59] In their scientific contribution, the effect of cisplatin in the secondary structure of proteins have been demonstrated: at high drug concentrations, the number of α-helix in the cisplatin-HSA adduct (with respect to the native HSA) was decreased, and the percentage of β-sheets was increased.

Since the discovery of the molecular effect of cisplatin in proteins many efforts have been done to understand the binding process of the platinum centre to the proteic host, studying the relative stability of the cisplatin-protein adducts depending on the amino acid which is coordinated to the metal. To have some numbers, the binding constant (K) of cisplatin to the HSA is $8.52 \cdot 10^2$ M^{-1} (ΔG = -4.00 kcal/mol)[59] which is significantly lower than the corresponding to the cisplatin-Guanine interaction (-18.3 kcal/mol for the first substitution, -23.4 kcal/mol for the second).[60]

Other example of the molecular effect of cisplatin in proteins is the case of the Na^+-K^+-ATPase.[61] The biological function of this type ATPase is transporting the Na^+ and K^+ ions through a negative electrochemical gradient. Cisplatin stabilizes the secondary chain of the protein by means of a H-bond interaction with the CO, CN and NH groups present in the biological system.

Cisplatin not only affects the structure of the proteins, also its function can be altered. This is the case of the transferrin,[62,63] an iron carrier protein. Cisplatin coordinates to the transferrin close to the binding site, avoiding Fe to enter in the active site and modifying the behaviour of the protein.

One would expect that cisplatin coordinates to proteins in the same way than to DNA: by the substitution of one or more of its ligands with a residue (or nucleotide, in the case of DNA) maintaining its expected square planar geometry. This is not what we observed in two X-ray structures found in the PDB database, where cisplatin loses one of its ligands and bounds to the protein forming an unusual trigonal planar structure. The first explanation for such singular coordination of cisplatin is the difficulty to identify water molecules in X-ray techniques, specially if there is a heavy metal close to area where it is expected to find the water molecule. Nonetheless, reasonable doubts exist on the possibility of tri-coordinated cisplatin-protein adducts.

The main goal of this section is to assess the existence of tricoordinate platinated structures in biological media. To do so, the relative stability of tri and tetra-coordinated cisplatin structures is deeply investigated by computational means.

6.3.1 Tri-coordinated cisplatin compounds. Trick or treat?

The most common geometry in Pt(II) compounds is the square planar. This geometry is given by the electronic state of the metal, which has a d^8 configuration. Moreover, some tri-coordinated

compounds, in T-shape form, can be seen in some special cases:[64]

- True T-shaped compounds: the first pure T-shaped compound was characterized by Braunschweig in 2005.[65] This system corresponds to a 14-electron T-shaped Pt(II) boryl complex. The stabilization of the 14-electron structure is given by the presence of bulky ligands which protect the vacancy in the forth position.

- Agostic T-shaped compounds: this is the case were a 3-centre-2-electron interaction between the metal and a C-H bond present in a coordinated ligand appears. One example of this kind of T-shaped compounds is given by Orpen and coworkers, who published one example of β-agostic interaction which fills the vacant site.[66] The Pt-H interaction is sufficiently strong to cleave the C-H bond, yielding to a chemical equilibrium between the Pt-H and the C-H states.

- Counterion masked T-shaped compounds: the nature of the T-shaped Pt(II) compounds is cationic in most of the cases. Then, it is possible than anionic ligands could reach the vacant site on order to fulfil the electronic requirements of the metal. In this case, the existence of a T-shaped compound is alleged by the labile character of the counteranion.[67]

- Solvent masked T-shaped compounds: similar to the precedent case. A σ-interaction can appear between the Pt(II) and a solvent molecule.[68]

In the following section, the stability of tricoordinated cisplatin-protein adducts is compared with the solvent masked T-shaped systems with the aim of determining the veracity of the X-ray structures shown in the literature.

6.3.2 Study of the relative stability of tri and tetra-coordinated cisplatin-protein adducts

Only few X-ray structures of cisplatin-protein adducts can be found in the Protein Data Bank. Some of them present an unusual coordination of the drug, with unexpected geometries from the inorganic point of view. This is the case of the two cisplatin-protein adducts crystallized and resolved by Casini and coworkers, which corresponds to an hen egg lysozyme (PDB code: 2I6Z)[69] and a bovine Cu, Zn superoxide dismutase (PDB code: 2AEO).[70] In both structures, cisplatin is coordinated to a histidine to complete a planar trigonal structure (Figure 6.18). In the X-ray characterization of both systems, no clear signal of the fourth ligand is appreciated, being really weak in the 2I6Z structure and absent in the 2AEO one. However, both research groups are in agreement in the fact that the fourth ligand could be a labile water molecule. So then, the possibility of the existence of tricoordinated cisplatin species in proteins is the main focus of this part of the thesis.

The initial models are constructed following the next rule: the cisplatin ligand is represented by the Pt-NH_3 segment (the one observed in the hen egg lysozyme). The third coordination is given by one of the selected amino acids (histidine, lysine, cysteine, cysteinate, metionine, tyrosine,

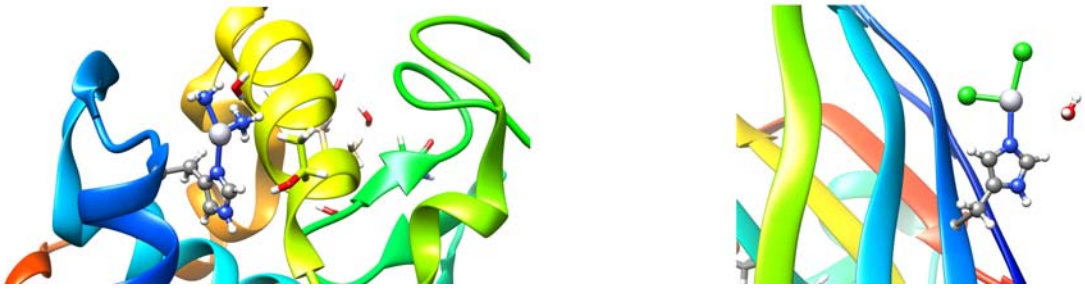


Figure 6.18: Unexpected geometry of the hen cisplatin-protein adducts. 2I6Z on the left, 2AEO on the right.

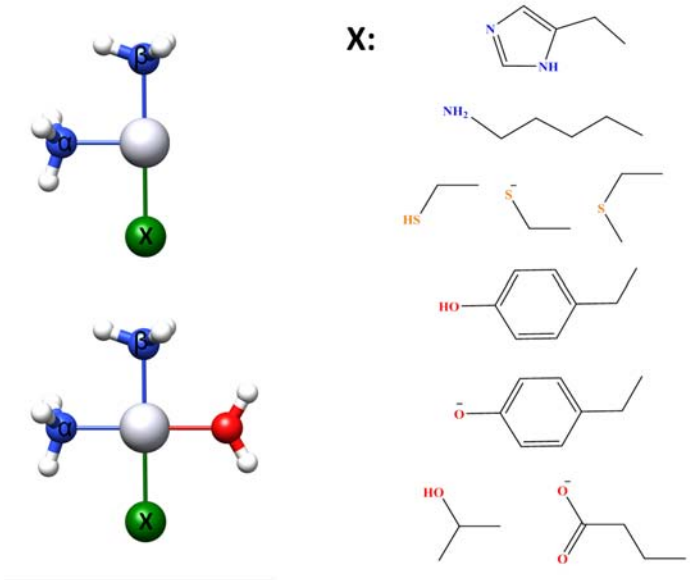


Figure 6.19: Labels used in this part of the study.

tyrosinate, threonine and glutamate) or a water molecule. In the case of tetracoordinated structures, the fourth site is filled by a water molecule (Figure 6.19).

The procedure followed to go ahead with this comparative study has four different steps:

1. Geometry optimization of both tri and tetracoordinated set of structures.
2. Decomposition of the strength of the cisplatin-residue bond.
3. Dehydration process of tetracoordinated platinated compounds.
4. Explicit solvation of the second coordination sphere of cisplatin in hen egg lysozyme.

Geometry optimization

The first step is to characterize the minimum energy geometry of trigonal planar cisplatin compounds coordinated to different amino acids: histidine, lysine, cysteine, cysteinate, methionine, tyrosine,

tyrosinate, threonine, glutamate and also a water molecule (Figure 6.19). The different compounds, with formula $[Pt(NH_3)_2X]^{+/2+}$ where X represents the coordinating ligand, have been minimized by means of Gaussian09 calculations carried out at the DFT/B3LYP level of theory. The chosen basis sets were 6-31+G* for the main group elements and LANL2DZ for the platinum atom and its pseudopotential. Energies were recalculated increasing the size of the basis set of the main group elements to 6-311+G*. Also some tests with the LANL2TZ basis for Pt and its pseudopotential were performed, but the results do not show remarkable improvements comparing them with the double-ζ basis set. To ensure the quality of the geometry minimization, frequency calculations were further carried out in standard conditions (298.15 K and 1 atm of pressure). All the amino acids are in the standard protonation state (i.e. histidine residue corresponds to the neutral histidine δ) or in the state that has been attributed to the metal binding (i.e. tyrosinate). For glutamate, two possible coordination modes were found, depending on the chelation of the residue which can be bonded by one oxygen or by both yielding to a final tetracoordinated geometry. The later is referred as η^2-glutamate.

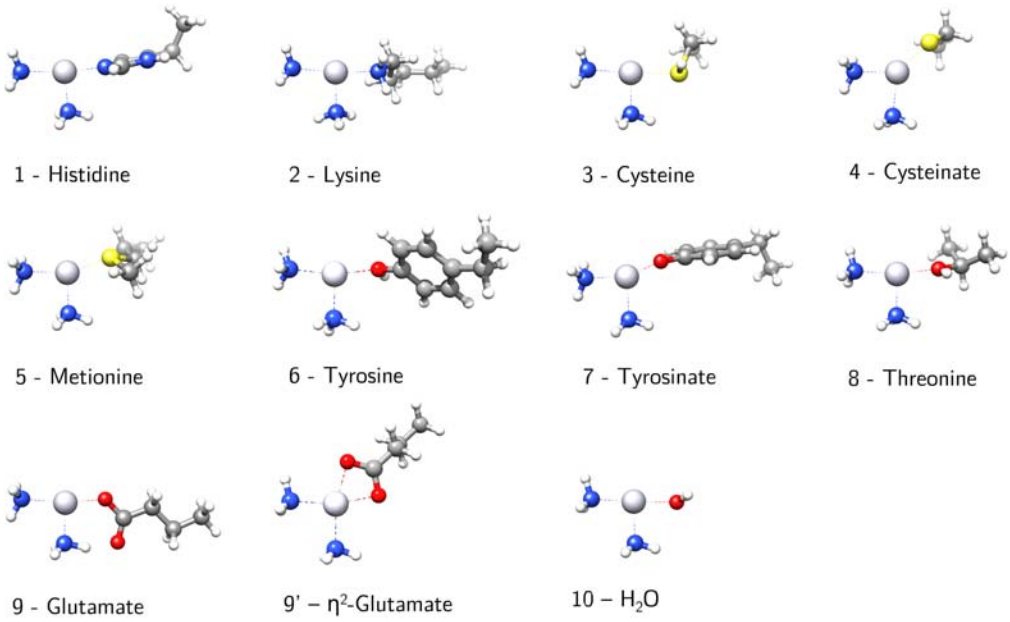


Figure 6.20: Geometry optimized tricoordinated structures.

Minimized geometries show the expected results in almost all the cases, as it can be seen in Table 6.5 and in Figure 6.20. Pt-N_β distances are quite similar in all the compounds, being 0.03Å the maximum difference in the bond length. Sulphur containing ligands, as well as the negatively charged ones presents longer bond distances. η^2-glutamate differs from the η^1-glutamate form due to the double chelation.

The N_α-Pt-X bond give us some clues about the T-shape of the different complexes. For water, histidine, lysine, cysteine, tyrosine, threonine and glutamate deviations from a perfect T-shaped

geometry do not exceed 13°. Methionine and tyrosinate present a higher deviation of 20 and 21° respectively. The ideal Y-shape is almost reached by the cysteinate compound, with a N_α-Pt-X angle of 137°.

Table 6.5: Geometrical parameters for tricoordinated Pt(II) complexes

	Pt-N_α	**Pt-N_β**	**Pt-X**	**N_α-Pt-X**
1 - His	2.13	2.03	2.01	173.1
2 - Lys	2.13	2.04	2.09	173.5
3 - Cys	2.14	2.03	2.38	176.3
4 - Cys^-	2.15	2.15	2.17	137.3
5 - Met	2.15	2.05	2.28	159.5
6 - Tyr	2.09	2.09	2.09	167.1
7 - Tyr^-	2.15	2.13	1.93	158.7
8 - Thr	2.09	2.03	2.05	169.0
9 - Glu	2.13	2.01	1.97	169.8
9' - η^2-Glu	2.10	2.10	2.07	167.1
10 - H_2O	2.06	2.02	2.11	177.4

Tetracoordinated structures does not show many differences with the tricoordinated ones. In Table 6.6 one can see the geometry variables which can be modified by the presence of a water molecule, or other ligand, in the vacant site, which are the Pt-X bond distance and the N_α-Pt-X (remember that the N_α is in trans position with respect to the X amino acid).

Pt-X bond lengths are, in the majority of the cases, slightly longer (less than 0.1Å) in tetracoordinated structures. Some exceptions are the cysteinate and tyrosinate, which differences from the tricoordinated systems are 0.16Å and 0.31Å, respectively. N_α-Pt-X angles are, in all the cases, no more than 10° away from the expected 180°, being 170.9°, which corresponds to the glutamate complex, the most distorted one. As it can be seen in Figure 6.21, the loose of a water molecule in the studied square planar complexes does not implies highlighting structural rearrangements.

In this section, the possibility of existence of trigonal planar structures in a T-shaped configuration has been proved in gas phase conditions. Also, an unique Y-shaped compound corresponding to the cysteinate complex appears stable. The next step is to evaluate the strength of the Pt-X bond in order to find the any energetic clue to understand the existence of the reported T-shaped cisplatin-protein adducts.

Energy Decomposition of the bond between cisplatin and amino acids

An efficient way to sort the strength of the interaction of cisplatin with protein residues is performing an Energy Decomposition Analysis (EDA).[71,72] EDA calculates the interaction between two

Table 6.6: Highlighted differences between tri and tetracoordinated structures

	Pt-X		**N_α-Pt-X**	
	tricoordinated	**tetracoordinated**	**tricoordinated**	**tetracoordinated**
1 - His	2.01	2.04	173.1	177.2
2 - Lys	2.09	2.09	173.5	177.1
3 - Cys	2.38	2.38	176.3	172.7
4 - Cys^-	2.17	2.33	137.3	175.1
5 - Met	2.28	2.36	159.5	178.0
6 - Tyr	2.09	2.11	167.1	176.3
7 - Tyr^-	1.93	2.14	158.7	178.8
8 - Thr	2.05	2.08	169.0	174.6
9 - Glu	1.97	2.02	169.8	170.9
10 - H_2O	2.11	2.13	177.4	179.1

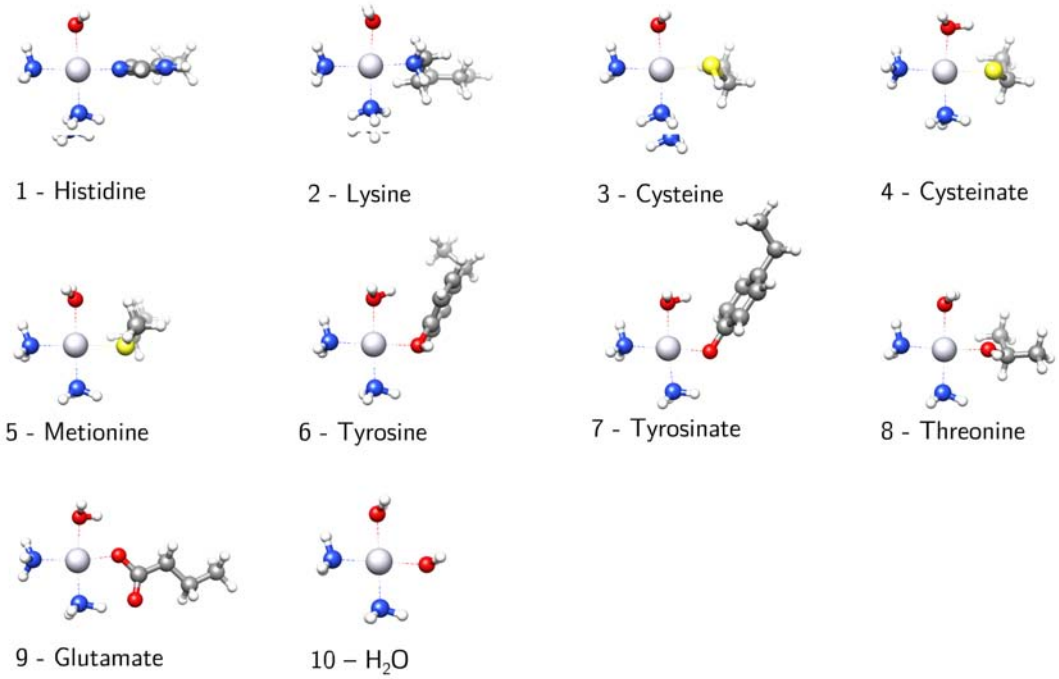


Figure 6.21: Geometry optimized tetracoordinated structures.

fragments of the same molecule, and then decomposes the energy of this interaction in different terms: electrostatic energy, orbital interaction, Pauli repulsion, and a combination of two or more of the mentioned terms (more details in the Material and Methods chapter of this dissertation). In this section, one fragment corresponds to the drug moiety being either $Pt(NH_3)_2$ for the tricoordinated species or $Pt(NH_3)_2H_2O$ for tetracoordinated ones. The second fragment is the bonded amino acid. In this case study, the employed amino acids were the same than in the previous section: histidine,

lysine, cysteine, cysteinate, metionine, tyrosine, tyrosinate, threonine and glutamate (the last in its η^1 and η^2 forms).

Calculations have been performed under the EDA scheme as implemented in the ADF 12.0 package,[73] at DFT B3LYP level of theory using TZP (core double zeta, valence triple zeta, polarized) as basis set, adding the zero order regular approximation (ZORA)[74–76] to take into account the relativistic corrections of the platinum center.[77]

Regarding the absolute energy values of the bonds formed between the metal and the residue (X), for tricoordinated compounds, the Pt-X bond energies (Table 6.7), are strongest (about 300 kcal/mol) for cysteinate, glutamate and tyrosinate. Then, histidine, metionine and lysine present bond energies around 100 kcal/mol, followed closely by cysteine, tyrosine and threonine (89, 81 and 78 kcal/mol, respectively). The Pt-H_2O bond shows the lowest value: 59 kcal/mol. The EDA analysis of these bonds show that the bond energy terms that dominate are the electrostatic energy and orbital interaction terms on the charged residues over the neutral ones. η^2-Glutamate, due to its bidentate form, presents the most stable Pt-X interaction but being not so far from the cysteinate compound.

Table 6.7: Decomposed binding energies of cisplatin to the selected residues for tricoordinated compounds. Energies in kcal/mol.

	-$B_{\Delta E}$	Electrostatic en.	Pauli rep.	Orbital int.	ΔE preparation
1 - His	-119.8	-164.6	137.3	-98.0	5.4
2 - Lys	-99.2	-134.3	116.4	-85.7	4.5
3 - Cys	-89.4	-103.1	106.3	-95.7	3.1
4 - Cys^-	-331.6	-414.1	249.5	-175.3	8.3
5 - Met	-102.5	-120.2	126.0	-114.4	6.0
6 - Tyr	-81.2	-72.6	64.1	-78.9	6.1
7 - Tyr^-	-304.5	-304.1	147.8	-157.1	7.9
8 - Thr	-78.4	-95.4	80.5	-70.3	6.7
9 - Glu	-306.0	-334.8	149.1	-124.6	4.3
9' - η^2-Glu	-335.6	-368.6	174.2	-147.7	5.6
10 - H_2O	-59.2	-73.5	58.2	-45.1	1.1

For tetracoordinated compounds the order presented above is mainly mantained (Table 6.8 and Figure 6.22): the most stable bonds correspond to negatively charged amino acids (cysteinate, glutamate and tyrosinate) with an average value of 280 kcal/mol; followed by histidine, lysine and methionine (103, 85 and 84 kcal/mol, respectively); then cysteine, threonine and tyrosine (76 kcal/mol for cys, and 65 kcal/mol for thr and tyr); finally the lower value corresponds to the bound Pt-H_2O (50 kcal/mol). In these systems, the electrostatic and orbital terms dominate the bonding energy, which is identical to the tendency observed in the tetracoordinated systems. The presence of a water molecule in *cis* position with respect to the residue destabilizes the Pt-X bond from 30 kcal/mol (for negatively charged residues) to an average of 15 kcal/mol for the neutral ones. The lack of a fourth ligand in tricoordinated compounds is compensated by a stronger Pt-X bond.

Table 6.8: Decomposed binding energies of cisplatin to the selected residues for tetracoordinated compounds. Energies in kcal/mol.

	-$B_{\Delta E}$	Electrostatic en.	Pauli rep.	Orbital int.	ΔE preparation
1 - His	-103.2	-151.2	126.9	-83.2	4.37
2 - Lys	-84.6	-130.6	117.8	-78.5	6.7
3 - Cys	-76.1	-98.7	100.9	-81.6	3.3
4 - Cys^-	-292.2	-360.6	182.2	-125.5	11.7
5 - Met	-83.6	-108.6	108.1	-90.8	7.7
6 - Tyr	-64.5	-73.1	66.5	-64.0	6.1
7 - Tyr^-	-273.1	-286.5	120.4	-119.3	12.4
8 - Thr	-64.7	-82.6	69.7	-56.3	4.5
9 - Glu	-276.6	-316.6	133.2	-104.8	11.6
10 - H_2O	-49.5	-67.0	55.3	-39.1	1.4

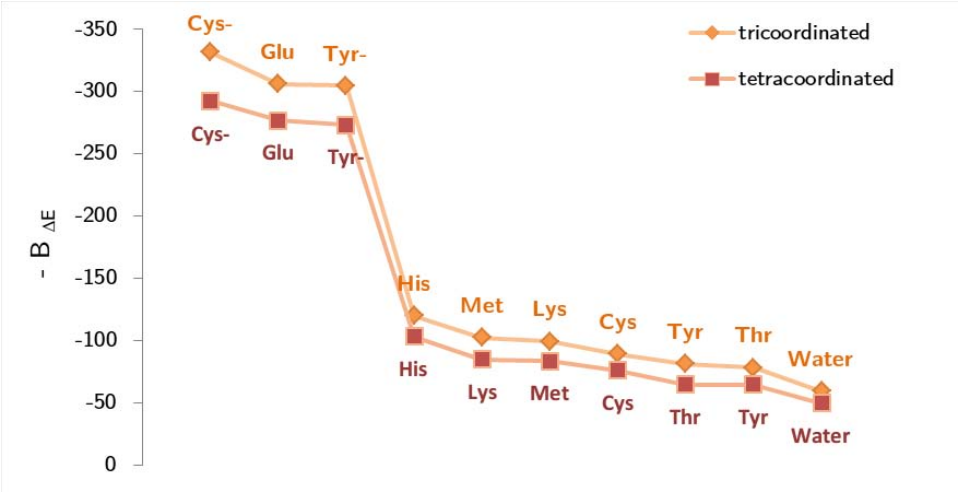


Figure 6.22: Bond energies of both tri and tetracoordinated compounds.

From the energy decomposition analysis, the energy differences for the cisplatin-amino acid bonds in neutral compounds are not so large (about 10kcal/mol in favour of the tricoordinated compounds. Larger differences are found for negatively charged species (See Table 6.7 and Table 6.8). Despite of this, the information reported here does not prove the formation of these compounds in physiological conditions and its thermodynamic availability, which it will be investigated in the next section.

From square to trigonal planar geometries

One of the X-ray tricoordinated structures reported in the initial part of this study shows a weak signal in the vacant site, being described with the authors as a labile water molecule. In order to study the stability of the formed tricoordinated compound with respect to the original square planar one, the Gibbs energy of the dehydration process has been calculated in different media where the cisplatin drug can be present once its administration: water, blood plasma, protein surface

and inside the protein matrix. To accurately compute the energy of dehydration in all the media, calculations have been performed using continuum solvent models. All the simulations have been carried out in the same way than in the previous section, at DFT B3LYP level of theory and using TZP (core double zeta, valence triple zeta and polarized) as basis set, adding the zero order regular approximation (ZORA), using the ADF 12.0 package. Solvent effects were computed using the conductor-like screening model (COSMO[78–80]) also implemented in ADF 12.0. To mimic the environment as well as possible, four different dielectric constants (ε) have been used depending on the simulated media: water (ε=78.4), blood plasma (ε=58.0),[81] protein surface or accessible site (ε=9.0), and inside the protein matrix (ε=2.0).[82,83]

Additionally, two corrections were further applied in the calculation of Gibbs energies in order to standardize the solvent conditions .[84–86] The first one corrects the concentration of 1 mol of ideal gas (24.46 L/mol) to 1 M (Equation 6.1):

$$\Delta G^{0\rightarrow *} = RTln(22.46) \tag{6.1}$$

The later refers to the concentration of one solvent (water) molecule, which is the case in this section, in a pure water solution (55.34 M) (Equation 6.2):

$$\Delta G^{*\rightarrow 1} = RTln(55.34) \tag{6.2}$$

This second correction is only applied in the ε=78.4 and ε=58 cases because both corresponds to aqueous media.

The available quantum mechanics software does not directly calculate the ΔG^* in solvent, being mandatory the use of thermodynamic cycles to obtain that value (Figure 6.23). Results are presented in Table 6.9.

Table 6.9: Gibbs energies of the dehydration (in kcal/mol).

	ΔG			
	ε=78.4	ε=58	ε=9.0	ε=2.0
1 - His	17.3	17.6	17.6	25.5
2 - Lys	18.1	18.2	18.4	27.3
3 - Cys	18.9	19.1	19.1	27.3
4 - Cys^-	11.2	11.5	9.8	11.9
5 - Met	20.3	20.5	20.1	26.5
6 - Tyr	12.6	13.1	14.4	26.4
7 - Tyr^-	8.2	9.3	10.0	17.5
8 - Thr	19.7	19.6	19.7	28.2
9 - Glu	16.0	16.1	15.5	21.4
9' - η^2-Glu	-57.1	-56.7	-58.5	-56.9
10 - H_2O	20.0	21.3	21.5	30.9

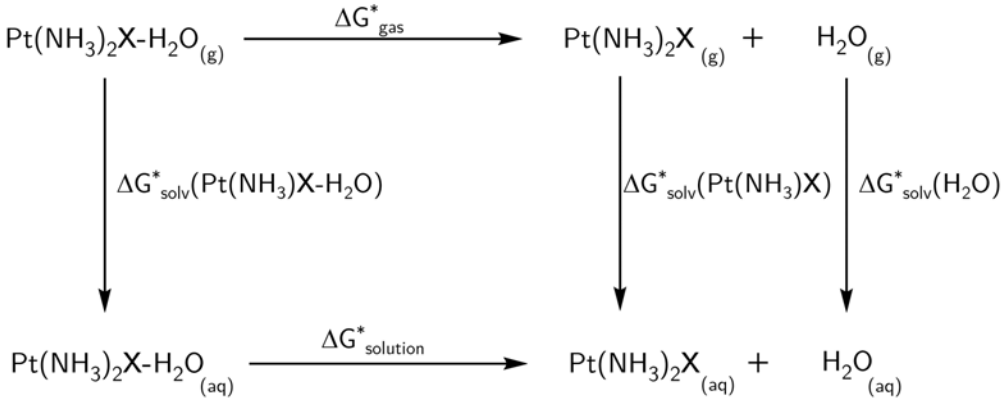


Figure 6.23: Thermodynamic cycle used in this study.

With the exception of the η^2-glutamate complex, all the systems present an endergonic Gibbs energy of dehydration, which means that under physiological conditions, the release of the coordinated water molecule is not spontaneous. The species with lower Gibbs energy of dehydration are cysteinate (4), tyrosine (6) and tyrosinate (7), with a range of values between 8.2 to 26.4 kcal/mol. Histidine, lysine, cysteine and glutamate present slightly higher ΔG_{sol} values, with a range of 16.0 to 27.3 kcal/mol. The higher ΔG_{sol} values correspond to the water, metionine and threonine complexes, with a range of 19.6 to 30.9 kcal/mol.

Independently of the dielectric constant of the medium, the general tendency is maintained. The variation of ε from 78.4 to 9 does not alter significantly the Gibbs energy of dehydration. However, for ε=2, which represents the inner binding site environment, the Gibbs dehydration energy increases between 6 and 12 kcal/mol, due to the high hydrophobicity of the media. Two exceptions can be observed here, cysteinate and η^2-glutamate, which ΔG_{sol} energy do not vary.

In this part of the study, the impossibility of the presented models to loose spontaneously the coordinated water molecule has been demonstrated. Despite of this, the existence of tricoordinate cisplatin-protein adducts have been reported in the literature. The only unexplored variable which can explain this fact is the effect of the protein environment, which will be analysed in the next section.

Explicit solvation of the second coordination sphere of cisplatin in hen egg lysozyme

The effect of the protein environment has been explicitly analysed in this section using the QM/MM approach implemented in the ONIOM scheme. The quantum mechanical region is calculated at DFT B3LYP level of theory, and includes the $Pt(NH_3)_2$ moiety, the His15 of the hen egg lysozyme and

four water molecules placed nearby the inorganic ligand. The rest of the system has been modelled using the AMBER forcefield. The entire system contains 2548 atoms, 32 described at QM level. Atoms in a sphere of 5 Å around the metal centre (a total of 723) were allowed full flexibility (Figure 6.24).

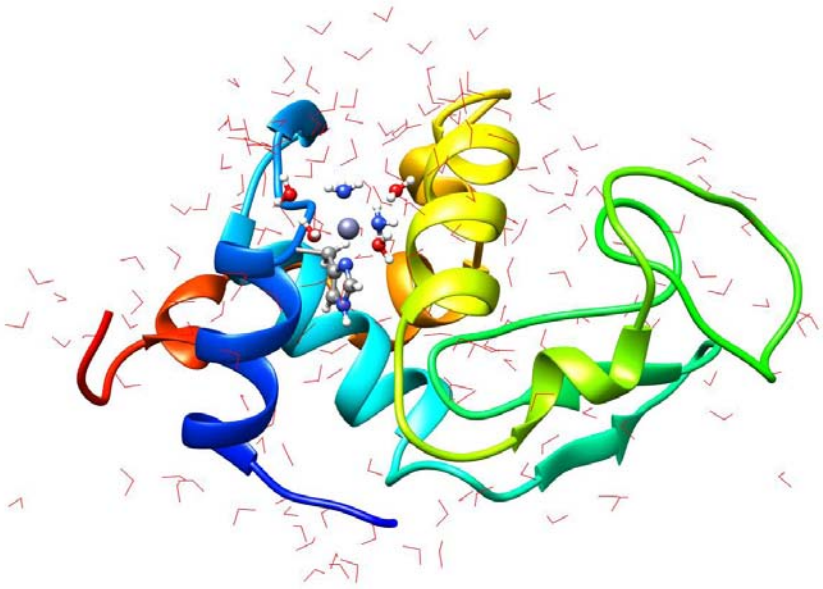


Figure 6.24: QM/MM partition of the system. QM atoms are shown in a ball and strick representation. MM part is composed by the wire atoms and the rest of the protein. Water molecules from the solvent media are coloured in red.

The QM/MM optimized geometry presents a high similarity with the experimental one characterized by Casini and coworkers[70] (see Table 6.10) with an unique, but important, difference: the calculated structure presents a coordination of a water molecule in the vacant site.

Table 6.10: Structural parameters of both synthesized and calculated cisplatin-hen egg lysozime adducts. Distances in Å and angles in degrees.

	Experimental	Calculated
Pt-N_{α}	2.02	2.02
Pt-N_{β}	2.02	2.02
Pt-N_{His}	2.13	2.08
Pt-OH_2	-	2.13
N_{α}-Pt-N_{β}	89.7	89.9
N_{β}-Pt-OH_2	-	173.7
N_{β}-Pt-N_{His}	94.2	91.5

Calculations do not show any local effect of the protein environment which can stabilize the T-shaped coordination of platinum in hen egg lysozime, avoiding the coordination of a water molecule present in the second coordination sphere. Nevertheless, the system presents a second water molecule at 2.6Å away from the metal centre (see Figure 6.25). The minimal difference between the platinum

centre and the two closer water molecules suggest a possible equilibrium, where the two solvent molecules exchange their positions by means of a SN2 mechanism.[87] This equilibrium can explain the lack of density in the fourth vacancy of the metal. Nevertheless, as the authors of the crystallographic structure pointed out, the experimental tri-coordinated geometry of the complexes can be an artefact of the unavailability of X-ray techniques to identify water molecules near transition metals.

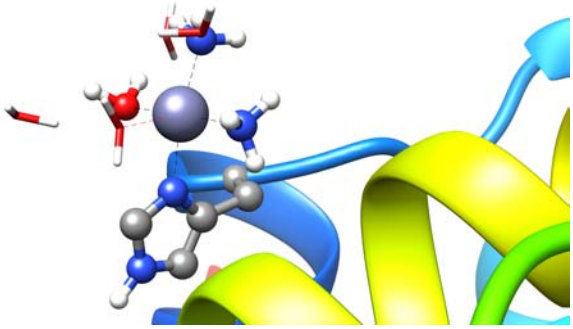


Figure 6.25: Representation of the QM part of the simulation.

6.4 Conclusions

The interaction of cisplatin with DNA and proteins has been the focus of this chapter. In the first part, the interaction of cisplatin drug with DNA was analysed. Once cisplatin reaches the DNA, the coordination deforms the double strand in different ways, depending on the relative position of the two guanine bases coordinated to the cisplatin drug in the DNA double strand. Performing an exploration of the relative position of both guanine bases in the cisplatin-GG adduct, one can almost predict the kind of adduct which is formed in the binding process. However, a local study of the cisplatin-GG compound is not enough, being necessary adding the long-range effects of the binding of cisplatin to DNA to have a broad description of the molecular effects of cisplatin in DNA.

The second part is focused on answering a specific question: the possibility of tricoordinated cisplatin adducts with proteins. This study has been motivated by the surprising structure of two crystallographic structures, reported by Casini and coworkers, which presents a tricoordinated cisplatin-protein adduct. Calculations on cluster models shows the possibility of existence of T-shaped (and one case of Y-shaped) platinum complexes in gas phase. Deeper studies on the bond energy between platinum and different amino acids reveal that the strength of this bond is higher than in the square-planar ones, to supply the lack of the fourth ligand. However, Gibbs energy of dehydration calculations demonstrate that tricoordinated compounds can not be formed under physiological conditions. Finally, the possibility of the protein environment to stabilize the tricoordinated geometry has been discarded due to the presence of two water molecules at less than 3Å away from the metal.

In this chapter, four different computational methods have been used to describe the molecular effect of the coordination of cisplatin drug to biological hosts, being a clear example of how the conjunction of different techniques can be crucial in the description of bioinorganic processes.

6.5 References

[1] Peyrone, M. *Liebigs Ann. Chem.* **1844**, *51*, 1–29.

[2] Rosenberg, B.; Van Camp, L.; Krigas, T. *Nature* **1965**, *205*, 698–699.

[3] Rosenberg, B.; Vancamp, L. *Nature* **1969**, *222*, 385–386.

[4] Sherman, S. E.; Lippard, S. J. *Chem. Rev.* **1987**, *87*, 1153–1181.

[5] Kelland, L. *Nat. Rev. Cancer* **2007**, *7*, 573–584.

[6] *PLATINOL- cisplatin injection, powder, lyophilized, for solution*, http://dailymed.nlm.nih.gov/dailymed/lookup.cfm?setid=cd53d182-3d90-4a67-81d0-c61353aff4f6, Accessed: 2010-10.

[7] *Cisplatin Injection*, http://www.nlm.nih.gov/medlineplus/druginfo/meds/a684036.html, Accessed: 2011-10-15.

[8] Gibson, D. *Dalton T.* **2009**, 10681–10689.

[9] Eastman, A. *Biochemistry* **1986**, *25*, 3912–3915.

[10] Fichtinger-Schepman, A. M. J.; van Oosterom, A. T.; Lohman, P. H.; Berends, F. *Cancer Res.* **1987**, *47*, 3000–3004.

[11] Baik, M.-H.; Friesner, R. A.; Lippard, S. J. *J. Am. Chem. Soc.* **2003**, *125*, 14082–14092.

[12] Mantri, Y.; Lippard, S. J.; Baik, M.-H. *J. Am. Chem. Soc.* **2007**, *129*, 5023–5030.

[13] Mandal, R.; Kalke, R.; Li, X.-F. *Chem. Res. Toxicol.* **2004**, *17*, 1391–1397.

[14] Anderson, M. E.; Naganuma, A.; Meister, A. *FASEB J.* **1990**, *4*, 3251–3255.

[15] Greggi Antunes, L. M.; Darin, J. D.; Bianchi, M. d. L. P. *Pharmacol. Res.* **2001**, *43*, 145–150.

[16] Carsey, T.; Boudreaux, E. *Chem.-Biol. Interac.* **1980**, *30*, 189–201.

[17] Lipiński, J. *Inorg. Chim. Acta* **1988**, *152*, 151–157.

[18] Lipiński, J. *J. Mol. Struc-Theochem* **1989**, *201*, 295–305.

[19] Basch, H.; Krauss, M.; Stevens, W. J. *Inorg. Chem.* **1985**, *24*, 3313–3317.

[20] Carloni, P.; Andreoni, W.; Hutter, J.; Curioni, A.; Giannozzi, P.; Parrinello, M. *Chem. Phys. Lett.* **1995**, *234*, 50–56.

[21] Burda, J. V.; Leszczynski, J. *Inorg. Chem.* **2003**, *42*, 7162–7172.

[22] Carloni, P.; Sprik, M.; Andreoni, W. *J. Phys. Chem. B* **2000**, *104*, 823–835.

[23] Spiegel, K.; Rothlisberger, U.; Carloni, P. *J. Phys. Chem. B* **2004**, *108*, 2699–2707.

[24] Robertazzi, A.; Platts, J. A. *Chem. Eur. J.* **2006**, *12*, 5747–5756.

[25] Gkionis, K.; Platts, J. A. *J. Biol. Inorg. Chem.* **2009**, *14*, 1165–1174.

[26] Gkionis, K.; Platts, J. A. *Comp. Theor. Chem.* **2012**, *993*, 60–65.

[27] Gkionis, K.; Mutter, S. T.; Platts, J. A. *RSC Adv.* **2013**, *3*, 4066–4073.

[28] Bhattacharyya, D.; Ramachandran, S.; Sharma, S.; Pathmasiri, W.; King, C. L.; Baskerville-Abraham, I.; Boysen, G.; Swenberg, J. A.; Campbell, S. L.; Dokholyan, N. V. et al. *PloS ONE* **2011**, *6*, e23582.

[29] Coste, F.; Malinge, J.-M.; Serre, L.; Leng, M.; Zelwer, C.; Shepard, W.; Roth, M. *Nucleic Acids Res.* **1999**, *27*, 1837–1846.

[30] Huang, H.; Zhu, L.; Reid, B. R.; Drobny, G. P.; Hopkins, P. B. *Science* **1995**, *270*, 1842–1845.

[31] Wu, Y.; Bhattacharyya, D.; King, C. L.; Baskerville-Abraham, I.; Huh, S.-H.; Boysen, G.; Swenberg, J. A.; Temple, B.; Campbell, S. L.; Chaney, S. G. *Biochemistry* **2007**, *46*, 6477–6487.

[32] Coste, F.; Shepard, W.; Zelwer, C. *Acta Crystallogr. D* **2002**, *58*, 431–440.

[33] Yang, D.; van Boom, S. S.; Reedijk, J.; van Boom, J. H.; Wang, A. H.-J. *Biochemistry* **1995**, *34*, 12912–12920.

[34] Garderen, C. J.; Houte, L. *Eur. J. Biochem.* **1994**, *225*, 1169–1179.

[35] Ohndorf, U.-M.; Rould, M. A.; He, Q.; Pabo, C. O.; Lippard, S. J. *Nature* **1999**, *399*, 708–712.

[36] Gelasco, A.; Lippard, S. J. *Biochemistry* **1998**, *37*, 9230–9239.

[37] Todd, R. C.; Lippard, S. J. *Chem. Biol.* **2010**, *17*, 1334–1343.

[38] Todd, R. C.; Lippard, S. J. *J. Inorg. Biochem.* **2010**, *104*, 902 – 908.

[39] Takahara, P. M.; Rosenzweig, A. C.; Frederick, C. A.; Lippard, S. J. *Nature* **1995**, *377*, 649–652.

[40] Frisch, M. J. et al. *Gaussian 09 Revision D.01*, sian Inc. Wallingford CT 2009.

[41] Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 1372–1377.

[42] Lee, C.; Yang, W.; Parr, R. G. *Phys. Rev. B* **1988**, *37*, 785–789.

[43] Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 5648–5653.

[44] Hehre, W. J.; Ditchfield, R.; Pople, J. A. *J. Chem. Phys.* **1972**, *56*, 2257–2261.

[45] Hariharan, P. C.; Pople, J. A. *Theor. Chim. Acta* **1973**, *28*, 213–222.

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

[47] Dapprich, S.; Komáromi, I.; Byun, K. S.; Morokuma, K.; Frisch, M. J. *J. Mol. Struc-Theochem* **1999**, *461*, 1–21.

[48] Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. *J. Am. Chem. Soc.* **1995**, *117*, 5179–5197.

[49] Wang, J.; Wang, W.; Kollman, P. A.; Case, D. A. *J. Mol. Graph. Model.* **2006**, *25*, 247 – 260.

[50] Kumar, D.; Altun, A.; Shaik, S.; Thiel, W. *Faraday discuss.* **2011**, *148*, 373–383.

[51] Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926–935.

[52] Pettersen, E. F.; Goddard, T. D.; Huang, C. C.; Couch, G. S.; Greenblatt, D. M.; Meng, E. C.; Ferrin, T. E. *J. Comput. Chem.* **2004**, *25*, 1605–1612.

[53] Barone, V.; Cossi, M. *J. Phys. Chem. A* **1998**, *102*, 1995–2001.

[54] Cossi, M.; Rega, N.; Scalmani, G.; Barone, V. *J. Comput. Chem.* **2003**, *24*, 669–681.

[55] Cui, Q.; Bahar, I. *Normal mode analysis: theory and applications to biological and chemical systems*; CRC press, 2005.

[56] Hinsen, K. *J. Comput. Chem.* **2000**, *21*, 79–85.

[57] Scheeff, E. D.; Briggs, J. M.; Howell, S. B. *Mol. Pharmacol.* **1999**, *56*, 633–643.

[58] Chou, K.-C. *Biochem. J* **1984**, *221*, 27–31.

[59] Neault, J.; Tajmir-Riahi, H. *BBA-Protein Struct. M* **1998**, *1384*, 153–159.

[60] Raber, J.; Zhu, C.; Eriksson, L. A. *J. Phys. Chem. B* **2005**, *109*, 11006–11015.

[61] Neault, J.; Diamantoglou, S.; Nafisi, S.; Tajmir-Riahi, H. *J. Photoch. Photobio. B* **2008**, *91*, 167–174.

[62] Wally, J.; Halbrooks, P. J.; Vonrhein, C.; Rould, M. A.; Everse, S. J.; Mason, A. B.; Buchanan, S. K. *J. Biol. Chem.* **2006**, *281*, 24934–24944.

[63] Luo, L.-Z.; Jin, H.-W.; Huang, H.-Q. *J. Proteomics* **2012**, *77*, 237–250.

[64] Ortuño, M. A.; Conejero, S.; Lledós, A. *Beilstein J. Org. Chem.* **2013**, *9*, 1352–1382.

[65] Braunschweig, H.; Radacki, K.; Rais, D.; Scheschkewitz, D. *Angew. Chem. Int. Edit.* **2005**, *117*, 5796–5799.

[66] Mole, L.; Spencer, J. L.; Carr, N.; Orpen, A. G. *Organometallics* **1991**, *10*, 49–52.

[67] Brainard, R. L.; Nutt, W. R.; Lee, T. R.; Whitesides, G. M. *Organometallics* **1988**, *7*, 2379–2386.

[68] Butts, M. D.; Scott, B. L.; Kubas, G. J. *J. Am. Chem. Soc.* **1996**, *118*, 11831–11843.

[69] Casini, A.; Mastrobuoni, G.; Temperini, C.; Gabbiani, C.; Francese, S.; Moneti, G.; Supuran, C. T.; Scozzafava, A.; Messori, L. *Chem. Commun.* **2006**, 156–158.

[70] Calderone, V.; Casini, A.; Mangani, S.; Messori, L.; Orioli, P. L. *Angew. Chem. Int. Edit.* **2006**, *45*, 1267–1269.

[71] Morokuma, K. *J. Chem. Phys.* **1971**, *55*, 1236–1244.

[72] Ziegler, T.; Rauk, A. *Theor. Chim. Acta* **1977**, *46*, 1–10.

[73] Te Velde, G.; Bickelhaupt, F. M.; Baerends, E. J.; Fonseca Guerra, C.; van Gisbergen, S. J.; Snijders, J. G.; Ziegler, T. *J. Comput. Chem.* **2001**, *22*, 931–967.

[74] van Lenthe, E.; Baerends, E.-J.; Snijders, J. G. *J. Chem. Phys.* **1993**, *99*, 4597–4610.

[75] van Lenthe, E.; Baerends, E.-J.; Snijders, J. G. *J. Chem. Phys.* **1994**, *101*, 9783–9792.

[76] van Lenthe, E.; Ehlers, A.; Baerends, E.-J. *J. Chem. Phys.* **1999**, *110*, 8943–8953.

[77] Torrent, M.; Sola, M.; Frenking, G. *Chem. Rev.* **2000**, *100*, 439–494.

[78] Pascual-Ahuir, J.; Silla, E.; Tomasi, J.; Bonaccorsi, R. *J. Comput. Chem.* **1987**, *8*, 778–787.

[79] Klamt, A.; Schüürmann, G. *J. Chem. Soc.* **1993**, 799–805.

[80] Klamt, A. *J. Phys. Chem.-US* **1995**, *99*, 2224–2235.

[81] Cook, H. *Brit. J. Appl. Phys.* **1952**, *3*, 249–255.

[82] Dudev, T.; Lim, C. *J. Am. Chem. Soc.* **2000**, *122*, 11146–11153.

[83] Dudev, T.; Lim, C. *J. Phys. Chem. B* **2000**, *104*, 3692–3694.

[84] Pliego Jr, J. R. *Chem. Phys. Lett.* **2003**, *367*, 145–149.

[85] Bryantsev, V. S.; Diallo, M. S.; Goddard Iii, W. A. *J. Phys. Chem. B* **2008**, *112*, 9709–9719.

[86] Alí-Torres, J.; Rodríguez-Santiago, L.; Sodupe, M. *Phys. Chem. Chem. Phys.* **2011**, *13*, 7852–7861.

[87] Cooper, J.; Ziegler, T. *Inorg. Chem.* **2002**, *41*, 6614–6622.