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**

# 7 ⋯ Enantioselective formation of artificial metallopeptides

Related paper:
**Appendix A.5:** Programmed stereoselective assembly of
DNA-binding helical metallopeptides
**Appendix A.6:** The folding of a metallopeptide

*I feel very adventurous.*
*There are so many doors to be opened,*
*and I'm not afraid to look behind them.*

**Elizabeth Taylor**

## 7.1 Using metallopeptides at the interface between biology and chemistry

One can not try to describe the word "peptide" without thinking only in the biological world. Having in mind that the minimal proteic entity is the amino acid, one can consider that two amino acids linked via peptide bond can form the smallest representation of a peptide. Despite of its small size, peptides are involved in a vast number of biological processes, such as cell apoptosis,[1] anti-microbial protection[2] and membrane disruption.[3] Apart from their biological influence, peptides have a vast array of applications. As an example, peptides can be used as biosensors;[4] some peptide sequences act as substrates for specific enzymes, being key in assays of enzymatic activity and screening of enzymatic inhibitors. An other application is the synthesis of biomaterials to create polymers[5] to be applied in medicine or in tissue engineering.[6]

Despite of it small size, peptides present a high flexible structure, basically provided by the rotations about the N-C$^\alpha$ and C$^\alpha$-C bonds from the side chain of the amino acids of the peptide chain. However, not all the peptides are equally flexible. Some residues can hinder the rotation of the mentioned bonds, reducing their flexibility. As an example, glycine is the residue with allows more flexibility to the system, and proline the one which forces a more fixed structure.[7] Peptides also tend to aggregate. There is not a clear explanation of why and how peptides aggregate but one of the reasons is due to their flexibility. Additionally, the electrostatic repulsion between charged and/or polar residues can affect the aggregation,[8] but also the hydrophobicity of peptides plays a crucial role.[9] An other feature of peptides is their interaction with metals. Polar and charged amino acids of the peptide chain are susceptible to coordinate metals easily. This coordination can also favour the aggregation of peptides[10] and the interaction between peptides and macromolecules;[11] indeed this chemical process is one of the causes of some neurological disorders such as Alzheimer and Parkinson diseases.[12] The folding process of peptides is still nowadays a mystery and remains as one of the main challenges in peptide research. Peptides can fold towards lots of different structures due to the vast conformational space done by the high flexibility of the system. In the particular case of metallopeptides, metals tend to bind peptides in different modes of coordination, interacting with the peptide or with external ligands, such as solvent molecules. However, the major driving force for peptide folding is done by the hydrophobicity of peptides and also its interaction with solvent molecules[13] and the relative contribution of the metal is not clear.

The high flexibility, the tendency to aggregate and the easy coordination to the metal, in conjunction with its small size, make peptides attractive templates for the design of biologically friendly coordination compounds that integrate precise structural and biofunctional properties encoded by the amino acid sequence of the peptide ligands.[14] Moreover, the same statements make the design and study of peptides in general, and metallopeptides in particular, a great challenge. The control of the residues that can participate in the first and the second coordination sphere of the metal is extremely difficult. A promising approach to solve this limitation consists in incorporating chemical groups with strong metal affinity in the same peptide sequence (i.e. bipyridine residues). With these artificial peptides, the first coordination sphere of the metal is easily defined and the researcher can manipulate the asymmetric environment of the metal using the rest of the peptidic sequence. Additionally, using chiral amino acids is possible to control the stereoselectivity of a synthetic metallopeptide.[15]

The use of computation in the study of peptide systems does not follow a straight line. Commonly with other fields, the size of the system is the bottleneck in the application of modelling techniques in peptides, being more complicated in the case of metallopeptide research. In the case of big systems, containing thousands of atoms, molecular mechanics methods can describe the macroscopic properties of the system, but can not represent the electronic changes occurring during the design process. Quantum mechanics based techniques are able to accurately calculate the electronic properties of the system, but the computational cost can be unaffordable if the system is big. The use of Quantum mechanics/Molecular mechanics (QM/MM) hybrid methods can take the best part from both molecular and quantum mechanics fields.[16]

An other property that hinders peptide design is the vast number of structural conformations they can reach in the folding process. The complexity of the system makes difficult having an accurate description of the conformational space of short chains at quantum mechanical level, being needed methodologies close to the complete basis set extrapolation and CCSD(T) corrections to have the correct results in small chains such as dipeptide systems.[17] However, other computational approaches at lower level of theory, such as homology modelling,[18] QSAR[19] or statistical studies[20] are also great tools to be applied in the field of peptide research. One example is the use of an integrative protocol composed by quantum mechanical calculations and homology modelling to study of the aggregation process of the amyloid-$\beta$ peptide in presence of metals as a possible cause of the Alzheimer disease.[18,21]

Here we present a *symbiotic* work that encompasses experimental and theoretical efforts to discover the clues on the design of enantioselective artificial metallopeptides. All the results presented here were performed in collaboration with the experimental group of E. Vázquez and M. Vázquez from the *Universidade de Santiago de Compostela*. First, we will introduce the system under study and the origin of this work. Then, the study of the energetic and structural keys on the enantioselective formation of artificial metallopeptides is presented. Finally, one example of the applicability of the enantioselective formation of helical peptides is shown.

## 7.2 A benchmarking system for predicting the enantioselectivity on chiral metallopeptides

In this section, we briefly introduce the metallopeptide under study and the nomenclature followed during all the chapter. Its peptide chain is composed by a set of six amino acids featuring three metal-binding bipyridine units connected by two short loops. The peptide sequence of the loops is the central part of our study, being the responsible of the chiral properties of the peptide chain.

As a systematic nomenclature of the system under study we will enumerate the different parts of the metallopeptide following the description drawn in Figure 7.1. The first terminal position is called $N_{ter}$ (from N-terminal) side, which is directly coordinated with the first bipyridine (the blue one). Linking the first and the second (in green) bipyridines is the N-Loop, which contains the sequence Gly-X-Gly. The C-Loop links the second bipyridine to the last one (in red), and is composed by the sequence Gly-Y-Gly. The other terminal position is called $C_{ter}$ (Figure 7.1).

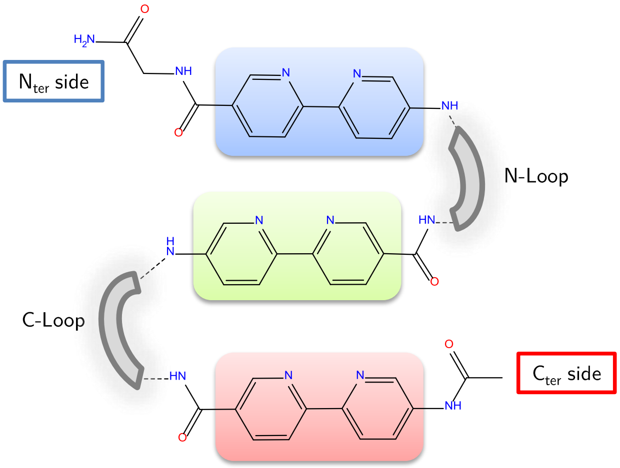


**Figure 7.1:** Schematic representation of the peptide used in this study with labels.

Loops present a $\beta$-turn promoting the sequence that directs the folding of peptide chains into discrete mononuclear species and encodes the chirality of their resulting complexes. In the initial system, both loops contain a chiral proline amino acid. The D or L conformation of this residue drives the orientation of the $\beta$-turn, affecting the position of the coordinating bipyiridines and controlling the $\Delta\Lambda$-isomerism of the metal centre.

### 7.2.1 Previous studies on chiral metallopeptides

A preliminary investigation on the stereoselective synthesis of chiral metallopeptides based on the system before described was performed in 2012 as a collaborative work between the group of E. Vázquez and M. Vázquez from the *Univerdidade de Santiago de Compostela* and J.-D. Maréchal from the *Universitat Autònoma de Barcelona*.[15]

The system under study was a peptide chain formed by 3, 4 and 5 amino acids (one proline and the rest, glycines) with a 2,2'-bipyridine ligand on each end. Then, this peptide was coordinated to a metal (in this specific case Co(II), Ni(II) or Zn(II)), allowing the formation of the final metallopeptide. In the case of Co(II) and Ni(II) systems, the metal prefers to adopt a octahedral geometry, fulfilling the vacant sites with two water molecules. Depending on the coordination of the 2,2'-bipyridine ligands, the metal centre presents a Δ or Λ conformation. For Zn(II) complexes, the metal centre prefers the tetrahedral geometry, yielding the *R* or *S* isomer, depending also in the coordination of the 2,2'-bipyridine parts (Figure 7.2).

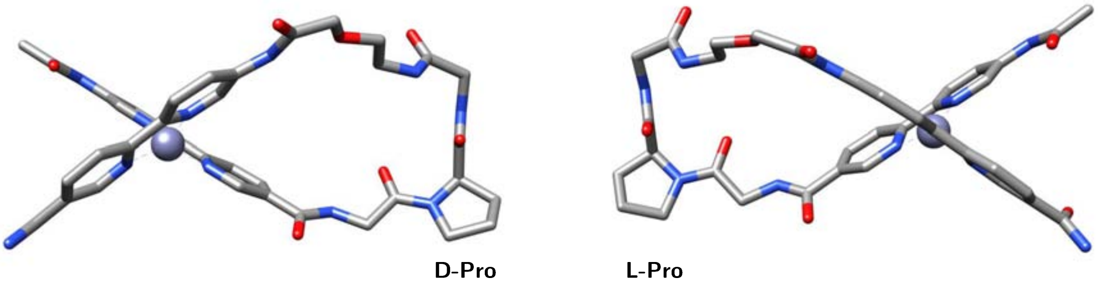


**Figure 7.2:** Structural representation of Zn metallopeptides in both D and L configurations of Proline residue.

The influence of the chirality of the proline residues on the metal stereocentre has been studied applying Circular dichroism (CD) technique.[a] Results show mirror-image spectra for the D-Proline and the L-Proline metallopeptides, which confirms that the conformation of the metal centre depends on the chirality of the proline present in the peptide chain.

To further investigate the structural features of the D and L-Proline peptides and identify the relationship between the chirality of the proline residue and the metal centre, molecular modelling studies have been carried out. The first step was an exploration of the conformational space of the unbound peptide chain using Monte Carlo samplings coupled with low-energy mode displacements.[22] The lowest energy conformations shown the first evidence of the effect of chiral prolines in the peptide chain: D-Proline and L-Proline peptides present an asymmetry in the final structure of the peptide, being the former folded in an anticlockwise configuration and the other in a clockwise way (similar than in Figure 7.2). Structures of the Zn(II) complexes were obtained from the lowest energy conformation of the isolated ligands, applying a constrained minimization that forces the system to have a tetrahedral geometry of the nitrogen atoms of the bipyridine groups. Latter Quantum mechanics/Molecular mechanics calculations on the D-Proline and L-Proline bounded to Zn(II) have been performed. The lower energy structures present a high symmetry between both systems, obtaining the R metal isomer in the D-Proline system, and the S isomer in the other,

---

[a] CD technique measures the difference in the absorption of left-handed circularly polarised light and right-handed circularly polarised light over a range of wavelengths. This methodology is useful to study the conformational changes of a given macromolecule.

something that fitted with the experimental results, but also clearly identified which proline leads to which chiral state of the metal.

## 7.3 Computational description of the enantioselective formation of metallopeptides

This project has been performed hand by hand with the experimental group of *Universidade de Santiago de Compostela* (Spain). All the efforts have been focused in the description of the enantioselective formation process of the metallopeptides synthesized by the experimental group.

### 7.3.1 In quest of the key structural variables

The number of structural variables involved on the enantioselective formation of a metallopeptide is quite big to perform an exhaustive exploration. Therefore, we focused our attention in the exploration of all the possible metal stereoisomers depending on the 2,2'-bipyridine coordination and its effect on the peptide chain.

#### Nomenclature

To make clear the description of the system and the latter presentation of the results, the nomenclature used to identify the isomers follows the next rules:

- The structure of the metal and the coordinated 2,2'-bipyridines are described using the configuration of the stereogenic centre of the metal ($\Delta$ or $\Lambda$) and then a number which identify the isomer, accordingly with Figure 7.3. Example: $\Delta$1.
- Peptides are identified by the configuration of the stereogenic centres of both containing prolines with a two letter code, being first the configuration of the proline placed in the N-Loop, and then the one of the C-Loop (in accordance with Figure 7.1). Examples: DD, DL, LD, LL.
- The identification of the entire metallopeptide is given by first identifying the structure of the metal and then, the configuration of the peptide. Examples: $\Delta$1-DD and its enantiomer, $\Lambda$1-LL.

#### The metal and the 40 stereoisomers

In our case study, the metal centre (Fe(II)) interacts with the lone pairs of the N atoms of the 2,2'-bipyridine ligands bonded to the peptide, achieving an octahedral geometry. The number of possible stereoisomers to take into account in the binding of the metal to the prepared peptide is big, being increased by the fact that 2,2'-bipyridine ligands are not symmetric in our case study due to the different residues coordinated in their *para* position. Despite of that, from the 8 possible structural isomers, only 5 are not forbidden by the structural restrictions prompted by the peptide: the coordination of two consecutive bipyridines have to be in *cis* because the loop involved in the $\beta$-turn is no long enough to allow a *trans* coordination of the second nitrogen of one bipyridine and the first nitrogen atom of the consecutive ligand. In Figure 7.3 a schematic representation of the binding of the three bipyridine ligands (in blue, in red and in green) is given, showing the final arrangement of the chelates around the metal. Any different coordination of the metal to the peptide has been taken into account, due to their high affinity to the 2,2'-bipyridine complexes.

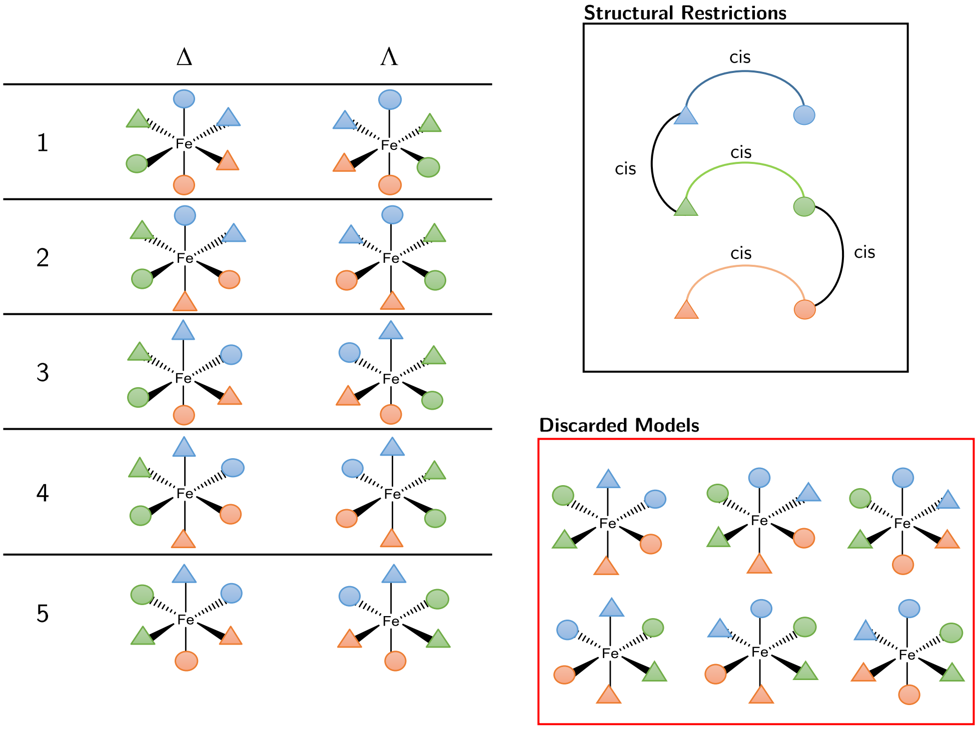


**Figure 7.3:** Schematic representation of the different metal-peptide binding modes. The color of the drawings are in accordance with the color of the bipyridine residues shown in Figure 7.1.

In coordination chemistry, the binding of at least two bichelate ligands to the metal yields to the formation of the $\Delta$ and $\Lambda$ optical isomers (Figure 7.4). In the next section the formation of one or other stereoisomer depending on the chirality of the proline residues is further analysed.

### 7.3.2 Energetic study of the relative stability of Fe-metallopeptides

Once the entire set of Fe-metallopeptides are built, the next step is to calculate the energy differences between the diverse stereoisomers. To do so, a step-by-step analysis has been performed, improving the complexity of the model at each stage. A first approximation, gas phase calculations, has been performed to check the electronic state of the metal, then solvent effects, electronic embedding and dispersion effects have been accounted for.

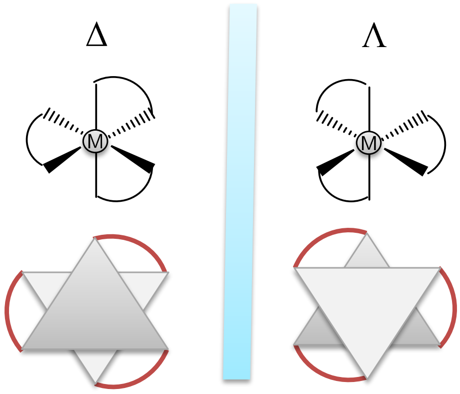


**Figure 7.4:** Structural differences on Δ and Λ enantiomers.

To perform the modelling, the two-layered ONIOM hybrid method[23] has been chosen. This approach takes into account the shape of the entire system without enlarging the computational cost. The high level layer was described at DFT level (B3LYP[24]) and includes the Fe atom and the three coordinated 2,2'-bipyridine ligands. The used basis sets are aug-cc-pVTZ[25,26] for the metal centre and the 6-311G*[27,28] one for the main group elements. The lower level layer has been modelled with the Amber forcefield and includes the peptide chains (Figure 7.5).

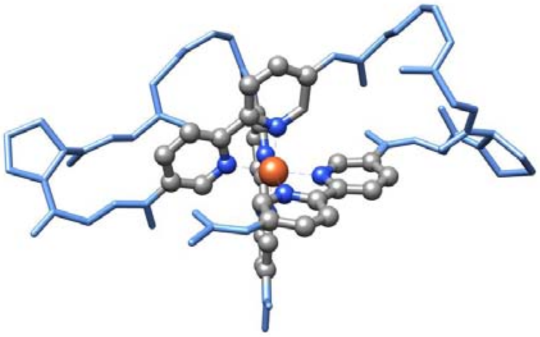


**Figure 7.5:** QM/MM partition: ball and stick atoms represent the QM part, and blue wires represent the MM one. Hydrogens have been removed to a better understanding of the image.

### Gas phase calculations: a first approximation

The most stable electronic state of the metal centre has first to be investigated. The oxidation state of the metal was known, and corresponds to +2, but the most stable spin state for the Fe(II) ion has to be determined comparing the energy at low, intermediate and high spin configurations. In this previous step, only the first coordination sphere of the system has been taken into account (the QM partition shown in Figure 7.5), that corresponds to the iron centre and the three 2,2'-bipyridine ligands. Results show a preference for the low spin configuration (2.5 kcal/mol and 13 kcal/mol more stable than the high spin and intermediate spin configurations, respectivelly), which is in agreement with similar systems found in the literature.[29,30]

After being determined the electronic structure of the Fe(II) atom, calculations in gas phase on the entire system have been performed. Results show the presence of only one minimum for the DD peptide (Δ3-DD). The second most stable structure is 9.0 kcal/mol over the minimum. In the case of the DL peptides, there is also only one minimum (Λ3-DL), being closely followed by other 5 compounds at less than 5kcal/mol: Δ2-DL, Δ3-DL (its metal-ligand enantiomer), Λ3-DL, Λ2-DL and Λ1-DL (Figure 7.6). This small energy difference suggest an equilibrium in the early stage of formation of the four systems, that corresponds to the experimental observations at short time.

For LL, the minimum corresponds to the Λ3-LL system. In the case of LD, the minimum is the Δ3-LD one.

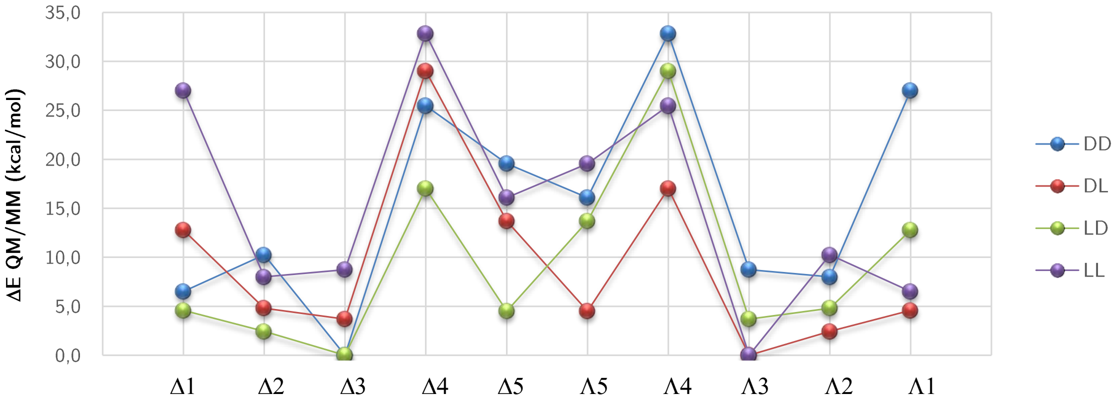


**Figure 7.6:** Energy diagram for all the isomers studied in this study. Calculations performed in gas phase.

### Approximation to the reality

Having a realistic model is key to reproduce the experimental results obtained in the laboratory. One clear example is the use of continuum solvent models to accurately reproduce the environment where the experimental study is taking part. Regarding peptides, the group of Bernèche and coworkers (University of Basel) recently demonstrated the importance of the solvent in the folding process of a peptide.[31] They found that a driving force that increases the folding propensity of peptides containing aromatic residues arises from the lack of hydration of the carbonyl and amide groups on both sides of the bulky hydrophobic side chain. They postulate that, for aromatic residues, bulky side chains limit the access of water molecules near carbonyl and amide groups. Because of that, it would be energetically preferable for these backbone functional groups to interact with each other, forming intramolecular hydrogen bonds and favouring the peptide folding.

In our specific case, experimental work has been carried out using water as a solvent. To be in agreement with the experimental conditions, calculations were performed using a implicit solvent model of water: the Conductor-like Polarizable Continuum Model (CPCM).[32,33] Calculation of the solvent effects shows a high stabilization of the Δ4 and its specular image (Λ4) for all four sets of peptides by more than 10 kcal/mol (Figure 7.7).

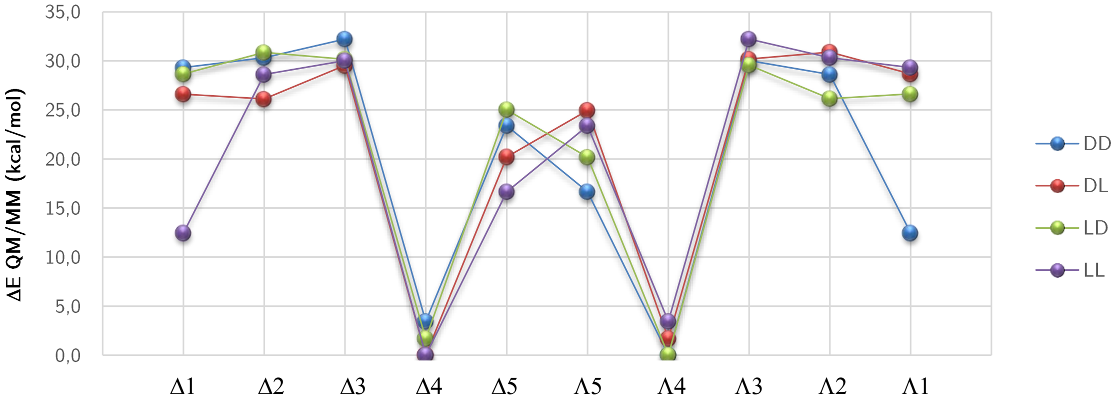


**Figure 7.7:** Diagram of the effect of the solvent in each set of metallopeptides.

The improvement of the model went a little bit further. Other step forward to the real system is enhancing the DFT part by adding a dispersion correction to the result, which has an stabilizing effect overall the system. In our case, B3LYP was the chosen functional to be included in the QM part of the ONIOM framework. This functional does not include any dispersion term, being almost mandatory its inclusion to correct the effect of the London forces. To calculate the dispersion term, the gd3 web application created by the group of S. Grimme was employed [b].[34] Other way to include this term is choosing a dispersion connected DFT functional, but we did not follow this approach to keep the methodology employed in the previous studies (Section 7.2.1).

There is an other correction concerning the ONIOM method. By default, ONIOM treats separately both QM and MM regions. Using electronic embedding, the partial charges of the MM region are incorporated into the quantum mechanical Hamiltonian, which provides a better description of the electrostatic interaction between both regions and allows the QM region to be polarized.

The total energy of the different systems, shown in Figure 7.8, has been corrected following the statements presented before. The diagram presents a shape which differs from the Figure 7.6 in the disposition of the local minima. For the DD peptide, the most stable isomer is the Λ4-DD, followed closer by Δ4-DD (2.7 kcal/mol higher in energy). In experimental conditions, this small difference in energy between both will not explain the stereoselective formation of one or other metallopeptide short time after the reaction between the metal and the peptide starts. In the case of DL peptide, Λ4-DL is the preferred geometry, being the next (Λ3-DL) quite far in energy (5.5 kcal/mol) to allow a thermodynamic equilibrium. The energy difference between the minimum and its metal enantiomer, Δ4-DL is higher enough to allow the stereoselective formation of the Λ4-DL compound

For LL and LD peptides, results correspond to the specular image of that described before. For LL, the minimum corresponds to the Δ4-LL system. In the case of LD, the minimum is the Δ4-LD

---

[b] http://toc.uni-muenster.de/DFTD3/

one. Here we first noticed the dependence of the configuration of the stereogenic centre of the proline placed in the N-loop in the formation of the Λ or Δ complex.

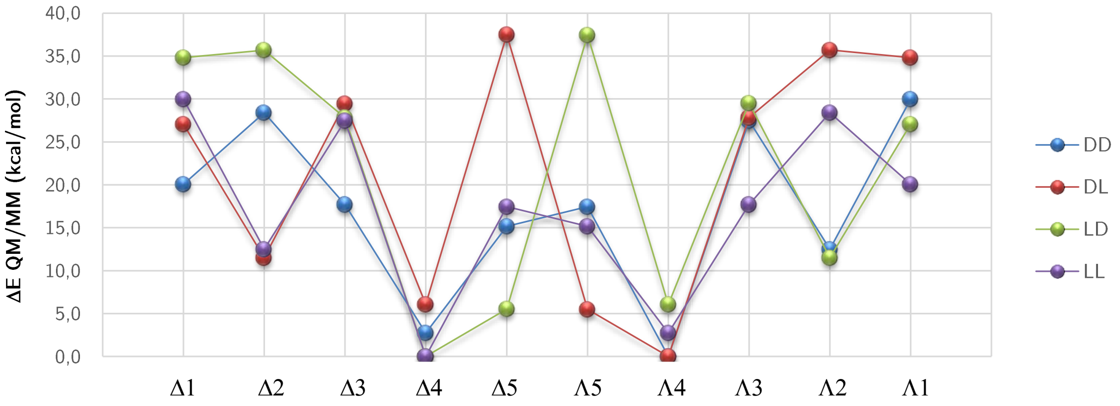


**Figure 7.8:** Energy diagram for all the isomers studied in this study. Calculations performed taking into account implicit solvation, electronic embedding and dispersion effects.

Including the entropic effects by calculating the Gibbs energy does not have any substantial consequences on the overall shape of the energy diagram, only smoothing the differences in energy between all the compounds (Figure 7.9).

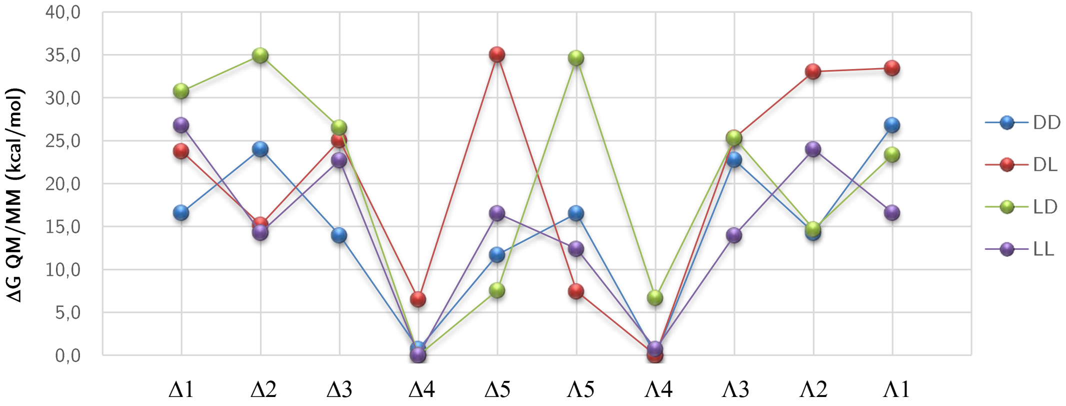


**Figure 7.9:** Gibbs energy diagram for all the isomers studied in this study. Calculations performed taking into account implicit solvation, electronic embedding and dispersion effects.

In Figure 7.10, the effect of the solvent (green), dispersion (red) and electronic embedding (purple) over the total energy of the system, as well as all the factors together (black) is shown (Note: the embedding energy includes the energy of the system in gas phase (cyan)). The effect of the solvent is not negligible at all, leading to the preference for the Λ4 system over the local minimum found in gas phase (Δ3). Dispersion effects also stabilize the Λ4 complexes. Lastly, including electronic embedding has a lower but noticeably consequence on the final disposition of the different isomers.

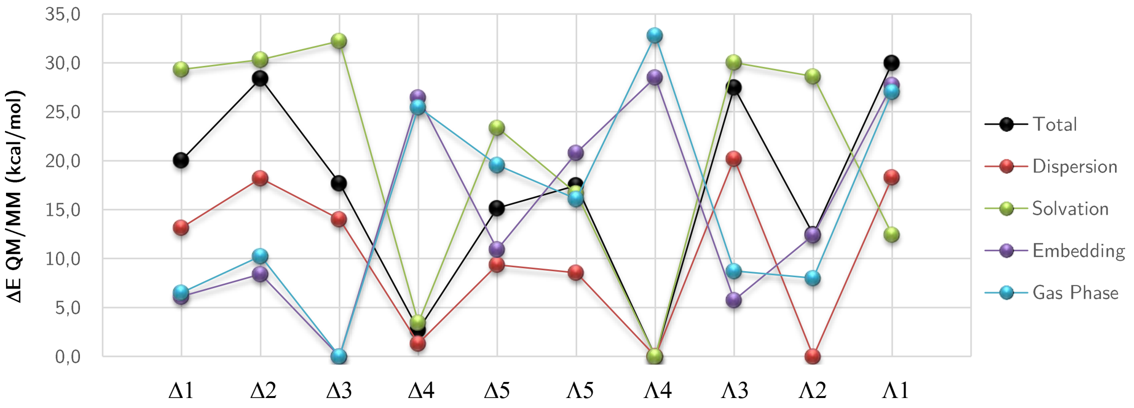


**Figure 7.10:** Descomposition of the energy terms for the DD peptide.

As a conclusion, each improvement of the methodology represents a step forward on the reliability of the system. Including the solvent has a big impact on the relative stability of the complexes, changing the most stable structure. Other improvements such as dispersion and entropic effects, and electronic embedding, do not cause substantial changes on the relative energy of the different stereoisomers but have a stabilizing effect on the total energy of the systems.

In agreement with the experimental results, the group of the *Universidade de Santiago de Compostela* found the same minima than the one computationally determined. Additionally, they also found a relation between the configuration of the stereogenic centre of the proline residue of the N-loop and the chirality of the metal centre. The structural variables which can describe this connection will be analysed.

### 7.3.3 Structural analysis of the Fe-metallopeptides

In this section some structural variables are investigated to better analyse the ground of the stereoselective formation of the metallopeptides under study. First, the effect of prolines will be tested, by means of swapping the proline residue by a glycine, to skip the effect of the chirality of the residue. Then, the QM/MM energy breakdown is studied, in order to find who imposes the configuration of the metal centre. Finally, a closer look on the configuration of the proline residues is given.

#### The role of proline residues

From the energetic study of the system we found that the possibility of the stereoselective formation of the Δ or the Λ compound can be ruled by the configuration of the stereogenic centre of the proline placed in the N-loop. Replacing proline residues with other non-chiral amino acids allows isolating the effect of each loop separately, which can help us to understand the consequences of the configuration on the stereocentre of the proline in the binding process. The nomenclature for the glycine containing peptides follows the same rules than the inial one, changing the D or L letter by a G in the loop where the glycine is placed (for example, if the glycine is in the C-loop and the proline in the N-loop is in D configuration, the name of the peptide is NG).

The energy diagram of the glycine mono-substituted peptides can be seen in Figure 7.11. For DG and LG peptides, results are totally in agreement with the DD and LL ones, being the minima the Λ4 compound for DD and DG peptides, and the Δ4 for the LL and LG ones. Additionally, the energy difference between the minimum and its enantiomer is the same for proline-proline peptides and proline-glycine ones (2.7 kcal/mol). If the substitution takes place in the N-loop, both Δ4 and Λ4 metallopeptides are in equilibrium. The energy difference between them is only 0.2 kcal/mol.

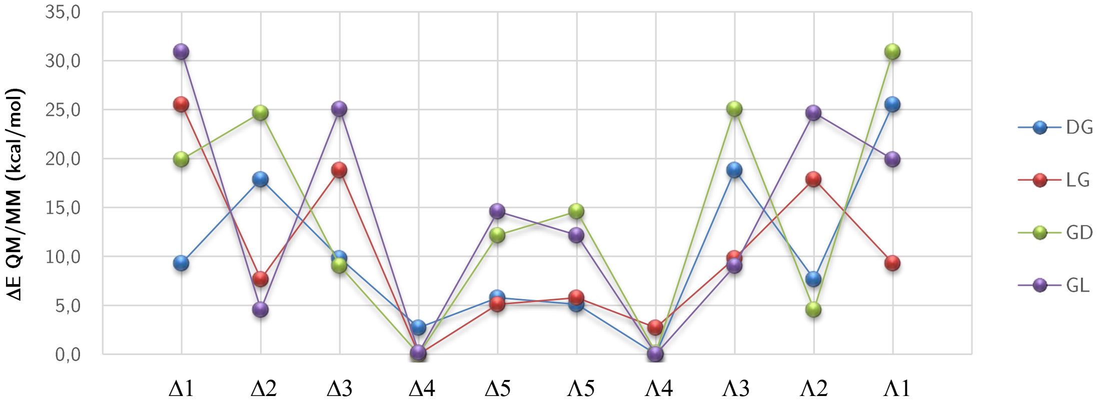


**Figure 7.11:** Energy diagram of the glycine mono-substituted metallopeptides.

Additionally, a di-substitution of both prolines by glycines is also evaluated, showing a preference for the Λ4 compound, which agrees with the results for the DD, DL and DG metallopeptides (Figure 7.12).

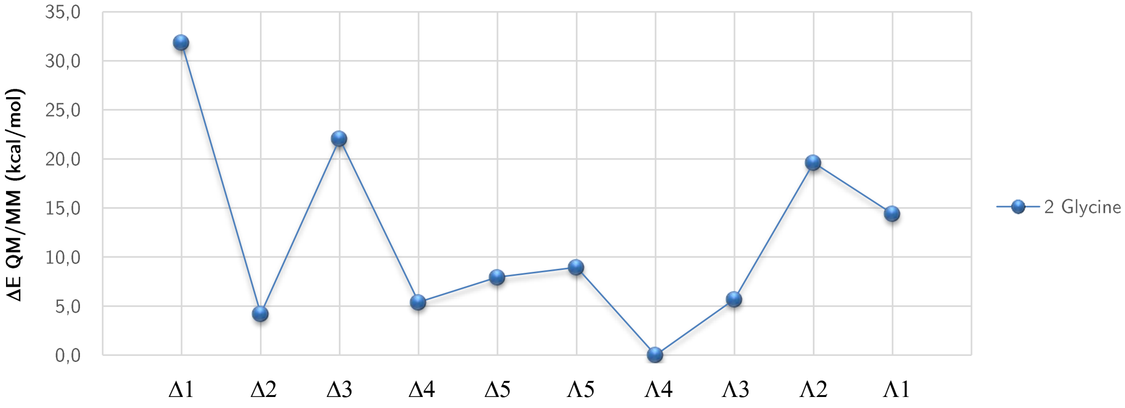


**Figure 7.12:** Energy diagram of the glycine di-substituted metallopeptides.

The step-by-step removing of the proline residues on the peptides clarify the dominance of the N-loop on the enantioselective formation of the studied metallopeptides. Results in DG and LG compounds are in agreement with the ones found in the DX and LX (where X is D or L) systems, still presenting stereoselectivity. However, if the substitution is placed in the N-loop position, two minima are found which means that the stereoselective formation of metallopeptides is not given by these complexes.

### What imposes the isomerism of the metal? first coordination sphere vs peptide

Since now, the relative participation of the first coordination sphere of the metal against the peptide part (QM and MM partitions applied on the ONIOM calculation, respectively) has not been evaluated. In Figure 7.13, the total energy of the system (including solvent effects and electronic embedding) has been decomposed in its QM, MM and QM/MM terms. This can allow us to separate the effect of the first coordination sphere and the peptide in two different energy variables, and weight it with the total energy of the system.

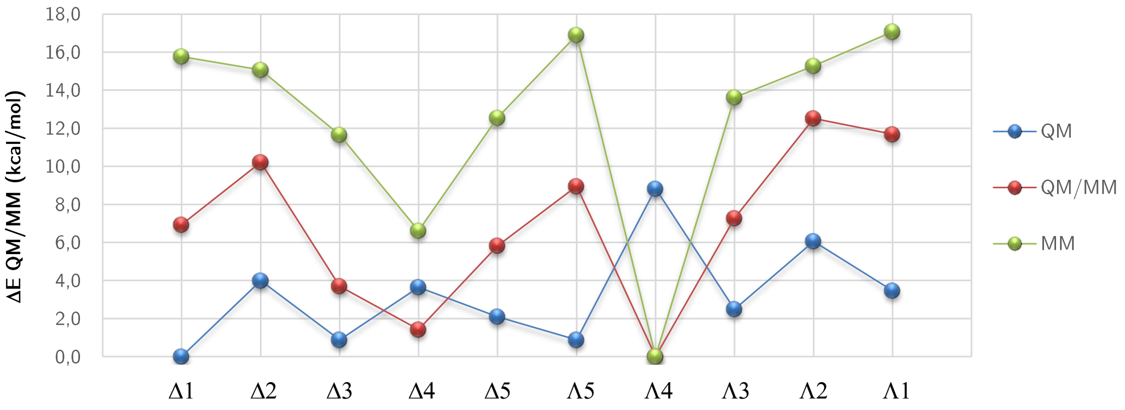


**Figure 7.13:** Decomposition of the total QM/MM energy (red) in QM (blue) and MM (green) contributions for the DD set of metallopeptides.

Both MM and QM/MM diagrams present the minimum at the same point, the Λ4-DD system. The stereoselective formation of that minimum or its metal enantiomer (Δ4-DD) is favoured by the MM term, with a difference in energy of 6.5 kcal/mol.

Moreover, the energy diagram of the MM region presents an almost equal shape than the one for the total QM/MM energy, differing only in the Δ1-DD and the Λ1-DD systems. That fact demonstrates the dominance of the MM region in the total energy of the system, which is smoothed by the energetic effect of the QM part.

## Effect on the cis/trans configuration of prolines

To find a clue on the high stability of the MM part on the Λ4 (or Δ4) a closer look on the geometry of the loops of the different systems has been done. In particular, the *cis/trans* configuration of prolines (Figure 7.14) has been studied.

The lower energy structures (Λ4 for DD and DL complexes and Δ4 for the LL and LD ones) present a *cis* configuration in both proline amino acids. In contrast, the other complexes have a *trans*

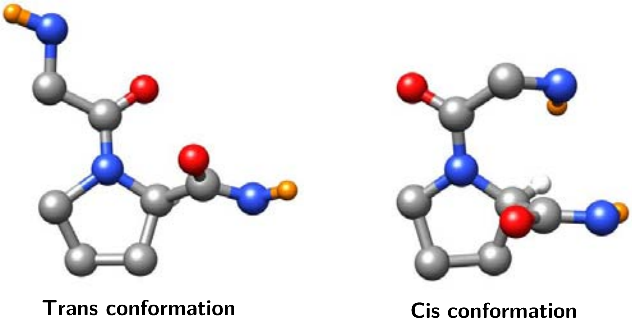


**Figure 7.14:** Trans (left) and cis (right) configurations of D-Proline residues

configuration. These observations surprised both theoretical and experimental groups, because of 1) the initial trans configuration of proline in the unbound peptide and 2) the preference of proline to adopt a trans configuration.[35] A *trans-cis* isomerization of the proline groups has been proposed as a last step on the folding process, something in agreement with folding processes in proteins,[36] which is supported by the *cis* configuration on the minimum energy structures. However, there is a in creasing number of evidences that points at the *cis*/*trans* isomerization of prolines as a major driving force to folding processes. In fact, CD measurements performed on this particular system and then interpreted by computational meals shows that this is indeed a major contribution.[c]

## 7.4 Using helical metallopeptides for DNA recognition

Helicates are defined as polymetallic helical double-strand complexes.[37] One of their features is the possibility to bind to DNA, being potential anticancer drugs.[38] However, the enantioselective synthesis of helicates is not efficient enough to be an alternative to the traditional DNA-binding agents.[39] An alternative approach to synthesize DNA-binding helicates has been proposed in this study by the experimental group of E. Vázquez and M. Vázquez from the *Universidade de Santiago de Compostela*.[40] The structural viability of the design proposed by the experimental group has been tested by means of molecular modelling methods, which will be discussed in this section.

The structure corresponding to the helicate synthesized in this work is quite more complex than the previous one. It contains two Gly-Pro-Gly loops, two metal centres and six 2,2'-bipyridine ligands (Figure 7.15). Only the D-Pro/D-Pro and the L-Pro/L-Pro compounds have been intensively studied here. From here, the first ones will be called DD-Helicate and the L-Pro/L-Pro ones, LL-Helicate.

Molecular modelling have been performed on the LL-Helicate and DD-Helicate complexes. Both systems have been prepared performing molecular dynamics simulations on the loops and clustering the solutions. A latter QM/MM refinement on the most representative clusters takes into account the effect of the solvent (in this case, water) and including electronic embedding effects. The lower

---

[c] This paper can be read in Appendix A.6.

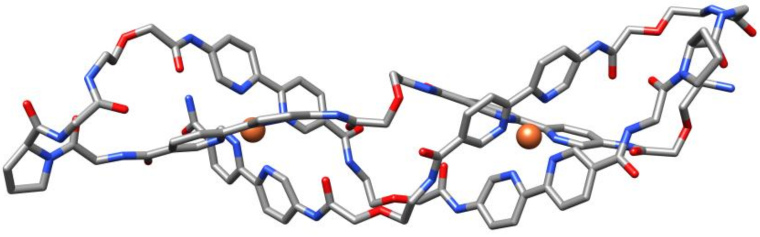


**Figure 7.15:** Structural representation of the ΛΛ-Fe(LL-Helicate).

energy systems present a helical geometry (seen previously in Figure 7.15) with the metal centre in octahedral geometry binding each one of the three bipyridine residues. These structures provided support to the experimental group to go ahead with their synthesis.

DD-Helicate and its enantiomer LL-Helicate give rise to mirror image Circular Dichroism spectra with two bands of opposite signs. The signs are consistent with the Δ-configuration for the DD-Helicate, and also with the Λ-configuration for the LL-Helicate. This means that the chiral proline residues does not only direct the folding peptide chain into discrete dinuclear helicate, but also encode its chirality.

The resulting peptide helicates have been later tested in their cell-internalization. The experimental group incubated Vero cells with the ΛΛ-Fe(LL-Helicate) for 12h. After washing, they observed that the compound was efficiently internalized into the cells, concentrating in endocytic vesicels in the cytoplasm.

## 7.5 Conclusions

A computational description of metallopeptides needs special care due to the high complexity of their structure. In the case presented on this chapter, despite of the size of the system does not represents a challenge for the actual QM software, its huge flexibility complicates the modelling process, in particular if our purpose is performing an extensive search of all the possible isomers resulting from the different manners to fold. Additionally, molecular mechanics techniques are not an option, because an accurate description of the metal was needed. In this case, the hybrid QM/MM was the employed method, including the metal and its coordinated 2,2'-bipyridine ligands in the QM part, and leaving the rest of the system represented at MM level of theory. In order to have accurate results, some corrections where taken into account. First, solvent effects were included to reproduce the laboratory environment. Then, the DFT functional (B3LYP in this case) was corrected including the dispersion effects using the gd3 web application created by the group of S. Grimme. Lastly, the charges of the MM part were included in the QM region setting up the electronic embedding feature of the ONIOM method. All these improvements increase the reliability of the system, approaching the model to the reality.

The final results display a preference for the Λ4 isomer for DD and DL systems, and Δ4 for LD and LL. Then, we noticed the preference on the formation of one or other metal isomer depending on the configuration of the proline residue present in the N-loop by means of isolating the effect of each proline residue.

Decomposing the total QM/MM energy on its QM and MM parts, shows that the metal is not governing the stereoselective formation of the metallopeptide. Results show a clear dominance of the MM region, which includes both loops that link the coordinating 2,2'-bipyridine residues, in front of the influence of the first coordination sphere of the metal, modelled in the QM part. A closer look to the loops shows that the *cis* configuration for proline residues is shown only in the mininum energy structures.

Altogether, here we presented an example of symbiosis between experimental an theoretical groups. Unifying efforts from both sides by using the best methods from theoretical and experimental areas we successfully describe the enantioselective formation of artificial metallopeptides. However, not all the work is done, being lots of applications on these system still to discover and apply.

## 7.6 References

[1] Wu, C.-F.; Bishopric, N. H.; Pratt, R. E. *J. Biol. Chem.* **1997**, *272*, 14860–14866.

[2] Harder, J.; Bartels, J.; Christophers, E.; Schroder, J. *Nature* **1997**, *387*, 861–861.

[3] Papo, N.; Shai, Y. *Biochemistry-US* **2003**, *42*, 458–466.

[4] Liu, Q.; Wang, J.; Boyd, B. J. *Talanta* **2015**, *136*, 114–127.

[5] Sanghvi, A. B.; Miller, K. P.-H.; Belcher, A. M.; Schmidt, C. E. *Nat. Mater.* **2005**, *4*, 496–502.

[6] Holmes, T. C. *Trends Biotechnol.* **2002**, *20*, 16–21.

[7] Huang, F.; Nau, W. M. *Angew. Chem. Int. Edit.* **2003**, *115*, 2371–2374.

[8] Norouzy, A.; Assaf, K. I.; Zhang, S.; Jacob, M. H.; Nau, W. M. *J. Phys. Chem. B* **2015**, *119*, 33–43.

[9] Jelinek, R. *Cellular and Biomolecular Recognition*; John Wiley & Sons, 2009.

[10] Hutchens, T. W.; Nelson, R. W.; Allen, M. H.; Li, C. M.; Yip, T.-T. *Biol. Mass Spectrom.* **1992**, *21*, 151–159.

[11] Hart, B. R.; Shea, K. J. *J. Am. Chem. Soc.* **2001**, *123*, 2072–2073.

[12] Barnham, K. J.; Bush, A. I. *Curr. Opin. Chem. Biol.* **2008**, *12*, 222–228.

[13] Nicholls, I. A.; Ramström, O.; Mosbach, K. *J. Chromatogr. Sci.* **1995**, *691*, 349–353.

[14] Albrecht, M.; Stortz, P. *Chem. Soc. Rev.* **2005**, *34*, 496–506.

[15] Rama, G.; Ardá, A.; Maréchal, J.-D.; Gamba, I.; Ishida, H.; Jiménez-Barbero, J.; Vázquez, M. E.; Vázquez López, M. *Chem-Eur. J.* **2012**, *18*, 7030–7035.

[16] Monari, A.; Rivail, J.-L.; Assfeld, X. *Accounts chem. res.* **2012**, *46*, 596–603.

[17] Kang, Y. K.; Park, H. S. *Chem. Phys. Lett.* **2014**, *600*, 112–117.

[18] Alí-Torres, J.; Mirats, A.; Maréchal, J.-D.; Rodríguez-Santiago, L.; Sodupe, M. *AIP Adv.* **2015**, *5*, 092402.

[19] Du, Q.-S.; Huang, R.-B.; Chou, K.-C. *Curr. Protein Pept. Sc.* **2008**, *9*, 248–259.

[20] Keller, A.; Nesvizhskii, A. I.; Kolker, E.; Aebersold, R. *Anal. Chem.* **2002**, *74*, 5383–5392.

[21] Alí-Torres, J.; Maréchal, J.-D.; Rodríguez-Santiago, L.; Sodupe, M. *J. Am. Chem. Soc.* **2011**, *133*, 15008–15014.

[22] Kolossváry, I.; Guida, W. C. *J. Am. Chem. Soc.* **1996**, *118*, 5011–5019.

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

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

[25] Kendall, R. A.; Dunning Jr, T. H.; Harrison, R. J. *J. Chem. Phys.* **1992**, *96*, 6796–6806.

[26] Woon, D. E.; Dunning Jr, T. H. *J. Chem. Phys.* **1993**, *98*, 1358–1371.

[27] McLean, A.; Chandler, G. *J. Chem. Phys.* **1980**, *72*, 5639–5648.

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

[29] Lawson Daku, L. M.; Vargas, A.; Hauser, A.; Fouqueau, A.; Casida, M. E. *Chemphyschem* **2005**, *6*, 1393–1410.

[30] Swart, M. *J. Chem. Theory Comput.* **2008**, *4*, 2057–2066.

[31] Bignucolo, O.; Leung, H. T. A.; Grzesiek, S.; Bernèche, S. *Journal of the American Chemical Society* **2015**, *137*, 4300–4303.

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

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

[34] Grimme, S.; Antony, J.; Ehrlich, S.; Krieg, H. *J. Chem. Phys.* **2010**, *132*, 154104.

[35] Zimmerman, S. S.; Scheraga, H. A. *Macromolecules* **1976**, *9*, 408–416.

[36] Wedemeyer, W. J.; Welker, E.; Scheraga, H. A. *Biochemistry-US* **2002**, *41*, 14637–14644.

[37] Lehn, J.-M.; Rigault, A.; Siegel, J.; Harrowfield, J.; Chevrier, B.; Moras, D. *P. Natl. Acad. Sci. USA* **1987**, *84*, 2565–2569.

[38] Hotze, A. C.; Kariuki, B. M.; Hannon, M. J. *Angew. Chem. Int. Edit.* **2006**, *118*, 4957–4960.

[39] Terazzi, E.; Guénée, L.; Varin, J.; Bocquet, B.; Lemonnier, J.-F.; Emery, D.; Mareda, J.; Piguet, C. *Chem.-Eur. J.* **2011**, *17*, 184–195.

[40] Gamba, I.; Rama, G.; Ortega-Carrasco, E.; Maréchal, J.-D.; Martínez-Costas, J.; Vázquez, M. E.; López, M. V. *Chem. Commun.* **2014**, *50*, 11097–11100.

# 8 ⋯ Conclusions

*The only way to do great work is to love what you do.*

**Steve Jobs**

This dissertation was focused on the description of the most important (and visible) effects of the binding of transition metals to biological hosts. All the study was performed using a vast variety of molecular modelling tools, which have been chosen following the next criteria: the chemical interaction under study, the size of the system and affordability of the selected method.

The principal issues studied in this Ph. D. Thesis, can lead to the following conclusions:

1. **Inert biometallic binding.**

   In many cases, the interaction between an organometallic compound and a biological partner is not prompted by the coordination of the metal to the host. There the metal have a little contribution in the binding and only has a structural role, holding its coordinated ligands in a specific geometry. In other words, the binding occurs between the second coordination sphere of the organometallic compound and the host, not influencing the electronic and structural properties of the metal.

   When the original structure is known, information about second coordination sphere events can be obtained using QM/MM approaches, but does not allow wide exploration if the starting point is unknown. However, an accurate description of the metal centre is not mandatory. In this dissertation we tested the capability of protein-ligand docking methods, which are far less time consuming, to reproduce the binding of an organometallic ligand to a protein, obtaining encouraging results. Neither changes in the protein-ligand docking algorithm have been performed nor addition of parameters have been done to perform our simulation, being only needed the specification of the metal ligand bonds.

2. **Active biometallic binding.**

   One of the major questions in bioinorganic chemistry is how the coordination of transition metals constrains the biological hosts affecting its structure, specially the conformation of the directly bonded region. But, what happens when the binding of these organometallic compounds to biological hosts actually involves changes in the $1^{st}$ coordination sphere of the metal? Fine electronic representation of the model is needed at some points. In this dissertation, two quantum mechanical based techniques are used to describe the local effects of metal coordination. The first one is an energy exploration of the conformational space of the artificial *cd***HO** and its activation mechanism using QM/MM geometry optimization algorithms. This yields to a final potential energy surface that allows us to find that the $1^{st}$ coordination sphere participates in the conformational change that the metalloenzyme has to suffer to being activated.

   Other way to proceed is by the depiction of the bonding terms, studied using the energy decomposition analysis (EDA). This method was used to find the key electronic variables that explain the preference of the metal to bind one or other residue when cisplatin bins to protein.

3. **Long range effects of the binding of an organometallic compound to a biological host.**

Taking a look only to the first coordination sphere events does not provide us information of the structural changes of the host out of the binding site. Constrained systems, such as the DNA double strand system, are susceptible to being affected by the changes occurred in a small region. QM/MM approaches can give us an idea of the second coordination sphere events, but are not able to go further because long range effects need an unreachable time scale with that technique. Molecular modelling tools are the best choice to model the entire system. In this Ph.D. Thesis, Normal Mode Analysis was the chosen methodology to describe the global consequences of the binding of metals to the biological host, in particular cisplatin to DNA. Results also suggests that the specificity of cisplatin binding comes from a balance between $1^{st}$ and $2^{st}$ coordination sphere events and mechanical consequences. This is in agreement with the butterfly-effect: a small change in a specific region can affect the overall system.

4. **The balance between the first and second coordination sphere processes.**

   When the conformational space of the host is wide, not only the metal affects its structure, also the opposite effect can be shown. The last chapter of this dissertation show the effect of the conformation of the peptide chain in the folding process of a metallopeptide. In flexible systems, where a single change can affect the conformation of the entire molecule, taking into account the best number of variables is key. In this case, we considered the different conformations of the stereogenic centre of a certain residue present in a peptide chain. The model was studied with QM/MM approaches, describing accurately the first coordination sphere and leaving the rest of the system in the molecular mechanics part. Splitting the energy terms of the QM/MM method we find that the molecular mechanics region was the one that dominates the enantioselectivity of the folding process.

# A ··· Publications related to the Ph.D. Thesis

# 1 What can molecular modelling bring to the design of artificial inorganic cofactors?

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

View Article Online / Journal Homepage / Table of Contents for this issue
PAPER www.rsc.org/faraday_d | Faraday Discussions

# What can molecular modelling bring to the design of artificial inorganic cofactors?†

**Victor Muñoz Robles, Elisabeth Ortega-Carrasco,**
**Eric González Fuentes, Agustí Lledós and Jean-Didier Maréchal***

***Received 23rd March 2010, Accepted 28th April 2010***
**DOI: 10.1039/c004578k**

In recent years, the development of synthetic metalloenzymes based on the insertion of inorganic catalysts into biological macromolecules has become a vivid field of investigation. The success of the design of these composites is highly dependent on an atomic understanding of the recognition process between inorganic and biological entities. Despite facing several challenging complexities, molecular modelling techniques could be particularly useful in providing such knowledge. This study aims to discuss how the prediction of the structural and energetic properties of the host–cofactor interactions can be performed by computational means. To do so, we designed a protocol that combines several methodologies like protein–ligand dockings and QM/MM techniques. The overall approach considers fundamental bioinorganic questions like the participation of the amino acids of the receptor to the first coordination sphere of the metal, the impact of the receptor/cofactor flexibility on the structure of the complex, the cost of inserting the inorganic catalyst in place of the natural ligand/substrate into the host and how experimental knowledge can improve or invalidate a theoretical model. As a real case system, we studied an artificial metalloenzyme obtained by the insertion of a Fe(Schiff base) moiety into the heme oxygenase of *Corynebacterium diphtheriae*. The experimental structure of this species shows a distorted cofactor leading to an unusual octahedral configuration of the iron with two proximal residues chelating the metal and no external ligand. This geometry is far from the conformation adopted by similar cofactors in other hosts and shows that a fine tuning exists between the coordination environment of the metal, the deformability of its organic ligand and the conformational adaptability of the receptor. In a field where very little structural information is yet available, this work should help in building an initial molecular modelling framework for the discovery, design and optimization of inorganic cofactors. Moreover, the approach used in this study also lays the groundwork for the development of computational methods adequate for studying several metal mediated biological processes like the generation of realistic three dimensional models of metalloproteins bound to their natural cofactor or the folding of metal containing peptides.

---

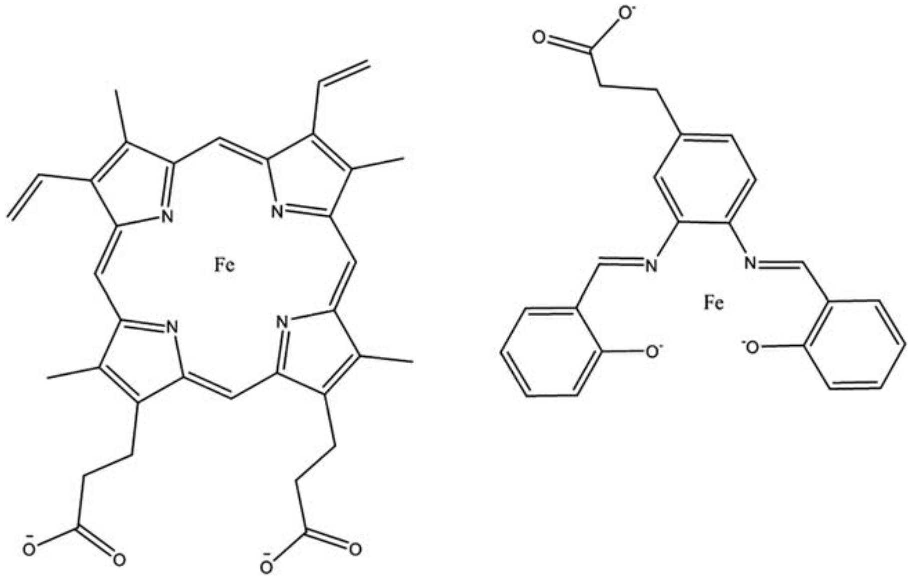
*Departament de Química, Universitat Autònoma de Barcelona, Edifici C.n., 08193 Bellaterra, Barcelona, Spain. E-mail: jeandidier.marechal@uab.es; Fax: +34 93 581 2920; Tel: +34 93 581 4936*

† Electronic supplementary information (ESI) available: Further structural analysis on experimental and computed structures are provided as figures and data tables. See DOI: 10.1039/c004578k

## Introduction

The development of manmade metalloproteins and metalloenzymes has nowadays risen to an unprecedented level of development.[1] In this particular area, the design of artificial species resulting from the insertion of an inorganic catalyst into a protein has become a vivid field of research.[2] The design of these hybrids generally follows the same principles that govern naturally occurring enzymes like hemoenzymes: the inorganic moiety—the artificial cofactor—provides the chemical reactivity while the protein—the host—provides most of the substrate selectivity and catalytic regiospecificity. Different systems have already been used as receptors of artificial cofactors like apo-myoglobin,[3,4] heme-oxygenase,[5] avidin[6] and streptavidin,[7] lyzosyme,[8] human serum albumin,[9,10] xylanase[11] and papain.[12] These receptors have been inserted into inorganic complexes, such as metal(Schiff bases),[3,5] Rh diphosphines[6] and Ru diamines,[7] Mn(corrole)[9] and Co(phthalocyanine)[10] and Fe(porphyrin)[11], and the range of chemical reactions performed by the resulting composites is vast, including hydrogenation, sulfoxidation or even Diels–Alder reactions. To the date, the insertion of the inorganic moiety into its protein receptor is performed by three major approaches including the covalent linkage of the inorganic complex to the natural ligand of the host,[6,7] the covalent linkage of the inorganic complex to one or several specific amino acids of the receptor[4,12] or the direct insertion of the inorganic complex into a vacant cavity of the host.[3,5,9–11] Because of the unnatural kind of interaction between the host and the inorganic system, several optimization steps are necessary to produce an efficient bio-inspired catalyst. However, the first step in the design of these composites concentrates on providing a reasonably good binding affinity between the cofactor and the host. This part of the process involves several key bioinorganic considerations on the cofactor–protein complementarities like: what is/are the possible orientation(s) of the cofactor into the host? What residues of the host could be part of the first coordination sphere of the metal? How does the binding of the metallic cofactor into its host modulate the shape of the final complex (metal mediated induced fit)?

The system designed by Ueno *et al.* resulting from the incorporation of a Fe(Schiff base) complex into the heme cavity of *Corynebacterium diphtheriae* heme oxygenase (*cd***HO**) (Scheme 1) clearly shows the importance of these considerations.[5] The



**Scheme 1** Side to side representation of the protoporphyrin IX—the natural substrate of *cd***HO**—and the artificial Fe(Schiff base) cofactor considered in this study.

particular artificial cofactor used in this design presents a chemical scaffold which has substantial similarities with the heme although its aromatic part represents approximately three quarters of the heme and the macrocycle is not closed. For maintaining an efficient electron pathway, the inorganic moiety is substituted by a propionate; one of the substituents of the natural cofactor. Heme oxygenases (**HO**) are all-$\alpha$ enzymes that catalyse the first step of the auto-oxidation of the heme.[13,14] Prior to this oxidative process, the heme binds at the core of the unbound **HO** and generates the resting state of the holo-enzyme. At this stage, the Fe(III) is coordinated by the 4 nitrogen atoms of the heme in equatorial and by the $N_\varepsilon$ of a proximal histidine (His20 in *cd***HO**) and an external distal ligand (water or hydroxyl) in axial (Fig. 1a). The catalytic process is initiated by the reduction of the metal from Fe(III) to Fe(II), followed by the removal of the external ligand from the coordination sphere of the metal, and leads to a square pyramidal configuration of the iron. Subsequently, the fixation of the oxygen on the iron can take

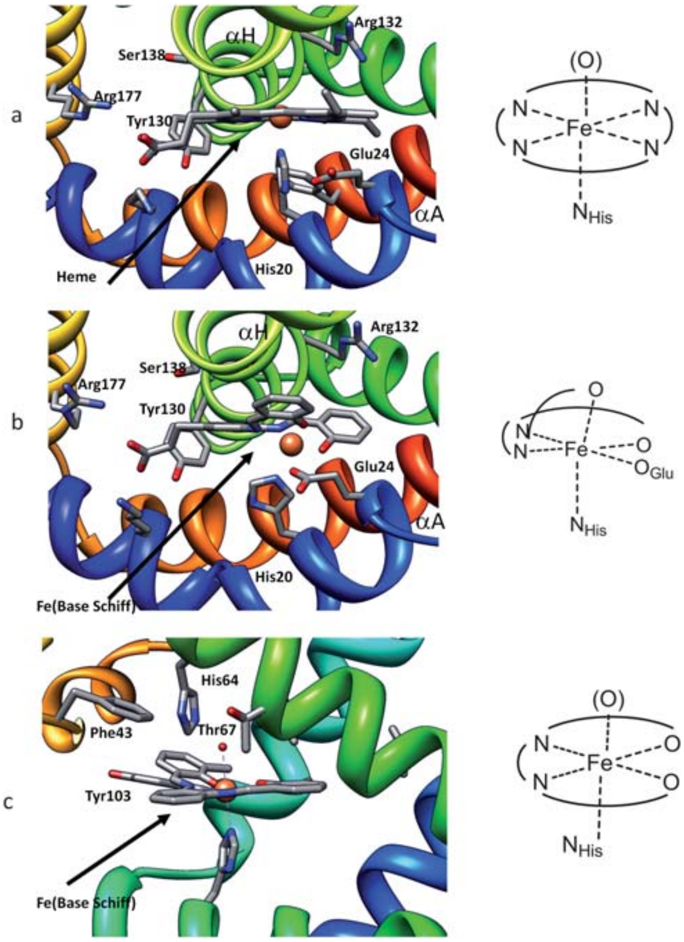


**Fig. 1** Geometries of the binding site of several crystallographic structures of proteins bound to an inorganic complex. Three dimensional views are given on the left and a schematic representation of their metal environment on the right. Panel (a) corresponds to the structure of the **heme** bound to *cd***HO** (PDB code 1iw0), panel (b) to the **Fe(Schiff base)** compound bound to *cd***HO** (PDB code 1wzd) and under investigation in this work and finally panel (c) corresponds to the structure of a similar **Fe(Schiff base)** compound bound to **apoMb** (PDB code 1ufj). In the schematic views of the metal environment, the coordination sites mentioned in between parenthesis are occupied by an external ligand ($O_2$, OH, $OH_2$, *etc.*).

place on the distal side of the heme. **HO**s are quite flexible proteins that accept the heme,[15] its first metabolite[16] or even large imidazole based inhibitors that are able to bind the iron at the distal site.[17] By incorporating an Fe(Schiff base) in place of the heme in the binding site of the *Corynebacterium diphtheriae* heme oxygenase, Ueno *et al.*[5] generated an artificial enzyme that activates $O_2$ to $O^-{}_2$ using the same natural electron partner as *cd***HO**. This system represents a major step forward in the design of artificial metalloenzymes both because of its ability to maintain naturally occurring biological interactions of the receptor and because it opens new avenues in biomedicine and electronics. Nevertheless, the X-ray structure of the **Fe(Schiff base)·HO** composites, available at the Protein Data Bank,[18] presents intriguing structural features of the binding of the cofactor[5] (Fig. 1b) and is quite different from the hexacoordinated and pentacoordinated structures observed in **Fe(Schiff base)·apo-Mb**[3] (Fig. 1c) or **heme·OH** complexes (Fig. 1a). Indeed, in this composite, the iron displays a distorted octahedral configuration with no external ligand bound on the distal side of the metal. Instead, all the ligands of the octahedron are strictly provided by the Schiff base and the receptor. This geometry is possible because of a substantial distortion of the organic part of the cofactor that allows the Schiff base to coordinate the iron with 3 atoms in equatorial positions (2 nitrogen and 1 oxygen atoms) and one in an axial position (the remaining oxygen atom occupies the distal coordination site of the iron). The remaining positions of the octahedron are occupied by the $N_\varepsilon$ of proximal histidine 20 (axial) and $O_\varepsilon$ of glutamate 24 (equatorial). This unexpected geometry clearly shows that a fine tuning between inorganic and biological rules controls the recognition process between the host and the cofactor and should be taken into account for the design of this kind of artificial metalloenzymes. The crystallization of the **Fe(Schiff base)·HO** species allowed the authors to generate a new cofactor with higher affinity for *cd***HO**;[19] an optimization that would have been difficult without sufficient molecular knowledge.

Thinking in molecular modelling techniques to provide with the relevant amount of atomic information for the design and optimization of synthetic metalloenzymes with artificial cofactors seems quite natural. Computational techniques are nowadays broadly used in many areas of chemistry and biology to understand and predict the atomic behaviour of molecules. In fact, the use of computational techniques for the design of *de novo* proteins[20] or metal binding proteins[1,21,22] has already showed interesting applications. Because of the large conformational and chemical spaces to consider in this particular field, most of the approaches used so far are based on simplified molecular models and geometrical rules. However, when dealing with synthetic cofactors containing transition metal ions, an accurate representation of the system could significantly improve the prediction of the cofactor–host interactions. To the date, no study has focused in providing such level of accuracy. This is mainly due to several challenges that molecular modelling techniques face in this particular area.

Accurate simulations of transition metal ion and more particularly those with accessible open shell configurations (*i.e.* Fe(II), Fe(III), Mn(II), Cu(II), *etc.*) are achieved by quantum based calculations being *ab initio* or functional density theory methods. Whether by pure Quantum Mechanical (QM) calculations on small models of the enzyme active site or hybrid Quantum Mechanical/Molecular Mechanics (QM/MM) simulations on the full macromolecular entity, those methods have been widely applied in bioinorganic systems.[23–26] Both approaches have lead to major breakthroughs on the catalytic mechanisms of metalloenzymes including those containing a metallic cofactor like cytochromes P450.[27–29] However, those techniques by themselves are not convenient for studying the binding process of a metallic cofactor to its host. The extensive conformational exploration required for such simulations is indeed far too expensive with conventional *ab initio* based approaches. In fact, the characterization by computational means of the complexes formed by two molecular partners relies on the computational family of protein–ligand docking techniques. Those methods are intended to identify the protein–ligand complexes with the best possible affinities (low energy complexes) and have been successfully applied in biomedical fields like drug design and drug metabolism. In these techniques, the extensive exploration of the conformational space is accessible thanks to approximated binding energy calculations (scoring) allied to random based geometrical searches. A large number of protein–ligand docking algorithms are nowadays available which differences mainly raise from the way the conformational space is explored and how the binding energies are calculated.[30] Under a major effort of part of the computational chemistry community, the accuracy of protein–ligand docking prediction has substantially improved in the recent years. For example, most protein–ligand docking programs have been upgraded with approaches for dealing with the flexibility of specific amino acids, the presence of structural water molecules and the conformational variability of ring groups.[31] Nonetheless, the presence of a transition metal ion in the binding cavity of a protein remains an additional complexity that only a few programs, like Gold, FlexX and Glide,[32,33,34] have partially overcome. These optimizations greatly improve the prediction of metalloproteins interacting with organic ligands but are still limited when dealing with the fixation of a metal-containing ligand to a protein host. Several variables are the reasons for this limitation like the complexity to generate parameters for metal containing systems with empirical approaches (*i.e.* the possibility to take into account fine electronic effects like those associated to spin properties), the difficulty of reproducing the flexibility of the first coordination sphere of the metal in particular for those systems that contain organic macrocycles or the availability of a reduced number of experimental data on inorganic complexes bound to their host (reducing the possibility of finding a convenient general benchmark for new parameterizations and implementations). One way to produce accurate simulations of the interaction of metal containing ligands with their host is to combine protein–ligand docking with quantum mechanics based techniques. In the recent years, several developments and applications of combined QM/MM and protein–ligand docking approaches have shown substantial successes.[35,36,37] To date, however, such combinations have mainly been dedicated to refine the energetic of non bonded interactions and improve of the scoring functions leading, in between others, to better prediction of the interaction of organic ligands with metalloproteins.[37] We therefore foresee that an extended molecular modelling framework including current state-of-the-art quantum mechanical techniques and protein–ligand dockings should be a convenient tool for the prediction of protein–inorganic cofactor interactions.

In recent years, our group has been dedicating increasing efforts to the particular field of artificial metalloenzymes.[9] This prompted us to develop new tools and generate new protocols. In trying to identify the maximum number of variables that should be taken into account for the prediction of the interactions of inorganic complexes with a protein by computational means, we found that **Fe(Schiff base)·HO** would represent of the best possible benchmark because, on one hand, the metal-inserted Schiff bases are some of the most common alternatives to porphyrinic systems representing an archetypical synthetic cofactor[3,5,38] and, on the other, the experimental structure of **Fe(Schiff base)·HO** crystallizes all the possible challenging questions molecular modelling techniques could encounter in the design of this kind of bio-inspired catalysts. Here, we discuss how an integrative computational approach could predict **inorganic compound·protein** interactions by a four step process including: (1) the characterization of the most stable conformations of the cofactor previous to its binding to the host and taking into account the electronic properties of the metal (oxidation and spin state); (2) the prediction of the most likely orientation(s) of the cofactor in the binding site of the host, (3) the determination of the amino acids that can participate to the first coordination sphere of the metal; and (4) the refinement of the resulting hypothetical models by accurate calculations.

## Materials and methods

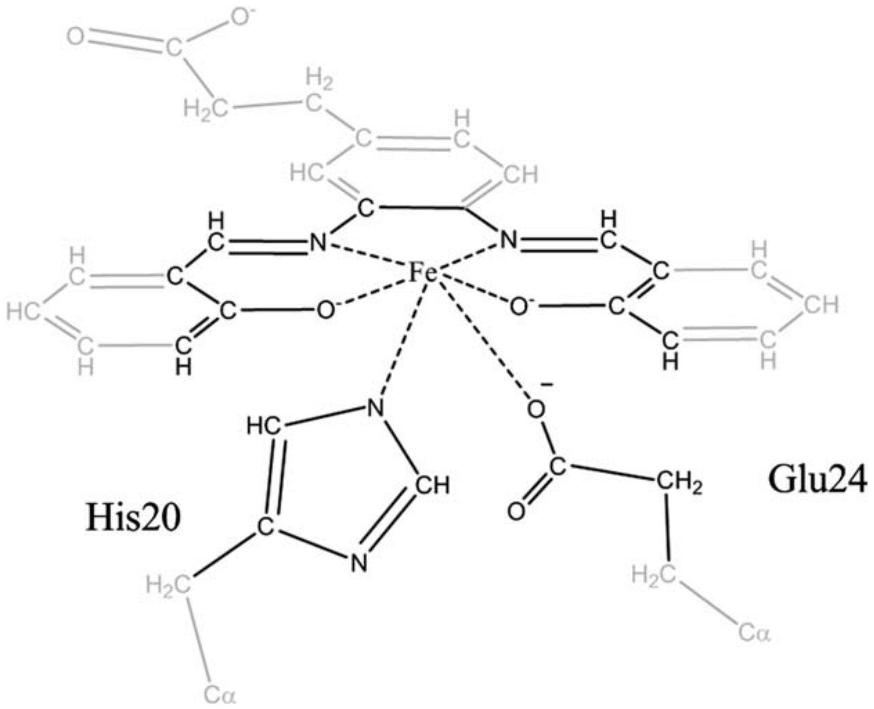
To predict the structural and energetic properties of an inorganic catalyst bound to its host, our protocol integrates current state-of-the-art approaches from several molecular modelling and structural bioinformatics fields including pure quantum mechanical calculations, metal binding prediction, protein–ligand docking and quantum mechanical/molecular mechanics calculation. All along the process, molecular visualization, set up, analysis and in house implementations have been generated in the UCSF chimera environment.[39]

The structures of the unbound cofactors have been prepared by pure quantum mechanical calculations. Geometry optimizations have been performed with the density functional theory B3LYP[40] as implemented in the program Gaussian 03.[41] The basis set used for C, N, H and O was the split valence 6-31g**[42] and for the metal ion the double-$\zeta$ basis set LANL2DZ.[43] The pseudo potential associated to this basis has also been applied to the metal.

Protein–ligand dockings have been performed with the program Gold[44] (version 4.1) and the Chemscore scoring function.[45] Because very few crystal structures of metallic artificial cofactors bound to proteins are available nowadays, the generation of an accurate scoring function for such species is not a trivial task. However, Gold is one of the few programs that provide optimized parameters for the interaction of organic ligands with a metal containing protein.[46] Therefore, for the purpose of this study, we generated a new atom type for an iron atom contained in a ligand (cofactor) that acts as a hydrogen-bond donor in order to mimic the propensity of the metal to interact with Lewis bases. For the numerical values of the Fe(cofactor)–host interactions we adapted the existing metal–ligand terms in Gold. All dockings have been performed with the default parameters of Gold for the conformational search and all rotatable bounds of the propionate were explored during the docking process. However, the inorganic core remains fixed. In our protocol, the induced deformations of the cofactor and the receptor are treated by post-docking analysis, in particular QM/MM minimizations. Twenty solutions have been generated for each docking experiment. The binding cavity has been defined as a 20 Å sphere around the $N_\varepsilon$ atom of the Histidine 20. The structures of the receptor were prepared with the UCSF chimera environment. The assessment of the exact degree of the receptor flexibility in protein–ligand docking simulations remains an issue in current state-of-the-art algorithms even if substantial improvements have been obtained in the recent years.[31] Because numerous structures of heme oxygenase are available in the Protein Data Bank, in the present study, the global flexibility of the receptor is included by carrying out calculations on 4 conformations of *cd***HO**. The amino acid flexibility has been introduced using a soft potential for the residues of the binding site. Despite some limitations, both approximations generally lead to a better prediction of relevant protein–ligand complexes.[47,48]

For the determination of the residues that are able to coordinate the metal and the generation of post-docking refined structures, we used an in house code developed in the UCSF Chimera package and aimed to predict metal binding environments. This approach (to be published) shares some common grounds with recently published algorithms for metal binding prediction.[49] In brief, this protocol relies on the fact that the ability of a residue to coordinate a metal depends on two main spatial features (1) the closeness of its C$\alpha$ to the metal and (2) the possibility of its side chain to adopt at least one orientation compatible with the coordination to the metal and does not generate bad contacts with the atoms in its vicinity. Based on statistics carried out on more than 400 structures of iron containing proteins, a Fe–$C_\alpha$ distance of 7 Å has been used as a structural criterion to characterize those amino acids that are correctly positioned to bind the metal ion. Having performed an initial analysis to detect those residues, the resulting list of candidates is reduced based on simple chemical rules (*i.e.* hydrophobic residues are not considered as possible candidates). For each remaining candidates, all possible rotameric states are generated using the backbone dependent libraries of Dunbrack *et al.*[50] as implemented in the UCSF chimera program. Those residues that present at least one rotamer with distances lower than 3.5 Å with the metal and do not generate important bad contacts with their close environment are retained as possible metal coordinating species. For this latest group of residues, all the possible combinations of rotamers are finally generated and the resulting models are ranked in function of their interactions with their close environment. Those better ranked are selected as initial models for QM/MM refinements. This latest part of the process of predicting possible metal binding environments is very similar to those approaches used in homology modelling but intensively focuses on the first coordination sphere of the metal. Moreover, it is to notice that, having characterized the residues that could bind the metal, a possible alternative of the approach presented here is to use flexible dockings as implemented in several docking programs. However, to maintain the consistence with our hypothesis of work and avoid possible pitfalls due to bad electrostatic representations of metal–protein interactions in modern docking programs, we stayed with our approach based on geometric criteria followed by QM based calculations.

For each model obtained from the previous step, QM/MM optimizations have been performed using the ONIOM[51] scheme as implemented in Gaussian 03.[41] Because both oxidation and spin states could impact on the stability of the different systems and are also involved in the mechanism of the **Fe(Schiff base)**·*cd***HO** artificial enzyme, calculations were carried out on Fe(II) and Fe(III) species in both low and high spin configurations. Minimizations were undertaken using the B3LYP[40] density functional with a reduced basis set consisting in 6-31g for C, O and N, sto-3g for H and Lanl2DZ for Fe. Single point B3LYP calculations have been performed on optimized structures with the same extended basis set as for calculation on isolated systems and repeated with the M05[52] and B3PW91[53] functionals. Neither the change in basis nor the change of functional produced chemically relevant



**Scheme 2** QM/MM partition used in this study.

changes in the results and are therefore not commented in this discussion. Regarding the QM/MM partition used for these calculations, the quantum mechanical part includes the iron and the atoms of the Schiff base the closest to it, the imidazole ring of the histidine 20 and the propionate moiety of the Glutamate 24 (Scheme 2). All remaining atoms have been included in the molecular mechanics region of the system for which the AMBER force field has been used.[54] Standard hydrogen link atoms have been applied for the frontier between QM and MM parts. It is worth noticing that this partition has been tested on the crystallographic structure of **Fe(Base Schiff)·HO** system and leads to a very good structural agreement. The calculation of the charges has been done using the Antechamber[55] program as implemented in UCSF chimera. The entire cofactor and the region of the protein that contains the possible chelating residues have been allowed full flexibility (corresponding mainly to the helix A). For some initial models, the distance between the metal and the residues that are candidate for coordination are slightly too large. In this case, a first part of the minimization has been performed with a distance constraint of 2.5 Å between the metal and the residue. This constraint was released in a second step of the minimization leading to the energy minima described in this work.

## Results and discussion

### Pure quantum mechanical calculations on the isolated cofactor

The structure of the isolated Fe(Schiff base) has been optimized with pure quantum mechanical calculations using the B3LYP functional and an extended basis set (see Materials and methods). The calculations have been carried out on the tetracoordinated species where the Schiff base provides an equatorial environment for the metal of the N2O2 kind. To check how the oxidation and spin states impact on the structure of the system, calculations have been carried out on low spin and high spin Fe(II) and Fe(III) species. The resulting geometries of the four systems are in very good agreement with the experimental structures available for similar Fe(Schiff base) compounds at the Cambridge Structural Database[56] (Table S1†). This shows that density functional approaches represent a convenient tool to generate the initial structures of the cofactors. Regarding the general shape of the Fe(Schiff base) group—a critical point for posterior protein–ligand dockings—high spin species present a subtle doming while low spin ones remain mainly planar. Moreover, the iron atom is slightly displaced out of the average plane of the cofactor for high spin configurations while remaining inside the plane for low spin ones. Nonetheless, the overall structural deviation between the different theoretical systems is rather small (root mean square deviations (rmsd) differences lower than 0.3 Å, Fig. S1†) and we therefore hypothesize that a unique structure of the cofactor should be sufficient for proceeding with the subsequent dockings. Interestingly, the optimized geometries of the isolated cofactor are also in very good agreement with the experimental structures of a similar Fe(Schiff base) inserted in apo-Mb[3] but substantially different from those observed in the experimental structure of the **Fe(Schiff base)·HO** complex (rmsd approx. 0.7 Å). Therefore, a computational protocol able to correctly predict the interaction between the cofactor and its host should be able to reproduce the unusual conformation of the inorganic group of the **Fe(Schiff base)·HO** composite. To do so, the characterization of the protein environment and the ability to alter the shape of the inorganic moiety is necessary. *Could molecular modelling approaches give insight on where and how the cofactor binds?*

### Determining possible binding modes of the Fe(Schiff base)·*cd*HO

An initial run of protein–ligand dockings has been performed using the program Gold (version 4.1). In order to discuss the predictiveness of our protocol, the calculations have been carried out on conformations of the receptor different from the crystallographic structure of the **Fe(Schiff base)**·*cd***HO** composite. In the simulations, all rotatable bonds have been allowed the maximum degree of freedom but the structure of the inorganic core has been remained fixed; our hypothesis is that deformations of the inorganic moiety induced upon binding can be produced in post docking calculations. The flexibility of the receptor has been introduced both by applying a soft potential for residues of the binding cavity as well as performing the docking on several crystallographic structures of the heme oxygenase of *Corynebacterium diphtheriae.* These structures correspond to different liganded states of the enzyme including heme free (PDB code 1ni6),[57] heme bound (PDB code 1iw1),[15] oxy-heme bound (PDB code 1v8x)[58] and heme bound with a large inhibitor bound to the iron (PDB code 3czy).[17] For the 3 latest structures, both the heme and the inhibitor were removed from the binding site of the protein in order to carry out the dockings. It is worth noting that structures 1ni6 and 3czy correspond to human forms of heme oxygenase and homology models of *Corynebacterium diphtheriae* were therefore generated for the consistence of our study. For each protein–ligand docking, twenty runs have been performed leading to twenty low energy putative **Fe(Schiff base)·HO** complexes (here also referred to as solutions).

Because of the difficulty of correctly reproducing metal–protein interactions by actual scoring functions, we analysed the entire series of **Fe(Schiff base)·HO** solutions generating during the docking process. To do so, the root mean square deviations between all simulated complexes have been calculated using the structure of the inorganic catalyst as a criterion of comparison. The rmsd values obtained show substantial variation and range from 0.42 to 6.7 Å. Despite this apparent structural dispersion, closer analysis highlights four distinctive families (clusters) of binding modes referred thereafter as binding modes A, B, C and D (Fig. 2).

In the first subset (A), the cofactor is oriented such that its central aromatic ring points towards the polar patch constituted by Tyrosine 130, Arginine 177 and Lys 13 (Fig. 3a). In this orientation, the carboxylate group of the propionate of the cofactor presents a strong polar interaction with the same patch of residues. In the second group (B), the central aromatic group points toward the core of the binding site cavity and represents an inversion of about 180° with respect to binding mode A (Fig. 3b). In this orientation, a strong polar interaction between the propionate and the side chain of Arginine 132 is observed. Interestingly, in all the structures of **HO** available in the Protein Data Bank, this residue presents its iminium group hanging at the centre of the binding site cavity. In the third group (C), the Fe(Schiff base) moiety moves from its position in A by a rotation of about 90° around the axis passing through the iron atom and perpendicular to the average plane of the cofactor (Fig. 3c). The rotation takes place in the direction of the solvent leading

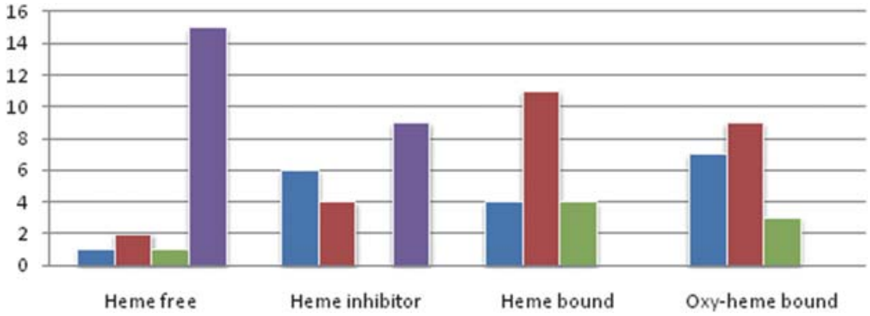


**Fig. 2** Evolution of the number of docking solutions consistent with one of the binding modes A, B, C or D. Blue bars correspond to orientations consistent with binding mode A, red bars with binding mode B, green bars with binding mode C and purple bars with binding mode D. The height of each bar corresponds to the number of solutions consistent with the relevant orientation. Results are given for four different conformations of the receptor.

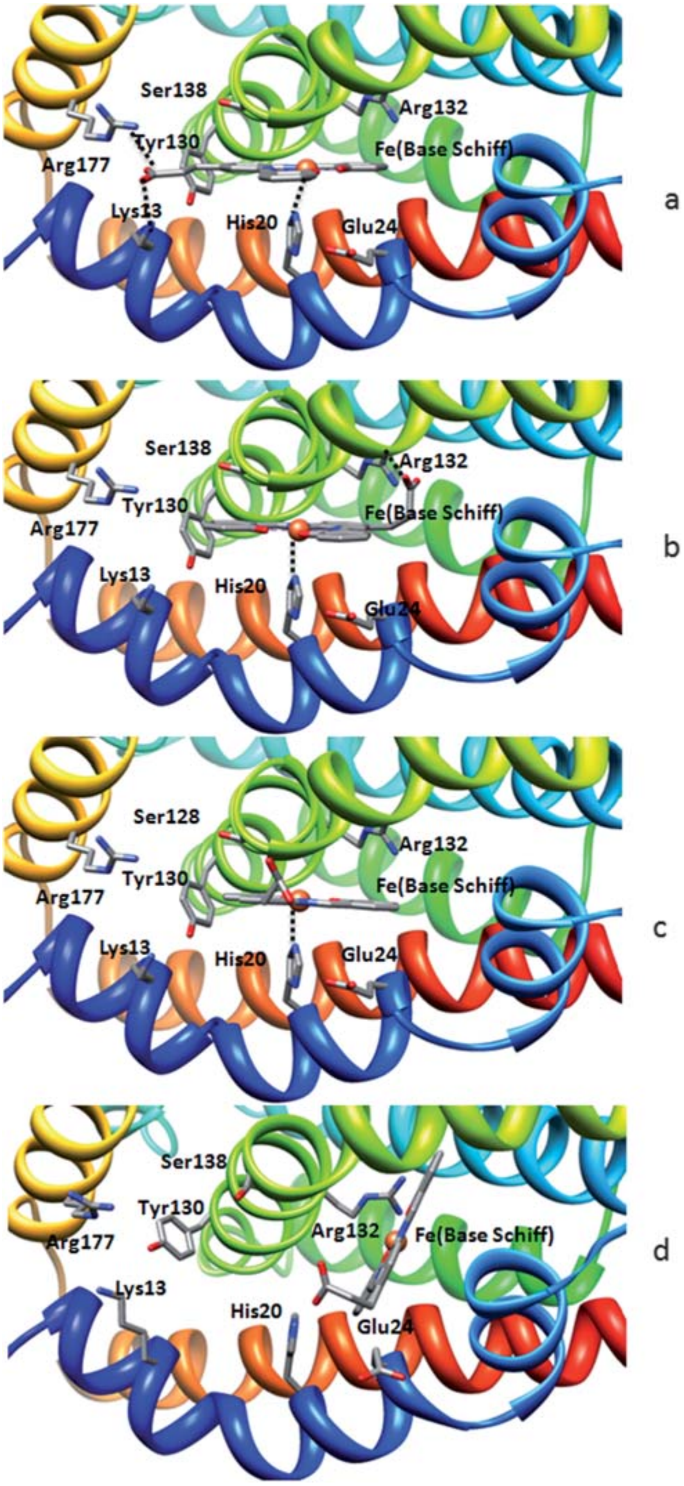


**Fig. 3** Structural representations of one representative solution of the binding modes A (a), B (b), C (c) and D (d).

the central aromatic moiety and carboxylate moiety of the propionate to become more accessible to the aqueous medium. Finally, in the fourth binding orientation (D), the entire cofactor is included into the same cavity that occupies the Arg132 side chain; a cavity generally occupied by iron chelating inhibitors[17] (Fig. 3d). It is to notice that none of the structures of heme·HO complexes present the natural cofactor of heme oxygenases at this site.

The conformation of the receptor affects on the prediction of the structure of the **Fe(Schiff base)·*cd*HO** complex (Fig. 2). When the dockings are performed on a structure of the receptor corresponding to the heme bound or oxy-heme bound HO, orientations A, B and C are observed but not the orientation D. Moreover, the binding modes B and C are slightly more populated than binding mode A. When the dockings are performed on protein scaffolds corresponding to heme-free or heme-inhibited *cd***HO**, the repartition of the binding modes is different. The number of the binding modes B and C decreases noticeably or even disappears while the number of orientations of the group A remains stable or even increases. Docking solutions of the mode D can only be generated for these particular conformations of the receptor. These observations are the results of the quite wider binding site cavity at the proximity of the Arg132 displayed by these two structures.

As expected, energetic considerations does not allow the discrimination between these four binding modes at this stage. All the calculated energies represent reasonably good binding affinities with Chemscore values close to $-30$ kJ mol$^{-1}$ (Table 1). However, the energy differences obtained between each the binding orientations are generally below 4 kJ mol$^{-1}$ and never higher than 6.5 kJ mol$^{-1}$. The lack of accurate energetic and structural reproduction of metal–protein interaction suggests that the ranking of the different orientations obtained at this stage should be considered with caution. However, experimental knowledge, statistical interpretation and structural considerations allow us to discard some of these binding modes. First, no residue in the vicinity of the iron in the binding mode D could possibility coordinate the iron while at least His20 is in the close environment of the iron for orientations A, B and C. Second, this mode is only observed for conformations of the receptor that present the lesser degrees of complementarities with the heme. Since the protoporphyrin IX and the Schiff base share a quite similar chemical scaffold, good structural overlap should be expected in their binding modes. Based on these two considerations, orientation D should be discarded. Regarding modes A, B and C, all the cofactors are anchored in between the alpha helices αA and αH in agreement with the experimental structures of **heme·HO** systems. In all cases, short distances (below 3.5 Å) between the $N_\varepsilon$ of His20 and the iron are observed. These three orientations of the inorganic group actually correspond to different possible matches between the three aromatic rings of the Schiff base and the different pyrrole containing quadrants of the heme. The presence of such modes is actually consistent with experimental observations of multiple heme binding orientations of apo-myoglobin[59] as well as heme oxygenase mutants.[60] It is therefore possible that orientations A, B and C could actually be populated in solution. Nonetheless, binding mode C, which propionate group is exposed to the solvent and has weaker polar complementarities with the receptor than binding modes A and B, should be expected to be less stable. It would be therefore reasonable to hypothesise that only binding orientations A and B are biologically relevant.

**Table 1** Chemscore values in kJ mol$^{-1}$ for the lowest energy solution of each clusters of the binding modes A, B, C and D with four different conformations of the receptor

| | Heme bound | Heme inhibitor bound | Heme free | Oxy-heme bound |
|---|---|---|---|---|
| A | 32.7 | 31.5 | 28.2 | 32.9 |
| B | 31.4 | 30.8 | — | 28.8 |
| C | 32.7 | 24.9 | — | 29.2 |
| D | — | 31.0 | 30.5 | — |

Regarding the prediction ability of our protocol, the rmsd values between the complexes generated by the simulations and their experimental counterpart range from 0.8 to 6.5 Å (Fig. S2†). The docking solutions displaying the better structural overlap with the crystallographic structure of the **Fe(Schiff base)·HO** composite are all members of the group A (rmsd values lower than 1.5 Å). This very good structural agreement between computed and experimental complexes shows that, despite some limitations, current state-of-the-art protein–ligand dockings can already provide relevant binding orientations of an inorganic compound into its host. Interestingly, a major amount of binding modes consistent with the experiment is obtained when the dockings are performed on conformations of the receptor extracted from heme containing *cd***HO** structures. This result is consistent with the actual knowledge of virtual screening procedures. Indeed, it has been shown that for the docking of a given chemical, the quality of the results generally improves for those conformations of the receptor that are obtained for a structure bound to similar compounds.[47] Despite the encouraging results obtained by the docking procedure, none of the resulting solutions present either the conformational deformation of the cofactor or the hexacoordinated environment of the iron associated with it, a phenomenon that should have a major weight in discriminating between the different binding orientations and is mainly linked to the coordination properties of the iron. *Could computational tools allow the characterization of those residues that can directly interact with the first coordination sphere of the metal?*

## Identification of *cd*HO residues as possible chelating groups of the iron

This question relies on the general problem of *in silico* prediction of metal binding sites. We carried out this part of the work by using a new approach developed in our group (to be published). Similar to current state-of-the-art algorithms in this field, this approach is adapted to the analysis of the dockings ensemble for metal containing ligands (see Materials and methods). First, statistics on the all solutions have been performed to detect residues with an adequate position of its $C_\alpha$ for coordinating the metal. Calculations over $C_\alpha$–Fe distances highlight 7 residues as possible chelating agents: Histidine 20, Glutamate 21, Glutamate 24, Arginine 177, Aspartate 136, Serine 138 and Asparagine 204 (Fig. 4). The results are quite similar when the conformation of the receptor corresponds to heme containing

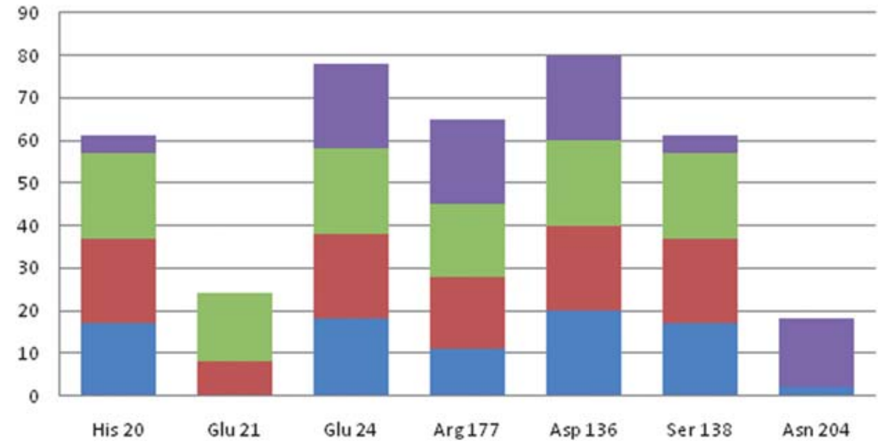


**Fig. 4** Number of hits for the possible coordination to the metal of seven residues in the *cd***HO** binding site. Only the residues that, in a given docking solution, display distances of less than 7 Å between its alpha carbon and the metal are considered. Statistics are performed on each individual docking solutions for four conformations of the receptor: heme free (purple), oxy-heme bound (green), heme bound (red) and heme inhibited (blue) conformations of *cd***HO**.

structures of the heme oxygenase (*i.e.* Glu24 or His20 present 20 hits for heme-oxy and heme bound structure) but present a different profile when dealing with the heme free scaffolds of HO (*i.e.* His20 is found 5 times and Glu24 20 times). This result is due to the higher number of docking solutions of the group D that predicts the location of the cofactor far from Histidine 20. When comparing with the experimental structures of **heme·HO** and **Fe(Schiff base)·HO** species, the results obtained on heme containing conformations appear more reliable. This newly highlights on the possible pitfalls in predicting cofactor–host interaction when the conformation of the receptor is not compatible enough with the chemical on which the simulation are performed.

Very few hits are obtained for Glu21 and Asn204 which makes their elimination from the list of possible chelating residues a natural step. For the five remaining candidates, the only location of the $C_\alpha$ is not a sufficient criterion for metal binding and a more detailed analysis is required. In order to evaluate if the side chain of these residues could adopt a conformation adequate for the coordination of the metal, the generation of all rotameric states of these amino acid has first been performed. This conformational search shows that Arg132, Asp136 and Ser138 are not valid candidates because of the numerous bad contacts observed for the rotameric states that provide distances with the metal adequate for coordination. For Glutamate 24, however, rotamers with no bad contacts with the atoms of its vicinity whether have the side chain pointing to the solvent (*i.e.* rotamer E1) or the $O_\varepsilon$ atom at an acceptable distance for coordinating the metal (only two candidates E2 and E3) (Fig. 5). Interestingly, when comparing with the heme bound structure of *cd***HO**, one can see that the rotameric states E2 and E3 are not compatible with the fixation of the cofactor. Indeed, the extraquadrant of the heme with respect to the Fe(Schiff base) sterically avoids the rotation of the glutamate towards the binding site. Therefore, the bigger size of the heme, the natural substrate of heme oxygenase, and its stronger rigidity explain why Glutamate 24 is not able to bind the iron in the **heme·HO** complex. Finally, Histidine 20 is systematically predicted to be in the first coordination sphere of the metal and the analysis of all its possible rotameric states shows that two orientations are compatible for the coordination of the metal (rotamers H1 and H2) (Fig. 5).

At the final stage of the structural refinement of the metal environment of the binding modes A and B, we generated the 6 possible combinations of His20 and Glu24 rotameric states. Taking into account the number of clashes between these residues and their environment as well as common knowledge on coordination

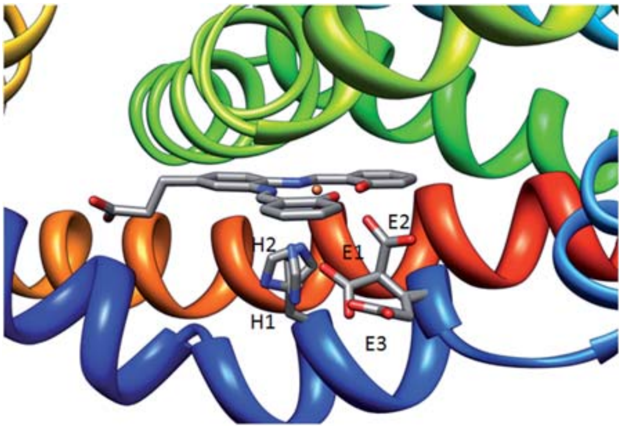


**Fig. 5** Representation of the individual rotameric states of Histidine 20 (H1 and H2) and Glutamate 24 (E1, E2 and E3) with possible impact on the coordination of the iron of the cofactor.

chemistry, we were able to reduce the conformational space to 2 candidates which correspond to the combination E1-H1 (referred to hereafter as model 1) and E2-H2 (referred to hereafter as model 2). The resulting environments of the metal obtained in these models correspond to hypothetical hexacoordinated (models A1 and B1) and pentacoordinated (models A2 and B2) configurations of the iron. Unfortunately, this part of the protocol does not lead to significant energetic differences to discuss the relative stability of the different complexes. To do so, methods able to provide an accurate representation of the electronic properties of the full molecular system are required. *Could QM/MM calculations provide us with such information?*

### QM/MM refinement and final models

In order to judge the chemical relevance of the four putative binding modes of the Fe(Schiff base) group into the cdHO binding site (A1, A2, B1 and B2), QM/MM optimizations have been undertaken for each orientation. In order to reduce the computational cost, these calculations have been limited to the four **Fe(Schiff base)·*cd*HO** complexes obtained from the dockings performed on only one of the heme bound conformations of the receptor.

Regarding the binding mode A, all the optimized geometries of the model A1 lead to distorted octahedral configurations with at least one of the atoms of the first coordination sphere of the metal displaced about 13° (on average) from the ideal value of 90° of a perfect octahedron (Fig. 6a and Table S2†). All the metal–ligand distances are compatible with iron coordination and display values of about 2.0 Å on average. Independently of the oxidation or spin states, in all these complexes, the overall scaffold of the Schiff base is distorted with deviation of the planarity from 0.3 to 0.5 Å. The metal ion is displaced out of the average plane of the prosthetic group with values ranging from 0.76 to 1.50 Å (minimum value for Fe(III) low spin and maximum value for Fe(II) high spin respectively). In this conformation, the Schiff base provides with 4 coordinating atoms and both $N_\varepsilon$ of His20 and $O_\varepsilon$ of Glu24 complete the octahedral structures. These optimized structures are in very good agreement with their experimental counterparts with values generally closer to Fe(III) high spin system; the expected populated state in the experimental conditions (Table S2†).

All optimized geometries of the model A2 leads to **Fe(Schiff base)·HO** complexes with square pyramidal geometries (Fig. 6b and Table S2†). The Schiff base provides with the equatorial N2O2 environment and the $N_\varepsilon$ of the histidine 20 represents the proximal ligand. The conformation of the macrocycle is generally more planar for low spin than high spin systems and the metal lays closer to the average plane of the Schiff base in the low spin (0.3 Å) than in the high spin species (0.56 Å). The glutamate remains out of the coordination sphere of the metal with distances between the metal and the $O_\varepsilon$ of Glu24 depending on the oxidation state of the metal. While Fe–$O_\varepsilon$ distances in optimized Fe(II) high spin and low spin systems are 4.7 and 6.0 Å respectively, they are 4.1 and 4.4 Å in optimized Fe(III) high spin and low spin respectively. This shows that, despite not being coordinated to the metal, the glutamate gets closer to the metal in the structures with the iron in the oxidation state +III. The higher electrostatic interaction between Fe(III) and the glutamate tends to attract the negatively charged residue to the immediate environment of the metal.

Beyond the distortion of the cofactor and the change of coordination of the metal, the comparison between the models A1 and A2 provides other pieces of molecular information. First, for Fe(II) species, the pentacoordinated systems show shorter coordination bonds that the hexacoordinated ones (1.93 *versus* 2.01 Å and 2.07 *versus* 2.17 Å for Fe(II) low spin and high spin respectively). Such behaviour is not observed for the Fe(III) species; the change of the first coordination sphere is almost null in this case (1.99 *versus* 1.96 Å for Fe(III) low spin species and 2.07 *versus*

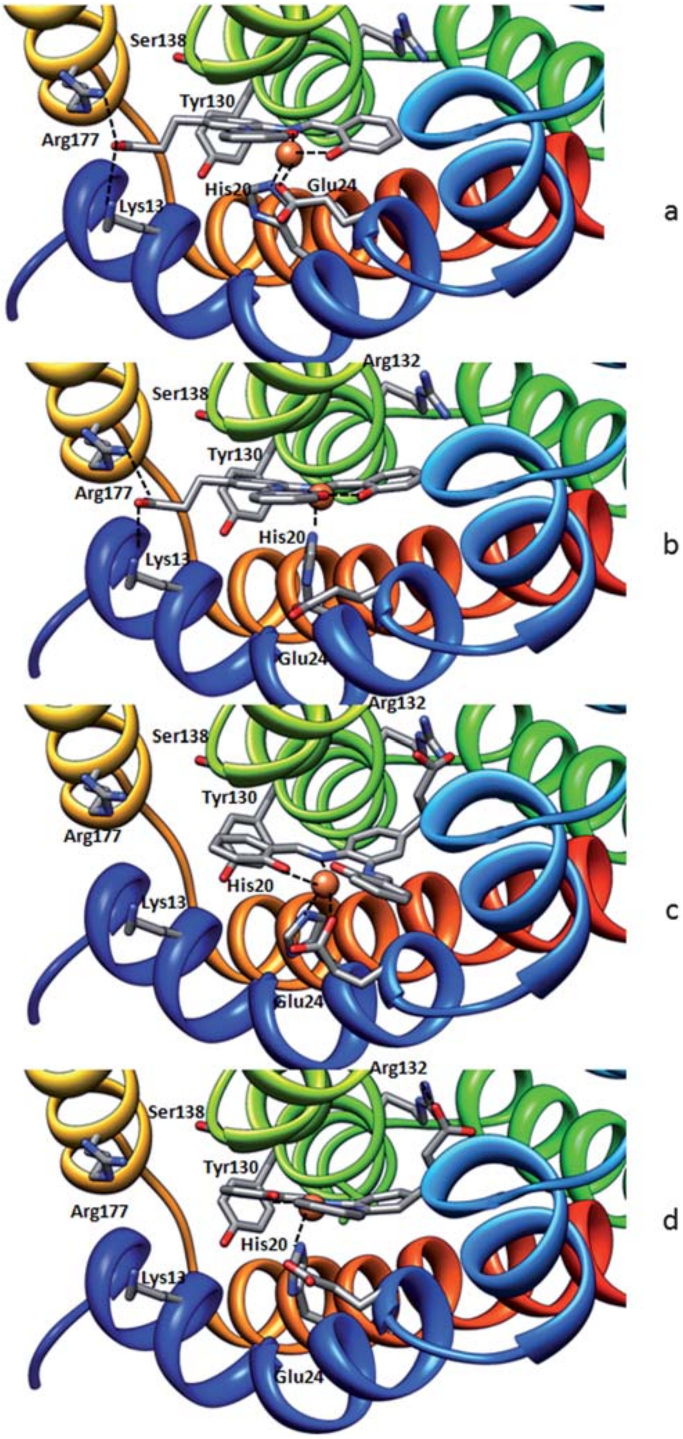


**Fig. 6** Structures of the B3LYP/AMBER optimized systems of the **Fe(Schiff base)**·*cd***HO** complex for (a) binding mode A1, (b) binding mode A2, (c) binding mode B1, and (d) binding mode B2.

2.08 Å for Fe(III) low spin). Such differences between both oxidation states can be due to the electronic nature of the systems but undoubtedly has some relation with the stronger electrostatic interaction between the Glutamate 24 and Fe(III).

**Table 2** QM, MM and QM/MM energies of ONIOM B3LYP/AMBER minimized structures of models A1, A2, B1 and B2 in different electronic configurations. Energies are given in kJ mol$^{-1}$

| | Model | $\Delta E_{\mathrm{QM}}$ | $\Delta E_{\mathrm{MM}}$ | $\Delta E_{\mathrm{QM/MM}}$ |
|---|---|---|---|---|
| Fe(II) LS | Hexacoordinate A1 | 0.0 | 0.0 | 0.0 |
| | Pentacoordinate A2 | −107.0 | 70.3 | −36.7 |
| | Hexacoordinate B1 | X | X | X |
| | Pentacoordinate B2 | −24.8 | 613.8 | 588.9 |
| Fe(II) HS | Hexacoordinate A1 | 0.0 | 0.0 | 0.0 |
| | Pentacoordinate A2 | −25.1 | 35.9 | 10.9 |
| | Hexacoordinate B1 | 11.2 | 632.5 | 643.7 |
| | Pentacoordinate B2 | 57.5 | 725.9 | 783.4 |
| Fe(III) LS | Hexacoordinate A1 | 0.0 | 0.0 | 0.0 |
| | Pentacoordinate A2 | 92.4 | 45.4 | 137.8 |
| | Hexacoordinate B1 | 85.0 | 626.2 | 711.2 |
| | Hexacoordinate B2 | 167.1 | 802.7 | 969.8 |
| Fe(III)HS | Hexacoordinate A1 | 0.0 | 0.0 | 0.0 |
| | Pentacoordinate A2 | 45.6 | −43.3 | 2.3 |
| | Hexacoordinate B1 | 70.3 | 519.6 | 589.9 |
| | Pentacoordinate B2 | −18.2 | 735.4 | 717.3 |

Another interesting phenomenon appears in the change of the conformation of the receptor in function of the configuration of the metal. Indeed, the relative position of the αA helix is substantially closer to the cofactor in the hexacoordinated state than in the pentacoordinated one. This motion is actually consistent with the structural differences observed between **heme·HO** and **Fe(Schiff base)·HO** complexes. This shows that our calculations are able to reproduce a fine tuning between the change of coordination of the metal, the shape of the overall cofactor and the conformation of the macromolecule (Fig. 7).

For the model B1, the optimized geometries obtained with our QM/MM approach are very variable. The minimization of both Fe(II) and Fe(III) high spin systems leads to an octahedral configuration of the iron while the Fe(III) low spin system presents a square pyramidal one (Table S2†). Unfortunately, we were not

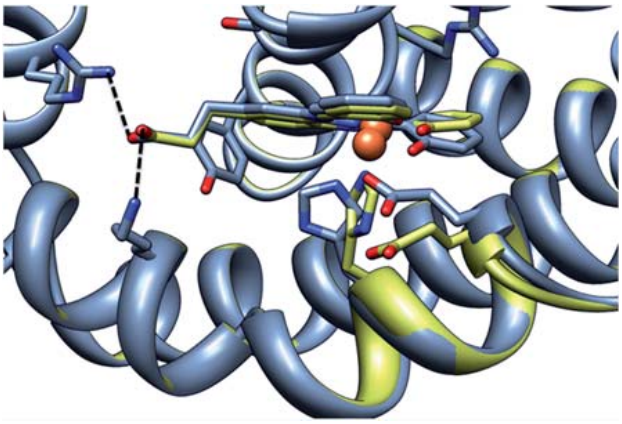


**Fig. 7** Metal induced deformation of the helix αA observed between B3LYP/AMBER optimized hexacoordinated (model A1 in grey) and pentacoordinated species (model A2 in yellow) of **Fe(Schiff base)**·*cd***HO**. The overlap has been performed on the Cα of the entire molecular system.

able to find a stable minimum for the B1 model in the Fe(II) low spin configuration. With regard to the high spin systems, the structural tendencies of the close environment of the metal are similar to those observed in the optimized structures of the model A1: the iron is bound to the four ligands provided by the Schiff base, the $N_\varepsilon$ of the Histidine 20 and the $O_\varepsilon$ of the Glutamate 24. All bond lengths of the first coordination sphere of the metal are about 2.0 Å. Importantly, added to a distortion of the cofactor similar to those observed for the model A1, the Fe(Schiff base) also presents a large deviation of the central aromatic ring of Schiff base out the average plane (Fig. 6c). The resulting deformation is never observed either in gas phase minimized structures of the Fe(Schiff base) compound or in the structural database of small molecules or proteins. It is most likely that such deformation represents an important additional constraint on the cofactor. A displacement of helix αA is also observed for these species and is more important than those obtained in the corresponding minimized structures of the model A1. Finally, for the Fe(III) low spin system, the initial hexacoordinated configuration is not maintained during the minimization and the final structure presents a pentacoordinated environment with the side chain of the Glu24 out of the coordination sphere of the metal (Fig. 6d). It is most likely that the constraints suffered by the cofactor as described earlier added to the general tendency of low spin systems to adopt a planar conformation provoke the removal of the sixth ligand for the first coordination sphere of the metal and the subsequent change in geometry. Interestingly, the structural features of the species are actually similar to those of the model B2. Precisely, the optimized structures of the model B2 present a square pyramidal conformation with the same coordination environment than reported for the model A2. The distances are about 2.0 Å for the first sphere of coordination and, in this conformation, the Glutamate 24 presents its side chain pointing towards the solvent. Like for the optimized geometries of the model B1, the minimized structures of the model B2 presents severe distortions of the cofactor. In this case though, this deformation corresponds to the out-of-plan bending of the central aromatic ring. The differences observed between models B1 and B2, are due to the exclusion of the side chain of the Glutamate 24 from the vicinity of the iron that allows a slight displacement of the cofactor in B1 towards the solvent but increases the contacts on the central aromatic moiety with the αG and αH helices.

The apparently unfavourable distortions observed in all the optimized structures resulting from the binding mode B are consistent with the calculated QM/MM energies (Table 2). Both models B1 and B2 present differences in energy higher than 650 kJ mol$^{-1}$ when comparing with the less stable optimized structures obtained for the binding mode A. The main contribution in such energetic differences is associated to the MM part of the system. This sustains that the large distortions of both the helix and the aromatic ring of the Schiff base have a destabilizing impact on the system. Importantly, this high difference in energy allows us to discard binding orientation B as a relevant candidate for the prediction of the binding of the Fe(Schiff base) to *cd***HO** and let the optimized structures of binding mode A be unique possible candidates for the formation of the complex; a result in agreement with the experimental knowledge.

Regarding binding mode A, the differences in energy between the hexacoordinated (A1) and pentacoordinated (A2) models are generally subtle and do not overcome 132 kJ mol$^{-1}$. Since comparing different spin and oxidation states of the iron by density functional theory remains a complex question, we prefer to discuss the relative stability of the different conformations for a given spin state. For the Fe(II) low spin systems, one can see that the pentacoordinated square pyramidal conformation (A2) is preferred to the hexacoordinated one (A1) by 36.7 kJ mol$^{-1}$ (Table 2). Most of the difference is provided by the quantum mechanical part of the system showing that the electronic properties dominate in this stabilization. In Fe(II) high spin systems, the hexacoordinated configuration is slightly more stable by 10.9 kJ mol$^{-1}$ but this low difference in energy suggests a readily conversion between both configurations in solvent condition. A similar behaviour should be expected for Fe(III) high spin systems because the difference in QM/MM energy is lower than 2.5 kJ mol$^{-1}$. However, such conversion is most likely impossible for the low spin Fe(III) since the pentacoordinated system is 137.8 kJ mol$^{-1}$ less stable than its hexacoordinated counterpart. Interestingly, this state corresponds to the expected electronic configuration of the crystallographic species. Based on the energetic trends obtained from our QM/MM calculations, transition between pentacoordinated and hexacoordinated structures in solution are most likely to occur for other electronic configurations that the high spin Fe(III).

From a more fundamental point of view, we stress that, with the exception of the Fe(III) low spin, the energetic tendencies obtained in the QM region are largely counterbalanced by those of the MM region (*i.e.* respectively −25.1 and 35.9 kJ mol$^{-1}$ for Fe(II) high spin systems). This clearly highlights the importance of finding a method which is able to accurately reproduce the fine electronic effects of the accurate predictions of transition metal–protein interactions, including the real size system.

The QM/MM calculations performed on the different models generated by refinement of the docking solutions of the models A and B of the **Fe(Schiff base)·*cd*HO** provide key molecular information. First, such calculations allow discarding models B1 and B2 as possible binding orientations based on energetic and structural considerations. QM/MM calculations are also able to provide with three dimensional models of the hexacoordinated **Fe(Schiff base)·*cd*HO** systems in very good agreement with the experiment. This result clearly highlights the good predictive ability that the kind of integrative protocol presented here can offer. Interestingly, our calculations also provide three dimensional models of pentacoordinated species of the **Fe(Schiff base)·*cd*HO** and sustain the idea that these conformations should be accessible in experimental conditions; the transition between penta- and hexacoordinated configurations should be dependent on the electronic configuration of the metal. *But what is the real significance of these pentacoordinated structures? Could they have chemical sense or even provide molecular insight on the design, redesign or optimization of artificial cofactors?*

## Computational insights on Fe(Schiff-base)·HO binding mechanism

The **Fe(Schiff-base)·HO** composite designed by Ueno *et al.*[5] represents a major step forward in the design of synthetic metalloenzymes. The possibility that this composite gains its activity *via* an electron transfer with the natural partner of ***cd*HO** represents a new conceptual framework for applications in biomedicine and electronics. This composite is also one of the few with a resolved crystallographic structure and represents an excellent model system for testing the prediction of the interaction of inorganic complexes with a protein host by computational means.

From our simulations, the distorted conformation of the cofactor appears as the results of several molecular variables. The lack of a "fourth quadrant" of the Fe(Schiff base) cofactor with respect to the heme means one equatorial side of the metal is more accessible to the environment than for the natural substrate of ***cd*HO**. The local flexibility of Glutamate 24, the global flexibility of αA helix and the positive charge of the iron stabilize the entrance of the glutamate in the first coordination sphere of the metal. The flexibility of the Schiff base on its part allows the sufficient displacement of the iron out-of-the-plane of the cofactor so that the Glutamate 24 can bind the metal at the remaining equatorial positions of the octahedron. The unexpected conformation of the cofactor is therefore the result of the relative energetic costs between the distortion of the Schiff base, the formation of a new coordination bond and the displacement of the helix.

Our calculations also show that the conversion from a hexacoordinated system with His20 and Glu24 as coordinating amino acids of the iron to a pentacoordinated system with the His20 as the unique protein ligand should be accessible in experimental conditions. This provides interesting information on a quite struggling piece of molecular event. Indeed, **Fe(Schiff-base)·HO** systems activate oxygen from $O_2$ to $O^{-}{}_2$. This process is related to the change of oxidation state of the iron and the transition from Fe(III) to Fe(II); the latest corresponds to the electronic state that allows the fixation of the oxygen. Intriguingly, the octahedral configuration of the iron, as observed in the experimental structures, appears quite inadequate for performing such fixation. Indeed, the iron atom has its first coordination sphere crowded and no room is available for oxygen binding. The calculation of the accessible volume around the iron in the experimental system shows that very little room is available at this site (about 6.1 A$^3$) (Fig. 8a). Therefore, one wonders how the oxygen could bind onto this hexacoordinated structure.

The analysis of the optimized pentacoordinated complexes for Fe(II) high spin shows that a wider space is accessible on the top of the iron (about 70 A$^3$) with a volume similar to the one observed in crystallographic complexes of HO bound with oxy-heme (66 A$^3$) (Fig. 8b). As one would expect, the formation of the pentacoordinated structure should initiate the catalytic process of this artificial enzyme. Therefore, the 3 dimensional model of the pentaccordinated systems, and in particular the high spin one, could provide interesting information for the catalytic

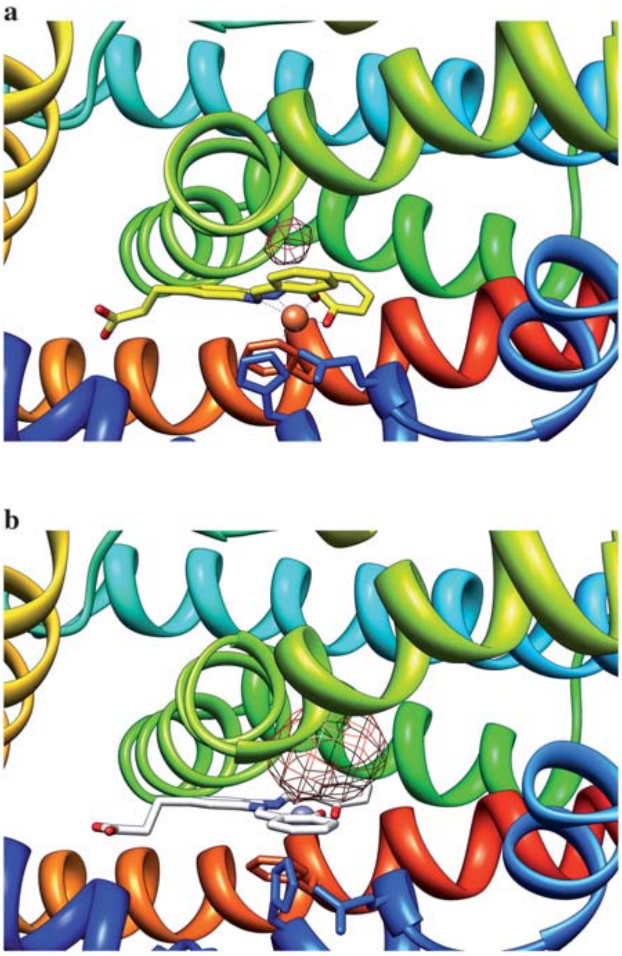


**Fig. 8** Comparison of the available space above the iron moiety in (a) the experimental structure of the **Fe(Schiff base)·HO** composite (octahedral configuration) and (b) the theoretical **Fe(Schiff base)·HO** system obtained with the ONIOM B3LYP/AMBER calculations (square pyramidal configuration).

properties of this family of artificial enzymes. However, the study of this particular part of the design goes further than the scope of the actual discussion.

### Computational techniques for predicting the interactions of inorganic species with biological systems

The structure of the first composite obtained by the insertion of a Fe(Schiff base) into the binding site of the heme oxygenase of *Corynebacterium diphtheriae* represents one of the most challenging systems to predict by molecular modelling techniques. This is mainly due to the number of bioinorganic variables that are to be taken in account for an accurate prediction of the interaction between the inorganic compound and its host: plasticity of the first coordination sphere of the metal, electrostatic interactions between both partners, changes in oxidation and spin states, flexibility of the macrocyclic inorganic moiety, plasticity of the receptor, *etc*. In fact, these different aspects could be summarised into one of the fundamental questions in the recognition process of bioinorganic systems: how the metallic species **induces** conformational changes on the protein and how the protein scaffold **controls** the coordination sphere of the metal. Nowadays, no unique molecular modelling technique is able to deal accurately with these two aspects. Developments of more accurate empirical functions able to deal with fine electronic effects of transition metals or the development of faster *ab initio* or DFT codes will undoubtedly lead to major improvements and breakthroughs in this field. To date, however, only protocols that combine several techniques together represent reliable alternatives to study the binding of metallic species to biological macromolecules.

In this study, an initial integrative computational protocol is applied to a unique system with major implications for the design of artificial enzymes. The possibility of determining several energetically relevant coordination states of a metal when interacting with its host as well as the characterization of the impact of metal binding on the structure of the receptor opens new avenues in bio-inorganic chemistry. Indeed, very similar approaches that combine Quantum Mechanical steps (for an accurate representation of the metal environment) with approximated computational approaches (for an extensive search on the conformational space of the entire molecular system) should substantially help the prediction of the native structures of metalloproteins. For example, these approaches should allow the generation of three dimensional models of structures of metalloproteins bound to their natural cofactor when only the apo-structure of the protein is available (being the apo-structure from experimental nature or provided by homology modelling calculations). These tools should particularly useful since the release of the structure apo-structures of metalloproteins in the Protein Data Bank is frequent and this phenomenon is likely to increase because of the growing number of structural genomic projects. Several ways of developments can be imagined for upgrading the general framework of homology modelling procedures by quantum mechanical extensions and its is to expect that they would lead to more realistic structural and energetic properties of the resulting three dimensional models. Another possible application would be the integration of QM steps for the *ab initio* simulations of the folding process of proteins, peptides or DNA upon the binding of metallic species. In this case, quantum mechanical based approaches should be coupled with Molecular Dynamics methodologies. Such approaches can potentially bring major steps forward in biomedicine and more particularly in neurodegenerative diseases like Alzheimer or Parkinson.

## Conclusion

"What could bring molecular modelling techniques into the design of artificial metal containing cofactor?" is a vast question. The present study shows that individual computational chemistry techniques could provide some molecular insights in this field but that a protocol integrating several of them can be highly predictive. In particular, this work shows that standard protein–ligand dockings already provide relevant orientations of the cofactor into the host cavity when adequate receptor structures are selected for the process, but lack the necessary accuracy for the first coordination sphere of the metal in order to discriminate between all possible solutions. Statistical treatment and structural modelling on the docking results allows the correct characterization of those residues that could bind to the metal and to generate a first series of structural models. QM/MM calculations thanks to an accurate representation of the metal and its environment allow discriminating between hypothetic binding modes and stable coordination states. The integrative approach presented here leads to structural models in an extremely good agreement with the challenging **Fe(Schiff base)·HO** structure of the composite designed by Ueno *et al.*[5] and shows a very interesting degree of predictiveness. Moreover, for this particular system, our calculations also predict that the transition between square pyramidal and octahedral configurations of the iron should have a reduced energetic cost, partially explaining the activation step of the system. Although substantial improvements are still required, this study clearly demonstrates that integrative computational approaches are good predictive tools for the determination of low energy **metal-cofactor·protein** complexes. Natural extensions of the present work outside the artificial metalloenzymes framework includes the improvement of homology modelling techniques for the prediction of native structures of metalloproteins bound to their natural metallic cofactor or the prediction of metal induced folding of proteins and peptides. We are currently exploring several of these aspects.

## Acknowledgements

We are grateful to the Spanish “Ministerio de Ciencia e Innovación” (MICINN (Projects) CTQ2008-06866-C02-01 and Consolider Ingenio 2010 CSD2007-00006) and to the Generalitat de Catalunya (2009/SGR/68) for financial support. V. M. R. thanks the MICINN for a FPI fellowship. J.-D. M. thanks Dr Laura Masgrau for useful discussions on several parts of the manuscript.

## References

1 Y. Lu, N. Yeung, N. Sieracki and N. M. Marshall, *Nature*, 2009, **460**, 855–862.

2 T. R. Ward, *Bio-inspired Catalysts, Top. Organomet. Chem.*, 2009, 25.

3 T. Ueno, T. Koshiyama, M. Ohashi, K. Kondo, M. Kono, A. Suzuki, T. Yamane and Y. Watanabe, *J. Am. Chem. Soc.*, 2005, **127**, 6556–6562.

4 J. R. Carey, S. K. Ma, T. D. Pfister, D. K. Garner, H. K. Kim, J. A. Abramite, Z. Wang, Z. Guo and Y. Lu, *J. Am. Chem. Soc.*, 2004, **126**, 10812–10813.

5 T. Ueno, N. Yokoi, M. Unno, T. Matsui, Y. Tokita, M. Yamada, M. Ikeda-Saito, H. Nakajima and Y. Watanabe, *Proc. Natl. Acad. Sci. U. S. A.*, 2006, **103**, 9416–9421.

6 M. E. Wilson and G. M. Whitesides, *J. Am. Chem. Soc.*, 1978, **100**, 306–307.

7 M. Creus, A. Pordea, T. Rossel, A. Sardo, C. Letondor, A. Ivanova, I. LeTrong, R. E. Stenkamp and T. R. Ward, *Angew. Chem., Int. Ed.*, 2008, **47**, 1400–1404.

8 I. W. McNae, K. Fishburne, A. Haptemariam, T. M. Hunter, M. Melchart, F. Y. Wang, M. D. Wlakinshaw and P. J. Sadler, *Chem. Commun.*, 2004, 1786–1787.

9 A. Mahammed and Z. Gross, *J. Am. Chem. Soc.*, 2005, **127**, 2883–2887.

10 M. T. Reetz and N. Jiao, *Angew. Chem.*, 2006, **118**, 2476–2478.

11 (*a*) R. Ricoux, R. Dubuc, C. Dupont, J.-D. Maréchal, A. Martin, M. Sellier and J.-P. Mahy, *Bioconjugate Chem.*, 2008, **19**, 899–910; (*b*) R. Ricoux, M. Allard, R. Dubuc, C. Dupont, J.-D. Maréchal and J.-P. Mahy, *Org. Biomol. Chem.*, 2009, **7**, 3208–3211.

12 L. Panella, J. Broos, J. Jin, M. W. Fraaije, D. B. Janssen, M. Jeronimus-Stratingh, B. L. Feringa, A. J. Minnaard and J. G. de Vries, *Chem. Commun.*, 2005, 5656–5658.

13 D. J. Schuller, A. Wilks, P. R. Ortiz de Montellano and Thomas L. Poulos, *Nat. Struct. Biol.*, 1999, **6**, 860–867.

14 M. Rivera and Y. Zeng, *J. Inorg. Biochem.*, 2005, **99**, 337–354.

15 S. Hirotsu, G. C. Chu, M. Unno, D. S. Lee, T. Yoshida, S. Y. Park, Y. Shiro and M. Ikeda-Saito, *J. Biol. Chem.*, 2004, **279**, 11937–11947.

16 L. Lad, J. Friedman, H. Li, B. Bhaskar, P. R. Ortiz De Montellano and T. L. Poulos, *Biochemistry*, 2004, **43**, 3793–3801.
17 M. N. Rahman, J. Z. Vlahakis, W. A. Szarek, K. Nakatsu and Z. Jia, *J. Med. Chem.*, 2008, **51**, 5943–5952.
18 F. C. Bernstein, T. F. Koetzle, G. J. Williams, E. E. Meyer Jr., M. D. Brice, J. R. Rodgers, O. Kennard, T. Shimanouchi and M. Tasumi, *J. Mol. Biol.*, 1977, **112**, 535.
19 N. Yokoi, T. Ueno, M. Unno, T. Matsui, M. Ikeda-Saito and Y. Watanabe, *Chem. Commun.*, 2008, 229–231.
20 V. Nanda and R. L. Koder, *Nat. Chem.*, 2010, **2**, 15–24.
21 O. Maglio, F. Nastri, R. T. M. Martin de Rosales, M. Faiella, V. Pavone, W. F. DeGrado and A. Lombardi, *C. R. Chim.*, 2007, **10**, 703–720.
22 C. M. Summa, M. M. Rosenblatt, J.-K. Hong, J. D. Lear and W. F. DeGrado, *J. Mol. Biol.*, 2002, **321**, 923–938.
23 R. J. Deeth, *Struct. Bonding*, 2004, **113**, 37–69.
24 F. Nesse, *JBIC, J. Biol. Inorg. Chem.*, 2006, **11**, 702–711.
25 P. E. M. Siegbahn and F. Himo, *JBIC, J. Biol. Inorg. Chem.*, 2009, **14**, 643–651.
26 H. Martin Senn and W. Thiel, *Angew. Chem., Int. Ed.*, 2009, **48**, 1198–1229.
27 (*a*) C. M. Bathelt, A. J. Mulholland and J. N. Harvey, *J. Phys. Chem. A*, 2008, **112**, 13149–13156; (*b*) C. M. Bathelt, J. Zurek, A. J. Mulholland and J. N. Harvey, *J. Am. Chem. Soc.*, 2005, **127**, 12900–12908.
28 J. C. Schöneboom, S. Cohen, H. Lin, S. Shaik and W. Thiel, *J. Am. Chem. Soc.*, 2004, **126**, 4017–4034.
29 V. Guallar, M.-H. Baik, S. J. Lippard and R. A. Friesner, *Proc. Natl. Acad. Sci. U. S. A.*, 2003, **100**, 6998–7002.
30 R. T. Kroemer, *Curr. Protein Pept. Sci.*, 2007, **8**, 312–328.
31 M. A. Miteva, J.-D. Maréchal, C. R. Robert and David Perahia in *In silico lead discovery*, Bentham Science Publishers, ed. Maria A Miteva, 2010, in press.
32 G. Jones, P. Willett, R. C. Glen, A. R. Leach and R. Taylor, *J. Mol. Biol.*, 1997, **267**, 727–748.
33 B. Seebeck, I. Reulecke, A. Kämper and M. Rarey, *Proteins: Struct., Funct., Bioinf.*, 2008, **71**, 1237–1254.
34 R. A. Friesner, J. L. Banks, R. B. Murphy, T. A. Halgren, J. J. Klicic, D. T. Mainz, M. P. Repasky, E. H. Knoll, Mee Shelley, J. K. Perry, D. E. Shaw, P. Francis and P. S. Shenkin, *J. Med. Chem.*, 2004, **47**, 1739–1749.
35 P. Fong, J. P. McNamara, I. H. Hillier and R. A. Bryce, *J. Chem. Inf. Model.*, 2009, **49**, 913–924.
36 A. E. Cho, V. Guallar, B. J. Berne and R. Friesner, *J. Comput. Chem.*, 2005, **26**, 915–931.
37 A. E. Cho and D. Rinaldo, *J. Comput. Chem.*, 2009, **30**, 2609–2616.
38 A. Pordea, D. Mathis and T. R. Ward, *J. Organomet. Chem.*, 2009, **694**, 930–936.
39 E. F. Pettersen, T. D. Goddard, C. C. Huang, G. S. Couch, D. M. Greenblatt, E. C. Meng and T. E. Ferrin, *J. Comput. Chem.*, 2004, **25**, 1605–1612.
40 (*a*) A. D. Becke, *J. Chem. Phys.*, 1993, **98**, 5648–5652; (*b*) C. Lee, W. Yang and R. G. Parr, *Phys. Rev. B: Condens. Matter*, 1988, **37**, 785–789.
41 M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, J. A. Montgomery, Jr., T. Vreven, K. N. Kudin, J. C. Burant, J. M. Millam, S. S. Iyengar, J. Tomasi, V. Barone, B. Mennucci, M. Cossi, G. Scalmani, N. Rega, G. A. Petersson, H. Nakatsuji, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, M. Klene, X. Li, J. E. Knox, H. P. Hratchian, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. Ochterski, P. Y. Ayala, K. Morokuma, G. A. Voth, P. Salvador, J. J. Dannenberg, V. G. Zakrzewski, S. Dapprich, A. D. Daniels, M. C. Strain, O. Farkas, D. K. Malick, A. D. Rabuck, K. Raghavachari, J. B. Foresman, J. V. Ortiz, Q. Cui, A. G. Baboul, S. Clifford, J. Cioslowski, B. B. Stefanov, G. Liu, A. Liashenko, P. Piskorz, I. Komaromi, R. L. Martin, D. J. Fox, T. Keith, M. A. Al-Laham, C. Y. Peng, A. Nanayakkara, M. Challacombe, P. M. W. Gill, B. G. Johnson, W. Chen, M. W. Wong, C. Gonzalez and J. A. Pople, *GAUSSIAN 03 (Revision C.02)*, Gaussian, Inc., Wallingford, CT, 2004.
42 W. J. Hehre, R. Ditchfield and J. A. Pople, *J. Chem. Phys.*, 1972, **56**, 2257–2261; P. C. Hariharan and J. A. Pople, *Theor. Chim. Acta*, 1973, **28**, 213–222.
43 P. J. Hay and W. R. Wadt, *J. Chem. Phys.*, 1985, **82**, 299–310.
44 M. L. Verdonk, J. C. Cole, M. J. Hartshorn, C. W. Murray and R. D. Taylor, *Proteins: Struct., Funct., Genet.*, 2003, **52**, 609–623.
45 M. D. Eldridge, C. W. Murray, T. R. Auton, G. V. Paolini and R. P. Mee, *J. Comput.-Aided Mol. Des.*, 1997, **11**, 425–445.

46 S. B. Kirton, C. W. Murray, M. L. Verdonk and R. D. Taylor, *Proteins: Struct., Funct., Bioinf.*, 2005, **58**, 836–844.

47 M. Rueda, G. Bottegoni and R. Abagyan, *J. Chem. Inf. Model.*, 2010, **50**, 186–193.

48 D. A. Gschwend, A. C. Good and I. D. Kuntz, *J. Mol. Recognit.*, 1996, **9**, 175–186.

49 M. Babor, S. Gerzon, B. Raveh, V. Sobolev and M. Edelman, *Proteins: Struct., Funct., Bioinf.*, 2008, **70**, 208–217.

50 R. L. Dunbrack Jr., *Curr. Opin. Struct. Biol.*, 2002, **12**, 431–440.

51 S. Dapprich, I. Komáromi, K. S. Byun, K. Morokuma and M. J. Frisch, *THEOCHEM*, 1999, **461–462**, 1–21.

52 Y. Zhao, N. E. Schultz and D. G. Truhlar, *J. Chem. Phys.*, 2005, **123**, 161103–161103.

53 J. P. Pardew and Y. Wang, *Phys. Rev. B: Condens. Matter*, 1986, **33**, 8822.

54 W. D. Cornell, P. Cieplak, C. I. Bayly, I. R. Gould, K. M. Merz Jr., D. M. Ferguson, D. C. Pellmeyer, T. Fox, J. W. Caldwell and P. A. Kollman, *J. Am. Chem. Soc.*, 1995, **117**, 5179–5197.

55 J. Wang, W. Wang, P. A. Kollman and D. A. Case, *J. Mol. Graphics Modell.*, 2006, **25**, 247–260.

56 F. H. Allen, *Acta Crystallogr., Sect. B: Struct. Sci.*, 2002, **58**, 380–388.

57 L. Lad, D. J. Schuller, H. Shimizu, J. Friedman, H. Li, P. R. Ortiz de Montellano and T. L. Poulos, *J. Biol. Chem.*, 2003, **278**, 7834–7843.

58 M. Unno, Toshitaka Matsui, G. C. Chu, M. Couture, T. Yoshida, D. L. Rousseau, J. S. Olson and M. Ikeda-Saito, *J. Biol. Chem.*, 2004, **279**, 21055–21061.

59 G. N. La Mar, H. Toi and R. Krishnamoorthi, *J. Am. Chem. Soc.*, 1984, **106**, 6395–6401.

60 Y. Zeng, R. Deshmukh, G. A. Caignan, R. A. Bunce, M. Rivera and Angela Wilks, *Biochemistry*, 2004, **43**, 5222–5238.

## 2 Computational insights on the possibility of tri-coordinated cisplatinated adducts with protein models

Ortega-Carrasco, E.; Cossio, F.P.; Lledós, A.; Maréchal, J.-D. *J. Inorg. Biochem.* **2012**, *117*, 230-236.

Journal of Inorganic Biochemistry 117 (2012) 230–236

Contents lists available at SciVerse ScienceDirect

Journal of Inorganic Biochemistry

journal homepage: www.elsevier.com/locate/jinorgbio

# Computational insights on the possibility of tri-coordinated cisplatinated adducts with protein models

Elisabeth Ortega-Carrasco [a], Fernando P. Cossío [b], Agustí Lledós [a], Jean-Didier Maréchal [a],*

[a] *Departament de Química, Universitat Autònoma de Barcelona, 08193 Bellaterra, Spain*
[b] *Departamento de Química Orgánica I-Kimika Organikoa I Saila, Kimika Fakultatea, Universidad del País Vasco-Euskal Herriko Unibertsitatea (UPV/EHU), P.K. 1072, 28080-San Sebastián Donostia, Spain*

A R T I C L E I N F O

*Article history:*
Received 25 June 2012
Received in revised form 20 September 2012
Accepted 21 September 2012
Available online 28 September 2012

*Keywords:*
Molecular modelling
Cisplatin
Protein interaction
Trigonal planar geometry
Dehydration energy

A B S T R A C T

In the organism, cisplatin binds to numerous proteins. These interactions can ultimately lead to the emergence of resistance and side effects. Little is known on these recognition processes with only few crystallographic structures of cisplatinated proteins released so far. Some of them, like the monoadduct of cisplatin with the hen egg white lysozyme, display unexpected structural features. Instead of the usual square planar configuration of the metal, an apparent T-shaped geometry is observed. This tri-coordinated structure could be a consequence of some crystallographic limitations. However, the increasing reports of tri-coordinated Pt(II) organometallic complexes questions whether it could also have some physiological relevance. Here, we present a computational study allying pure quantum mechanical and hybrid quantum mechanical/molecular mechanics methodologies to shed light on this particular question. Calculations on monoadducts of *cis*-diamminediaquo-platinum(II) with protein models show that square planar geometries are, as expected, the most stable ones. Dehydrations leading to trigonal geometries have Gibbs energies ranging from 8 to 31 kcal/mol and indicate that some of them may be possible in a proteic environment. Nonetheless, we also observed that such conditions are not afforded in the hen egg white lysozyme adduct.

## 1. Introduction

Cisplatin (*cis*-[Pt(NH$_3$)$_2$Cl$_2$]) represents a major success in anticancer therapy as well as a cornerstone in the fields of bioinorganics and metallodrug design [1]. Its therapeutic action is now well known and stands on a direct interaction with DNA strands. Despite some preferences to form 1,2 guanine intra-strands adducts, cisplatin is too unspecific to afford selective therapies. In the organism, the drug also interacts with many biomolecules and participates to a cascade of molecular events that ultimately lead to side effects and the emergence of resistance [2].

Cisplatin interacts with a broad range of extra and intracellular proteins with approximately 65 to 98% of the drug already bound to plasma proteins short after administration [3]. The list of cisplatin targets is constantly increasing and includes, for instance, transferrin [4], serum albumin [5–7], haemoglobin [3] or apo lipoproteins [7], ubiquitin [8] or metathioneins [9] or copper transport proteins [10]. Today, cisplatin–protein interactions are widely characterized by spectroscopic techniques (i.e. mass spectrometry) that allow to identify the binding site(s) of the drug but remain often ambiguous to describe its exact mode of coordination.

Only few X-ray structures of cisplatin–protein adducts have been released so far. Despite representing fundamental three dimensional models for further developments of platinum derivatives, some of them show unexpected geometries from an inorganic chemistry point of view. In particular, the complexes of cisplatin with the hen egg white lysozyme [11] (pdb code 2I6Z with a resolution of 1.90 Å) or with the bovine Cu, Zn superoxide dismutase [12] (pdb code 2AEO with a resolution of 1.80 Å) reveal unusual tri-coordinated coordination modes of the metal. In both cases, the drug binds at the surface of the protein via a Pt(II)–histidine coordination bond but the lack [11] or weak density [12] at the fourth coordination site of the metal is consistent with a T-shape geometry. Technical limitations could obviously be the origin of the lack of density of a fourth ligand (most likely a water molecule). Nonetheless, reasonable doubts exist on whether tri-coordinated species could be native.

The viability of trigonal planar Pt(II) systems are generally a centre of debate.[1] These 14-electron complexes have been shown to be primary actors in the catalytic cycles of homogeneous catalysts [13,14] and have been suggested to participate in an dissociative binding mechanism of cisplatin to DNA [15]. Today, a general consensus exists in which these species are considered unstable and only

---

* Corresponding author. Tel.: +34 93 581 49 36; fax: +34 93 581 29 20.
*E-mail address:* jeandidier.marechal@uab.cat (J.-D. Maréchal).

[1] Note that we refer to trigonal planar systems those in which one central atom is bound to three others in the same plane. T-shaped and Y-shaped geometries represent particular cases of this geometry [63].

http://dx.doi.org/10.1016/j.jinorgbio.2012.09.020

transitory in nature. Nonetheless, tri-coordinated Pt(II) organometallic complexes have been increasingly reported in the recent years [16–22]. Most of them adopt a T-shaped geometry that is stabilized by a weak agostic interaction between one of the ligand and the vacant site of the metal. When the ligands are hindered enough, Y-shaped geometries can also be observed [23]. All these tri-coordinated systems suggest that under an adequate second coordination sphere of the metal, trigonal planar geometries of Pt(II) complexes could be reached. Therefore, one could fairly wonder if such conditions are fulfilled for cisplatin–protein adducts in physiological conditions.

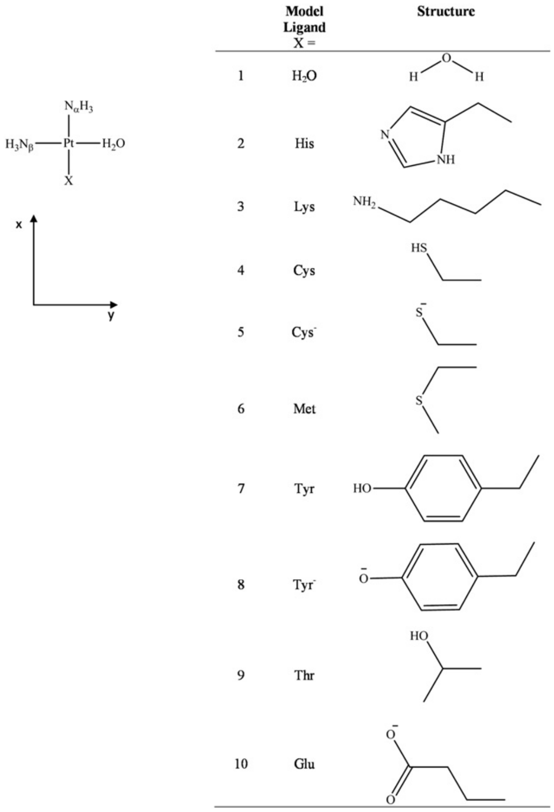
Theoretical works on the molecular mechanism of cisplatin are profuse. Most are centred on its interaction with DNA [24–26]. Only a reduced, but enlightening, amount of work has been dedicated to its binding to proteins. All of them have been carried out using pure quantum mechanical approaches on models that mimic the first coordination sphere of the metal. These studies have shed light on the molecular factors that allow the drug to reach its final target [27], explored the mechanism of hydration of the drug under physiological media [28] considering pH and chloride concentration [29] or provided with rationale of the kinetics over thermodynamics grounds for the selectivity of the drugs towards sulphur containing amino acids (cysteine and methionine) against N- and O-amino acid donors [30,31]. Regarding tri-coordinated complexes, calculations have also discarded them as intermediates in water exchange processes in tetraaquo paladium(II) and platinum(II) complexes [32], the activation of cisplatinum in water [33] and identified steric contributions in their stabilization against dimeric or square planar forms [34]. To our knowledge though, attempts to understand their role in the interaction of cisplatin with proteins have not been made.

Here, we report an analysis of cisplatin monoadducts with protein models in trigonal planar configurations. Calculations were carried out with pure quantum mechanical (QM) calculations on cluster models as well as with hybrid quantum mechanics/molecular mechanics (QM/MM) on the cisplatinated hen egg white lysozyme. After characterizing the geometries of $[Pt(II)(NH_3)_2X]^{+/2+}$ adducts (where X stands for water and 8 different amino acid models and water) using pure density functional theory (DFT) calculations, we analysed three different variables that could be crucial in defining the mode of binding of the drug with proteins in a tri-coordinated mode: 1) the energy of the Pt–X bond, 2) the cost of the transition from square planar monoaquo $[Pt(II)(H_2O)(NH_3)_2X]^{+/2+}$ to its trigonal counterpart and 3) the impact of the bulk solvent on this transition. The role of the anisotropic environment created by the protein on the metal and the accessibility of discrete water molecules in its neighbourhood were finally investigated with the ONIOM (DFT:AMBER) scheme on the X-ray system.

## 2. Methods

### 2.1. Cluster models

First, calculations were carried out on structures derived from *cis*-diamminediaquo-platinum(II) interacting with proteins as well as the initial diaquo complex. They correspond to the trigonal planar $[Pt(II)(NH_3)_2X]^{+/2+}$and square planar $[Pt(II)(H_2O)(NH_3)_2X]^{+/2+}$systems where X is water (**1**), histidine (**2**), lysine (**3**), cysteine (**4**), cysteinate (**5**), methionine (**6**), tyrosine (**7**), tyrosinate (**8**), threonine (**9**) and glutamate (**10**) (Scheme 1). Geometry optimizations and Gibbs energy calculations were performed with the Gaussian09 package [35] at the DFT B3LYP level [36,37]. Calculations were carried out with a mixed basis set combining the effective core potential LANL2DZ [38] for Pt and the 6-31+G* [39] for the main group elements. Energy minimizations were performed in gas phase. These energies were improved by means of single point calculations increasing the size of the basis set for the main group elements to 6-311+G* [40]. In the case of the metal, increasing the size of the effective core potential to a LANL2TZ did not result in any significant changes of the energy. Frequency calculations were performed in order to calculate the Gibbs energy at standard conditions (at 298.15 K and 1 atm of pressure).



**Scheme 1.** Labelling and two dimensional representations of the model systems $[Pt(NH_3)_2(H_2O)X]^{+/+2}$ and $[Pt(NH_3)_2X]^{+/+2}$ used in this study with X an amino acid model.

Solvation effects were computed on the optimized structures with the conductor-like screening model (COSMO [41–43]) as implemented in ADF 12.0 package [44]. Defining the exact dielectric constant for a proteic environment is a complex process [45,46] and as a general consensus four dielectric constants (ε) were used to simulate different media: water (ε = 78.4), blood plasma (ε = 58.0) [47], and protein environment with high (i.e. surface or relatively accessible sites) (ε = 9.0) or low (i.e. deep anchored in the protein matrix) (ε = 2.0) dielectric constants [48–50].

Two corrections were further applied. One corresponds to the change of Gibbs energies from standard state of gas (1 atm) to liquid (M). Assuming an ideal gas behaviour, this correction is:

$$\Delta G = \Delta nRT1n(24.46)$$

where Δn refers to the change of the number of species in the reaction. The second corresponds to the Gibbs energy change associated to *n* moles of water from 1 to 55.34 M (its concentration in pure liquid water) [51–53]:

$$\Delta G = nRT1n(55.34).$$

To study the strength of the coordination bond between the metal and individual amino acids energy decomposition analysis (EDA) was performed as implemented in the ADF 12.0 package [54–56]. EDA calculates the bond energy ($-B_{\Delta E}$) among two fragments A and B following the next formula:

$$-B_{\Delta E} = \Delta E_{\text{int}} + \Delta E_{prep}$$

where $\Delta E_{int}$ corresponds to the sum of the classical electrostatic interaction between both fragments ($\Delta V_{elst}$), the Pauli repulsion ($\Delta E_{pauli}$) and the associated orbital interaction energy ($\Delta E_{oi}$)[44]:

$$\Delta E_{int} = \Delta V_{elst} + \Delta E_{pauli} + \Delta E_{oi}.$$

Additionally, the term $\Delta E_{prep}$ is the total energy that is necessary to distort and electronically excite all fragments to the state which A and B are in the molecule AB [57].

### 2.2. QM/MM calculations

In the last part of this work, QM/MM calculations were undertaken using the ONIOM [58] scheme implemented in Gaussian09 on the cisplatinated hen egg white lysozyme adduct [11]. The QM part of the system was modelled with the B3LYP functional while the MM part was treated with the AMBER force field [59]. The total number of atoms of the system is 2548 with 32 described in the QM part and the rest in the MM region. 160 water molecules were added to solvate the entire protein. Atoms in a sphere of 5 Å around the metal were allowed full flexibility while the rest (1825 atoms) was frozen. To avoid artefacts between QM and MM representation of water molecules, those in the nearby region of the drug (4 in total) were computed at the quantum mechanical level. The boundary between QM and MM was described with the link-atom scheme.

## 3. Results and discussion

### 3.1. Trigonal planar geometries of $[Pt(II)(NH_3)_2X]^{+/2+}$ models

This part of the study aims at characterizing the geometry of trigonal planar geometries of cisplatin adducts with amino acids. Geometry optimizations of $[Pt(II)(NH_3)_2\mathbf{X}]^{+/2+}$ with **X** being models of amino acids (water (**1**), histidine (**2**), lysine (**3**), cysteine (**4**), cysteinate (**5**), methionine (**6**), tyrosine (**7**), tyrosinate (**8**), threonine (**9**) and glutamate (**10**)) were carried out with Gaussian 09 considering usual standard protonation states or those, less frequent, that have been attributed to metal binding (i.e. tyrosinate) (Scheme 1).

For each $[Pt(II)(NH_3)_2\mathbf{X}]^{+/2+}$ systems, reasonable metal–ligand bond lengths were obtained with Pt–NH$_3$ distances between 2.0 and 2.2 Å and Pt–X distances from 1.9 to 2.4 Å (Table 1). For the latest, shorter distances were observed for nitrogen and oxygen donors over sulphur containing ones and negatively charged residues over their neutral counterparts. Bond lengths of the first coordination sphere of the metal were found to be similar to those obtained for square planar geometries (with differences lower than 0.1 Å, see Table S1 in Supporting information) and suggest that the removal of the water molecule from the system has little structural impact on the M–L bond. Only the cysteinate containing complexes displays a noticeable difference of 0.16 Å between its tri-coordinated and tetra-coordinated configurations.

The $N_\alpha$–Pt–X angle ranges from 137.3 to 177.4° (Table 1) that shows that most of calculated complexes are T-shaped in nature (Table 1 and Fig. 1). For water (**1**), histidine (**2**), lysine (**3**), cysteine (**4**), tyrosine (**7**), and threonine (**9**) deviations from a perfect T-shaped geometry do not exceed 13°. A very good agreement is obtained between complex **2** and the experimental structure of the histidine bound cisplatinated hen egg white lysozyme [11]. The deviation is slightly more pronounced for the tri-coordinated complexes with methionine (**6**) and tyrosinate (**8**) ($N_\alpha$–Pt–X of 159.5 and 158.7° respectively) reaching to an almost ideal trigonal planar situation for the cysteinate (**5**) ($N_\alpha$–Pt–X of 137.3°).

Two minima were characterized for $[Pt(II)(NH_3)_2(\mathbf{Glu})]^+$ (**10** and **10′**) that mainly differ in the interaction of the metal with the carboxylic group. In **10**, the amino acid binds the metal in a monodentate fashion with the remaining carboxyl oxygen hydrogen bonded to the β amino group (a hydrogen bond already present in the square planar geometry as shown in Fig. S1). In **10′**, the carboxylate acts as a bidentate ligand and leads to a slightly distorted square planar geometry (Nα –Pt–X = 167.1°). This complex was computed to be ca. 25 kcal/mol more stable than its monodentate counterpart. To our knowledge, these distinctive mono- and bidentate binding modes of the glutamate with cisplatin have not been reported so far. It might have though some implications in the distinctive binding profiles experimentally observed for cisplatin interaction with oxygen, nitrogen and sulphur donors [7,8,31] and should be further investigated.

Until now, our results indicate that for most amino acids, including histidine, stable trigonal planar geometries with T-shaped configuration exist as minima on the gas phase potential energy surface. With the exception of the monodentate complex $[Pt(II)(NH_3)_2(\mathbf{Glu})]^+$, no interaction of the second coordination sphere is required for the formation of this tri-coordinated species. Only one Y-shaped structure has been characterized and corresponds to the binding of the deprotonated cysteine to the drug.

### 3.2. Strength of the cisplatin-X bond

The M–X bond energy was studied for complexes **1** to **10′** (Scheme 2). Calculations were carried out using the energy decomposition analysis implemented in ADF 12.0 (Table 2). This scheme provides with the total bond energy as well as discrete energetic terms associated with chemically meaningful concepts such as orbital interactions, Pauli repulsion and electrostatic energies.

Regarding the total bond energies, lower values were obtained for cysteinate, tyrosinate and glutamate (about −300 kcal/mol in average for **5**, **8** and **10** respectively) followed by histidine, lysine and methionine (in a range of −120 to −100 kcal/mol for **2**, **3** and **6**). Finally, threonine, cysteine and tyrosine present the weaker bonds though still substantially stronger than the one obtained for a water molecule (−78.4, −89.4 and −81.2 respectively in front of −59.2 kcal/mol). Although the EDA analysis does not point towards a unique energetic component to account for the differences observed, it first shows that electrostatic interactions are dominant in this trend

**Table 1**
Relevant geometric parameters for tri-coordinated Pt(II) complexes **1** to **10′**. Distances are in Å and angles in degrees.

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

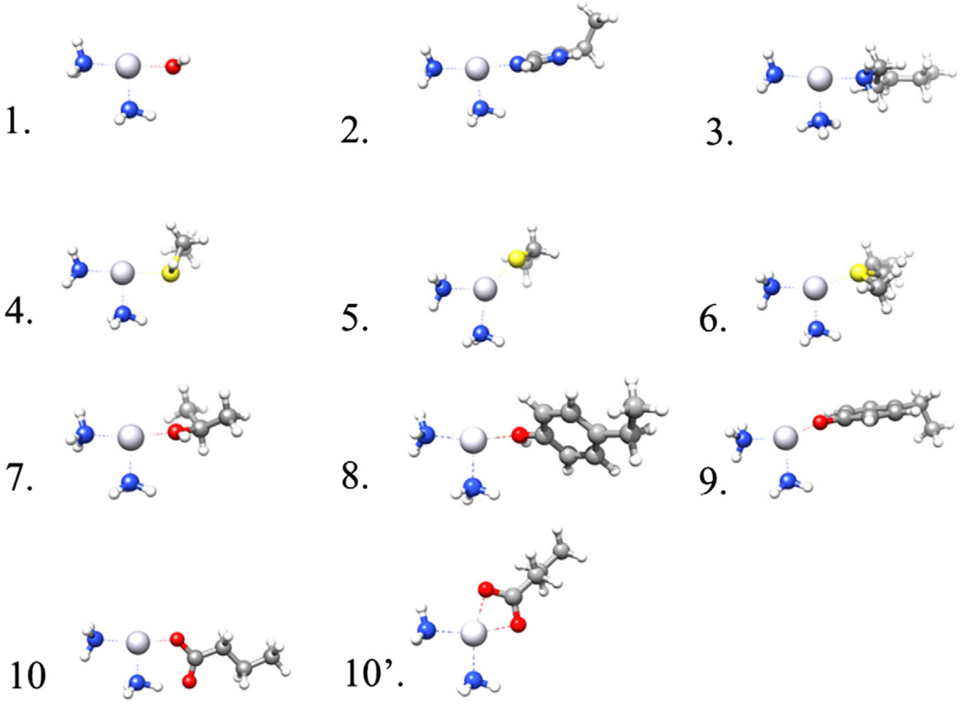


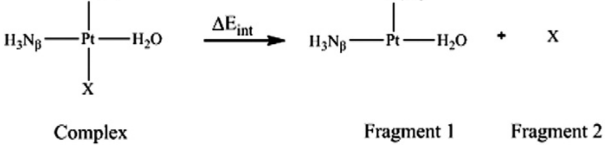
**Fig. 1.** Representative geometries of optimized $[Pt(NH_3)_2X]^{+/+2}$ systems. Atom colouring code: blue/nitrogen, red/oxygen, white, hydrogen, grey/carbon, light grey/platinum, yellow/sulphur.

being up to 4 times higher for negatively charged residues than for neutral ligands (Table 2). For neutral species, smaller bond energies for threonine, tyrosine and cysteine versus methionine, lysine and histidine are due to weaker electrostatic and/or orbital interactions.

Not surprisingly, the bond energy associated to the glutamate containing systems **10** and **10′** are the greatest of the series (ca. −300 kcal/mol). Moreover, its bidentate mode is favoured by 30 kcal/mol over the monodentate one. The EDA shows that this is related to a better orbital interaction in the former. This is mainly due to the additional Pt–O bond formed in **10** with respect to **10′**. Interestingly, the Pt–cysteinate bond energy in **5** is of the same order of magnitude than the Pt–glutamate one in **10′** despite presenting monodentate coordination mode. Natural bond orbital analysis shows that a σπ bonding between the sulphur atom and the metal occurs in this system. The additional π orbital results from the combination of the $3p_x$ of the sulphur and a mixture of $7p_x$, $6d_{xy}$ and $6d_{xz}$ of the metal. This double bond is responsible for the shorter M–X distance observed in **5** as well as the deviation of the system towards a Y-shaped geometry.

When comparing trigonal and square planar geometries, one can see that the overall trend of bond energies between amino acids is preserved (Table S1 and Lau et al. [50]). Nonetheless, bond energies are systematically stronger for trigonal than square planar systems from about 15 (**1**, **2**, **3**), 20 (**4**, **10**, **6**) or even 40 (**5**) kcal/mol. The lack of a fourth ligand is naturally compensated by stronger M–X interactions.

This part of the study only focuses on the nature of the metal–heteroatom bond and does not consider other potentially relevant variables of the binding process (including entropic contributions, solvation or anisotropic effects of a protein matrix). This has been the objective of the following sections.



**Scheme 2.** Schematic representation of the meaning of EDA calculations taking square planar complex as an example.

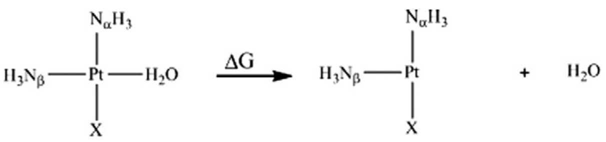
### 3.3. Energetic cost of the transition from square to trigonal planar geometries

Taking aquocisplatin as a study system, the transition from square planar $[Pt(II)(NH_3)_2(H_2O)\mathbf{X}]^{+/2+}$ to trigonal planar geometries $[Pt(II)(NH_3)_2\mathbf{X}]^{+/2+}$ consists in the removal of one water molecule from the first coordination sphere of the metal (Scheme 3). In order to better evaluate the contribution of tri-coordinated geometries in the binding mechanism of cisplatin to proteins, the Gibbs energies of dehydration of the monoaquo square planar adducts, ΔG, were calculated. Since solvation effects could have impact on these results, calculations were carried out using continuum solvent models. We used dielectric constants to simulate pure water ($\varepsilon = 78.8$), plasma ($\varepsilon = 58$), relatively solvent exposed region of proteins or protein with relatively high permissivity ($\varepsilon = 9$) or regions deeply anchored in the proteic matrix environment ($\varepsilon = 2$) (Table 3).

For all the systems considered in this study, the Gibbs energies of dehydration were found to be positive. The removal of the aqueous ligand from the square planar complexes is predicted not to be spontaneous. Different orders of magnitude of ΔG are observed though. For continuum models that mimic strong to limited solvent exposed environments (dielectric constant from 78.8 to 9) three subsets can be defined. The subset 1 contains systems **8** (tyrosinate), **5** (cysteinate) and **7** (tyrosine) for which ΔG varies from about 8 to 14.5 kcal/mol. The subset 2 contains systems **1**–**4** (containing a second water molecule, histidine, lysine and a cysteine), **6** (methionine) and **9** (threonine) with

**Table 2**
Gas-phase binding energies of amino acid models to cisplatin in trigonal and square planar geometries obtained from energy decomposition analysis [51,52] as implemented in ADF 12 [44]. Energies are reported in kcal/mol.

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 10′ |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | $H_2O$ | His | Lys | Cys | Cys- | Met | Tyr | Tyr- | Thr | Glu | Glu |
| $-B_{\Delta E}$ (trigonal) | −59.2 | −119.8 | −99.2 | −89.4 | −331.6 | −102.5 | −81.2 | −304.5 | −78.4 | −306.0 | −335.6 |
| Electrostatic energy | −73.5 | −164.6 | −134.3 | −103.1 | −414.1 | −120.2 | −72.6 | −304.1 | −95.4 | −334.8 | −368.6 |
| Pauli repulsion | 58.2 | 137.3 | 116.4 | 106.3 | 249.5 | 126.0 | 64.1 | 147.8 | 80.5 | 149.1 | 174.2 |
| Orbital interaction | −45.1 | −98.0 | −85.7 | −95.7 | −175.3 | −114.4 | −78.9 | −157.1 | −70.3 | −124.6 | −147.7 |
| ΔE preparation | 1.13 | 5.4 | 4.5 | 3.1 | 8.3 | 6.0 | 6.1 | 7.9 | 6.74 | 4.3 | 5.6 |
| $-B_{\Delta E}$ (square planar) | −49.5 | −103.2 | −84.6 | −76.1 | −292.2 | −83.6 | −64.5 | −273.1 | −64.7 | −276.6 | −276.6 |
| Electrostatic energy | −67.0 | −151.2 | −130.6 | −98.7 | −360.6 | −108.6 | −73.1 | −286.5 | −82.6 | −316.6 | −316.6 |
| Pauli repulsion | 55.3 | 126.9 | 117.8 | 100.9 | 182.2 | 108.1 | 66.5 | 120.4 | 69.7 | 133.2 | 133.2 |
| Orbital interaction | −39.1 | −83.24 | −78.5 | −81.64 | −125.5 | −90.8 | −64.0 | −119.3 | −56.3 | −104.8 | −104.8 |
| ΔE preparation | 1.4 | 4.37 | 6.7 | 3.3 | 11.7 | 7.7 | 6.1 | 12.4 | 4.5 | 11.6 | 11.6 |



**Scheme 3.** Schematic representation of the dehydration mechanism considered in this study.

values from about 17.0 to 21.5 kcal/mol. The latest subset only accounts for the glutamate bound cisplatin in monodentate (**10**) and bidentate (**10′**) modes.

The lower endoergonicities of the dehydration process of subset 1 over subset 2 are due to different factors. In the cysteinate containing system (**5**), the presence of an additional bond is the major component of the small ΔG value as it leads to a smaller gap between trigonal and square planar complexes. For tyrosine containing complexes, small ΔG values are obtained because of weaker desolvation penalties. This result is to consider with caution since **7** and **8** are substantially larger than the other systems of the series. The comparison of the solvation energy of systems with different charges and size still remains a challenge in computational chemistry and represents a limitation of cluster models [60].

Interestingly, the Gibbs energy of dehydration varies drastically between **10** and **10′** (−55 and 16 kcal/mol respectively). In fact, the removal of the water to form a bidentate adduct is the only spontaneous process observed in this study. This suggests that this binding mode could be substantially favoured in the binding of cisplatin with a protein.

Whatever the dielectric constant, the general trend of dehydration energies is maintained. Only inversion between the dehydration of $[Pt(II)(NH_3)_2(\mathbf{Cys\text{-}})]^+$ **(5)** and $[Pt(II)(NH_3)_2(\mathbf{Tyr\text{-}})]^+$ **(8)** as the most favourable processes is observed. Moreover, the variation of the dielectric constant from 78.8 to 9 does not alter the Gibbs energy more than 2 kcal/mol for each complexes. However, when reaching $\varepsilon = 2$ ΔG increases from 6 to 12 kcal/mol. Apparently, the removal of fourth ligand is less likely to occur in highly hydrophobic media (like a deep protein core) than in more solvated environments (i.e. the bulk solvent, a protein with a low dielectric constant or at the interface between the protein and the water media). This is consistent with the experimental T-shaped structures where the drug interacts with at the surface of the hen egg white lysozyme. Surprisingly, the dehydration energies of $[Pt(II)(NH_3)_2\mathbf{(Cys\text{-})}]^+$ (**5**) and $[Pt(II)(NH_3)_2\mathbf{Glu}]^+$ **(10′)** are not affected by the change of the continuum conditions.

The comparison between the Gibbs energy of dehydration of complexes **2** to **10′** with the one calculated for the initial diaquo system (**1**) allows one to study how the amino acids affect the binding of the fourth ligand. For complexes of subset 2, the differences range from 1 to 3 kcal/mol which represents a weak alteration of the Pt–$OH_2$ interaction. For those of the subset 1 differences are from 8 to 10 kcal/mol. In these systems, the binding of cisplatin to the protein should noticeably weaken the interaction with the remaining water molecule. Finally, in the formation of the bidentate adduct with glutamate (**10′**), the impact on the Pt–$OH_2$ raises up to 86 kcal/mol. These tendencies remain mostly unchanged for dielectric constants from 78.8 to 9 but are more pronounced for $\varepsilon = 2$. However, for the cysteinate containing system **5** the Pt(II)–$OH_2$ interaction is substantially weakened independently of dielectric properties of the medium. These results suggest that such tri-coordinated adducts may equally occur in different media.

The results of this part of the study provide some clues on the cost of the removal of the water molecule from square planar monoadducts of cisplatin with amino acids. They remain though partial regarding the hen egg white lysozyme adduct. The cluster models show that histidine can form a T-shaped adduct with a good Pt–N bond energy and slightly increases the weakness of the leaving water. However, the dehydration

**Table 3**
Energy terms associated to the dehydration of cis-$[Pt(II)(NH_3)_2(H_2O)X]^{+/2+}$. $\Delta E_0$ is the difference in potential energy, ΔZPE the zero point energy correction, $\Delta E_{TOT}$ the sum of the two previous term, $\Delta G_{solv}$ the solvation energy calculated with the COSMO model and ΔG the total Gibbs energy of dehydration. Energies are reported in kcal/mol.

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 10′ |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | $H_2O$ | His | Lys | Cys | Cys- | Met | Tyr | Tyr- | Thr | Glu | Glu |
| $\Delta E_0$ | 52.7 | 45.3 | 47.7 | 47.3 | 25.5 | 44.3 | 27.9 | 34.0 | 48.4 | 12.2 | 37.8 |
| ΔZPE | −3.2 | −3.5 | −3.3 | −3.5 | −3.3 | −3.4 | −3.5 | −3.8 | −3.5 | −2.6 | −3.6 |
| $\Delta E_{TOT}$ | −2.8 | −3.0 | −3.0 | −3.0 | −2.8 | −3.0 | −3.1 | −3.0 | −3.1 | −2.3 | −3.3 |
| $\Delta Gsolv(\varepsilon=78.8)$ | −24.1 | −18.2 | −20.6 | −19.1 | −5.1 | −14.3 | −2.6 | −14.8 | −19.6 | −12.7 | −3.7 |
| $\Delta Gsolv(\varepsilon=58)$ | −22.8 | −17.9 | −20.4 | −18.9 | −4.8 | −14.1 | −2.4 | −13.8 | −19.4 | −12.6 | −3.4 |
| $\Delta Gsolv(\varepsilon=9)$ | −20.3 | −15.7 | −18.0 | −16.6 | −4.1 | −12.2 | −2.3 | −10.7 | −17.2 | −10.9 | −2.9 |
| $\Delta Gsolv(\varepsilon=2)$ | −10.9 | −7.7 | −9.1 | −8.4 | −2.1 | −5.8 | −1.6 | −3.2 | −8.7 | −5.0 | −1.4 |
| $\Delta G(\varepsilon=78.8)$ | 20.0 | 17.3 | 18.1 | 18.9 | 11.2 | 20.3 | 12.6 | 8.2 | 19.7 | 16.0 | −57.1 |
| $\Delta G(\varepsilon=58)$ | 21.3 | 17.6 | 18.2 | 19.1 | 11.5 | 20.5 | 13.1 | 9.3 | 19.6 | 16.1 | −56.7 |
| $\Delta G(\varepsilon=9)$ | 21.5 | 17.6 | 18.4 | 19.1 | 9.8 | 20.1 | 14.4 | 10.0 | 19.7 | 15.5 | −58.5 |
| $\Delta G(\varepsilon=2)$ | 30.9 | 25.5 | 27.3 | 27.3 | 11.9 | 26.5 | 26.4 | 17.5 | 28.2 | 21.4 | −56.9 |

**Table 4**
Structural parameters of the experimental and calculated cisplatinated adduct of hen lysozyme. Distances are in Å and angles in degrees.

| | Experimental. | Calculated. |
|---|---|---|
| Pt–$N_\alpha$ | 2.02 | 2.02 |
| Pt–$N_\beta$ | 2.02 | 2.02 |
| Pt–$OH_2$ | – | 2.13 |
| Pt–$N_\delta$(His15) | 2.13 | 2.08 |
| $N_\alpha$–Pt–$N_\beta$ | 89.7 | 89.9 |
| $N_\beta$–Pt–$OH_2$ | – | 173.7 |
| $N_\beta$–Pt–$N_\delta$(His15) | 94.2 | 91.5 |

process is predicted to be ca. 17 kcal/mol and casts doubts on the possibility of this process to occur spontaneously. How the protein environment could impact on the stabilization of the tri-coordinated adduct was therefore further analysed.

### 3.4. QM/MM calculations on cisplatin–hen egg white lysozyme adduct

To test the possibility of the protein environment to stabilize tri-coordinated geometries, QM/MM calculations with the ONIOM scheme were carried out on the experimental structure of the cisplatinated hen egg white lysozyme. Because of the number of water molecules present in the crystal structure is limited in the neighbourhood of the metal, the system has been previously solvated with the AmberTools[61] implemented in UCSF chimera [62]. To avoid possible inconsistencies between the quantum mechanical region of the system and the molecular mechanical one, all the water molecules at 5 Å of the metal were included in the QM region.

The QM/MM optimized structure leads to an excellent agreement with its experimental counterpart. Both the amino groups of the drug and the histidine remain in similar positions. The structural differences between calculated and experimental structures do not overcome 0.05 A and 3° for the first coordination sphere of the metal (Table 4). However, their coordination numbers are different since the optimized structure presents a clear square planar geometry and the experimental structure a T-shaped one. During the optimization process, one of the solvent molecules in the quantum mechanical part of the QM/MM partition reaches the metal and binds on its vacant site (Fig. 2). Despite multiple attempts, no true minimum structure was characterized with a T-shaped geometry. This result conditioned that no dehydration energy could be calculated.

These calculations show that no local constraints, steric or electrostatic, in the vicinity of the metallodrug prevent the binding of an additional water molecule to the metal. As the authors of the crystal structure suggest, the experimental tri-coordinated geometry likely results from technical difficulties in identifying the water molecules in the surrounding of the drug. The origin of the lack of density at the fourth site is not quantifiable from our calculations.

## 4. Conclusion

This work is based on the reasonable doubts that recent crystallographic data offer on the stability of tri-coordinated cisplatin–protein adducts. Calculations on cluster models show that trigonal planar geometries exist on the potential energy surface in gas phase. Most are T-shaped with little structural differences with respect to their corresponding square planar complexes. Stronger bond energies between the metal and the protein models are predicted to compensate the loss of the fourth ligand. However, for all amino acids, the Gibbs energies of dehydration are systematically endergonic. The calculations suggest that the formation of trigonal planar complexes is not spontaneous and is increasingly prohibitive as the drug deepens in the protein scaffold. They also highlight interesting features on the binding of cisplatin to cysteinate, which apparently leads to low energy Y-shaped geometries, and to glutamate, which provides with a strong bidentate adduct. Regarding the crystal structure of cisplatinated hen egg white lysozyme, this study shows that a T-shaped geometry of the cisplatin–histidine adduct exists on the potential energy surface in gas phase. Its dehydration cost is though, quite prohibitive and QM/MM calculations show that no interaction of the drug with the nearby atoms of the proteins prevents the formation of a square planar adduct. In agreement with the authors

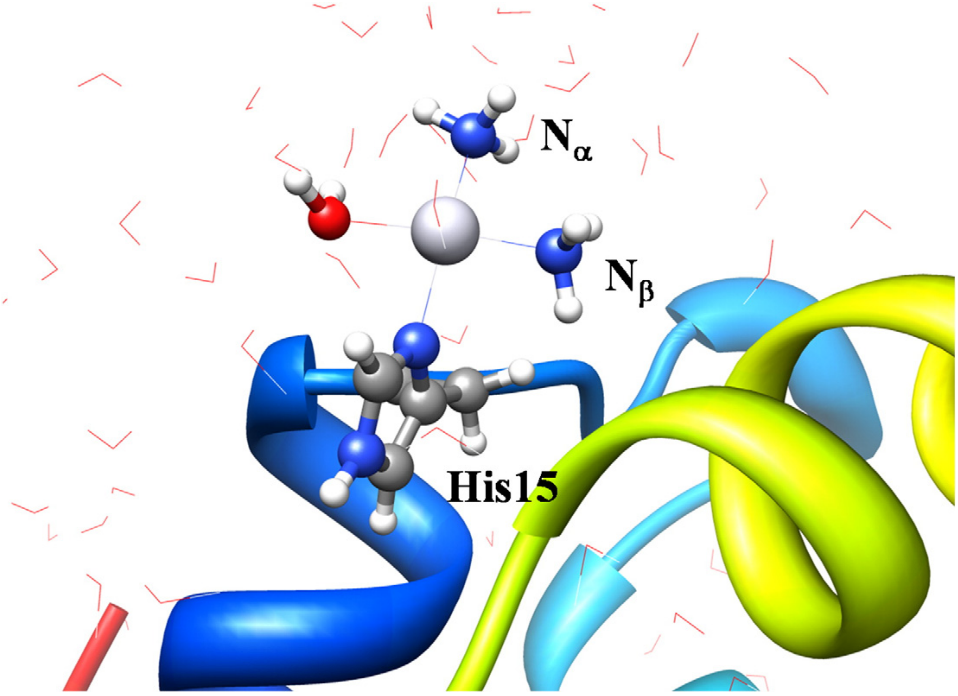


**Fig. 2.** Three dimensional representations of the ONIOM (B3LYP/AMBER) optimized geometry of the histidine bound cisplatin hen egg white lysozyme. [11].

of the crystal structure, the T-shaped configuration observed in the experimental system is likely to be due to crystallographic limitations in identifying the fourth ligand of the metal. The true physiological adduct is most likely square planar in this case.

Taken together this study provides insights on the nature of cisplatinated adducts with proteins and the energetic requirement necessary to stabilize tri-coordinated complexes. Nonetheless, only the diaquo form of the drug has been considered in this study and other species (i.e. dichloro ones) have not been considered. Dynamical considerations could also help in quantifying the life time of tri-coordinated configurations of Pt(II) systems and its transition to square planar ones. Expanding the number of cisplatinated precursors as well as performing quantum based molecular dynamics calculations should be taken into account in future works.

## Acknowledgements

We are particularly thankful to the Spanish “Ministerio de Economía y Competitividad” for financial support through projects CTQ2011-23336, CCTQ 2010-16959 and ORFEO Consolider- Ingenio 2010 Programme (Grant CSD2007-00006), the Generalitat de Catalunya through project 2009SGR68 and E.O.-C. Thanks to the Universitat Autònoma de Barcelona for a scholarship (UAB-PIF). E. O.-C. and J.-D. M. are thankful to Pr. Luis Rodríguez Santiago for its enriching discussions.

## Appendix A. Supplementary data

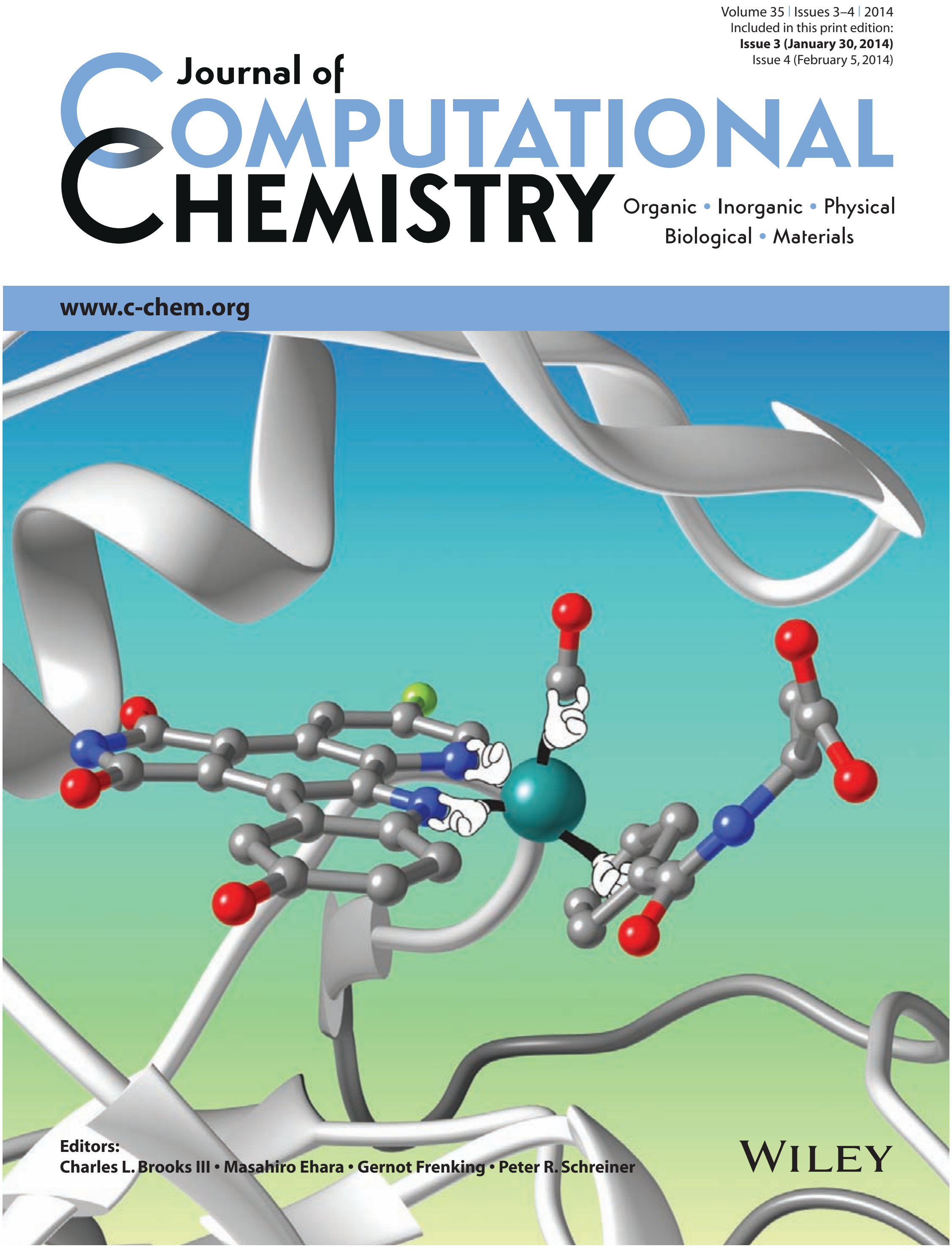
Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.jinorgbio.2012.09.020.

## References

[1] J.C. Dabrowiak, Metals in Medicine, Wiley, 2009.
[2] L. Galluzzi, L. Senovilla, I. Vitale, J. Michels, I. Martins, O. Kepp, M. Castedo, G. Kroemer, Oncogene 31 (2012) 1869–1883.
[3] R. Mandal, R. Kalke, X.-F. Li, Chem. Res. Toxicol. 17 (2004) 1391–1397.
[4] I. Khalaila, C.S. Allardyce, C.S. Verma, P.J. Dyson, ChemBioChem 6 (2005) 1788–1795.
[5] I. Ivanov, J. Christodoulou, J. Parkinson, K.J. Barnham, A. Tucker, J. Woodrow, P.J. Sadler, J. Biol. Chem. 273 (1998) 14721–14730.
[6] J.F. Neault, H. Tajmir-Riahi, Biochim. Biophys. Acta Bioenerg. 1384 (1998) 153–159.
[7] J. Will, D. Wolters, W.S. Sheldrick, ChemMedChem 3 (2008) 1696–1707.
[8] T. Peleg-Shulman, Y. Najajreh, D. Gibson, J. Inorg. Biochem. 91 (2002) 306–311.
[9] A.V. Karotki, M. Vasák, J. Biol. Inorg. Chem. 14 (2009) 1129–1138.
[10] F. Arnesano, S. Scintilla, G. Natile, Angew. Chem. Int. Ed. 119 (2007) 9220–9222.
[11] A. Casini, G. Mastrobuoni, C. Temperini, C. Gabbiani, S. Francese, G. Moneti, C.T. Supuran, A. Scozzafava, L. Messori. (Cambridge, U. K.) Chem. Commun. (2007) 156–158.
[12] V. Calderone, A. Casini, S. Mangani, L. Messori, P.L. Orioli, Angew. Chem. Int. Ed. 45 (2006) 1267–1269.
[13] U. Frey, L. Helm, A.E. Merbach, R. Romeo, J. Am. Chem. Soc. 111 (1989) 8161–8165.
[14] J.A. Casares, P. Espinet, G. Salas, Chem. Eur. J. (2002) 4843–4853.
[15] C.D. Hubbard, R. van Eldik, J. Coord. Chem. 60 (2007) 1–51.
[16] A. Doppiu, E. Herdtweck, Synthesis 8 (2003) 105–108.
[17] O. Rivada-Wheelaghan, M.A. Ortuño, J. Díez, A. Lledós, S. Conejero, Angew. Chem. Int. Ed. 51 (2012) 3936–3939.
[18] W. Baratta, C. Mealli, E. Herdtweck, A. Ienco, S. Mason, P. Rigo, J. Am. Chem. Soc. 126 (2004) 5549–5562.
[19] H. Braunschweig, K. Radacki, K. Uttinger, Chem. Eur. J. 14 (2008) 7858–7866.
[20] S.H. Crosby, G.J. Clarkson, J.P. Rourke, J. Am. Chem. Soc. 131 (2009) 14142–14143.
[21] M.J. Ingleson, M.F. Mahon, A.S. Weller, Chem. Commun. (2004) 2398–2399.
[22] O. Rivada-Wheelaghan, B. Donnadieu, C. Maya, S. Conejero, Chem. Eur. J. 16 (2010) 10323–10326.
[23] G. Berthon-Gelloz, B. de Bruin, B. Tinant, I.E. Markó, Angew. Chem. Int. Ed. 48 (2009) 3161–3164.
[24] P. Carloni, M. Sprik, W. Andreoni, J. Phys. Chem. B 104 (2000) 823–835.
[25] K. Spiegel, U. Rothlisberger, P. Carloni, J. Phys. Chem. B 108 (2004) 2699–2707.
[26] K. Gkionis, J.A. Platts, J. Biol. Inorg. Chem. 14 (2009) 1165–1174.
[27] D.V. Deubel, J. Am. Chem. Soc. 124 (2002) 5834–5842.
[28] J.V. Burda, M. Zeizinger, J. Leszczynski, J. Comput. Chem. 26 (2005) 907–914.
[29] T. Zimmermann, J. Leszczynski, J.V. Burda, J. Mol. Model. 17 (2011) 2385–2393.
[30] D.V. Deubel, J. Am. Chem. Soc. 126 (2004) 5999–6004.
[31] J. Reedijk, Proc. Natl. Acad. Sci. U.S.A. 100 (2003) 3611–3616.
[32] R.J. Deeth, L.I. Elding, Inorg. Chem. 2 (1996) 5019–5026.
[33] J.K.-C. Lau, B. Ensing, Phys. Chem. Chem. Phys. 12 (2010) 10348–10355.
[34] S. Moncho, G. Ujaque, A. Lledós, P. Espinet, Chem. Eur. J. 14 (2008) 8986–8994.
[35] M.J. T., G. W. Schlegel, H. B. Scuseria, G. E. Robb, M. A. Cheeseman, J. R. Scalmani, G.; Barone, V. Mennucci, B. Petersson, G. A. Nakatsuji, H. Caricato, M. Li, X. Hratchian, H. P. Izmaylov, A. F. Bloino, J. Zheng, G. Sonnenberg, J. Frisch, Gaussian, Inc., Wallingford CT, (2009).
[36] A.D. Becke, J. Chem. Phys. 98 (1993) 1372–1377.
[37] C. Lee, W. Yang, R.G. Parr, Phys, Phys. Rev. B Condens. Matter Mater. Phys. 37 (1988) 785–789.
[38] P.J. Hay, W.R. Wadt, J. Chem. Phys. 82 (1985) 270–283.
[39] G. Fogarasi, X. Zhou, P.W. Taylor, P. Pulay, J. Am. Chem. Soc. 114 (1992) 8191–8201.
[40] R. Krishnan, J.S. Binkley, R. Seeger, J.A. Pople, J. Chem. Phys. 72 (1980) 5639.
[41] A. Klamt, G. Schuurmann, J. Chem. Soc., Perkin Trans. 2 2 (1993) 799–805.
[42] A. Klamt, J. Phys. Chem. 99 (1995) 2224–2235.
[43] J.L. Pascual-Ahuir, E. Silla, J. Tomasi, R. Bonaccorsi, J. Comput. Chem. 8 (1987) 778–787.
[44] G. Te Velde, F.M. Bickelhaupt, E.J. Baerends, C. Fonseca Guerra, S.J.A. van Gisbergen, J.G. Snijders, T. Ziegler, J. Comput. Chem. 22 (2001) 931–967.
[45] C.N. Schutz, A. Warshel, Genetics 417 (2001) 400–417.
[46] T. Simonson, C.L. Brooks, J. Am. Chem. Soc. 7863 (1996) 8452–8458.
[47] H.F. Cook, J. Appl. Phys. 3 (1952) 249–255.
[48] T. Dudev, C. Lim, J. Am. Chem. Soc. (2000) 11146–11153.
[49] T. Dudev, C. Lim, J. Phys. Chem. B 104 (2000) 3692–3694.
[50] J.K.-C. Lau, D.V. Deubel, Chem. Eur. J. 11 (2005) 2849–2855.
[51] J.R. Pliego, Chem. Phys. Lett. 367 (2003) 145–149.
[52] V.S. Bryantsev, M.S. Diallo, W.A. Goddard, J. Phys. Chem. B 112 (2008) 9709–9719.
[53] J. Alí-Torres, L. Rodríguez-Santiago, M. Sodupe, Phys. Chem. Chem. Phys. 13 (2011) 7852–7861.
[54] W.-J. van Zeist, F.M. Bickelhaupt, Org. Biomol. Chem. 8 (2010) 3118–3127.
[55] K. Morokuma, J. Chem. Phys. 55 (1971) 1236–1245.
[56] T. Ziegler, A. Rauk, Theor. Chim. Acta 46 (1977) 1–10.
[57] M. von Hopffgarten, G. Frenking, WIREs Comput. Mol. Sci. 2 (2012) 43–62.
[58] S. Dapprich, I. Komáromi, K.S. Byun, K. Morokuma, M.J. Frisch, J. Mol. Struct. 461–462 (1999) 1–21.
[59] W.D. Cornell, P. Cieplak, C.I. Bayly, I.R. Gould, K.M. Merz, D.M. Ferguson, D.C. Spellmeyer, T. Fox, J.W. Caldwell, P.A. Kollman, J. Am. Chem. Soc. 117 (1995) 5179–5197.
[60] O. Gutten, I. Beššeová, L. Rulíšek, J. Phys. Chem. A 115 (2011) 11394–11402.
[61] J. Wang, W. Wang, P.A. Kollman, D.A. Case, J. Mol. Graphics Modell. 25 (2006) 247–260.
[62] E.F. Pettersen, T.D. Goddard, C.C. Huang, G.S. Couch, D.M. Greenblatt, E.C. Meng, T.E. Ferrin, J. Comput. Chem. 25 (2004) 1605–1612.
[63] T. Ziegler, J. Am. Chem. Soc. 105 (1983) 7543–7549.

# 3 Assessing protein-ligand docking for the binding of organometallic compounds to proteins

Ortega-Carrasco, E.; Lledós, A.; Maréchal, J.D. *J. Comput. Chem*, **2014**. *35*, 192-198.



# Assessing Protein–Ligand Docking for the Binding of Organometallic Compounds to Proteins

Elisabeth Ortega-Carrasco, Agusti Lledós, and Jean-Didier Maréchal*

Organometallic compounds are increasingly used as molecular scaffolds in drug development projects; their structural and electronic properties offering novel opportunities in protein–ligand complementarities. Interestingly, while protein–ligand dockings have long become a spearhead in computer assisted drug design, no benchmarking nor optimization have been done for their use with organometallic compounds. Pursuing our efforts to model metal mediated recognition processes, we herein present a systematic study of the capabilities of the program GOLD to predict the interactions of protein with organometallic compounds. The study focuses on inert systems for which no alteration of the first coordination sphere of the metal occurs upon binding. Several scaffolds are used as test systems with different docking schemes and scoring functions. We conclude that ChemScore is the most robust scoring function with ASP and ChemPLP providing with good results too and GoldScore slightly underperforming. This study shows that current state-of-the-art protein-ligand docking techniques are reliable for the docking of inert organometallic compounds binding to protein. © 2013 Wiley Periodicals, Inc.

DOI: 10.1002/jcc.23472

## Introduction

Since the early success of cisplatin in anticancer therapies, organometallic compounds have been increasingly involved in drug design[1] and elements like ruthenium, rhodium, and gold are now part of the toolbox of metal based medicine.[2] Several properties of transition metals are interesting for medicinal applications but the most important one is the possibility to access structural arrangements that are unreachable in pure organic architectures (i.e., square planar or octahedral). These geometries substantially increase the chemical space available for molecular design and can optimize the complementarity of a drug with its biological target.

Metallodrugs can interact with their target in two major modes of action; by involving an exchange of ligands between the drug and the receptor, in which case the interaction is mainly driven the first coordination sphere of the metal, or by remaining completely inert upon binding, in which case the protein–ligand interaction takes only place throughout the second coordination sphere of the metal. In the recent years, a major focus of attention has been given to this latest kind of systems. Numerous groups have obtained encouraging results including Sadler and Melchart[3] in their investigations on ruthenium half-sandwich complexes or Metzler-Nolte and coworkers[4] in their work with Fe and Ru complexes as inhibitors of protein kinases. In this field, the work of Meggers and coworkers is particularly interesting as it both leads to highly specific candidates for kinases involved in cancer as well as provides with numerous X-ray structures of the resulting complexes.[5]

Protein–ligand dockings have long gained a central position in drug design projects. Despite being based on the simplistic estimation of the energy of interaction, their predictiveness is now well recognized. To date, the most commonly used docking software like GOLD,[6] Glide,[7] or FlexX[8] offer some optimized parameters to deal with metals. Nonetheless, these parameters have been developed for systems where the metal is part of the receptor. Little has been done to benchmark how they actually work when the metal is part of the ligand; a variable far more difficult to model because of the structural and electronic properties of the system. We recently showed that protein–ligand docking coupled with quantum mechanical/molecular mechanics (QM/MM) allows faithful predictions of the binding of organometallic compounds to protein for cases in which a ligand exchange occurs.[9] In these systems, an accurate calculation of the electronic properties of the transition metal is necessary to reproduce metal induced effects. However, for compounds with inert scaffold like the ones we are interested here, the metal mainly has a structural role and does not coordinate to the target. It is to expect that such precise treatment of electronics is not necessary and that standard protein–ligand docking methodologies could already provide with good predictions.

The aim of this study is to assess the quality of one of the most successful docking program, GOLD (version 5.1), to predict the interaction of inert organometallic metallodrugs with

---

*E. Ortega-Carrasco, A. Lledós, J.-D. Maréchal*
*Departament de Química, Universitat Autònoma de Barcelona, 08193 Bellaterra, Spain*
*E-mail: jeandidier.marechal@uab.cat*

Contract grant sponsor: Ministerio de Economía y Competitividad; Contract grant number: CTQ2011-23336; Contract grant sponsor: ORFEO Consolider-Ingenio 2010 Programme; Contract grant number: CSD2007-00006; Contract grant sponsor: Generalitat de Catalunya; Contract grant number: 2009SGR68; Contract grant sponsor: Universitat Autònoma de Barcelona (E. Ortega-Carrasco)

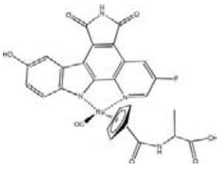
proteins. Because the number of crystal structures of proteins bounds to this kind of metallodrugs is relatively small and lower again are those for which binding constants have been reported, this benchmark has been performed on the systems designed at Meggers' laboratory.

## Material and Methods

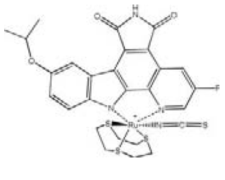
Docking calculations have been performed on 16 complexes released by Meggers' laboratory and corresponding to 15 half-sandwich compounds (14 ruthenium and 1 osmium containing systems) and one which corresponds to an iron schiff base bound to different drug targets. The Protein Data Bank (PDB) reference of each system and their associated reference code used in this work (**in parenthesis**) are 2JLD (**1**),[10] 2YAK (**2**),[11] 3FXZ (**3**),[12] 3CST (**4**),[13] 3Q4C (**5**),[14] 2BZI (**6**),[12] 3BWF (**7**),[15] 2BZH (**8**), 2IWI (**9**),[16] 3CSF (**10**),[13] 2BZJ (**11**), 3M1S (**12**),[17] 3FY0 (**13**),[12] 3PUP (**14**),[11] 2OI4 (**15**),[18] and 1G3C (**16**)[19] (Table 1). A total of eight different kinases were considered (PIM-1, PIM-2, PI3K,GSK-3, BRAF, DAPK1, PAK1) as well as Beta Trypsin.

Calculations have been performed using four different protocols that diverge on which initial structure of the ligand has been used and which amount of flexibility of the receptor has been allowed. Regarding the ligand, dockings have been

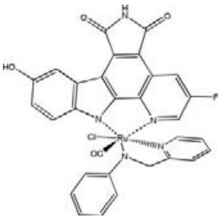
**Table 1.** Structures of the organometallic ligands of this study.



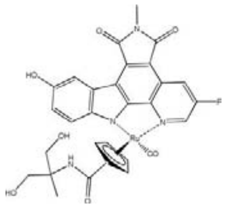
1, 2JLD, GSK-3
*0.04* nM



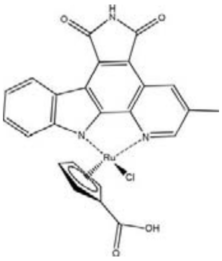
2, 2YAK, DAPK1
*2*



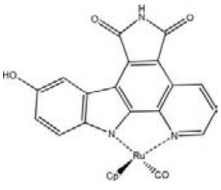
3, 3FXZ, PAK1
*130*



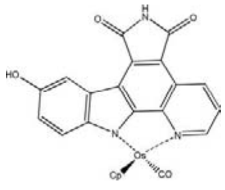
4, 3CST, PI3K
*0.04*



5, 3Q4C, BRAF
*370*

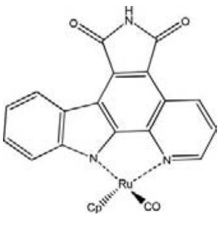


6, 2BZI, PIM1
*0.2*

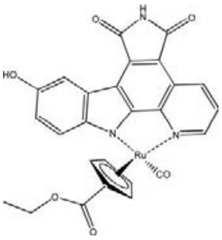


7, 3BWF, PIM1
*0.2*

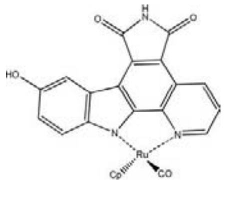
**10 – 3CSF – PIK3**



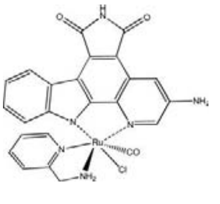
8, 2BZH, PIM1
9, 2IWI, PIM2



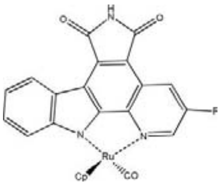
11, 2BZJ, PIM1



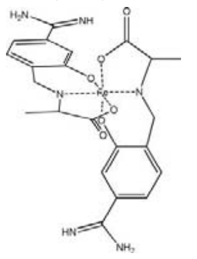
12, 3M1S, GSK3
13, 3FY0, PAK1



14, 3PUP, GSK3



15, 2OI4, PIM1



16, 1G3C, BetaTrypsin

Labels correspond to the reference used in this study, the PDB access code, the receptor name. $IC_{50}$ values when known are reported in italic.

carried out with the geometry of their corresponding crystallographic structure or with an optimized geometry obtained after minimization with the density functional theory (DFT) B3LYP.[20,21] For the receptor, calculations have been undertaken first with a rigid scaffold as it stands in the PDB and second with most of the residues of the binding site flexible. Finally, docking experiments have been carried out using the four built-in scoring functions available in GOLD: ChemScore,[22,23] GoldScore,[24] ChemPLP,[25] and ASP.[26] In total, 16 protein–ligand docking series have been carried out. Performing the calculations under these distinctive schemes was expected to allow the assessment three key players in docking accuracy: the quality of the scoring functions, the impact of predocking process of a ligand by QM minimization (something particularly relevant for screening exercises) and the impact of the flexibility of the receptor in predicting the interaction of metallodrugs with proteins. Docking results have been compared with the experimental systems in structural and energetic terms.

The same protein–ligand dockings have been performed with the GOLD5.1 suite of programs. GOLD is based on an optimized genetic algorithm that provides with extensive search of the protein–ligand conformational space and allow the characterization of low energy complexes on the potential energy surface. As most docking algorithms, GOLD allied its conformational exploration with simplified but accurate enough estimation of the binding energy. These scoring functions are built on a series of energetic components for which individual interactions between ligand and protein atoms have been accurately benchmarked. GOLD accounts with 4 built-in scoring functions including the well-known ChemScore [eq. (1)].[22] Each of them have shown to be adequate than others in function of the systems under investigation.

$$\text{ChemScore} = \Delta G_{\text{binding}} + P_{\text{Clash}} + C_{\text{internal}} P_{\text{internal}} \qquad (1)$$

$$\Delta G_{\text{binding}} = \Delta G_0 + \Delta G_{\text{hbond}} + \Delta G_{\text{metal}} + \Delta G_{\text{lipo}} + \Delta G_{\text{rot}} \qquad (2)$$

Where ChemScore is the final score and $\Delta G_{\text{binding}}$ the estimated free energy of binding corrected by weighted ($C$) penalty contributions ($P$) like those accounting for bad contacts and unfavorable internal conformations of the ligand. $\Delta G_{\text{binding}}$ is obtained by linear combination of each physical components that participate to the binding energy like hydrogen bonds ($\Delta G_{\text{hbond}}$), lipophilicity ($\Delta G_{\text{lipo}}$), coordination bonds ($\Delta G_{\text{metal}}$) or entropic loss due to rotational restrictions upon binding($\Delta G_{\text{rot}}$) [eq. (2)].

To test our hypothesis that the metal center has a limited role in the interaction of the metalodrug with its receptor, its parameters in GOLD were not optimized. To satisfy GOLD requirements although, the metal was dealt like a dummy atom and the connectivity with its coordinating neighbors was specified leadings to a rigid region of the first coordination sphere during docking. Concretely, standard single bonds between the metal and all its coordinating atoms were added. In the case of aromatic coordination groups (i.e., Cp in systems like **1** to **13**—Table 1), individual bonds were also assigned between the metal and each atoms of the aromatic system. All remaining rotational bonds have been allowed free flexibility.

For each system, 20 docking solutions have been generated. Calculations have been performed using a cavity consistent with the location of the ligand in the experimental structure and involving a 20 Å radius sphere radius centered on an atom close to the center of mass of the binding site. For calculations performed with a flexible receptor, 5 to 8 residues of the cavity were allowed to rotate accordingly to the library of rotamer available in the GOLD5.1 suite. These residues correspond to those that show at least one of their atoms located below 5 Å of the ligand. The list of flexible residues for each system is reported in the Supporting Information (Supporting Information Table S1).

Optimized geometries of the drugs have been obtained by the DFT B3LYP[20,21] approach as implemented in Gaussian09.[27] Calculations were carried out using a mixed basis set with 6-31G*[28] for the main group elements and the LANL2DZ[29] one for transition metal. The pseudopotential associated to this basis set was also applied to this atom.

## Results and Discussions

### Predocking considerations: Structure of the ligand

A common aspect in protein–ligand dockings is the preprocessing of the receptor and the ligands. For the latest, a geometry optimization is generally carried out. For pure organic compounds, calculations are performed via force field or low cost QM approaches. However, in the case of transition metal complexes, high level QM calculations (and more particularly DFT) appear more adequate.

In this study, DFT/B3LYP optimizations of all the metallodrugs have been carried out and the minimized structures have been compared with their X-ray counterparts. In general terms, the rmsd values between optimized and experimental conformations are lower than 0.9 Å (Supporting Information Table S2). The only exception is observed for compound **4** that displays a deviation of 3.2 Å (Fig. 1). This low match is due to a different conformation of a di-*tert*-butyl-hydroxylated moiety of the Cp substituent. Nonetheless, this difference is expected to have little impact on the posterior dockings since this group is flexible and its conformation optimized during these calculations.

When only the atoms of the first coordination sphere of the metal is considered, experimental and theoretical structures are in very good agreement. The overall rmsd values generally drops below 0.2 Å (see Supporting Information Table S2). Moreover, the analysis of the distances and angles of the first coordination sphere of the metalodrugs show discrepancies that rarely exceed 0.2 Å and 5°, respectively, between experiment and theory. Two exceptions are remarkable though. On one side, the metalodrug of system **2** (pdb code 2YAK) presents a distance between the metal and the sulphur of its coordinated heterocycle 0.4 Å larger in the optimized geometry (2.39 vs. 1.97 Å). On the other, the metalodrug of system **5**

Figure 1. Best possible overlap between theoretical (black) and experimental (grey) structure of the isolated ligand of 2BZI (**6**) (left) and 3CST (**4**) (right).

(pdb code 3Q4C) shows discrepancies in the angles of the first coordination sphere of the metal that are substantially higher than for the other systems (in particular the relative position of the Cp substituent changes of about 20°—see Supporting Information Table S3). Although these structural differences are not dramatic, they may be linked to the relatively packed geometries of the drugs in the binding site of their receptors; an amount of pressure absent in the gas phase condition of the DFT calculations.

This initial part of the study shows that a DFT minimization of the organometallic ligands is apparently convenient for their predocking process since no major changes of the geometry of the metalodrugs are observed. It also highlights that the first coordination sphere of the transition metal is mostly unaffected by the protein environment and suggests that our working hypothesis should be valid. The prediction of protein–ligand complexes may be more sensitive on the initial structure of the metalodrug in two cases.

### Structural considerations of the docked complexes

A total of 16 different docking experiments with different protocols have been carried out with the suite of programs Gold 5.1. They differ on the geometry of the ligand (extracted from the X-ray structures of the metalodrug-receptor complex or obtained after optimization with the B3LYP functional), the degree of flexibility of the receptor (retaining the geometry of the receptor as in the X-ray structure or allowing full flexibility of the residues of binding site) and the scoring function (ChemScore, ASP, GoldScore, and ChemPLP) selected for the calculations. The structural quality of these 16 experiments was assessed calculating the root mean square deviations of the heavy atoms of the metalodrugs between the calculated binding orientations and their experimental counterparts. Rmsd values have been reported for the lowest energy solution as well as the one of the 20 generated solutions that presents the best match. A cut-off of 2.5 Å has been established to differentiate structures with good and bad overlap. Calculations are herein discussed based on three major behaviors: (1) the lowest energy solutions produce rmsd values below 2.5 Å (Fig. 2A), (2) one of the 20 predicted binding modes matches the crystal structure below this cut-off (Fig. 2B) or (3) none the calculated complexes matches the X-ray structure (Fig. 2C). Individual results of each docking calculations are reported in Supporting Information Tables S4–S7 and their overall summary in Table 2. General tendencies are first discussed and followed by more detailed comparisons.

In general terms, experiments performed with ChemScore, ASP, and ChemPLP have a success rate of about 75% (3 to 4 misfits) for the lowest energy solutions to reproduce the X-ray structure below a rmsd of 2.5 Å (Table 2). These values are often even lower than 1.5 Å and in these cases, almost perfect overlaps between calculated and experimental structures can be predicted (Fig. 2A). Such accuracy fits in the average success rate of GOLD to predict protein–ligand binding poses and shows that, overall, dockings can behave adequately when the ligands are metalodrugs interacting with their receptor throughout their second coordination sphere.[30] In general, the structural quality of the predictions worsens when the calculations are performed with GoldScore or when introducing flexibility to the receptor.

The systems for which the predictions are the less reliable are **4, 5, 10**, and **13**. In these cases, the combination of a high exposure of the binding site to the solvent, a low number of hydrogen bonds between the ligand and the protein and the presence of major Van der Waals contacts (Supporting Information Table S8) lead to a drug candidate slightly displaced outside of the binding site of their target or to partially fills different subpockets of the binding site (Fig. 2C).

When considering the ensemble of 20 solutions generated in each docking experiment, the rate of success to find a solution below the cut-off reaches 82 to 94% for ASP, ChemPLP, and ChemScore; an impressive result considering that no optimization has been done to deal with the metal (Table 2 and Fig. 2B). Some improvement is also obtained for GoldScore although it still outperforms in front of the other scoring functions (4 incorrect poses). Interestingly, the best matching orientations rarely have a score higher than 2 units from the lowest energy one (Supporting Information Tables S4–S7). This suggests that the scoring functions or the way the first coordination sphere of the metal is dealt during the docking are just slightly underperforming.

Even considering the entire set of 20 docking solutions generated during the calculations, systems **4, 5, 10**, and **13** (pdb codes 3CST, 3Q4C, 3CSF, and 3FY0, respectively) still present the less reliable predictions. Nonetheless, some experiments punctually lead to good predictions for **4, 10**, and **13** which means that the docking scheme may have some impact on finding correct solutions in these struggling systems (i.e., calculations with the X-ray structure of system **4** in a flexible binding site—Supporting Information Table S6). However, calculations on system **5** (pdb code 3q4c) systematically fail (Supporting Information Tables S4–S7). It is interesting to note that this metalodrug is precisely the one for which the DFT geometry differs the most from its experimental counterpart. However, the failure of all the docking schemes even those considering the receptor and the metalodrug in their experimental conformations suggests that the first coordination sphere of the metal is not responsible of those discrepancies (Fig. 2C).

Detailed analysis of the success rates of the different docking strategies are also revealing (see Supporting Information Tables S4–S7). Considering only the experiments performed with a rigid conformation of the receptor, our results show no major impact of the optimization of the metalodrug to predict

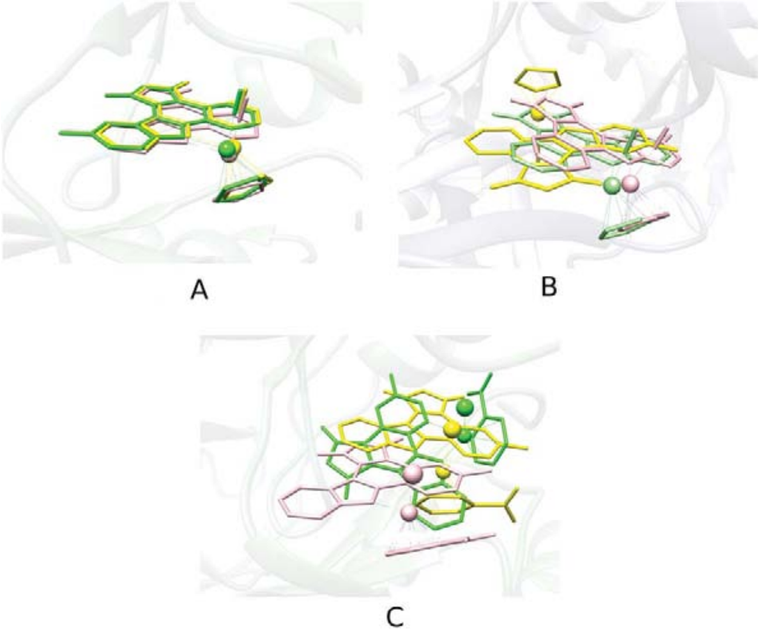


**Figure 2.** Structural comparison between the X-ray (in pink) and calculated structures of three illustrative cases. Panel A. Calculations on the system **12** (PDB code 3M1S) present a lowest energy solution (yellow) in very good agreement with the experiment. Only one significant cluster of solutions is encountered over the 20 calculated poses which also contain the best matching orientation (green). Panel B. Calculations on system **4** (PDB code 3CSF) present two major clusters of docking solutions. The one containing the lowest energy solution presents orientations inverted regarding the experimental structure (yellow). The second one contains the best matching solution that reproduces well the experimental system. Panel C. Calculations on system **5** (PDB code 3Q4C) presents two major clusters of solutions. The lowest energy one (yellow) only partially matches the aromatic region of the metalodrugs. The second one presents a similar orientation but bind deeper in the binding site cavity.

binding orientations (Fig. 3A). Calculations performed with the experimental geometry of the metalodrugs improve the rmsd values by only few decimals with respect to those performed with the optimized geometry (Supporting Information Tables S4 and S5). Major geometrical differences are only observed sporadically. For example, calculations on the complex 3CST (**4**) performed with ChemScore improves the quality of lowest energy solution when using the DFT geometry (from 7.6 Å to 3.7 Å for experimental and optimized geometries, respectively) but not the quality of the best matching one (1.5 to 3.3 Å for experimental and optimized geometries, respectively) (Supporting Information Tables S4 and S5 and Fig. 3B). The preprocessing of metalodrugs with DFT methodologies appears, therefore, as a valid solution for protein–ligand dockings and valuable for presynthetic screening experiments.

Providing the receptor with flexibility impacts at different degrees on the quality of the dockings. When calculations are performed with the ligand in its experimental conformation, the predictiveness is essentially of the same level of quality than those performed with a rigid receptor. It improves (ChemScore) or worsens (ASP and ChemPLP) by one only unit (Table 2, experiment A and C). This tendency is actually observed both for lowest energy and best matching solutions. However, Golscore is apparently more sensible than the other scoring functions and reaches up to prediction errors of the lowest solutions of eight systems.

When calculations are performed with DFT optimized geometries of the ligand, the influence of the flexibility of the receptor is more accentuated. The number of lowest energy solutions correctly predicted is only preserved with ChemScore while incorrect predictions with ASP and ChemPLP rise to 7 and 6, respectively (vs. 4–5 for the other experiments). GoldScore maintains seven wrongly predicted structures (Table 2, experiments B and D). Interestingly, the number of best matching solutions is not alter when performing the docking with optimized geometries of the metalodrug and is only affected by one unit in this case.

These results show that from a structural point of view the cumulative effect of a flexible receptor and a relaxed structure of the ligand by QM techniques decrease the possibility of the

**Table 2.** Number of predicted binding modes with RMSD values higher than 2.5 A between calculated and experimental structures.

| | ChemScore | | ASP | | ChemPLP | | GoldScore | |
|---|---|---|---|---|---|---|---|---|
| | Lowest energy solution | Best matching solution | Lowest energy solution | Best matching solution | Lowest energy solution | Best matching solution | Lowest energy solution | Best matching solution |
| A | 4 | 2 | 4 | 1 | 3 | 2 | 5 | 4 |
| B | 4 | 2 | 4 | 2 | 4 | 3 | 5 | 4 |
| C | 3 | 2 | 5 | 2 | 4 | 2 | 8 | 4 |
| D | 4 | 3 | 7 | 2 | 6 | 3 | 7 | 3 |

Results are reported considering the lowest energy solution obtained by the docking and those that display the lowest RMSD value. Experiment A corresponds to calculation with a rigid binding site and the geometry of the organometallic ligand provided from the X-ray structure, B corresponds to calculations with a rigid binding site and the B3LYP optimized structure of the ligand, C to calculations with a flexible binding site and the geometry of the organometallic ligand provided from the X-ray structure, and D to calculations with a flexible binding site and the geometry of the organometallic ligand provided from the B3LYP optimized structure of the ligand.

Figure 3. Comparison between docking solutions obtained with geometries of a metalodrug extracted from the X-ray structure and the optimized geometry obtained after B3LYP minimization. A. Lowest energy solutions of experimental (purple) and optimized (dark gray) geometries of the system 3 (PDB code 3FXZ). B. Best matching solution of the system 4 (PDB code 3CST) obtained with experimental (green) and optimized (orange) geometries of the metalodrug.

docking to find the lowest energy solution close to the native structure (at least under a run of 20 solutions). However, this barely affects the number of best hits when considering the entire set of 20 solutions generated in a docking experiment. Calculations performed with ASP, ChemScore, and ChemPLP have a good rate of success in reproducing the geometry of an organometallic compound bound to its receptor. ChemScore appears as the most robust scoring function with ASP and ChemPLP only slightly more sensitive. These three functions, behave equally well when providing at least one of the solutions below 2.5 Å of the experimental complexes. GoldScore appears as the worse scoring function for the objective of this work. As a general trend of structural prediction of organometallic compounds binding to a protein, the order of predictiveness is ChemScore$>$ASP $=$ChemPLP$\gg$GoldScore.

### Energetic considerations

Only a reduced set of metallodrugs bound to its receptor is available in the PDB. Less again are those for which the binding constant has been released for the same complex. Of the 16 systems studied so far, only 7 have their $IC_{50}$ reported and correspond to systems **1**–**7** (Table 1). In this section, these seven complexes were used as a training set to test the energetic predictiveness of GOLD. The quality of the docking prediction has been assessed by comparing the predicted binding affinity (score) with the experimental $IC_{50}$. To do so, only the score of lowest energy binding pose has been considered. The correlation between experimental and theoretical binding constants has been fitted through linear regression between the score and the logarithm of $IC_{50}$ (Figs. 4 and Supporting Information S1–S7).

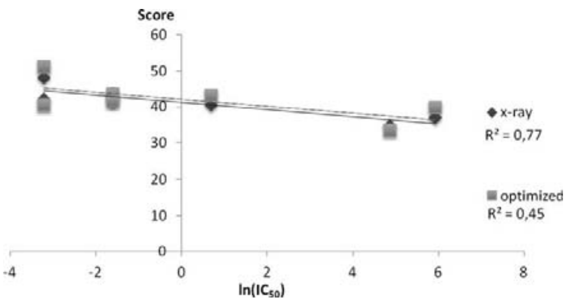


Figure 4. ASP versus ln($IC_{50}$) for optimized (black) and experimental conformation (grey) of the ligand and under a rigid conformation of the receptor.

The quality of the regressions is very variable and goes from the absence ($R^2$ = 0.03) to a very good ($R^2$ = 0.84) correlation between experiment and theory. The reduced dimension of the set is likely to be responsible for this variation; only few outliers are able to drastically impact on the regression trend. Based on this reduced amount of experimental binding affinities, only an estimation on how the docking schemes and scoring functions perform can be drawn.

First, the calculations undertaken with the structure of the organometallic moieties as they stand in the X-ray complexes perform better that those with the geometry previously optimized with DFT calculations. With the experimental geometry of the metalodrug, predictions are generally good with ASP ($R^2$ = 0.77), reasonable with ChemPLP ($R^2$ = 0.51), and poor with ChemScore ($R^2$ = 0.28). With the optimized structure of the metalodrug, the $R^2$ values drop drastically with 0.45 for ASP, 0.23 for ChemPLP, and 0.03 for ChemScore. This clearly highlights that despite the good structural overlaps observed between the calculated and experimental structures of the isolated metalodrugs and the protein–drug complexes, the structural rearrangements the metadrugs undergone during the DFT optimization process is still sufficient to alter the energetic predictiveness of the dockings.

As generally observed in docking experiments, the correlation between calculated and experimental binding energies systematically improved when calculations are performed with a flexible binding site. This is particularly obvious for dockings with ASP and ChemPLP scoring functions that both reach $R^2$ values of 0.5 to 0.8 depending on whether the structure of the ligand has been minimized or not. These results are surprising since these functions have a lower rate of success in the structural prediction of the 16 complexes than ChemScore. However, systems with the lower structural overlap obtained with ASP and ChemPLP have not their binding constants released and are not, therefore, present in this energetic training set.

Focusing on the quality of the scoring functions, the different scoring functions available in Gold clearly perform differently. In particular, predictions with GoldScore are those that display the worst correlations (highest value of $R^2$ = 0.49) and better regressions are generally obtained for the three other scoring functions. This agrees with the general knowledge that GoldScore performs better for the prediction of binding poses rather than binding affinities.[31] As a general tendency, we would expect the accuracy of the docking to follow the order: ASP $>$ChemPLP$>$ChemScore$\gg$GoldScore.

## Conclusions

The development of metallodrugs is an increasing field of research. Although modeling tools can be an interesting ally, the dynamics of the first coordination sphere of the metal during the binding process represent of the most challenging phenomenon protein–ligands docking have to deal with. Nonetheless, a large number of metallodrugs remain inert upon protein interactions. It means that little changes in the first coordination sphere of the metal are expected. This study shows that current state-of-the-art scoring functions and docking schemes, as those implemented in Gold, are efficient for predicting most of the structural features of the binding of organometallic compounds to proteins. Using the systems developed by Meggers and coworkers as a training set, docking calculations appear highly predictive from a structural point of view with a success between 75 and 94% in reproducing the correct orientation of the organometallic species in the binding site of their receptor. Despite the small set of experimental structures with binding energies reported nowadays, the scores obtained in the docking calculations are of good quality. The best energetic correlations between experiment and theory reach $R^2$ up to 0.8 and correspond to calculations performed with ASP, ChemPLP, and ChemScore. Of the scoring functions available in Gold, ChemScore appears as the most robust one both for structural and energetic prediction with ASP and ChemPLP closely following and being in certain circumstances more efficient. GoldScore presents the results with the poorest quality. Systematic errors in several systems are observed. In these cases, it is likely that the improvement of how dockings can explore the conformational changes related to the first coordination sphere of the metal would represent a major step forward. However, this study is, to our knowledge, the first extensive analysis on how the interaction of organometallic compounds can be predicted through docking approaches.

**Keywords:** protein–ligand dockings · metalodrugs · computational bioinorganics · kinase inhibition · drug design

How to cite this article: E. Ortega-Carrasco, A. Lledós, J.-D. Maréchal. *J. Comput. Chem.* **2014**, *35*, 192–198. DOI: 10.1002/jcc.23472

Additional Supporting Information may be found in the online version of this article.

[1] B. Rosenberg, L. Van Camp, T. Krigas, *Nature* **1965**, *205*, 698.
[2] Z. Guo, P. J. Sadler, *Angew. Chem. Int. Ed.* **1999**, *38*, 1512.
[3] M. Melchart, A. Habtemaiam, O. Novokova, S. A. Moggach, F. P. A. Fabbiani, S. Parsons, V. Brabec, P. J. Sadler, *Inorg. Chem.* **2007**, *46*, 8950.
[4] G. Gasser, N. Metzler-Nolte, *Curr. Opin. Chem. Biol.* **2012**, *16*, 84.
[5] E. Meggers, *Chem. Commun.* **2009**, 1001.
[6] M. L. Verdonk, J. C. Cole, M. J. Hartshorn, C. W. Murray, R. D. Taylor, *Proteins* **2003**, *52*, 609.
[7] R. A. Friesner, J. L. Banks, R. B. Murphy, T. A. Halgren, J. J. Klicic, D. T. Mainz, M. P. Repasky, E. H. Knoll, M. Shelley, J. K. Perry, et al. *J. Med. Chem.* **2004**, *47*, 1739.
[8] B. Seebeck, I. Reulecke, A. Kämper, M. Rarey, *Proteins* **2008**, *71*, 1237.
[9] V. M. Robles, E. Ortega-Carrasco, E. G. Fuentes, A. Lledós, J.-D. Maréchal, *Faraday Discusss.* **2011**, *148*, 137.
[10] G. E. Atilla-Gokcumen, N. Pagano, C. Streu, J. Maksimoska, P. Filippakopoulos, S. Knapp, E. Meggers, *Chembiochem.* **2008**, *9*, 2933.
[11] L. Feng, Y. Geisselbrecht, S. Blanck, A. Wilbuer, G. E. Atilla-Gokcumen, P. Filippakopoulos, K. Kräling, M. A. Celik, K. Harms, J. Maksimoska, et al., *J. Am. Chem. Soc.* **2011**, *133*, 5976.
[12] J. Maksimoska, L. Feng, K. Harms, C. Yi, J. Kissil, R. Marmorstein, E. Meggers, *J. Am. Chem. Soc.* **2008**, *130*, 15764.
[13] P. Xie, D. S. Williams, G. E. Atilla-Gokcumen, L. Milk, M. Xiao, K. S. M. Smalley, M. Herlyn, E. Meggers, R. Marmorstein, *ACS Chem. Biol.* **2008**, *3*, 305.
[14] P. Xie, C. Streu, J. Qin, H. Bregman, N. Pagano, E. Meggers, R. Marmorstein, *Biochemistry* **2009**, *48*, 5187.
[15] J. Maksimoska, D. S. Williams, G. E. Atilla-Gokcumen, K. S. M. Smalley, P. J. Carroll, R. D. Webster, P. Filippakopoulos, S. Knapp, M. Herlyn, E. Meggers, *Chemistry* **2008**, *14*, 4816.
[16] A. N. Bullock, S. Russo, A. Amos, N. Pagano, H. Bregman, J. E. Debreczeni, W. H. Lee, F. von Delft, E. Meggers, S. Knapp, *PloS one* **2009**, *4*, e7112.
[17] G. E. Atilla-Gokcumen, L. Di Costanzo, E. Meggers, *J. Biol. Inorg. Chem.* **2011**, *16*, 45.
[18] N. Pagano, J. Maksimoska, H. Bregman, D. S. Williams, R. D. Webster, F. Xue, E. Meggers, *Org. Biomol. Chem.* **2007**, *5*, 1218.
[19] E. Toyota, K. K. Ng, H. Sekizaki, K. Itoh, K. Tanizawa, M. N. James, *J. Mol. Biol.* **2001**, *305*, 471.
[20] C. Lee, W. Yang, R. G. Parr, *Phys. Rev. B* **1988**, *37*, 785.
[21] A. D. Becke, *J. Chem. Phys.* **1993**, *98*, 5648.
[22] M. D. Eldridge, C.W. Murray, T.R. Auton, G.V. Paolini, *J. Comput.-Aided Mol. Des.* **1997**, *11*, 425.
[23] C. W. Murray, T. R. Auton, M. D. Eldridge, *J. Comput.-Aided Mol. Des.* **1998**, *12*, 503.
[24] M. L. Verdonk, J. C. Cole, M. J. Hartshorn, C. W. Murray, R. D. Taylor, *Proteins* **2003**, *52*, 609.
[25] O. Korb, T. Stützle, T. E. Exner, *J. Chem. Inf. Model.* **2009**, *49*, 84.
[26] W. T. M. Mooij, M. L. Verdonk, *Proteins* **2005**, *61*, 272.
[27] M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G. A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H. P. Hratchian, A. F. Izmaylov, J. Bloino, G. Zheng, J. L. Sonnenberg, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, J. A. Montgomery, Jr., J. E. Peralta, F. Ogliaro, M. Bearpark, J. J. Heyd, E. Brothers, K. N. Kudin, V. N. Staroverov, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, N. Rega, J. M. Millam, M. Klene, J. E. Knox, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, R. L. Martin, K. Morokuma, V. G. Zakrzewski, G. A. Voth, P. Salvador, J. J. Dannenberg, S. Dapprich, A. D. Daniels, Ö. Farkas, J. B. Foresman, J. V. Ortiz, J. Cioslowski, D. J. Fox, Gaussian, Inc:, Wallingford CT, **2009**.
[28] G. Fogarasi, X. Zhou, P. W. Taylor, P. Pulay, *J. Am. Chem. Soc.* **1992**, *114*, 8191.
[29] P. J. Hay, W. R. Wadt, *J. Chem. Phys.* **1985**, *82*, 299.
[30] S. F. Sousa, P. A. Fernándes, M.J. Ramos, *Proteins* **2006**, *65*, 15.
[31] J. W. Liebeschuetz, J. C. Cole, O. Korb, *J. Comput.-Aided Mol. Des.* **2012**, *26*, 737.

Received: 21 June 2013
Revised: 25 September 2013
Accepted: 29 September 2013

# 4 Unravelling novel synergies between organometallic and biological partners: a quantum mechanics/molecular mechanics study of an artificial metalloenzyme

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

JOURNAL OF THE ROYAL SOCIETY Interface

rsif.royalsocietypublishing.org

Research

**Cite this article:** Ortega-Carrasco E, Lledós A, Maréchal J-D. 2014 Unravelling novel synergies between organometallic and biological partners: a quantum mechanics/molecular mechanics study of an artificial metalloenzyme. *J. R. Soc. Interface* **11**: 20140090.
http://dx.doi.org/10.1098/rsif.2014.0090

Received: 27 January 2014
Accepted: 17 April 2014

**Subject Areas:**
chemical biology, biochemistry, bioinformatics

**Keywords:**
computational bioinorganics, artificial metalloenzymes, activation process, resting state

**Author for correspondence:**
Jean-Didier Maréchal
e-mail: jeandidier.marechal@uab.cat

Electronic supplementary material is available at http://dx.doi.org/10.1098/rsif.2014.0090 or via http://rsif.royalsocietypublishing.org.

Royal Society Publishing

# Unravelling novel synergies between organometallic and biological partners: a quantum mechanics/molecular mechanics study of an artificial metalloenzyme

Elisabeth Ortega-Carrasco, Agustí Lledós and Jean-Didier Maréchal

Departament de Química, Universitat Autònoma de Barcelona, Bellaterra 08193, Spain

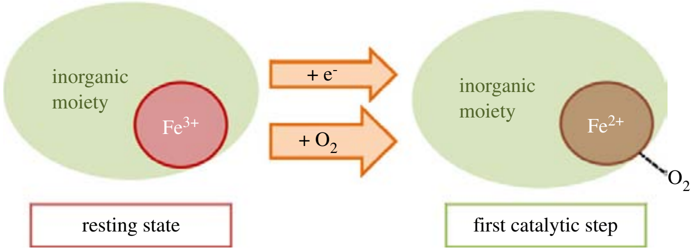
In recent years, the design of artificial metalloenzymes obtained by the insertion of homogeneous catalysts into biological macromolecules has become a major field of research. These hybrids, and the corresponding X-ray structures of several of them, are offering opportunities to better understand the synergy between organometallic and biological subsystems. In this work, we investigate the resting state and activation process of a hybrid inspired by an oxidative haemoenzyme but presenting an unexpected reactivity and structural features. An extensive series of quantum mechanics/molecular mechanics calculations show that the resting state and the activation processes of the novel enzyme differ from naturally occurring haemoenzymes in terms of the electronic state of the metal, participation of the first coordination sphere of the metal and the dynamic process. This study presents novel insights into the sensitivity of the association between organometallic and biological partners and illustrates the molecular challenge that represents the design of efficient enzymes based on this strategy.

## 1. Introduction

The interaction of organometallic compounds with biological macromolecules has become a major focus of attention in several areas of chemistry and its interfaces [1]. One of them is the design of artificial metalloenzymes obtained by the insertion of homogeneous catalysts into protein cavities; an approach with a tremendous potential in biocatalysis [2–4]. Numerous systems built on this concept have already been reported that include enzymes able to perform Diels–Alder reactions [5–6], transfer hydrogenation [7,8], sulfoxidation [9–11] and hydration [12] among others. In these hybrids, the molecular partnership between homogeneous and biological subsystems is reminiscent of those happening in natural haemoenzymes: the organometallic group affords the catalytic functionality, whereas the biomolecular one mainly dictates the substrate specificity and controls catalytic features such as enantioselectivity and/or regiospecificity [13].

Nowadays, the engineering of natural metalloenzymes is a widespread strategy. It mainly stands on the genetic alterations of the metal-binding site by a reduced number of amino acid substitutions, which leads to the modulation of their catalytic profiles [14–16]. However, the control of the structural and electronic properties of the metal centre using non-natural cofactors is more challenging because it involves a different chemobiological space from the one provided by the natural framework of evolution [1,17,18]. Unprecedented complementarities between organometallic and proteic moieties can be observed which could lead to binding and catalytic features different from the objectives pursued by enzyme designers. In other words, each partner may suffer molecular stresses which are absent in their conventional media such as unusual electronic states of the metal, metal–ligand exchanges upon binding of the catalyst to its receptor or even metal-mediated conformational changes.

One of the few artificial metalloenzymes for which the structure is reported in the Protein Data Bank (PDB code 1WZD) clearly illustrates the previous



**Scheme 1.** Initiation of the catalytic process for the artificial enzyme. (Online version in colour.)

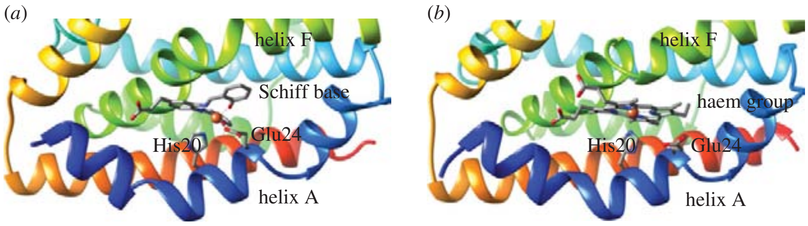


**Figure 1.** Representation of the crystallographic structures of the Fe(III) Schiff base · *cd***HO** (*a*) and the haem-bound form (*b*).

statement. This system, developed by Ueno *et al*. [19], was obtained by the substitution of haem by a Fe(Schiff base) (salophen) in the *Corynebacterium diphtheriae* haem oxygenase (*cd***HO**). Natural and artificial systems share important aspects of their catalytic mechanisms, i.e. both proceed throughout an oxido-reduction mechanism with the cytochrome P450 reductase as a partner, leading to the reduction of the iron atom of the cofactor from Fe(III) to Fe(II) and the subsequent coordination of a molecule of dioxygen (scheme 1). However, the natural system performs an entire oxidative process ending at the cleavage of the haem, whereas the artificial enzyme stops at the activation of the oxygen to form superoxide ions. No subsequent oxidation of the cofactor or any external substrate proceeds. This superoxidase activity differs drastically from the natural function of *cd***HO**; a surprising result considering the high chemical similarity between salophen Schiff bases and porphyrinic cofactors in terms of structure and reactivity.

The X-ray structure of the system provides some molecular hints on those differences. In this structure, the Fe(Schiff base) is highly distorted and has two residues of the receptor, His20 and Glu24, coordinating the metal (figure 1). Hypothesized as the resting state of the system, the iron atom is afforded a +3 oxidation state of unknown spin state and presents a distorted octahedral configuration with no room for the binding of the oxygen. This structure highly differs from those known for natural haemoenzymes including holo-*cd***HO**. Indeed, these species generally display resting states with either a square pyramidal high-spin (i.e. catalase [20]) or an octahedral low-spin configuration (i.e. peroxidases [21], cytochromes P450 [22] or haem oxygenase [23]) with only one residue of their apo-protein bound to the metal in the axial position and, eventually, one labile water molecule coordinated on the distal site of the haem to complete the octahedral configuration of the metal (figure 1). The strength of the coordination of this water molecule can be influenced by its environment and, in turn, conditions the transition from hexacoordinated resting states to active pentacoordinated species [24].

The experimental geometry of Fe(Schiff base) · *cd***HO** therefore suggests that both the resting state and the activation process could be different from those of naturally occurring Fe(III) haemoenzymes. Its divergence from the resting states of natural ones also questions the possibility of a crystallographic artefact; the known flexibility of apo-**HO** and salophens could be accentuated in the crystallographic media. A better understanding of the complementarities between the organometallic and the proteic moieties in Fe(Schiff base) · *cd***HO** could represent a major step forward in bioinorganics and further help in the design of artificial metalloenzymes and other bioinorganic architectures.

Computational chemistry is now widely used for the study of metalloenzymes. Most of its application focuses on unravelling the molecular grounds of their catalytic mechanisms [25]. Little has yet been done to decode the relative contributions of inorganic and biological partners and less still for enzyme design. This is due to the complexity of dealing with the binding of metal-containing systems to proteins by standard approaches [26]. Simulations involving large conformational samplings as well as accurate representation of the metallic environment are indeed necessary. We recently showed that a combination of structural bioinformatics, quantum-based methods (quantum mechanics (QM) and quantum mechanics/molecular mechanics (QM/MM)) and protein–ligand dockings represents an interesting solution in characterizing low-energy complexes of organometallic systems bound to proteins [27,28]. Here, we further investigate the molecular complementarity between artificial organometallic cofactors and proteic partners by performing an extensive QM/MM study on the experimental Fe(Schiff base) · *cd***HO** system, focusing on (i) the nature of its Fe(III) resting state, its comparison with those of natural haemoenzymes and its possible artefactual nature, (ii) the activation process, with the simulation of the full transition occurring between hexa- and pentacoordinated forms and considering the reduction of the metal centre, and (iii) the relative contributions of the inorganic and biological parts to the structural and energetic properties of the biocomposite. Additionally, this study illustrates the potential of QM/MM approaches to understand novel chemobiological processes.

## 2. Material and methods

Calculations have been performed on the entire Fe(Schiff base) · *cd***HO** system using the QM/MM ONIOM (DFT : AMBER) method implemented in Gaussian 09 [29]. The charges and protonation states of all titrable amino acids were automatically assigned using the interface provided by the UCSF Chimera package [30] with the exception of the iron-chelating histidine His20, which was manually set to be consistent with the coordination rules of the metal. Visual inspection was subsequently performed. The total charge of the system is $-8$ or $-7$ depending on the oxidation state of the iron. It can be divided by $-10$ for the isolated protein and $+2$ or $+3$ for the inorganic complex. The QM part in the QM/MM partition has a charge $+1$ in the Fe(II) species and $+2$ for the Fe(III) ones.

One of the main objectives of this work is to determine whether the experimental structure corresponds to a plausible electronic structure of the metal centre in a resting state configuration or not. Thus, we are interested in keeping the initial structure for our study as close as possible to the crystallographic one. With this purpose in mind, and considering that the absence of optimum force field parameters of the iron in the different coordination spheres of the salophen under consideration may lead molecular dynamics towards less realistic equilibrium structures, no such experiments were performed on the initial X-ray structure [31]. Moreover, as we were also interested in analysing the transition between hexa- and pentacoordinated forms of the Fe(Schiff base) · *cd***HO** complex, an accurate electronic representation of the first and second coordination spheres of the metal was mandatory, and QM/MM calculations were chosen as the best method for this purpose.

For the hexacoordinated system, calculations were carried out on the original crystallographic X-ray coordinates corresponding to this state (referred to as **6**, PDB reference: 1WZD). The equatorial environment of the metal results from an $N_2O$ chelation provided by the Schiff base and the O$\gamma$ of the Glu24 of the host; the remaining axial positions are occupied by the second oxygen of the cofactor on the distal side and the N$\varepsilon$ of residue His20 on the proximal one (figure 1). For the pentacoordinated systems with or without an axial water (referred to as **5**′ and **5**, respectively), the equatorial environment is entirely provided by the cofactor bound in an $N_2O_2$ fashion, and the axial proximal position is occupied by the His20. In these structures, the side chain of Glu24 has been removed from the first coordination sphere of the metal and moved towards the most probable rotameric state for this residue using the Dunbrack rotamer library of the UCSF Chimera package [32]. A fourth system, referred to as **6**′, was built to allow energetic comparison with **5**′. It consists of a similar system to **6** but with a water molecule added inside the binding site of the biocomposite in the nearest hole in the vicinity of the cofactor (between Lys13 and His25). In all these systems, the cofactor remains in a hydrophobic environment and packed between helices A and F of apo-*cd***HO**. The water molecules that could interfere with the mechanism would therefore be on the superficial region of the protein and would tremendously reorganize during the simulation of transition between penta- and hexacoordinated structures. This would lead to a highly unsmoothed potential energy landscape that is difficult to study and with relatively little improvement in the description of the chemical process at the metal site. Therefore, other molecules than those involved in the binding of the iron were removed from the calculations [31].

Two QM/MM partitions were used. The larger partition includes the metal, the side chain of the coordinating residues His20 and Glu24 up to their C$\alpha$ atoms and the entire aromatic part of the Schiff base (electronic supplementary material, scheme S1). The second partition contains the metal, the same coordinating residues of the protein up to the C$\alpha$ and a model of the first coordination sphere provided by the salophen where the aromatic moieties are replaced by simpler ethylenic groups (electronic supplementary material, scheme S2). In both cases, the rest of the cofactor was included in the MM part of the system and treated using the AMBER force field [33] for the amino acids and the GAFF [34] one for the atoms of the cofactor. In the first partition, 55 of the 3347 total atoms are included in the QM region, and in the second one, only 37. The smaller partition was used in the post-analysis and focused on the contribution of the first coordination sphere of the metal to the mechanism of the enzyme. To maintain the geometry of the system close to the experimental system, only atoms (QM and MM) within a sphere of about 15 Å from the iron where allowed to relax during the optimization process with the entire helix A included in this flexible scheme. In both partitions, the number of flexible atoms was 192 atoms and the rest remained fixed.

For each structure, QM/MM minimizations were undertaken for low-, intermediate- and high-spin Fe(II) (multiplicity values: 1, 3 and 5) and Fe(III) (multiplicity values: 2, 4 and 6) species. Optimizations have been performed with the B3LYP functional [35,36] using the 6–311+g* basis set [37,38] for the main group elements and the *aug*-cc-VTZ [39,40] one for the iron. Single point calculations on the optimized structures have been performed with the M06L [41], PBE [42] and B97D [43] functionals to identify possible drawbacks on the relative stability between different spin states [44–46]. For the same reason, calculations using electronic embedding have been also tested.

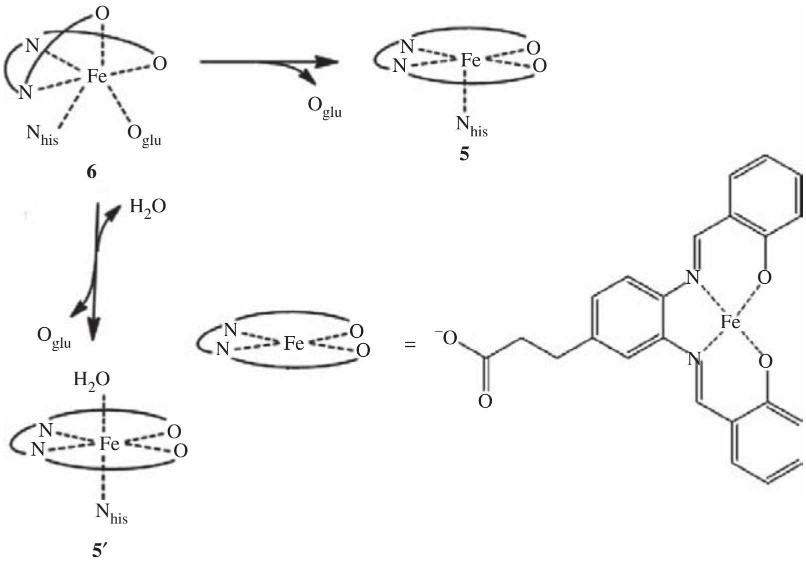
## 3. Results and discussion

### 3.1. Analysis of the Fe(III)(Schiff base)·HO resting state

Calculations have been carried out on the crystallographic system (referred to as **6**) and on configurations generally observed in natural haemoenzymes. The latter correspond to, on one side, the structure obtained by the removal of the glutamate from the first coordination sphere of the metal (leading to five coordinated iron and referred to as **5**) and, on the other, to the configuration obtained when substituting the glutamate by a water molecule (leading to an alternative six coordinated iron and referred to as **5**′) (scheme 2). All the optimizations were performed with the larger QM/MM partition described in the previous section.

Independently of their spin states, all optimized structures of the ferric systems **6** are in good agreement with their crystallographic counterpart (electronic supplementary material, table S1). The iron remains in an octahedral configuration with a distorted cofactor and coordinated to the protein by His20 and Glu24 (figure 2; electronic supplementary material, figures S1 and S2). In all cases, bond lengths and angles of the first coordination sphere of the metal are in good agreement between theory and experiment, but closer values are obtained for the Fe(III) high-spin system (generally lower by 0.1 Å with the exception of interaction Fe–$O_{glu}$, which increases by 0.2 Å for distances and 0.3° for angle values).

When removing the glutamate from the first coordination sphere (**5**), the system adopts a square pyramidal geometry (electronic supplementary material, table S1). The cofactor becomes more planar than in the experimental system with



**Scheme 2.** Representation of the experimentally observed and alternative resting states of the Fe(Schiff base)·**HO** considered in this part of the study. From left to right systems are referred to as **6**, **5** and **5′**.

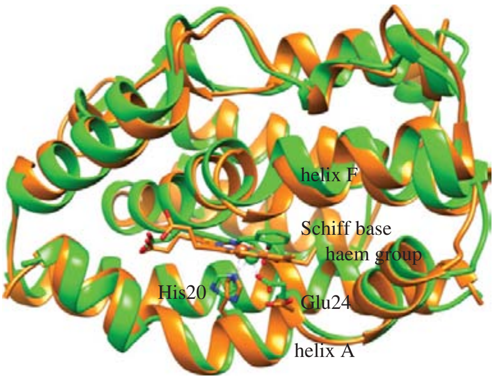


**Figure 2.** Superposition of haem-bound haem oxigenase (PDB code: 1IVJ) (orange) and the structure under study (green).

deviations from an ideal plane (OOP Fe values in the electronic supplementary material, table S1) of 0.2, 0.4 and 0.6 Å for low, intermediate and high spin, respectively, instead of the 0.7 Å observed for the experimental system. Overall, the same tendencies are also obtained for the substitution of the glutamate by a water molecule, being the cofactor slightly more planar in **5′**. Nonetheless, it must be noted that the Fe–($O_{water}$) bond is almost cleaved in the intermediate- and high-spin complexes, which is consistent with previous observations in cytochromes P450 (electronic supplementary material, figure S2 and table S2) [47].

In addition to the general changes in the first coordination sphere of the metal and the reorganization of the cofactor, the removal of the glutamate is also correlated with a reorganization of helix A. Optimized structures of **5** and **5′** show the $C_{ter}$ end of the helix pointing towards the solvent in a conformation similar to those observed in the haem-bound haem oxygenase structure [48]. This contrasts with the structures of haem-bound systems in which the macrocyclic nature of the cofactor prevents the Glu24 from getting close enough to the Fe(III).

Relative stabilities between the different optimized geometries of the Fe(III)(Schiff base) · *cd***HO** show a clear preference for the experimental configuration (**6**) with respect to the ferric **5** and **5′** structures (figure 4; electronic supplementary material, figure S7). Moreover, calculations performed with the (B3LYP : AMBER)-mechanical embedding (**ME**), (B3LYP : AMBER)-electronic embedding (**EE**), (M06L : AMBER)-**ME** and (B97D : AMBER)-**ME** level are consistent in predicting **6** to be about 5–10 kcal $mol^{-1}$ more stable in its high-spin configuration. Only the (PBE : AMBER)-**ME** scheme diverges, giving isoenergetic low- and high-spin wave functions in the case of configurations **6** and **5′**, a behaviour that has been associated with the weakness of this approach to discriminate between spin states of different multiplicity [49]. Finally, the differences in energy between **6** and **5** is comparable to those observed between **6** and **5′** but larger for the spin state of higher multiplicity. This has been associated with the weakening of the Fe–($O_{water}$) bond that could lead to a dissociation (electronic supplementary material, figure S2).

To further investigate the ferric system, the full transition paths between configurations **6** and **5** have been simulated. Because the use of electronic embedding does not present major improvement and taking into account previous considerations regarding the quality of the different functionals, the following part of the work was carried out under the (B3LYP : AMBER)-**ME** scheme.

The transition state structures have been characterized at Fe–$O_{glu}$ distances of 3.01, 2.55 and 3.37 Å for the low-, intermediate- and high-spin species, respectively (see figure 3 and electronic supplementary material, table S3, for more details). No major differences are observed in the overall structures of the different transition states and a unique negative vibrational mode has been identified. In the first coordination sphere of the metal, this mode shows major components in the breaking of the Fe–Glu24 bond as well as in the displacement of the His20 towards a perpendicular position relative to the cofactor and passing through the metal and the displacement of the glutamate towards the solvent (figure 3). Major

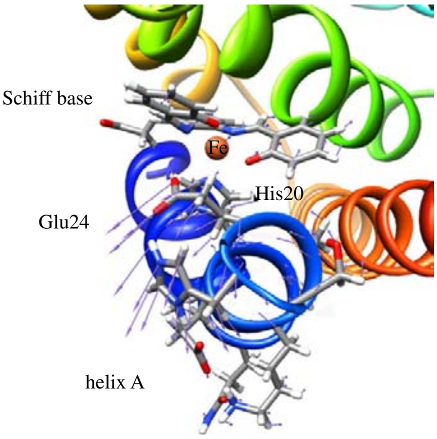


**Figure 3.** Representation of the negative vibrational mode of the transition state along the displacement from **6** to **5** configurations.

contributions are also observed for the cofactor, which bends to get close to planarity, as well as for helix A, whose $N_{ter}$ region gets closer to the Schiff base and whose $C_{ter}$ region gets more exposed to the solvent. This result clearly shows how the distortion of the cofactor, the modification of the first coordination sphere of the metal and the rearrangement of the structure of the receptor are intrinsically correlated and not individual phenomena.

The energy of the transition state is 20.7, 9.8 and 18.6 kcal mol$^{-1}$, above the absolute minimum of the system, for low-, intermediate- and high-spin systems, respectively (figure 4). They are sufficiently large to discard the possibility that the atomic motions at room temperature naturally allow the transition from glutamate bound (**6**) to glutamate unbound configurations (**5** and **5′**). Interestingly, the energetic breakdown in QM and MM contributions of the total QM/MM energy shows that the electronic (QM) term dominates in the relative stability of **6**, **5** and **5′** (electronic supplementary material, figure S3). Therefore, the changes in the structure of the receptor at helix A and the conformation of the salophen are mainly driven by electronic and geometric properties of the first coordination sphere of the metal.

Our calculations on this part of the study support that the experimental configuration of the Fe(Schiff base) · *cd***HO** observed in the crystallographic structure is the real resting state of the enzyme and that its electronic state corresponds to a high-spin configuration. The large differences in energy between the experimental configuration and other possible resting state geometries imply that dynamical equilibrium between them does not naturally take place in solution. In this, the constraint on the system is manly associated with the first coordination sphere of the metal.

As, in this configuration, neither labile ligand nor sufficient room is available on the top of the iron to coordinate the oxygen, the activation process of the enzyme most likely would be triggered by the reduction of the metal and would differ from the chemistry of haemoenzymes.

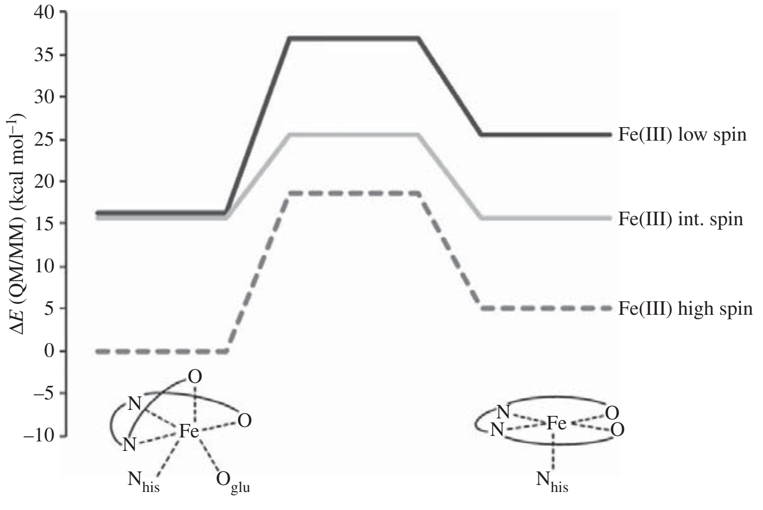
## 3.2. Electronic transition and activation mechanism

The previous part of the study reveals that the transition from glutamate bound (**6**) to glutamate unbound (**5** and **5′**) metal is unreachable for Fe(III) species. In this configuration, the oxygen cannot reach the iron and bind to it because of the absence of room in the upper position of the cavity placed in the axial position of the salophen. Therefore, the reduction of iron is necessary for the system to be activated and the reaction to proceed (scheme 1). To understand the activation mechanism of the enzyme, the entire transition paths from **6** to **5** have been simulated for Fe(II). The impact of the solvent molecule (**5′**) has only been estimated on the minima.

Optimizations of the Fe(II) systems lead to stable **6** and **5** conformations for all spin states. The general structural profiles of these species are similar to those observed in the ferric structures. Metal ligand distances of the first coordination sphere are slightly longer for the ferrous system. The position of the glutamate with respect to the metal is particularly affected, being approximately 0.1 Å closer for Fe(III) than Fe(II) systems at identical spin states. This is consistent with stronger electrostatic interactions in the former. Importantly, for **5′**, the water molecule remains coordinated to the iron in the axial position in the low-spin state Fe(II) only (see electronic supplementary material, table S3). For intermediate- and high-spin species, the water molecule leaves the first coordination sphere of the metal and lies at 3.33 and 3.56 Å, respectively. Such spontaneous removal of the solvent molecule is compatible with the creation of a vacant site on the top of the iron, as observed in haemoenzyme systems [36]. Therefore, simulation of transitions from **6** to **5** should be indicative enough of the activation mechanism.

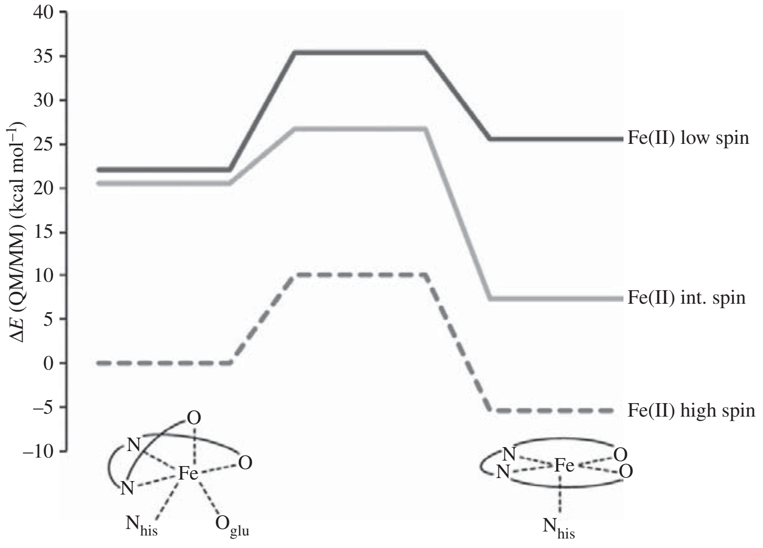
For (B3LYP : AMBER)-**ME** calculations, the transition between **6** and **5** is exothermic for intermediate and high spin and slightly endothermic for the low spin (figure 5; electronic supplementary material, figure S6). The high-spin state is substantially more stable than the two others, which suggests that the process should take place entirely in this spin state. These results are qualitatively supported by (M06L : AMBER)-**ME**, (B97D : AMBER)-**ME** and (B3LYP : AMBER)-**EE** (electronic supplementary material, figure S7). The (PBE : AMBER)-**ME** approach slightly diverges from the other functional by suggesting possible spin crossing between low- and intermediate-spin wave functions. Probably due to the weakness of this functional in dealing with the difference in energy between spin states, these results still have no significant impact on the chemistry of the system. For the transition from **6** to **5′**, the high-spin wave function remains the most probable spin state for the transition but the differences in energy observed are lower than for the transition between **6** and **5** (electronic supplementary material, figure S6). This is probably due to the presence of a water molecule in **5′** that can stabilize the system. Taken together, the stable geometries of penta- and hexacoordinated structures of the ferrous systems show that a conformation adequate for the binding of the oxygen is favoured after reduction of the iron and produce a stable high-spin configuration. This is strongly different from what is known of naturally occurring haemoenzymes. For example, in cytochromes P450, the entrance of the substrate in the binding site leads to the removal of the labile water ligand and the electron transfer to generate the ferrous species.

The transition path was therefore investigated. As the water chelating the iron in **5′** is dissociated in the high- and intermediate-spin function, further calculations were carried out only on the transition between **6** and **5**. Transition states



**6** **5**

**Figure 4.** Simplified representation of the potential energy surface for the transition from configuration **6** to configuration **5** Fe(III)(Schiff base) · *cd***HO** for low-, intermediate- and high-spin species. Calculations corresponding to the ONIOM(B3LYP : AMBER)-**ME**.



**6** **5**

**Figure 5.** Simplified representation of the potential energy surface for the transition from configuration **6** to configuration **5** Fe(II)(Schiff base) · *cd***HO** for low-, intermediate- and high-spin species. Calculations corresponding to the ONIOM(B3LYP : AMBER)-**ME**.

between these two structures occur at Fe–$O_{glu}$ distances at 3.44, 2.73 and 3.02 Å for low-, intermediate- and high-spin species, respectively. The general features of the lowest vibrational mode are similar to those of the ferric system as reported in the previous section. Additionally, the transition states for these different species are located at 13.4, 2.5 and 10.1 kcal mol$^{-1}$ above the reactant structure for low, intermediate and high spin, respectively. For the high-spin species, which represents the ground state of the system, this means a diminution of about 8 kcal mol$^{-1}$ and a value relatively accessible in standard conditions. It appears therefore that, when the system has been reduced, the transition between **6** and **5** has a low enough barrier to take place.

Enthalpic and kinetic considerations confirm that, once the electron has been transferred to the ferric ion, the removal of the glutamate from the first coordination sphere of the metal becomes favourable. The mechanism should take place via a high-spin state throughout the process and lead to stable square pyramidal structures compatible with oxygen binding at the vacant site of the metal. This would take place through a transition mechanism where a mixed contribution of the conformational change of the cofactor, the rearrangement of helix A and the change in coordination state of the metal occurs. Which of these features dominates in the process is yet to be determined but is fundamental to decoding the real extent of non-natural bioinorganic complementarities.

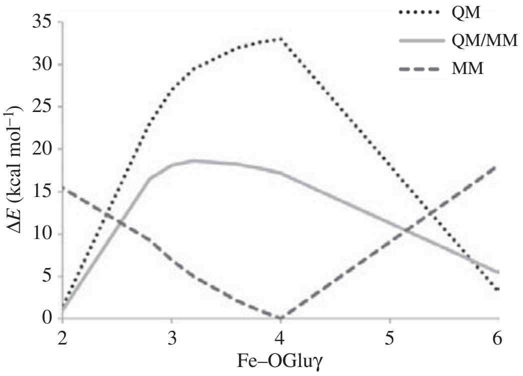


**Figure 6.** Relative contribution of the QM and MM region to the total energy of the complex (Fe(III) high-spin system).

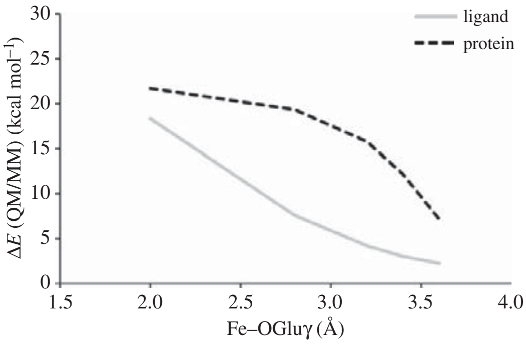


**Figure 7.** Example of the relative contributions of the ligand and protein systems in the total QM/MM energy associated with the transition path between **6** and **5** for the Fe(II) high-spin system.

## 3.3. First coordination sphere of metal versus protein–cofactor complementarities

The final question we aim to answer is to identify the molecular variables that condition the unexpected geometrical features of the Fe(Schiff base) · *cd***HO** system and so shed light on what are the energetic contributions that dominate in the formation of bio-organometallic systems. We therefore embarked on analysing the relative energy of the system, partitioning it in terms of (i) first coordination sphere of the metal versus the rest of the system and (ii) ligand versus protein (electronic supplementary material, scheme S2). To do so, we performed additional calculations where the QM/MM partition was constructed in such a way that the first coordination sphere of the metal was modelled in the quantum mechanics region, while the rest of the system was included in the MM one (electonic supplementary material, scheme S3). We first benchmark this novel partition by carrying on calculations on the transition from **6** to **5** and comparing these results with the larger one. Despite slight geometric and energetic nuances, the overall profiles are very similar (figure 6; electronic supplementary material, figure S4) and further analyses were therefore performed on these calculations.

Decomposition between QM and MM terms of the total QM/MM energy unambiguously shows that the QM term dictates the shape of the overall profile in all oxidation and spin states. The QM curves are only a few kcal mol$^{-1}$ higher than the QM/MM ones and both of them can almost overlap. The MM terms have a minor role and only modulate the shape of the QM/MM profile. However, it is still interesting to note that, in most cases, this part of the system remains almost unaffected (Fe(II) cases) or even stabilized (Fe(III) cases) for Fe–O$_{\mathrm{glu}}$ going from reactant to a few steps after the transition state. From then on, a destabilization of the system is observed, reaching up to *ca* 15 kcal mol$^{-1}$ (electronic supplementary material, figure S4). Structural analysis showed us that the first part of the transition implies a slight relaxation of helix A while the glutamate is removed. However, at longer distances some clashes at the hinge between the flexible region of our partition and the rigid one are observed, hence leading to higher energies. Therefore, our partition does not allow us to deal with the full extent of the molecular flexibility engendered by the relaxation of the first coordination sphere of the metal; something that definitely represents a major tour de force in molecular modelling. However, our results are still quite indicative. Once the Glu24 goes out of the first coordination sphere of the metal, the rearrangement of helix A can be energetically meaningful.

The larger partition allows us to consider the relative contribution of the cofactor and receptor in the total energy of the complex. In this case, a clear different pattern is observed. For all spin and oxidation states, the energy of both isolated cofactor and receptor decrease as a function of the distance between the glutamate and the metal (figure 7; electronic supplementary material, figure S5). The protein tends to stabilize to *ca* 10 kcal mol$^{-1}$ and the cofactor up to 14 kcal mol$^{-1}$. This means that a planar conformation of the Fe(Schiff base), the presence of the glutamate out of the first coordination sphere of the metal as well as a conformation of helix A with its C$_{\mathrm{ter}}$ end pointing towards the solvent would be preferred by considering the energetic properties of both individual species. All these statements are consistent with current knowledge on porphyrinic, salen and salophen moieties [36] as well as the crystallographic structure of haem-bound haem oxygenase [37].

This part of the study clearly shows that the structure observed experimentally corresponds to a stable Fe(III) distorted conformation of the cofactor with the Glu24 coordinating the metal. This structure can only be understood if considering that the coordination rules of the metal drive the tuning between the inorganic and proteic moieties. The Fe(III)Schiff base moiety, by being smaller and more flexible than the haem in the natural system, affords an additional coordination site that can be reached by a residue acting as a Lewis basis. Hence, it appears that all possible conformational changes in the cofactor or the proteic receptor that could satisfy the best coordination of the metal would take place even if both chemical and proteic systems should be penalized.

## 4. Conclusion

In this study, we embarked on the analysis of a novel resting state and investigated the nature of the activation process of a haemoenzyme to finally discuss the results in terms of physico-chemical properties governing the interaction between inorganic moieties and proteins. This study clearly shows that, despite the structural similarity of the artificial cofactor to the naturally occurring haem, the structure reported by

Ueno *et al*. represents indeed the conformation of the resting state of the enzyme and corresponds to a high-spin system. Despite the well-known tendency of salophen to adopt planar geometries and the apo-HO system to coordinate the haem through a unique histidine, the molecular rules that govern the structure of the resting state of the novel enzyme and its activation process are imposed by the first coordination sphere of the metal. The removal of the side chain of the Glu24 needs though a synergic mechanism driven by the first coordination sphere of the metal but implying an overall structural change in the relative position of the proximal helix and the conformation of the cofactor. These results show that foreseeing the degree of conformational variability of the inorganic cofactor and the proteic receptor and any possible ligand exchange that could take place in the coordination sphere of the metal is a significant variable for the design of novel biometallic hybrids. This work sheds light on the major impact of the metal in dictating the structure of bioinorganic composites and provides key information for the rational design of novel members of this family of catalysts.

Funding statement. We are particularly thankful to the Spanish 'Ministerio de Economía y Competividad' for financial support through projects CTQ2011-23336 and ORFEO Consolider-Ingenio 2010 Programme (grant no. CSD2007-00006), the Generalitat de Catalunya through project 2009SGR68, and E.O-C. thanks the Universitat Autònoma de Barcelona for her scholarship (UAB-PIF).

## References

1. Lu Y, Yeung N, Sieracki N, Marshall NM. 2009 Design of functional metalloproteins. *Nature* **460**, 855–862. (doi:10.1038/nature08304)
2. Steinreiber J, Ward TR. 2008 Artificial metalloenzymes as selective catalysts in aqueous media. *Coordin. Chem. Rev* **252**, 751–766. (doi:10.1016/j.ccr.2007.09.016)
3. Rosati F, Roelfes G. 2010 Artificial metalloenzymes. *ChemCatChem* **2**, 916–927. (doi:10.1002/cctc.201000011)
4. Ringenberg MR, Ward TR. 2011 Merging the best of two worlds: artificial metalloenzymes for enantioselective catalysis. *Chem. Commun.* **47**, 8470–8476. (doi:10.1039/c1cc11592h)
5. Talbi B, Haquette P, Martel A, de Montigny F, Fosse C, Cordier S, Roisnel T, Jaouen G, Salmain M. 2010 (η6-Arene) ruthenium(II) complexes and metallo-papain hybrid as Lewis acid catalysts of Diels–Alder reaction in water. *Dalton Trans.* **24**, 5605–5607. (doi:10.1039/c001630f)
6. Podtetenieff J, Taglieber A, Bill E, Reijerse EJ, Reetz MT. 2010 An artificial metalloenzyme: creation of a designed copper binding site in a thermostable protein. *Angew. Chem.* **122**, 5277–5281. (doi:10.1002/ange.201002106)
7. Dürrenberger M *et al*. 2011 Artificial transfer hydrogenases for the enantioselective reduction of cyclic imines. *Angew. Chem.* **123**, 3082–3085. (doi:10.1002/ange.201007820)
8. Creus M, Pordea A, Rossel T, Sardo A, Letondor C, Ivanova A, Letrong I, Stenkamp RE, Ward TR. 2008 X-ray structure and designed evolution of an artificial transfer hydrogenase. *Angew. Chem. Int. Edit.* **47**, 1400–1404. (doi:10.1002/anie.200704865)
9. Pordea A, Creus M, Panek J, Duboc C, Mathis D, Novic M, Ward TR. 2008 Artificial metalloenzyme for enantioselective sulfoxidation based on vanadyl-loaded streptavidin. *J. Am. Chem. Soc.* **130**, 8085–8088. (doi:10.1021/ja8017219)
10. Rousselot-Pailley P, Bochot C, Marchi-Delapierre C, Jorge-Robin A, Martin L, Fontecilla-Camps JC, Cavazza C, Ménage S. 2009 The protein environment drives selectivity for sulfide oxidation by an artificial metalloenzyme. *Chembiochem* **10**, 545–552. (doi:10.1002/cbic.200800595)
11. Ricoux R, Allard M, Dubuc R, Dupont C, Maréchal J-D, Mahy J-P. 2009 Selective oxidation of aromatic sulfide catalyzed by an artificial metalloenzyme: new activity of hemozymes. *Org. Biomol. Chem.* **7**, 3208–3211. (doi:10.1039/b907534h)
12. Bos J, Roelfes G. 2014 Artificial metalloenzymes for enantioselective catalysis. *Curr. Opin. Chem. Biol.* **19**, 135–143. (doi:10.1016/j.cbpa.2014.02.002)
13. Ward TR. 2009 *Bio-inspired catalysis*. Heidelberg, Germany: Springer.
14. Pratter SM, Konstantinovics C, Di Giuro CML, Leitner E, Kumar D, de Visser SP, Grogan G, Straganz GD. 2013 Inversion of enantioselectivity of a mononuclear non-heme iron(II)-dependent hydroxylase by tuning the interplay of metal-center geometry and protein structure. *Angew. Chem.* **125**, 9859–9863. (doi:10.1002/ange.201304633)
15. Lind MES, Himo F. 2013 Quantum chemistry as a tool in asymmetric biocatalysis: limonene epoxide hydrolase test case. *Angew. Chem.* **125**, 4661–4665. (doi:10.1002/ange.201300594)
16. Reetz MT, Peyralans JJ-P, Maichele A, Fu Y, Maywald M. 2006 Directed evolution of hybrid enzymes: evolving enantioselectivity of an achiral Rh-complex anchored to a protein. *Chem. Commun.* **2006**, 4318–4320. (doi:10.1039/b610461d)
17. Muñoz Robles V, Maréchal J-D, Bahloul A, Sari M-A, Mahy J-P, Golinelli-Pimpaneau B. 2012 Crystal structure of two anti-porphyrin antibodies with peroxidase activity. *PLoS ONE*, **7**, e51128. (doi:10.1371/journal.pone.0051128)
18. Cavazza C, Bochot C, Rousselot-Pailley P, Carpentier P, Cherrier MV, Martin L, Marchi-Delapierre C, Fontecilla-Camps JC, Ménage S. 2010 Crystallographic snapshots of the reaction of aromatic C-H with $O_2$ catalysed by a protein-bound iron complex. *Nat. Chem.* **2**, 1069–1076. (doi:10.1038/nchem.841)
19. Ueno T, Yokoi N, Unno M, Matsui T, Tokita Y, Yamada M, Ikeda-Saito M, Nakajima H, Watanabe Y. 2006 Design of metal cofactors activated by a protein–protein electron transfer system. *Proc. Natl Acad. Sci. USA* **103**, 9416–9421. (doi:10.1073/pnas.0510968103)
20. Sicking W, Korth H-G, Jansen G, de Groot H, Sustmann R. 2007 Hydrogen peroxide decomposition by a non-heme iron(III) catalase mimic: a DFT study. *Chemistry* **13**, 4230–4245. (doi:10.1002/chem.200601209)
21. Kuramochi H, Noodleman L, Case DA. 1997 Density functional study on the electronic structures of model peroxidase compounds I and II. *J. Am. Chem. Soc.* **119**, 11 442–11 451. (doi:10.1021/ja970574c)
22. Aissaoui H, Bachmann R, Schweiger A, Woggon W-D. 1998 On the origin of the low-spin character of cytochrome P450(cam) in the resting state: investigations of enzyme models with pulse EPR and ENDOR spectroscopy. *Angew. Chem. Int. Edit.* **37**, 2998–3002. (doi:10.1002/(SICI)1521-3773(19981116)37:21<2998::AID-ANIE2998>3.0.CO;2-P)
23. Sono M, Roach MP, Coulter ED, Dawson JH. 1996 Heme-containing oxygenases. *Chem. Rev.* **96**, 2841–2888. (doi:10.1021/cr9500500)
24. Balding PR, Porro CS, McLean KJ, Sutcliffe MJ, Maréchal JD, Munro AW, De Visser SP. 2008 How do azoles inhibit cytochrome P450 enzymes? A density functional study. *J. Chem. Phys. A* **50**, 12 911–12 918. (doi:10.1021/jp802087w)
25. Senn HM, Thiel W. 2009 QM/MM methods for biomolecular systems. *Angew. Chem. Int. Edit.* **48**, 1198–1229. (doi:10.1002/anie.200802019)
26. Kozísek M *et al*. 2008 Molecular design of specific metal-binding peptide sequences from protein fragments: theory and experiment. *Chemistry* **14**, 7836–7846. (doi:10.1002/chem.200800178)
27. Muñoz Robles V, Ortega-Carrasco E, González Fuentes E, Lledós A, Maréchal J-D. 2011 What can molecular modelling bring to the design of artificial inorganic cofactors? *Faraday Discuss.* **148**, 137–159. (doi:10.1039/c004578k)
28. Alí-torres J, Maréchal J-D, Rodríguez-Santiago L, Sodupe M. 2011 Three dimensional models of $Cu^{2+}$–Aβ(1–16) complexes from computational approaches *J. Am. Chem. Soc.* **133**, 15 008–15 014. (doi:10.1021/ja203407v)
29. Frisch MJ *et al*. 2009 *Gaussian 09, revision A.02.*
30. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE. 2004 UCSF Chimera? A visualization system for exploratory

research and analysis. *J. Comput. Chem.* **25**, 1605 – 1612. (doi:10.1002/jcc.20084)

31. Ke Z, Abe S, Ueno T, Morokuma K. 2012 Catalytic mechanism in artificial metalloenzyme: QM/MM study of phenylacetylene polymerization by rhodium complex encapsulated in apo-ferritin. *J. Am. Chem. Soc.* **37**, 15 418 – 15 429. (doi:10.1021/ja305453w)
32. Dunbrack RL. 2002 Rotamer libraries in the 21st century. *Curr. Opin. Struct. Biol.* **12**, 431 – 440. (doi:10.1016/S0959-440X(02)00344-5)
33. Cornell WD *et al*. 1995 A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *J. Am. Chem. Soc.* **118**, 2309. [Erratum in *J. Am. Chem. Soc.* **117**, 5179 – 5197] (doi:10.1021/ja955032e)
34. Wang J, Wolf RM, Caldwell JW, Kollman PA, Case DA. 2004 Development and testing of a general amber force field. *J. Comput. Chem.* **25**, 1157 – 1174. (doi:10.1002/jcc.20035)
35. Becke AD. 1993 Density-functional thermochemistry. III. The role of exact exchange. *J. Chem. Phys.* **98**, 5648 – 5653. (doi:10.1063/1.464913)
36. Lee C, Yang W, Parr RG. 1988 Development of the Colle – Salvetti correlation-energy formula into a functional of the electron density. *Phys. Rev. B* **37**, 785 – 789. (doi:10.1103/PhysRevB.37.785)
37. Krishnan R, Binkley JS, Seeger R, Pople JA. 1980 Contracted Gaussian basis sets for molecular calculations. I. Second row atoms, $Z = 11 – 18$. *J. Chem. Phys.* **72**, 5639 – 5644. (doi:10.1063/1.438980)
38. Chandrasekhar J, Spitznagel GW, Schleyer PVR, Clark T. 1983 Efficient diffuse function-augmented basis sets for anion calculations. III. The 3-21+G basis set for first-row elements, Li-F. *J. Comput. Chem.* **4**, 294 – 301. (doi:10.1002/jcc.540040303)
39. Dunning JTH. 1989 Gaussian basis sets for use in correlated molecular calculations. I. The atoms boron through neon and hydrogen. *J. Chem. Phys.* **90**, 1007 – 1024. (doi:10.1063/1.456153)
40. Dunning Jr TH, Harrison RJ, Kendall RA. 1992 Electron affinities of the first-row atoms revisited. Systematic basis sets and wave functions. *J. Chem. Phys.* **96**, 6796 – 6807. (doi:10.1063/1.462569)
41. Zhao Y, Truhlar DG. 2006 A new local density functional for main group thermochemistry, transition metal bonding, thermochemical kinetics, and noncovalent interactions. *J. Chem. Phys.* **125**, 1 – 18. (doi:10.1063/1.2370993)
42. Perdew JP, Burke K, Ernzerhof M. 1996 Generalized gradient approximation made simple. *Phys. Rev. Lett.* **77**, 3865 – 3868. (doi:10.1103/PhysRevLett.77.3865)
43. Grimme S. 2006 Semiempirical GGA-type density functional constructed with a long-range dispersion correction. *J. Comput. Chem.* **27**, 1787 – 1799. (doi:10.1002/jcc.20495)
44. Swart M. 2008 Accurate spin-state energies for iron complexes. *J. Chem. Theory Comput.* **4**, 2057 – 2066. (doi:10.1021/ct800277a)
45. Rydberg P, Olsen L. 2009 The accuracy of geometries for iron porphyrin complexes from density functional theory. *J. Phys. Chem. A* **113**, 11 949 – 11953. (doi:10.1021/jp9035716)
46. Strickland N, Harvey JN. 2007 Spin-forbidden ligand binding to the ferrous-heme group: ab initio and DFT studies. *J. Phys. Chem. B* **111**, 841 – 852. (doi:10.1021/jp064091j)
47. Meunier B, de Visser SP, Shaik S. 2004 Mechanism of oxidation reactions catalyzed by cytochrome P450 enzymes. *Chem. Rev.* **104**, 3947 – 3980. (doi:10.1021/cr020443g)
48. Rahman MN, Vlahakis JZ, Szarek WA, Nakatsu K, Jia Z. 2008 X-ray crystal structure of human heme oxygenase-1 in complex with 1-(adamantan-1-yl)-2-(1 H-imidazol-1-yl)ethanone: a common binding mode for imidazole-based heme oxygenase-1 inhibitors *J. Med. Chem.* **51**, 5943 – 5952. (doi:10.1021/jm800505m)
49. Harvey JN. 2006 On the accuracy of density functional theory in transition metal chemistry. *Ann. Rep. Sect. C Phys. Chem.* **102**, 203 – 226. (doi:10.1039/b419105f)

# 5 Programmed stereoselective assembly of DNA-binding helical metallopeptides

Gamba, I.; Rama, G.; Ortega-Carrasco, E.; Maréchal, J. D.; Martínez-Costas, J.; Vázquez, M. E.; López, M. V. *Chem. Commun.*, **2014**, *50*, 11097-11100.

Volume 50 | Number 76 | 4 October 2014 | Pages 11065–11244

ChemComm

Chemical Communications

www.rsc.org/chemcomm

ISSN 1359-7345

ROYAL SOCIETY OF CHEMISTRY

COMMUNICATION

M. Eugenio Vázquez, Miguel Vázquez López *et al.*

Programmed stereoselective assembly of DNA-binding helical metallopeptides

ChemComm

COMMUNICATION

View Article Online
View Journal

# Programmed stereoselective assembly of DNA-binding helical metallopeptides†

Cite this: DOI: 10.1039/c4cc03606a

Received 13th May 2014,
Accepted 4th July 2014

DOI: 10.1039/c4cc03606a

www.rsc.org/chemcomm

Ilaria Gamba,[a] Gustavo Rama,[a] Elizabeth Ortega-Carrasco,[b] Jean-Didier Maréchal,[b] José Martínez-Costas,[c] M. Eugenio Vázquez*[d] and Miguel Vázquez López*[a]

**A flexible and versatile synthetic approach for the construction of water-stable DNA-binding chiral peptide helicates based on the solid phase peptide synthesis (SPPS) methodology is reported.**

The development of non-natural agents that can reproduce the DNA recognition properties of natural transcription factors remains a major goal in biological chemistry,[1] as they might have a great impact on fundamental and applied biological research, and even lead to the development of gene-targeted therapies.[2] Most of the efforts in this area have been directed towards the development of small organic binders, but in the past few years there has been growing interest in the use of coordination and metallo-supramolecular compounds with unique structural and spectroscopic properties.[3]

Helicates have shown promising and unusual DNA-binding properties,[4] including major groove binding,[5] and recognition of DNA three-way junctions.[6] However, nearly 20 years after the pioneering studies by Prof. Jean-Marie Lehn,[7] helicates are still not viable alternatives to traditional DNA-binding agents. The slow development in the biological chemistry of metal helicates ultimately derives from the lack of efficient and versatile methodologies for their enantioselective synthesis,[8] which usually results in mixtures of isomers difficult to separate,[9] insoluble or unstable species in water,[10] and above all, lack of versatility for structural and functional optimization.[11] Prof. Scott recently proposed a new synthetic approach to obtain DNA-binding helicates based on the self-assembly of aldehydes and amines in the presence of M(II) ions that allows the formation of water-stable and optically pure helicates.[12] Although this method represents a great improvement over the traditional routes, it does not guarantee the straightforward access to structural variants to efficiently study the chemical and functional space.

Given our experience in the study of DNA recognition agents,[13] DNA-binding peptides,[14] metallopeptides,[15] and helicates,[10] we decided to investigate this problem by developing an alternative approach to synthesize DNA-binding chiral helicates. Herein we report our proposal that potentially combines the biocompatibility, modularity and structural control of peptides with the synthetic flexibility and versatility of solid-phase peptide synthesis (SPPS) methodology, which allows the straightforward construction and modification of the helicates and, therefore, the easy optimization of their biophysical properties.

2,2′-Bipyridine (Bpy) is a privileged chelator that leads to stable complexes with many metal ions that have been extensively used in coordination and supramolecular chemistry.[16] Thus, we decided to synthesize a Bpy analog appropriately modified for its application in SPPS as the basic component of our peptide helicates. Hence, the Bpy unit was derivatized with 5-amino-3-oxapentanoic acid (O1Pen) as a Fmoc-protected achiral amino acid (Fmoc-O1PenBpy-OH (**1**); Scheme 1 and ESI,† Scheme S1).[15]

Following the synthesis of the amino acid building block, we designed a peptidic ligand capable of folding into a single-stranded dinuclear hairpin helicate; the selected peptide sequence contains six O1PenBpy units, which satisfy the coordination requirements of two octahedral metal centers, arranged in three sets (or substrands) of two consecutive Bpy residues connected through two short loops. The loops, whose structures have been previously optimized,[15b] are key components of this design, as they include a β-turn promoting -[(D/L)-Pro]-Gly- sequence that pre-organizes the ligand and induces the folding of the peptidic chain into discrete dinuclear species, thus avoiding the formation of higher

---

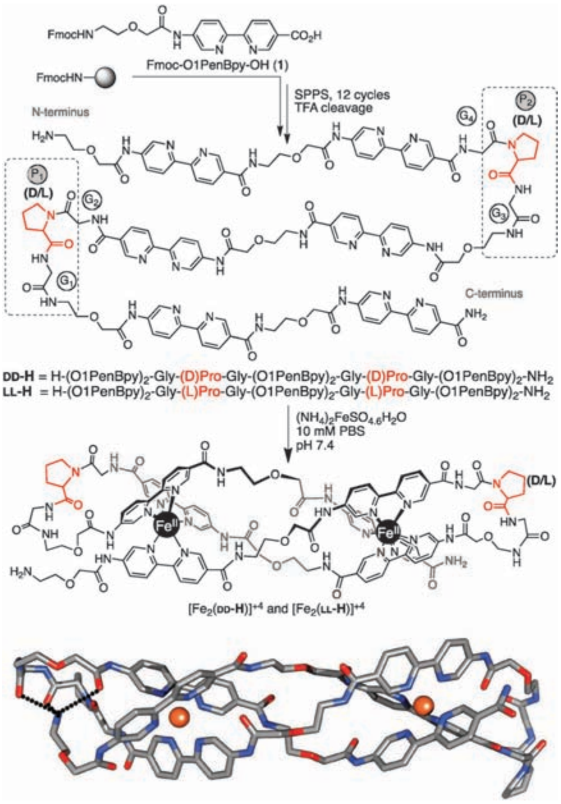
[a] Departamento de Química Inorgánica, Centro Singular de Investigación en Química Biolóxica e Materiais Moleculares (CiQUS), Universidade de Santiago de Compostela, 15782 Santiago de Compostela, Spain. E-mail: miguel.vazquez.lopez@usc.es

[b] Departament de Química, Universitat Autònoma de Barcelona, 08193 Cerdanyola, Spain

[c] Departamento de Bioquímica y Biología Molecular, Centro Singular de Investigación en Química Biolóxica e Materiais Moleculares (CiQUS), Universidade de Santiago de Compostela, 15782 Santiago de Compostela, Spain

[d] Departamento de Química Orgánica, Centro Singular de Investigación en Química Biolóxica e Materiais Moleculares (CiQUS), Universidade de Santiago de Compostela, 15782 Santiago de Compostela, Spain. E-mail: eugenio.vazquez@usc.es

† Electronic supplementary information (ESI) available: Synthetic procedures and characterization of peptide ligands and helicates, computational details, and spectroscopic studies (UV, CD and melting). See DOI: 10.1039/c4cc03606a



**Scheme 1** Solid-phase synthesis of the peptide ligands **LL-H** and **DD-H**. The L-Pro or D-Pro residues in the loops select a particular supramolecular chirality in the final helicates (ΛΛ – or ΔΔ –, respectively). Below, molecular representation of the most stable ΛΛ-$[Fe_2(\mathbf{LL\text{-}H})]^{4+}$ helicate highlighting the interaction of the ammonium group of the N-terminal end with the C-terminal loop.

order and polymeric complexes in the presence of metal ions (Scheme 1).

In order to gain some information about the structural viability of our design, molecular modeling was performed on the Fe(II) helicate derived from the peptide ligand **LL-H** (Scheme 1).[17] The system was optimized following a hierarchical protocol that consists of: (1) molecular dynamics simulation on the loops of an initial structure of the metallopeptide candidate; (2) clustering of the resulting MD, and (3) refinement by QM/MM geometry minimization of the most populated cluster representatives (see ESI,† Fig. S1). The final calculation was performed with solvent and electronic embedding effects. The lowest energy model presents a helical geometry (Scheme 1) with both Fe(II) ions in octahedral configurations and the six bipyridine rings ordered around the metal centers. The first coordination sphere of the metal ions is barely constrained by the entire peptide framework, indicating that an excellent complementarity between organic and inorganic moieties is achieved in this design.

Supported by the computational results, we synthesized two enantiomeric peptide ligands, each of them equipped with two L-Pro (**LL-H**) or D-Pro residues (**DD-H**) in their loops (Scheme 1). Both peptidic strands were obtained following standard Fmoc/*t*Bu solid-phase protocols,[18] and purified by reverse-phase HPLC (see ESI†). Incubation of ≈4.0 μM solutions of **LL-H** and **DD-H** in PBS buffer (pH = 7.4, 298 K) with increasing concentrations of Fe(II) resulted in a bathochromic shift of the bipyridine absorption band from 304 to 322 nm, as well as in the appearance of a new band centered at 545 nm, typical of octahedral trisbipyridine Fe(II) complexes (see ESI,† Fig. S2 and S4a).[19] The increase in the absorption intensity of the d–d band was used to calculate the binding constants, which were the same within the experimental error for both **LL-H** and **DD-H**,[20] thus for the first association $\beta_{1,1}(\mathbf{LL\text{-}H}) \approx \beta_{1,1}(\mathbf{DD\text{-}H}) \approx 12.9$ μM, and the overall formation constants $\beta_{2,1}(\mathbf{LL\text{-}H}) \approx \beta_{2,1}(\mathbf{DD\text{-}H}) \approx 18.9$ μM.[21,22] The successful assembly of the helicate, which can be considered as quantitative, was further confirmed by MALDI-TOF mass spectrometry showing a major peak at 2337.9 (see ESI†). To our knowledge, SPPS has never been used for the preparation of helicates. Moreover, fully-peptide ligands have not been used before as precursors of metal helicates, and there are only a handful of examples of organic strands equipped with oligopeptides as spacers or functional appendages.[23]

As expected, **LL-H** and its enantiomer **DD-H** give rise to mirror image CD spectra with two bands of opposite signs at approximately 290 and 324 nm (Fig. 1, and ESI†). The intensity of these bands decreases upon addition of Fe(II), while at the same time they display a bathochromic shift to 302 and 333 nm, respectively. The sign of the Cotton effect of the helicates is consistent with a Λ-configuration on each metal center for the **LL-H** derivative (ΛΛ-$[Fe_2(\mathbf{LL\text{-}H})]^{4+}$), and a Δ-configuration for the **DD-H** analog (ΔΔ-$[Fe_2(\mathbf{DD\text{-}H})]^{4+}$).[24] These data confirm that the enantiomeric peptide ligands give rise to the chiral helicates.[25] This means that the chiral proline residues do not only direct the folding of the peptide chain into a discrete dinuclear helicate, but also encode its chirality, which is selected under thermodynamic control at room temperature in water media upon incubation with Fe(II) ions. Controlling the metal center configuration in helicates usually requires demanding synthetic procedures for obtaining chiral organic ligands,[26] which complicates the access

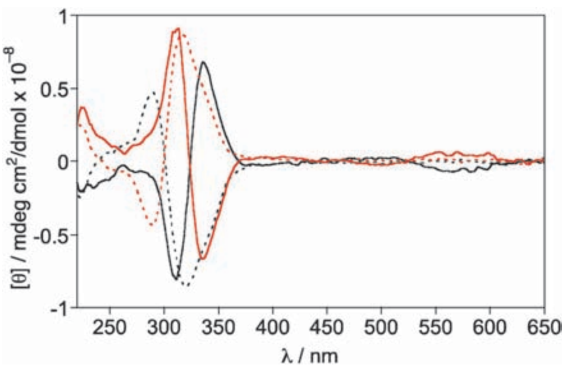


**Fig. 1** CD spectra of the **LL-H** and **DD-H** peptide ligands (50 μM) and their helicates. **LL-H** before (black dashed line) and after (black solid line, ΛΛ-$[Fe_2(\mathbf{LL\text{-}H})]^{4+}$) addition of 2 eq. of Fe(II) and **DD-H** before (red dashed line) and after (red solid line, ΔΔ-$[Fe_2(\mathbf{DD\text{-}H})]^{4+}$) addition of 2 eq. of Fe(II). All measurements were made in PBS buffer pH = 7.4, 298 K.

to multiple structural variants that are required for the systematic studies involved in the optimization of their biological properties. In our approach, however, the chirality of the peptide helicates arises naturally from two single Pro residues located in the loops connecting the three bisBpy substrands (Scheme 1).

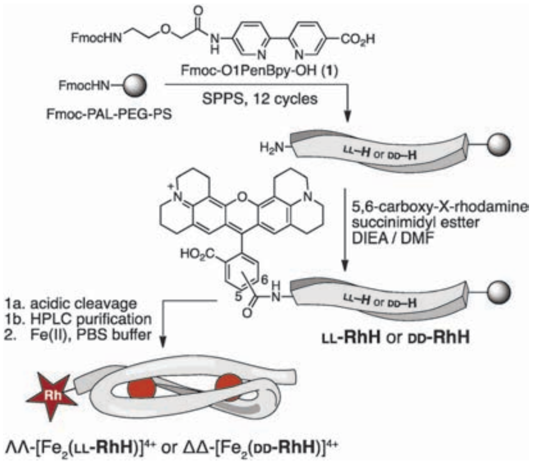
The chiral peptide helicates are very stable in water at room temperature, and no degradation or racemization was detected even after several weeks. Additionally, thermal denaturation experiments produced sigmoidal melting profiles for both helicates that were consistent with cooperative transitions in the decomplexation of the Fe(II) ions (see ESI,† Table S1, Fig. S5–S7). Both melting curves were qualitatively similar, with a $T_m$ temperature of ≈80.5 °C. Van't Hoff analysis showed that the unfolding process displays a positive entropic variation ($\Delta S \approx 447.1$ J mol$^{-1}$), consistent with the increase in molecular disorder resulting from the disassembly of the complexes; the unfolding process is also highly endothermic ($\Delta H \approx 157.9$ kJ mol$^{-1}$), in agreement with the large association constants obtained in the UV/Vis titrations and reported studies of related supramolecular metal complexes.[27] Alternatively, global analysis of the unfolding process using the Gibbs–Helmholtz equation provided qualitatively similar results.[28]

Having at hand the methodology for the efficient assembly of water-stable and chiral peptide helicates, we decided to explore the DNA binding properties of these species. In particular, we were interested in studying the affinity and chiral selectivity of their binding to the three-way DNA junction,[29,30] data that remain unknown until now.

Towards this end, we synthesized the N-terminal rhodamine labeled derivatives of the previously described LL-**H** and DD-**H** peptide ligands. In short, once the core peptide sequences were fully assembled, and still attached to the solid support, their N-terminal amines were reacted with 5-(and 6)-carboxy-X-rhodamine succinimidyl ester to yield the fluorescent peptidic ligands LL-**RhH** and DD-**RhH**, which were then cleaved from the resin and purified as described before, and incubated with 2 eq. of Fe(II) to form the corresponding fluorescent helicates ΛΛ-[Fe$_2$(LL-**RhH**)]$^{4+}$ and ΔΔ-[Fe$_2$(DD-**RhH**)]$^{4+}$ (Scheme 2).

Incubation of 1 μM solutions of the rhodamine-labeled helicates with increasing concentrations of the DNA three-way junction $Y_1Y_2Y_3$[31] (see ESI†) produced a concentration-dependent increase in the fluorescence anisotropy that could be fitted to a modified 1 : 1 binding mode, including higher order aggregates resulting from non-specific interactions (Fig. 2).[14] The ΛΛ-[Fe$_2$(LL-**RhH**)]$^{4+}$ peptide helicate displayed significantly higher affinity—almost 150 fold—for the $Y_1Y_2Y_3$ DNA junction than the enantiomeric ΔΔ-[Fe$_2$(DD-**RhH**)]$^{4+}$, with apparent dissociation constants for the 1 : 1 complexes of 0.25 ± 0.04 μM and 37.6 ± 1.0 μM, respectively.[32,33]

The structure of the peptide helicate obtained by molecular modeling is comparable to that of the simpler helicates previously described by Hannon *et al.* Specifically, the three-fold symmetry of the peptide strands folding around the metal ions creates a compact structure with the right diameter for insertion into the cavity at the center of the three-way junction. These similarities, together with the tight binding constants observed in the fluorescence titrations, support a similar binding model with the DNA junction (see ESI,† Fig. S9).[6]



Scheme 2 Synthesis of the rhodamine-labeled peptide ligands LL-**RhH** and DD-**RhH** by solid phase modification of the core peptides, and their corresponding helicates (ΛΛ-[Fe$_2$(LL-**RhH**)]$^{4+}$ or ΔΔ-[Fe$_2$(DD-**RhH**)]$^{4+}$ respectively). See Scheme 1 for complete sequences of LL-**H** and DD-**H**.

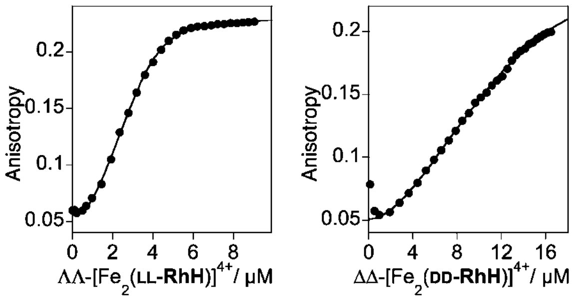


Fig. 2 Fluorescence anisotropy titrations of 1 μM solutions of peptide helicates with a DNA three-way junction ($Y_1Y_2Y_3$) in 10 mM PBS buffer, NaCl 10 mM; pH: 7.3, 20 °C. ΛΛ-[Fe$_2$(LL-**RhH**)]$^{4+}$ (left) and ΔΔ-[Fe$_2$(DD-**RhH**)]$^{4+}$ (right). Lines represent the best fit to 1 : 1 binding models including the contribution of nonspecific complexes. DNA oligo sequences: $Y_1$: 5′-CAC CGC TCT GGT CCT C-3′; $Y_2$: 5′-CAG GCT GTG AGC GGT G-3′; $Y_3$: 5′-GAG GAC CAA CAG CCT G-3′.

Once we characterized the basic properties of the peptide helicates, we decided to exploit the presence of the fluorescent rhodamine dye to study their cell-internalization. Thus, we incubated Vero cells with ΛΛ-[Fe$_2$(LL-**RhH**)]$^{4+}$ for 12 h, washed them with PBS, and observed the preparations without fixation. Gratifyingly, we found that the compound was efficiently internalized into the cells, concentrating in endocytic vesicles in the cytoplasm, which is consistent with the internalization pathway observed for other metal complexes,[34] and luminescent helicates (see ESI,† Fig. S10).[35]

In summary, we present a simple and versatile methodology for the enantioselective assembly of water-stable DNA-binding helicates. This novel approach relies on the modularity and synthetic flexibility of SPPS and on the preorganization of the peptide ligands provided by two single proline residues, which also encode the chirality of the helicates formed under thermodynamic control. We demonstrated the potential of this approach by synthesizing fluorescently-labeled helicates that show chiral discrimination and high selectivity for three-way DNA junctions and promising cell-internalization properties.

We are thankful for the support given by the Spanish grants SAF2010-20822-C02, CTQ2012-31341, CTQ2011-23336, CCTQ2010-16959 Consolider Ingenio 2010 CSD2007-00006, the Xunta de Galicia, GRC2010/12, GRC2013-041, PGIDIT08CSA-047209PR, and the Generalitat de Catalunya, 2009SGR68. Support of COST Action CM1105 is kindly acknowledged. G.R. thanks the INL for his PhD fellowship and E. O. the UAB for her PhD grant. M. V. L. dedicates this work to his newborn and third son, Iván.

## References

1 (*a*) E. Pazos, J. Mosquera, M. E. Vázquez and J. L. Mascareñas, *ChemBioChem*, 2011, **12**, 1958; (*b*) J. W. Højfeldt, A. R. Van Dyke and A. K. Mapp, *Chem. Soc. Rev.*, 2011, **40**, 4286.
2 R. Pollock, M. Giel, K. Linher and T. Clackson, *Nat. Biotechnol.*, 2002, **20**, 729.
3 *Metal Complex-DNA Interactions*, ed. N. Hadjiliadis and E. Sletten, Wiley, Chichester, UK, 2009.
4 (*a*) C. Piguet, G. Bernardinelli and G. Hopfgartner, *Chem. Rev.*, 1997, **97**, 2005; (*b*) M. Albrecht, *Chem. Rev.*, 2001, **101**, 3457.
5 I. Meistermann, V. Moreno, M. J. Prieto, E. Moldrheim, E. Sletten, S. Khalid, P. M. Rodger, J. C. Peberdy, C. J. Isaac, A. Rodger and M. J. Hannon, *Proc. Natl. Acad. Sci. U. S. A.*, 2002, **99**, 5069.
6 G. Oleski, A. G. Blanco, R. Boer, I. Usón, J. Aymamí, A. Rodger, M. J. Hannon and M. Coll, *Angew. Chem., Int. Ed.*, 2006, **45**, 1227.
7 J. M. Lehn, A. Rigault, J. Siegel, J. Harrowfield, B. Chevrier and D. Moras, *Proc. Natl. Acad. Sci. U. S. A.*, 1987, **84**, 2565–2569.
8 E. Terrazzi, L. Guénée, J. Varin, B. Bocquet, J.-F. Lemmonier, D. Emery, J. Mareda and C. Piguet, *Chem. – Eur. J.*, 2011, **17**, 184.
9 (*a*) M. J. Hannon, I. Meistermann, C. J. Isaak, C. Blomme, J. R. Aldrich-Wright and A. Rodger, *Chem. Commun.*, 2001, 1078; (*b*) N. Ousaka, Y. Takeyama and E. Yashima, *Chem. Sci.*, 2012, **3**, 466.
10 M. R. Bermejo, A. M. González-Noya, R. M. Pedrido, M. J. Romero and M. Vázquez, *Angew. Chem., Int. Ed.*, 2005, **44**, 4182.
11 L. Cardo, V. Sadovnikova, S. Phongtongpasuk, N. J. Hodges and M. J. Hannon, *Chem. Commun.*, 2011, **47**, 6575.
12 S. E. Howson, A. Bolhuis, V. Brabec, G. J. Clarkson, J. Malina, A. Rodger and P. Scott, *Nat. Chem.*, 2012, **4**, 31.
13 M. I. Sánchez, O. Vázquez, J. Martínez-Costas, M. E. Vázquez and J. L. Mascareñas, *Chem. Sci.*, 2012, **3**, 2383.
14 M. I. Sánchez, O. Vázquez, M. E. Vázquez and J. L. Mascareñas, *Chem. – Eur. J.*, 2013, **19**, 9923.
15 (*a*) I. Gamba, I. Salvadó, G. Rama, M. Bertazzon, M. I. Sánchez, V. M. Sánchez-Pedregal, J. Martínez-Costas, R. F. Brissos, P. Gámez, J. L. Mascareñas, M. Vázquez López and M. E. Vázquez, *Chem. – Eur. J.*, 2013, **19**, 13369; (*b*) G. Rama, A. Ardá, J.-D. Maréchal, I. Gamba, H. Ishida, J. Jiménez-Barbero, M. E. Vázquez and M. Vázquez López, *Chem. – Eur. J.*, 2012, **18**, 7030.
16 C. Kaes, A. Katz and M. W. Hossein, *Chem. Rev.*, 2000, **100**, 3553.
17 Fe(II) ions were selected due to the high stability exhibited by their bipyridine complexes in physiological media: J. L. Piere, M. Fontecave and R. R. Critchon, *BioMetals*, 2002, **15**, 341.
18 I. Coin, M. Beyermann and M. Bienert, *Nat. Protoc.*, 2007, **2**, 3247.
19 We have not observed the formation of precipitates during the titrations.
20 The overall formation constants are represented by the $\beta_{m,n}$ notation, where *m* and *n* refer to the stoichiometry of the metal ion and the ligand, respectively. Thus, $\beta_{m,n} = \beta(M_mL_n) = [M_mL_n]/([M]_m[L]_n)$.
21 R. Binstead and A. Zuberbühler, *SPECFIT/32 Global Analysis System*, TgK Scientific Limited.
22 These reported binding constants are only rough estimations, limited by the relatively high concentrations required for the UV/vis titrations: P. Thordarson, *Chem. Soc. Rev.*, 2011, **40**, 1305.
23 (*a*) M. Albrecht, O. Spiess, M. M. Schneider and P. Weis, *Chem. Commun.*, 2002, 786; (*b*) N. Ousaka, Y. Takeyama and E. Yashima, *Chem. Sci.*, 2012, **3**, 466; (*c*) E. C. Constable, C. E. Housecroft and S. Mundwiler, *J. Chem. Soc., Dalton Trans.*, 2003, 2112; (*d*) L. Cardo, V. Sadovnikova, S. Phongtongpasuk, N. J. Hodges and M. J. Hannon, *Chem. Commun.*, 2011, **47**, 6575.
24 M. Ziegler and A. von Zelewsky, *Coord. Chem. Rev.*, 1998, **177**, 257.
25 Studies with related mononuclear metallopeptides support the high stereoselectivity of the metal complexes derived from chiral peptides.
26 O. Mamula, A. von Zelewsky, P. Brodard, C.-W. Schläpfer, G. Bernardinelly and H. Stoeckli-Evans, *Chem. – Eur. J.*, 2005, **11**, 3049.
27 M. Ashram, M. Bqaeen and S. Mizyed, *J. Inclusion Phenom. Macrocyclic Chem.*, 2010, **67**, 81.
28 (*a*) N. J. Greenfield, *Nat. Protoc.*, 2006, **1**, 2527; (*b*) J. L. Neira, M. E. Vázquez and A. R. Fersht, *Eur. J. Biochem.*, 2000, **267**, 2859.
29 Q. Guo, M. Lu, M. E. Churchill, T. D. Tullius and N. R. Kallenbach, *Biochemistry*, 1990, **29**, 10927.
30 Previously reported data (ref. 6 and 33) indicate that three-way junctions are the preferred targets for these structures.
31 Control studies show that B-DNA is not a high-affinity target (see ESI,† Fig. S8).
32 The corresponding fluorescein analogs, $\Lambda\Lambda$-[Fe$_2$(LL-**FluH**)]$^{4+}$ and $\Delta\Delta$-[Fe$_2$(DD-**FluH**)]$^{4+}$, were also synthesized (see ESI†).
33 This is in agreement with the chiral discrimination displayed by other iron(II) helicates with DNA three-way junctions: J. Malina, M. J. Hannon and V. Brabek, *Chem. – Eur. J.*, 2007, **13**, 3871.
34 (*a*) W. Xu, J. Zuo, L. Wang, L. Ji and H. Chao, *Chem. Commun.*, 2014, **50**, 2123; (*b*) A. C. Komor and J. K. Barton, *Chem. Commun.*, 2013, **49**, 3617.
35 B. Song, C. D. B. Vandevyver, A.-S. Chauvin and J.-C. G. Bünzli, *Org. Biomol. Chem.*, 2008, **6**, 4125.

# 6 The folding of a metallopeptide

Gamba, I.; Rama, G.; Ortega-Carrasco, E.; Beradozzi, R.; Sánchez, V.M.; Di Bari, L.; Maréchal, J. D.; Vázquez, M. E.; López, M. V. *Submitted*

# The folding of a metallopeptide

Ilaria Gamba,[§] Gustavo Rama,[§] Elisabeth Ortega-Carrasco,[¶] Roberto Berardozzi,[∞] Víctor M. Sánchez Pedregal,[⌘] Lorenzo Di Bari,[∞] Jean-Didier Maréchal,*[,¶] M. Eugenio Vázquez*[,⌘] and Miguel Vázquez López*[,§]

[§]Departments of Inorganic Chemistry and [⌘]Organic Chemistry, Centro Singular de Investigación en Química Biolóxica e Materiais Moleculares (CiQUS), Universidade de Santiago de Compostela, 15782 Santiago de Compostela, Spain.

[¶]Department of Chemistry, Universitat Autónoma de Barcelona, 08193 Cerdanyola, Spain.

[∞]Dipartimento di Chimica e Chimica Industriale. Università di Pisa. 56124 Pisa, Italy.

**ABSTRACT:** We have applied solid-phase synthesis methods for the construction of tris(bipyridyl) peptidic ligands that coordinate Fe(II) ions with high affinity and fold into stable mononuclear metallopeptides. The main factors influencing the folding pathway and chiral control of the peptidic ligands around the metal ions has been studied both by experimental techniques (CD, UV-vis and NMR) and molecular modeling tools.

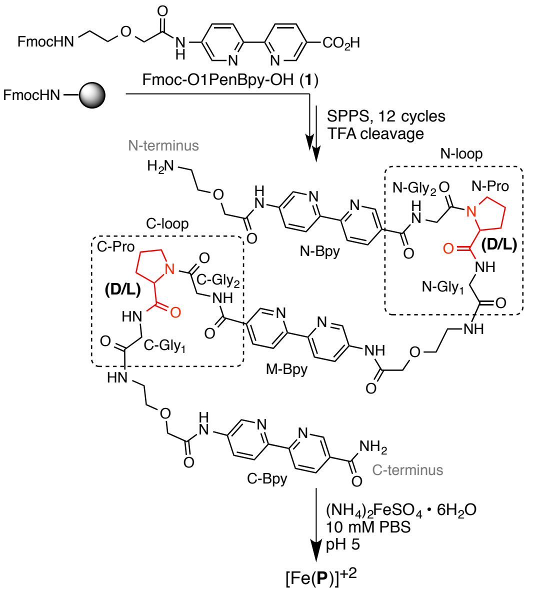
## INTRODUCTION

Understanding the way in which peptides and proteins fold into functional architectures in an autonomously-guided mechanism defined by the amino acid sequence, that is, the *protein folding problem*, has been a formidable scientific challenge since its identification more than 50 years ago. [1,2] Moreover, despite the increasing success in the prediction of the three dimensional structures of small proteins and peptides,[3] the study of metallopeptides—how they fold and misfold,[4] aggregate,[5] or interact with other molecules—is still in its infancy.[6,7] These studies are of great relevance, given the role of metal ions in the folding/misfolding of metalloproteins and also in relevant pathological processes, such as the induction of amyloid aggregation and precipitation in neurodegenerative diseases.[8] One of the main reasons for this underdevelopment is that the breakdown of the different energies involved in the folding of metallopeptides is very difficult to establish with peptides coded by the 20 natural amino acids, because too many physicochemical variables are involved when the donor atoms suitable for coordination are located in the side chains of the peptide sequence. In contrast, artificial metallopeptides, in which the metal-binding units are part of the main chain of the amino acid structure, represent excellent model systems for the study of metallopeptide folding, as they can be described with far less variables and offer much better coupling between the conformational preferences of the peptide chain and the coordinating properties of the metal ions.[9,10,11] Herein, we present a computational and experimental study of the folding and chiral control of a family of octahedral mononuclear metallopeptides containing metal-coordinating bipyridyl units as integral part of the peptide backbone.

## RESULTS

2,2'-bipyridine (Bpy) is a privileged metal chelator and, as such, has been extensively used in coordination and supramolecular chemistry.[12] We synthesized a Bpy analog appropriately modified for its application in solid phase peptide synthesis (SPPS) in which the Bpy unit was derivatized as a Fmoc-protected amino acid with 5-amino-3-oxapentanoic acid (Fmoc-O1PenBpy-OH, **1**, Scheme 1). Following the synthesis of the amino acid building block, we designed a set of six peptide ligands featuring three metal-binding bipyridine units connected by two short loops (**LL-P**, **DD-P**, **GD-P**, **GL-P**, **DL-P** and **LD-P**, Scheme1). The loops include a *β*-turn promoting sequence (Pro-Gly) that directs the folding of the peptide chains into discrete mononuclear species and encodes the chirality of their resulting complexes; [9] the N-terminal loop may contain a chiral proline (–Gly-(L)Pro-Gly– or –Gly-(D)Pro-Gly–), or an achiral sequence (–Gly-Gly-Gly–), while the C-terminal loop contains in all cases a chiral proline residue (–Gly-(L)Pro-Gly– or –Gly-(D)Pro-Gly–).[13] All the peptides were obtained using standard Fmoc/tBu solid-phase protocols,[14] and the final products were purified by reverse-phase HPLC and identified by MS (see ESI).

Following the synthesis of the peptide ligands, we first studied the thermodynamic stability of their metal complexes. Thus, incubation of low µM solutions of the peptide ligands **LL-P** (featuring homochiral Pro loops), **GL-P** (one chiral loop), and **DL-P** (heterochiral loops) at 298 K with Fe(II) ions resulted in qualitatively similar changes, most notably a clear bathochromic shift of the Bpy absorption band from 308 to approximately 316 nm due to the complexation processes (See ESI, Figures S1-S3). The binding constants derived from the UV/vis titrations indicate that all the complexes are very stable, so that the values of the formation constants are $\beta_{1,1} \approx 8$ for **LL-P** and **GL-P**, and $\beta_{1,1} \approx 9$ for the more stable heterochiral peptide complex with **DL-P** (ESI, Figures S4-6 and Table S1).[15] The assembly of the discrete mononuclear octahedral complexes was further confirmed by MALDI-TOF mass spectrometry (See ESI).[16]



| | N-loop | C-loop |
|---|---|---|
| **LL-P:** | –Gly–(L)Pro–Gly– | –Gly–(L)Pro–Gly– |
| **DD-P:** | –Gly–(D)Pro–Gly– | –Gly–(D)Pro–Gly– |
| **LD-P:** | –Gly–(L)Pro–Gly– | –Gly–(D)Pro–Gly– |
| **DL-P:** | –Gly–(D)Pro–Gly– | –Gly–(L)Pro–Gly– |
| **GL-P:** | –Gly–Gly–Gly– | –Gly–(L)Pro–Gly– |
| **GD-P:** | –Gly–Gly–Gly– | –Gly–(D)Pro–Gly– |

**Scheme 1.** Solid phase peptide synthesis of the tris(bipyridyl) peptide ligands **LL-P**, **DD-P**, **LD-P**, **DL-P**, **GL-P** and **GD-P**, as well as their corresponding Fe(II) mononuclear octahedral complexes.

As expected, the enantiomeric **LL-P** and **DD-P** ligands give rise to mirror image CD spectra characterized by two bands of opposite sign at 300 and 329 nm with a crossover at 314 nm, so that **LL-P** displays a positive couplet and **DD-P** a negative couplet. It is worth recalling that a CD couplet is defined a sequence of two bands of approximately equal amplitude and opposite signs, with crossover point close to the absorption maximum; the couplet is defined positive/negative according to the sign of its long-wavelength component. The intensity of these bands increases upon addition of 1 equivalent of Fe(II) ions and their increase is also accompanied by bathochromic shifts of both couplet components, respectively. Furthermore, upon iron chelation we observe the appearance of another broad bisignate feature allied to a charge transfer transitions around 550 nm. As observed in other cases before, this pair of bands in the *vis*-region have opposite sign sequence with respect to the UV-couplet.[17] The sign of the UV-couplet is consistent with a Λ configuration in the metal centre for [Fe(**LL-P**)]$^{2+}$ and with the opposite Δ configuration for [Fe(**DD-P**)]$^{2+}$ (Figure 1).[18] Interestingly, the CD spectra of the **LL-P**, **LD-P** and **DD-P** systems indicate that a single Pro residue in the sequence (namely, the N-terminal Pro) encodes the dominating chirality of the resulting metallopeptide. Thus, the CD spectra of **LD-P** and **LL-P**, as well as those of their corresponding Fe(II) complexes, [Fe(**LD-P**)]$^{2+}$ and [Fe(**LL-P**)]$^{2+}$ display a positive UV-couplet around 300 nm, and correspondingly a negative/positive sequence long-wavelength bands around 550 (Figure 1 and ESI, Figures S7-9). The presence of the UV couplet before complexation and the conservation of its sign (although with a significant increase in amplitude) after metal binding suggests that the bipyridyl moieties in the free peptides are to some extent preorganized to host the metal.

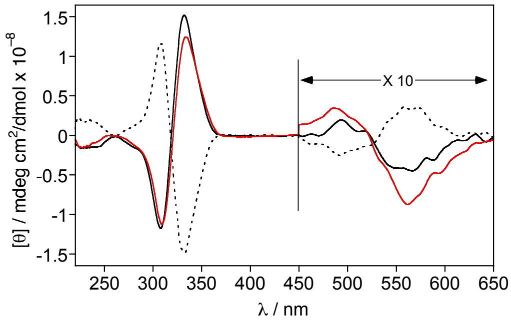


**Figure 1**. a) CD spectra of [Fe(**LL-P**)]$^{2+}$ (black line), [Fe(**LD-P**)]$^{2+}$ (red line), and [Fe(**DD-P**)]$^{2+}$ (dashed line). CD spectra were measured at 293 K, 10 mM PBS buffer, pH = 5.0.

While conducting these CD studies we became aware that the assembly of the Fe(II) metallopeptides is a kinetically slow process, and therefore we decided to monitor the folding of the peptide ligands in presence of Fe(II) ions. The homochiral ligands (**LL-P** and **DD-P**) displayed a monoexponential decrease (or increase) of their 332 nm CD signal at 40 °C upon addition of Fe(II), reaching a plateau after approximately 25 min (Figure 2a), consistent with a two-state process in which the unfolded peptides reach thermodynamic equilibrium as folded Fe(II)-metallopeptides ($U \rightleftarrows F$). The simple two-state process is also confirmed by the existence of a isodichroic point.[19] In contrast with this, the intensity of the CD signal at 332 nm of the peptides containing two proline residues with opposite chirality (**DL-P** and **LD-P**) displays a biphasic profile at 40 °C, and lacks the isodichroic point, so that a rapid increase (or decrease) in the CD intensity is followed by a slower exponential decay (or increase) (Figure 2b). In this case, the folding process cannot be simply described by the direct transition between an unfolded and a folded state ($U \rightleftarrows F$), but requires the consideration of an intermediate complex ($U \rightleftarrows I \rightleftarrows F$). Moreover, lowering the temperature from 40 °C to 20 °C resulted in an increase of the CD signal, but more importantly, it also induced a change in the kinetic profiles of the homochiral peptides, which become biphasic at 20 °C, thus suggesting the accumulation at this temperature of an intermediate complex similar to that observed for the mixed chirality peptides also at higher temperatures. In contrast with this, the mixed chirality peptide, **DL-P** as well as those with a single Pro residue, **GL-P** and **GD-P**, maintain their folding profiles, albeit with at significantly slower rate (see ESI for the kinetic constants, Table S2).

In all cases the observed changes in the CD only affect the intensity of the bands, but not their position, and these variations are not observed in the absorption spectra (ESI, Figures S13-21). This suggests that the variations in the CD intensity arise from changes in the relative position of the Bpy chromophores around the metal center as they rearrange to their most stable geometry through non-dissociative mechanisms, such as the Bailar or the Ray–Dutt twists isomerizations.[20]

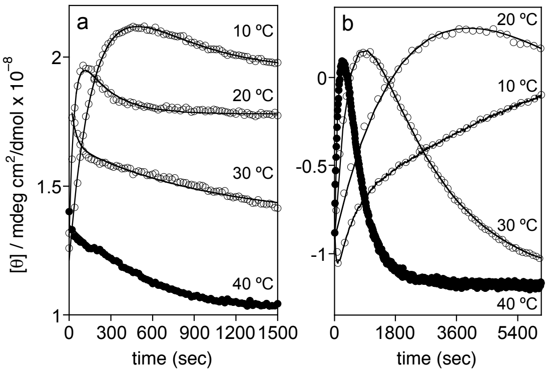


**Figure 2.** Evolution of the circular dichroism signal at 332 nm of **LL-P** (a), and **DL-P** (b) upon addition ot Fe(II) at various temperatures. **LL-P** Shows monoexponential decays at 40 and 30 °C, and the biphasic profiles only appear at lower temperatures. **DL-P** shows slower kinetics (note the timescale of the plots) with biphasic profiles at all temperatures.

At this point, we decided to further investigate the folding process of the metallopeptide assembly by experimental and theoretical tools. Surprisingly, HPLC analysis of the **DL-P** heterochiral peptide ligand in the presence of Fe(II) ions shows four peaks (ESI, Figure S22), suggesting the presence of four major isomers in solution. Moreover, these four complexes are in equilibrium, as evidenced by the observation of the same four peaks when each of the isolated peaks is injected back into the HPLC after a short equilibration time. **LL-P** and **GL-P** Fe(II) metallopeptides also show the same behavior. Moreover, the solution $^1$H-NMR spectrum of $[Fe(\mathbf{LL\text{-}P})]^{2+}$ shows a set of resonances compatible with four distinct isomers (ESI, Figures S23-28). We tried to assess whether these isomers are in chemical exchange equilibrium by variable temperature NMR in the 5-45 °C range (ESI, Figure S29), but no sign of chemical exchange was evident in this range. This is compatible with a slow rate of interconversion between the four isomers at the NMR timescale in the thermodynamic equilibrium suggested by the observations made by HPLC.

Molecular modeling and QM/MM optimization provided 3D models of the possible Fe(II) isomers derived from the **DL-P** ligand. These studies show that several (10) isomeric metallopeptide configurations differing in the folding of the peptidic chain around the metal center are energetically accessible (Figure 3 and ESI, Figures S30). It is worth noting that, for all isomers the lowest energy structures contain two *cis* prolines, which suggests that during the metallopeptide assembly process these residues isomerize from their more stable *trans* conformation in solution to satisfy the conformational requirements for metal coordination (see ESI Figure S31-33).[21] The energetic breakdown of the total QM/MM energy does not show major constraints on the first coordination sphere of the Fe(II) ion, and the fold of the rest of the peptide is mainly function of the energy of the loops (see ESI Figure S34). However, when considering the complete set of physical variables taken into account in our calculations, we observe that the major factor influencing the relative stability of the different isomers is the desolvation of the peptide chain that favors the more compact structures.[22]

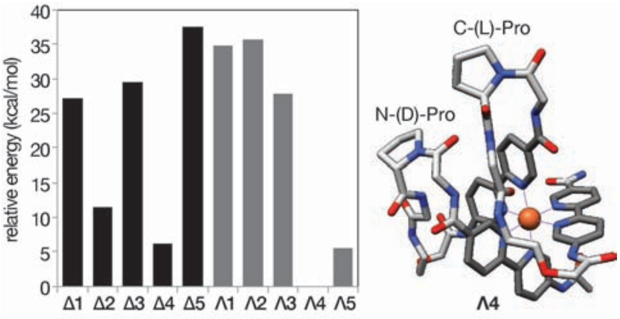


**Figure 3.** QM/MM relative energies of the possible Fe(II) complexes formed with the **DL-P** peptide ligand Δ isomers are shown in black, and Λ in light grey. The isomer labeled as Λ4, with the lowest energy of all the complexes is also represented on the right (bipyridines in darker shade of grey).

As we have seen, the CD spectra of the mixtures of isomers (Figure 1 and ESI, Figures S7-9) indicate an overall chiral selection governed by the chirality of the N-terminal Pro residue. In agreement with the experimental data, computation on the three metallopeptide systems (**LL-P/DD-P**, **GD-P/GL-P**, **DL-P/LD-P**) shows that the most stable geometry is independent of the chirality of the Pro residue in the C-terminal loop (ESI, Figures S31-33). We hypothesized that the relative position of the Pro residues (N-Pro or C-Pro) with respect to the central Bpy group (M-Bpy), as well as the asymmetry of the peptide ligands could be responsible of the dominance of the N-Pro residue in the selection of the overall chirality. An additional set of calculations was therefore performed with a 1ns MD of **DD-P** and **LL-P**. The analysis of the MD shows that the C-terminal Bpy unit (C-Bpy) is more flexible, and conformationally less defined than the N-Bpy group (Scheme 1), and also that the N-Bpy/M-Bpy pair is more preorganized than the M-Bpy/C-Bpy (ESI, Figure S35 and S36).

The ECD spectra for the 10 structures discussed above for the **DL-P** system (Figure 3; ESI, Figures S37-40) were simulated by means of time dependent DFT calculations (TD-DFT). To this end, we pruned the chromophores by replacing the chains at the sides of the amide groups with H-atoms. All the hydrogen atoms were reoptimized with standard DFT (CAM-B3LYP//SVP), clamping all the remaining atoms in their original positions. This truncation simplified the calculations, which otherwise would be excessively demanding, and it was justified by the fact that the chains connecting the bipyridyl amides provide only weak spectroscopic contributions. TDDFT at the same level used for geometry optimizations (CAM-B3LYP/SVP) provided the ECD spectra for the individual conformations shown in Figures S38 and S39. The proximity and dissymmetric orientation of the three chromophores is responsible for the typical exciton couplet structure, with D and L forms yielding positive and negative couplets, respectively.

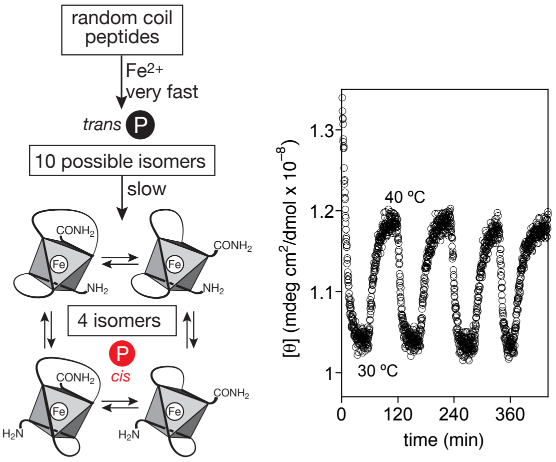


**Figure 4.** Left: Proposed complexation/folding mechanism of the metallopeptides involving initial coordination to the peptide ligands predominantly in *trans*-Pro configurations, and subsequent rearrangement and selection under thermodynamic control. Right: Circular dichroism of a 110 μM solution of $[Fe(\mathbf{LL\text{-}P})]^{2+}$ in 10 mM phosphate buffer pH 5.0 showing the reversible modification of the CD signal in response to changes in the temperature (30-40°C).

These theoretical predictions could be useful to explain the observations made during the CD *vs* time experiments above described. In fact, the timescale of the kinetic profiles suggests that the *trans* → *cis* isomerization of the proline residues is the underlying process promoting the reorganization of the peptide ligands around the metal ions.[23] This is in agreement with the known preference of Pro residues to present a *trans* conformation in short peptides in solution, and the molecular modeling studies showing that the M-Bpy/N-Bpy pair is significantly structured (and thus stabilizing an otherwise unstable *cis* proline in the N-loop). Moreover, Pro isomerization has also been proposed as the slow step in the denaturation/folding pathway of small proteins.[24] The folding of these Fe(II) trisbipyridyl metallopeptides could be explained as follows (Figure 4, left). In the initial state the free peptide ligands are poorly structured (particularly their C-terminal end), and have their Pro residues in *trans* configuration; the addition of Fe(II) ions to the solution of the peptide ligands gives rise to a number of isomeric metallopeptides in equilibrium, which progressively collapse into the observed set of four isomers with their Pro residues in the *cis* conformation. Interestingly, the position of the equilibrium between the selected isomers can be shifted by changes in the temperature. Thus for example, varying the temperature of a $[Fe(\mathbf{LL\text{-}P})]^{2+}$ solution between 30 and 40 ºC results in measurable changes in the CD spectrum, which can be reversibly switched between two extreme values (Figure 4, right).

## CONCLUSIONS

In summary, we describe the folding process of a family of trisbipyridyl Fe(II) mononuclear metallopeptides. The two-step folding process involves a rapid coordination of the ligand peptides to the Fe(II) ions and the formation of a number of isomers (up to 10 possible species can be formed); this mixture collapses into four well-defined isomers. Based on high-level theoretical calculations and kinetic data, we suggest that the key factor that drives the folding pathway is a trans → cis isomerization of the Pro residues located in the loops of the peptides connecting the three coordinating bipyridine units. Moreover, we have also observed that the position of the equilibrium between the four isomers can be reversibility switched by varying the temperature of the solution mixture. Finally, we have shown that there is an overall chiral selection in the mixture of isomers in the equilibrium, and that it is programmed by the chirality of the N-terminal Pro residue in the peptide ligand. We believe that these studies could help to improve the scarce knowledge about the folding mechanism in metallopeptides and metalloproteins.

## ASSOCIATED CONTENT

### Supporting Information

Synthesis and characterization of the coordinating residue, ligand peptides and metallopeptides; details and further data about the NMR, CD and UV-vis studies; details and further data about the molecular modeling and CD/DFT studies. This material is available free of charge via the Internet at http://pubs.acs.org.

## AUTHOR INFORMATION

### Corresponding Author

* miguel.vazquez.lopez@usc.es; eugenio.vazquez@usc.es; jean-didier.marechal@uab.cat

## ACKNOWLEDGMENT

Support of COST Action CM1105 is kindly acknowledged. G.R. We are thankful for the support given by the Spanish grants SAF2013-41943-R, CTQ2012-31341, CTQ2011-23336 and CTQ2013-49317-EXP; the ERDF and the European Research Council (Advanced Grant 340055); the Xunta de Galicia grants GRC2013-041 and PGIDIT08CSA-047209PR and the Generalitat de Catalunya grant 2009SGR68. Support of COST Action CM1105 is kindly acknowledged. G.R. thanks the INL for his PhD fellowship

## REFERENCES

(1) Kendrew, J. C.; Bodo, G.; Dintzis, H. M.; Parrish, R. G.; Wyckoff, H. *Nature* **1958**, *181*, 662–666.

(2) Anfisen, C. B. *Science* **1973**, *181*, 223–230; b) Dill, K. A.; MacCallum, J. L. *Science* **2012**, *338*, 1042–1046.

(3) Schweitzer-Stenner, R.; Uversky, V. *Peptide Folding, Misfolding & Nonfolding* (Ed.: R. Schweitzer-Stenner). Wiley-VCH. **2012**.

(4) Shortle, D. *FASEB J.* **1996**, *10*, 27–34.

(5) Aguzzi, A.; O'Connor, T. *Nat. Rev. Drug. Discov.* **2010**, *9*, 237–248.

(6) Lykourinou, V.; Ming, L.-J. in *Metallofoldamers: Supramolecular Architectures from Helicates to Biomimetics* (Eds.: G. Maayan, M. Albrecth), Wiley-VCH, **2013**, pp. 1–50.

(7) Ortega-Carrasco, E.; Lledós, A.; Maréchal, J.-D. *J. R. Soc. Interface* **2014**, *11*, 20140090

(8) a) House, E.; Collingwood, J.; Khan, A.; Korchazkina, O.; Berthon, G.; Exley, C. *J. Alzheimers Dis.* **2004**, *6*, 291–201; b) Exley, C. *J. Alzheimers Dis.* **2006**, *10*, 173–177; c) Permyakov, E. *Metallloproteomics*, John Wiley & Sons, Hoboken, New Jersey (USA), **2009**.

(9) Rama, G.; Ardá, A.; Maréchal, J.-D.; Gamba, I.; Ishida, H.; Jiménez-Barbero, J.; Vázquez, M. E.; Vázquez López, M. *Chem. Eur. J.* **2012**, *18*, 7030–7035.

(10) Gamba, I.; Salvadó, I.; Rama, G.; Bertazzon, M.; Sánchez, M. I.; Sánchez-Pedregal, V. M.; Martínez-Costas, J.; Brissos, R. F.;

Gamez, P.; Mascareñas, J. L.; Vázquez López, M.; Vázquez, M. E. *Chem. Eur. J.* **2013**, *19*, 13369–13375.

(11) Gamba, I.; Rama, G.; Ortega-Carrasco, E.; Maréchal, J.-D.; Martínez-Costas, J.; Vázquez, M. E.; Vázquez López, M. *Chem. Commun.* **2014**, *50*, 11097–11100.

(12) Kaes, C.; Katz, A.; Hossein, M. W. *Chem. Rev.* **2000**, *100*, 3553.

(13) We believe that the joint study of this set of peptide ligands will allow us to understand key elements in the folding of peptides and also to check the influence that the chiral Pro residues may have over the metal-directed peptide folding and chiral induction processes.

(14) Coin, I.; Beyermann, M.; Bienert, M. *Nature Prot.*, **2007**, *2*, 3247.

(15) The best fit to the experimental data suggests a significant proportion of the 1:2 ML species in the initial steps of the UV-vis titrations, when the peptidic ligand is in greater excess over the metal ion. Further increase in the metal-to-ligand ratios, progressively shifts the equilibrium towards the expected 1:1 complex later in the titration.

(16) The UV-vis spectra of the Fe(II) metallopeptides show intense bands centered at 543 nm. These data, together with the MALDI spectra, suggest octahedral coordination geometries of the metal centers in their respective complexes.

(17) Dragna, J.; Pescitelli, G.; Tran, L.; Lynch, V. M.; Anslyn, E.V.; Di Bari, L. M. *J. Am. Chem. Soc.* **2012**, *134*, 4398-4407.

(18) a) Mürner, H.; Belser, P.; Zelewsky, A. *J. Am. Chem. Soc.* **1996**, *118*, 7989–7994; b) Wang, P.; Miller, J. E.; Henling, L. M.; Stern, C. L.; Frank, N. L.; Eckerman, A. L.; Meade, T. J. *Inorg. Chem.* **2007**, *46*, 9853–9862.

(19) Pescitelli, G.; Di Bari, L.; Berova, N. M *Chem. Soc. Rev.* **2013**, *43*, 5211-5233.

(20) Rodger, A.; Johnson, B. F. G.; *Inorg. Chem.* **1988**, *27*, 3061–3062; Amati, M.; Lelj, F. *Theor. Chem. Account* **2008**, *120*, 447–457.

(21) Brandts, J. F.; Halvorson, H. R.; Brennan, M. *Biochemistry* **1975**, *14*, 4953–4963.

(22) Bignucolo, O.; Leung, H. T.; Grzesiek, S.; Bernèche, S. *J. Am. Chem. Soc.* **2015**, *137*, 4300–4303.

(23) a) Thomas, W. A.; Williams, M. K. *J. Chem. Soc., Chem. Commun.* **1972**, 994; b) Kern, D.; Schutkowski, M.; Drakenberg, T. *J. Am. Chem. Soc.* **1997**, *119*, 8403–8408.

(24) Brandts, J. F.; Halvorson, H. R.; Brennan, M. *Biochemistry* **1975**, *14*, 4953–4963.

# B … Publications non-related to the Ph.D. Thesis

# 1 Understanding the interaction of an antitumoral platinum (II) 7-azaindolate complex with proteins and DNA

Samper, K. G.; Rodríguez, V.; Ortega-Carrasco, E.; Atrian, S.; Maréchal, J. D., Cutillas, N.; Palacios, Ò. *BioMetals*, **2014**, *27*, 1159-1177.

Biometals (2014) 27:1159–1177
DOI 10.1007/s10534-014-9780-1

# Understanding the interaction of an antitumoral platinum(II) 7-azaindolate complex with proteins and DNA

**Katia G. Samper · Venancio Rodríguez · Elisabeth Ortega-Carrasco · Sílvia Atrian · Jean Didier Maréchal · Natalia Cutillas · Ana Zamora · Concepción de Haro · Mercè Capdevila · José Ruiz · Òscar Palacios**

Received: 22 May 2014 / Accepted: 25 July 2014 / Published online: 9 August 2014

**Abstract** The reactivity of the [Pt(dmba)(aza-N1)(dmso)] complex **1**, (a potential antitumoral drug with lower $IC_{50}$ than cisplatin in several tumoral cell lines) with different proteins and oligonucleotides is investigated by means of mass spectrometry (ESI-TOF MS). The results obtained show a particular binding behaviour of this platinum(II) complex. The interaction of **1** with the assayed proteins apparently takes place by Pt-binding to the most accessible coordinating amino acids, presumably at the surface of the protein -this avoiding protein denaturation or degradation- with the subsequent release of one or two ligands of **1**. The specific reactivity of **1** with distinct proteins allows to conclude that the substituted initial ligand (dmso or azaindolate) is indicative of the nature of the protein donor atom finally bound to the platinum(II) centre, *i.e.* N- or S-donor amino acid. Molecular modeling calculations suggest that the release of the azaindolate ligand is promoted by a proton transfer to the non-coordinating N present in the azaindolate ring, while the release of the dmso ligand is mainly favoured by the binding of a deprotonated Cys. The interaction of complex **1** with DNA takes always place through the release of the azaindolate ligand. Interestingly, the interaction of **1** with DNA only proceeds when the oligonucleotides are annealed forming a double strand. Complex **1** is also capable to displace ethidium bromide from DNA and it also weakly binds to DNA at the minor groove, as shown by Hoechst 33258 displacement experiments. Furthermore, complex **1** is also a good inhibitor of cathepsin B (an enzyme implicated in a number of cancer related events). Therefore, although compound **1** is definitely able to bind proteins that can hamper its arrival to the nuclear target, it should be taken into consideration as a putative anticancer drug due to its strong interaction with oligonucleotides and its effective inhibition of cat B.

**Keywords** Antitumoral compound · Platinum · DNA interaction · Protein interaction · Mass spectrometry

---

**Electronic supplementary material** The online version of this article (doi:10.1007/s10534-014-9780-1) contains supplementary material, which is available to authorized users.

K. G. Samper · E. Ortega-Carrasco ·
J. D. Maréchal · M. Capdevila · Ò. Palacios (✉)
Departament de Química, Facultat de Ciències,
Universitat Autònoma de Barcelona, Cerdanyola del
Vallès, 08193 Barcelona, Spain
e-mail: Oscar.Palacios@uab.cat

V. Rodríguez · N. Cutillas · A. Zamora ·
C. de Haro · J. Ruiz
Departamento de Química Inorgánica, Universidad
de Murcia, 30071 Murcia, Spain

V. Rodríguez · N. Cutillas · A. Zamora ·
C. de Haro · J. Ruiz
Instituto Murciano de Investigación Biosanitaria (IMIB),
Murcia, Spain

S. Atrian
Departament de Genètica, Facultat de Biologia,
Universitat de Barcelona, 08028 Barcelona, Spain

## Abbreviations

| | |
|---|---|
| aza | 7-Azaindolate |
| cat B | Cathepsin B |
| Cyt C | Cytochrome C |
| Dmba | *N,N*-dimethylbenzylamine-κ*N*,κ*C* |
| Dmso | Dimethylsulfoxide |
| DS | Double strand oligonucleotide |
| EB | Ethidium bromide |
| HSA | Human serum albumin |
| MT1 | $Zn_7$-MT complex of the recombinant mouse metallothionein isoform 1 |
| Myo | Myoglobin |
| SS | Single strand oligonucleotide |
| Tf | Transferrin |

## Introduction

Since the end of the 70 s, platinum-based complexes, such as cisplatin, carboplatin and oxaliplatin have resulted the most effective treatments against cancer (Lippert 1999; Jakupec et al. 2008; Harper et al. 2010). Their clinical success has promoted the design of further generations of Pt drugs aiming at overcoming several drawbacks (Barnes and Lippard 2004), like severe side-effects, the intrinsic resistance of some tumors, and the development of resistance induced after initial treatment (Jakupec et al. 2008). The mechanism of platinum-anticancer drugs involves their binding to DNA (Jamieson and Lippard 1999), which induces structural modifications on the double helix leading to apoptosis (Lippert 1999). The natural ability of cells to prevent toxicity promotes several responses: changes in the intracellular accumulation of the drug; increased production of intracellular thiols; increased capability of cells to repair platinum-induced DNA damage; and failure to initiate apoptosis in the presence of platinated DNA (Paolicchi et al. 2002). Due to the strong reactivity of platinum compounds toward S-donor molecules, which leads to the formation of very stable $Pt^{II}$-thiolate bonds, various kinds of intracellular thiol-rich molecules, as metallothioneins, account for conferring resistance to antitumor platinum drugs through their competition with DNA (Knipp et al. 2007). Also the interaction of the platinum drugs with other proteins may play crucial roles in their uptake and biodistribution processes, as well as in determining their toxicity profile, as they are the first potential interaction partners in the blood stream after intravenous administration (Barnes and Lippard 2004).

Therefore, decoding how platinum(II) complexes interact with proteins is key for designing new effective drugs. The contribution of electrospray ionization mass spectrometry (ESI–MS) to monitor such interaction has been proven to be extremely valuable when using, either nano-ESI sources, which are common in the study of the interaction of platinum drugs with proteins (Egger et al. 2008; Montero et al. 2007), or conventional ESI sources (Samper et al. 2012 and others), which, unlike the nano-ESI sources, allow to work under a wide set of experimental conditions.

The synthesis and X-ray crystal structure of a dmba-based platinum complex **1**, [Pt(dmba)(aza-N1)(dmso)], (dmba = *N,N*-dimethylbenzylamine-κ*N*, κ*C*; aza-N1 = 7-azaindolato-κ*N*1; dmso = dimethylsulfoxide-κ*S*), have been reported (Ruiz et al. 2010). Complex **1** not only contains a chelating ligand forming a very stable C–Pt bond (which accounts for the integrity of the Pt moiety), but also shows submicromolar activity both in A2780 and T47D cell lines ($IC_{50} = 0.34$ μM and 0.53 μM, respectively). Additionally, it exhibits very low resistance factors in the A2780 cell line, which has acquired resistance to cisplatin. Finally, the formation of adducts of this platinum(II) complex **1** with calf thymus DNA, followed by circular dichroism, has suggested important and promising modifications in the secondary structure of DNA, which point out the putative interest of this complex as an anticancer drug (Ruiz et al. 2010).

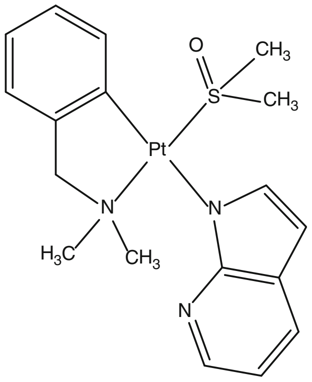
In this work we have studied the interaction of complex **1** with several proteins (albumin, transferrin, myoglobin -as a model for haemoglobin-, cytochrome C, and metallothionein -mammalian MT1) that can easily interact with metallic drugs after its administration in the blood stream and before their arrival to the putative target (*i.e.* the cellular nucleus). These interactions have been monitored by an ESI-TOF MS instrument equipped with a conventional ESI source and with a high resolution TOF analyzer. ESI–MS has also been used to monitor the interaction of complex **1** with a designed double-stranded oligonucleotide containing the GG Pt-binding site motif. To further investigate the binding mode between complex **1** and DNA, fluorescence competition experiments with ethidium bromide (EB) and Hoechst 33258 were carried out. The particular behaviour observed in the interaction of complex **1** with both kinds of biomolecules, proteins and oligonucleotides, made necessary the use of theoretical calculations in order to better rationalize the origin of its reactivity. The specific interaction of **1** with MT1 -a common target for $Pt^{II}$ compounds- has also been further investigated by optical spectroscopy. Finally, the inhibitory effect of complex **1** on cathepsin B (cat B) -an abundant and ubiquitously expressed cysteine peptidase whose increased expression and secretion have been related to tumoral cell migration and invasion capacity (Fernandez et al. 2001)- has been evaluated.

## Experimental section

### Synthesis and characterization of the Pt complex 1

The synthesis and full characterization of the [Pt(dmba)(aza-N1)(dmso)] complex used in this work, hereafter denoted as **1**, was already described (Ruiz et al. 2010). The 3D structure determined by X-ray diffraction allowed to prove the coordination environment of the $Pt^{II}$ centre shown in Fig. 1.

Due to the neutral nature of **1**, which is responsible of its relatively low solubility in water, the complex was dissolved in dimethylsulfoxide (dmso).



$C_{18}H_{23}N_3OSPt$

MW: 524.54 g/mol

**Fig. 1** Schematic representation of the [Pt(dmba)(aza-N1)(dmso)] complex used in this work, based on X-ray diffraction (from ref. (Castillo-Busto et al. 2009))

### Sample preparation and incubation previous to mass spectrometry analysis

Most of the proteins used in this work were purchased from Sigma-Aldrich: human serum albumin (A8763), transferrin (T3309), myoglobin (M6036) and cytochrome C (C3484). Mammalian metallothionein (mouse MT1 isoform) was recombinantly produced as a Zn-complex in *E. coli*, and fully characterized as previously reported (Cols et al. 1997). Several 1-mM solutions of each protein (except for MT1) were prepared in water by weighting the corresponding lyophilized protein. The purified recombinant preparations of MT1 consisted on $Zn_7$-MT1 complexes in a 50 mM Tris-$HClO_4$ solution at pH 7.0, and of a 0.275 mM protein concentration. Additionally, a 5-mM solution of **1** in dmso was prepared.

From all these starting solutions, 100-µL samples of 100-µM protein solutions were prepared by mixing the appropriate volume of each protein solution with the dmso solution of **1** to render preparations at the desired protein:Pt molar ratios (1:1, 1:5, 1:10) in 25 mM ammonium bicarbonate buffer at pH 7. The final solutions contained a 2 % concentration of dmso to keep the complex soluble Afterwards, the incubation of these mixtures was performed at 37 °C for 24 or 48 h in a stirring water bath.

The complementary single strand (SS) oligonucleotides used in this work, OP1 and OP2 (Table 1), were purchased from Eurofins MWG Synthesis GmbH (Ebersberg, Germany). In order to obtain the corresponding double-stranded (DS) oligonucleotide, equimolar quantities of each oligonucleotide (50 µM solutions in 25 mM ammonium bicarbonate buffer at pH 7.0) were incubated at 70 °C for 2 h and allowed to cool at room temperature overnight.

The SS and DS oligonucleotides were analyzed by ESI-TOF MS in negative mode as already described (Samper et al. 2012). These measurements confirmed the purity and identity of the single-stranded OP1 and OP2 oligonucleotides (Table 1), and allowed to observe the formation of the double-stranded oligonucleotide, which has been proved to be stable under the ESI–MS conditions assayed. The mass spectra obtained after incubation of equimolar amounts of the complementary OP1 and OP2 oligonucleotides showed a peak corresponding to the DS with similar intensity to those related to the single oligonucleotides, which suggested an efficiency of 50 % in their

**Table 1** Sequences and experimental molecular weight (MW) of the single (SS) and double (DS) stranded oligonucleotides determined by ESI–MS (Casini et al. 2008). The experimental error was always lower than 0.1 %

| | Name | Sequence | Experimental MW (Da) | Theoretical MW (Da) |
|---|---|---|---|---|
| SS | OP1 | 5′-CACTTCCGCT-3′ | 2,938.6 | 2,938.97 |
| | OP2 | 5′-AGCGGAAGTG-3′ | 3,116.6 | 3,117.10 |
| DS | | | 6,055.2 | 6,056.07 |

annealing reaction (Samper et al. 2012). This result was also confirmed by size exclusion-HPLC–UV.

Following an analogous procedure to that described above for the proteins, 100-μL samples, containing the appropriate amounts of the DS obtained and complex **1**, in order to reach a 20 μM concentration of DS and the desired DS:Pt molar ratios (1:1, 1:5, 1:10) were prepared. The mixtures were afterwards incubated at 37 °C for 24 or 48 h in a stirring water bath.

For the mass measurements, all the oligonucleotides-containing preparations were adequately diluted in order to reach the best S/N ratio in the mass spectra, at the chosen working conditions.

## Mass spectrometry conditions

Molecular mass determinations were performed by electrospray ionization mass spectrometry equipped with a time-of-flight analyzer (ESI-TOF MS) using a Micro Tof-Q Instrument (Bruker Daltonics GmbH, Bremen, Germany) calibrated with ESI-L Low Concentration Tuning Mix (Agilent Technologies), interfaced with a Series 1100 HPLC pump (Agilent Technologies) equipped with an autosampler, both controlled by the Compass Software.

The interaction of the platinum(II) complex **1** with proteins was analyzed in positive mode under the following experimental conditions: 20 μL of the sample were injected at 40 μL min$^{-1}$; the capillary-counterelectrode voltage was 4.5 kV; the desolvation temperature was 100 °C; dry gas at 6 L min$^{-1}$. Spectra were collected throughout a m/z range from 800 to 2500. The liquid carrier was a 85:15 mixture of 15 mM ammonium acetate and acetonitrile, pH 7.0.

The interaction of the platinum(II) complex **1** with the OP1, OP2 and DS oligonucleotides was analyzed in negative mode under the following experimental conditions: 10 μL of the sample were injected at 40 μL min$^{-1}$; the capillary-counterelectrode voltage was 3.9 kV; the desolvation temperature was 100 °C; dry gas at 6 L min$^{-1}$. Spectra were collected throughout a m/z range from 800 to 2500. The liquid carrier was a 90:10 mixture of 15 mM ammonium acetate and acetonitrile, pH 7.0.

All samples were injected at least in duplicate to ensure reproducibility.

## HPLC separation

HPLC was performed with a Series 1200 HPLC pump (Agilent Technologies, Santa Clara, CA, USA) equipped with an autosampler and a diode array detector, all controlled by the Compass Software. 100-μL aliquots of the samples were injected into a Superdex Peptide column (GE Healthcare, Fairfield, CT, USA) and eluted with 50 mM ammonium bicarbonate buffer at pH 7.5, at a flow rate of 550 μL min$^{-1}$. At the exit of the column, the absorbance was recorded at 210, 254, and 280 nm. Each of the peaks detected (due to DS, OP1, and OP2) were separately collected and all the fractions corresponding to a same peak were pooled and the sample homogenized before incubation with the Pt$^{II}$ complex.

## Optical spectroscopy

Circular dichroism (CD) spectroscopy was performed using a model J-715 spectropolarimeter (JASCO, Gross-Umstadt, Germany) equipped with a computer (J-700 software, JASCO). Measurements were carried out at a constant temperature of 25 °C maintained by a Peltier PTC-351 S apparatus (TE Technology Inc., Traverse City, MI, USA). Electronic absorption was measured on an HP-8453 diode-array UV–vis spectrophotometer (GMI Inc., Ramsey, MN, USA), using 1-cm capped quartz cuvettes, and correcting for the dilution effects by means of the GRAMS 32 software (Thermo Fisher Scientific Inc., Waltham, MA, USA). Fluorescence measurements were carried out with a Perkin-Elmer LS 55 50 Hz Fluorescence Spectrometer under different working conditions, depending on the experiment carried out.

### Ethidium bromide and Hoechst 33258 displacement experiments

In the ethidium bromide (EB) fluorescence displacement experiment, a 3-mL solution containing 10 μM calf thymus DNA (ct-DNA) and 0.33 μM EB (saturated binding levels (Barton et al. 1986)) in 50 mM Tris–HCl at pH 7.4 buffer was titrated with aliquots of a concentrated solution of complex **1**, thus producing solutions with varied molar ratios of **1** to ct-DNA. After each addition the solution was stirred at the appropriate temperature during 5 min before measurement. The fluorescence spectra of the solution were obtained by exciting at 520 nm and measuring the emission spectra at the 530–700 nm range, using 5 nm slits. The procedure was the same for the Hoechst 33258 reactions, using the following conditions: working solutions were 20 μM ct-DNA and 2 μM Hoechst 33258; $\lambda_{ex} = 338$ nm and $\lambda_{em} =$ 400–550 nm (with $\lambda_{max} = 464$ nm).

### Cathepsin B inhibition assay

Crude bovine spleen cathepsin B (cat B) was purchased from Sigma and used without further purification. The colorimetric cat B assay was performed in a solution containing 20 mM sodium acetate, and 1 mM EDTA at pH 5.1, using Z-L-Lys-ONp hydrochloride (Sigma) as the chromogenic substrate. In order for the enzyme to be catalytically functional, the cysteine residue present in the active site needs to be in a reduced form. Therefore, before using it, cat B was pre-reduced with dithiothreitol (DTT) to ensure that the majority of the enzyme is in a catalytically active form.

$IC_{50}$ determinations were performed in duplicate using a fixed enzyme concentration of 0.1 μM, and a fixed substrate concentration of 0.1 mM. Inhibitor concentrations ranged from 0.25 to 75 μM. The enzyme and inhibitor were co-incubated at 25 °C over a period of 24 h prior to the addition of substrate. Activity was measured over 1 min at 327 nm.

Cysteine reactivation was evaluated using an inhibitor concentration corresponding to $2 \times IC_{50}$. The enzyme was treated with an excess of DTT (Sigma D0632). After the activation, the enzyme and compound **1** were incubated at 25 °C for 24 h. Then, 1 mM L-cysteine was added and incubated at different times 1, 2, 3, 4, 6 and 24 h at 25 °C. Following incubation, the substrate was added and activity was assessed.

### Molecular modeling calculations

Quantum Mechanical (QM) optimizations have been performed on complex **1** and on several derivatives resulting from the substitution of the dmso or the azaindolate ligands with distinct models of amino acids. Calculations have been carried out with Gaussian 09 (2009) at the density functional theory (DFT) level using the B3LYP (Becke 1993; Lee et al. 1988); functional. The basis set used account for the 6–31 + G* (Hehre et al. 1972; Hariharan and Pople 1973; Spitznagel et al. 1982; Clark et al. 1983) for the main group elements and LANL2DZ (Hay and Wadt 1985) for Pt. LANL2DZ pseudopotential was also applied to the metal.

The strength of the interaction between the metal and its ligands has been evaluated using the energy decomposition analysis (EDA) as implemented in the ADF 2010 package (te Velde et al. 2001; Guerra et al. 1998; ADF 2010). Here, also calculations at DFT level have been performed with the becke (Becke 1988) and pw91c (Perdew et al. 1992) (for the correlation and exchange terms respectively) mixed functional. The basis set used in this part of the theoretical work was a TZP with a small frozen core for all the atoms, and ZORA formalism (van Lenthe et al. 1999) for the relativistic effects of the platinum centre.

## Results and discussion

The previous characterization of the platinum(II) complex **1** showed its activity in several tumoral cell lines as well as modifications in the secondary structure of DNA (Ruiz et al. 2010), but no further data were reported regarding this interaction. To advance in the identification of the species formed in the process, the number of Pt atoms attached to distinct proteins to which complex **1** could bind before reaching its target, and the nature of its possible interactions with proteins and DNA were investigated.

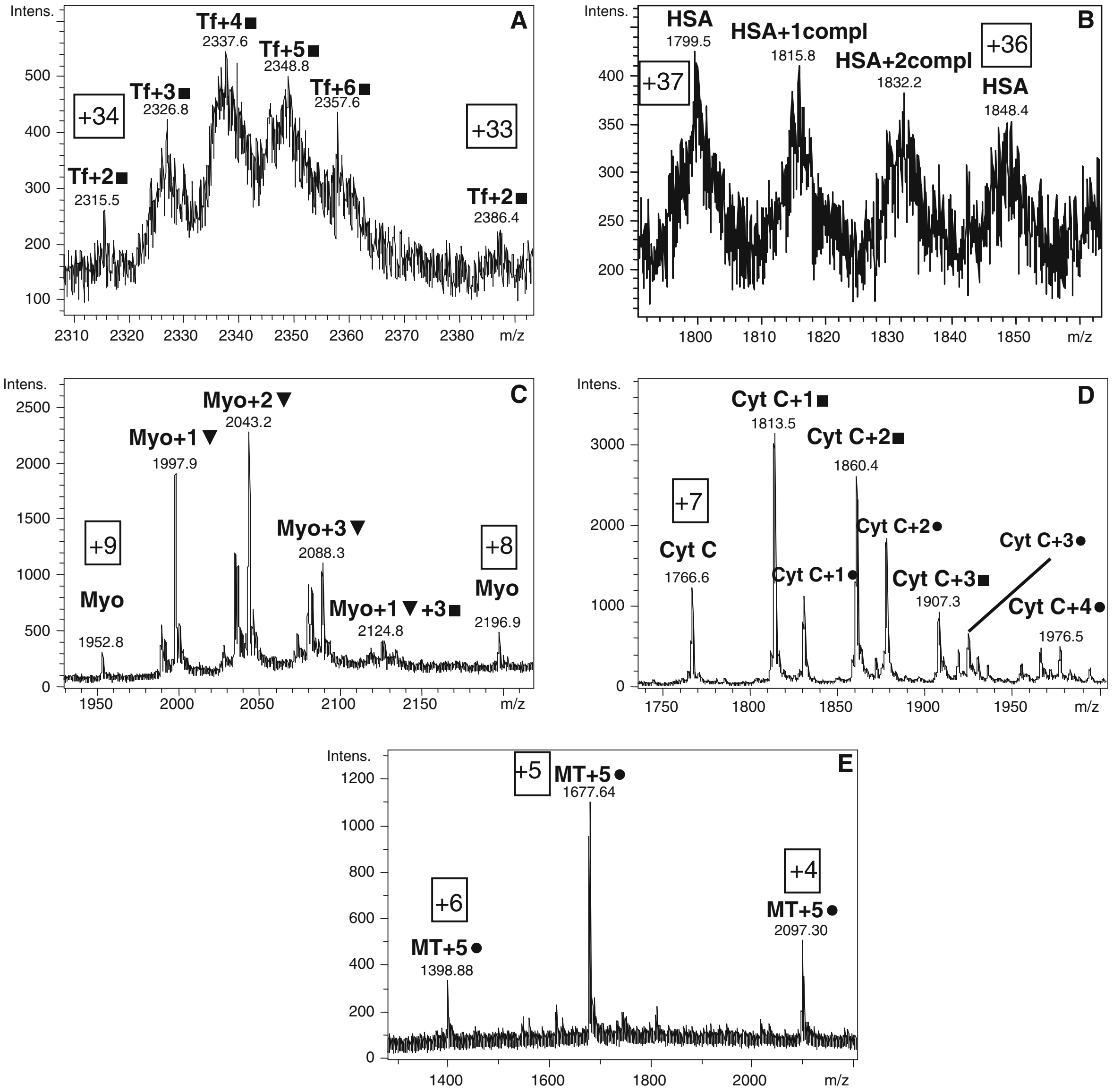


**Fig. 2** Mass spectra recorded after incubation (24 h at 37 °C) of complex **1** with **a** transferrin (Tf), **b** albumin (HSA), **c** myoglobin (Myo), **d** cytochrome C (Cyt C) and **e** mammalian MT1 at a 1:10 (Protein:Pt) molar ratios. The numbers preceded by the "+" symbol in the boxes denote the charge state of the peaks. The notation "*n* compl" denotes a mass increase corresponding to the addition of "*n*" whole complexes **1** (524.5 mass units at the corresponding charge state) to the protein. The notation "*n*▼" indicates a mass increase corresponding to the addition of "*n*" molecules of complex **1** to the protein after the elimination of the initially bound azaindolate ligand ($C_7H_5N_2$) (524.5–117.1 mass units at the corresponding charge state); the notation "*n*●" indicates a mass increase corresponding to the addition of "*n*" molecules of complex **1** after elimination of the initially bound dmso (524.5–78.1 mass units at the corresponding charge state); and the notation "*n*■" indicates a mass increase corresponding to the addition of "*n*" molecules of complex **1** after elimination of both initial ligands, azaindolate and dmso (524.5–195.2 mass units at the corresponding charge state)

Moreover, the capabilities of the complex to inhibit cathepsin B have also been determined in order to check its possible role in the control of tumor progression.

## Characterization of the proteins by ESI–MS

The proteins used in this work were previously analyzed by ESI-TOF MS in our labs by following a reported procedure (Samper et al. 2012) in order to determine their experimental molecular weight under our working conditions (**Table S1**). The analysis of each protein solution lead to the observation of several ionization states in the described m/z working range, allowing a very precise determination of each individual MW. The spectra recorded for most of the protein preparations showed single species, except for albumin and transferrin, which exhibited several peaks, frequently associated to acetylated or distinct glycosylated forms (Castillo-Busto et al. 2009). The experimental MW recorded for myoglobin (17567 Da), 615 Da higher than that expected for the apoprotein, confirms the presence of the haem group (616.45 Da).

## Interaction of the platinum(II) complex with the proteins

Complex **1** was incubated, at several molar ratios, with each of the chosen proteins (Tf, Myo, Cyt C, HSA and MT1). The whole set of mass spectra obtained allows to state that, despite complex **1** is able to react with all the assayed proteins, it does not show the same reactivity with all of them. For the sake of clarity, Table S2 contains all the information obtained (the species formed after the interaction with each protein and the relative intensity of their mass peaks), while Fig. 2 only shows the mass spectra recorded after incubation of **1** at the 1:10 protein-to-Pt molar ratios.

The ESI–MS data reveal that the species formed depend on two main factors: the nature of the protein and the assayed protein:Pt ratio. In the case of transferrin and albumin, a high amount of broad peaks, in low resolution spectra, were obtained (Figs. 2a, b respectively). However, within some experimental error, the corresponding species could be identified. Therefore, we can state that interaction of transferrin with **1** takes place through the elimination of both ligands, dmso and azaindolate (as deduced by the increase of multiples of 335 ± 7 mass units, *i.e.* Tf + n (compl-aza-dmso), *cf.* Fig 2 caption), from the coordination sphere of platinum (Fig. 2a). This is consistent with the observation that the composition of the distinct species was strongly dependent on the assayed protein-to-Pt ratios, being 6 the maximum of Pt moieties simultaneously attached to a single protein. Interestingly, no presence of the initial free transferrin was detected.

Very different are the results obtained with albumin, where the presence of peaks corresponding to the intact protein as the major species (Fig. 2b) denotes a poor interaction with **1**. The mass of the other observed peaks corresponds to the increase of multiples of ca. 550 ± 25 Da, suggesting that the Pt complex remains unaltered after protein binding, with a maximum ratio of two complexes per albumin molecule.

When the Pt complex **1** interacts with myoglobin (Fig. 2c) the main peaks observed inform about the binding of complex **1** after elimination of the azaindolate ligand (i.e. releasing of the ligand from the complex), although other minor peaks corresponding to the simultaneous release of both (aza + dmso) ligands were also observed (see **Table S2** for details). Oppositely, when **1** interacts with cytochrome C, the major peaks suggest the simultaneous elimination the azaindolate and the dmso ligands (i.e. releasing of both ligands), although minor peaks indicate the solely displacement of the dmso ligand (Fig. 2d).

Contrasting with literature data on the interaction of cisplatin with mammalian metallothioneins (Karotki et al. 2008), the incubation of **1** with the Zn-loaded form of MT1 rendered mainly the unreacted $Zn_7$-MT1 initial complex, together with very minor $Zn_xPt_y$-MT1 and $Pt_5$-MT1 species at the lowest Protein:Pt ratio assayed (**Table S2**). The partial or total substitution of the initially coordinated $Zn^{2+}$ ions can be understood either considering the presence of small amounts of impurities accompanying complex **1** or Pt-binding mechanisms similar to those reported in the literature where all ligands of cisplatin are displaced (Karotki and Vasak 2008). Interestingly, at Pt-to-protein molar ratios higher than 1:1, the main detected peak shows the formation of a species containing 5 Pt complexes in which **1** only releases the dmso ligand (Fig. 2e).

Thus, overall data led to the observation that complex **1** can interact with the assayed proteins in three different ways: (a) through elimination of the azaindolate group (probably due to the *trans* effect of the C–Pt bond); (b) through elimination of the dmso molecule, mainly behaving as a monofunctional complex in both cases; and (c) by release of both ligands from the Pt coordination sphere, thus probably behaving as a bifunctional complex. Interestingly, the elimination of the ligand when binding to a certain protein seems to depend on the precise nature of the protein. Thus, the binding of the platinum moiety to proteins apparently presents wider molecular patterns than expected.

In order to understand the basis of this differential reactivity, we took into consideration the amino acid sequences of the studied proteins, especially focusing on the presence of His and Cys, both amino acids normally associated to the coordination of $Pt^{II}$ ions in biomolecules (Ivanov et al. 1998). Myoglobin contains 154 amino acids (11 His and 0 Cys), with some of the His residues easily accessible on its surface (as shown by the corresponding 3D structure, ref *3rgk* in the Protein Data Bank). Contrarily, MT1 has 62 amino acids, among them 20 Cys and 0 His. On the one hand, it can be considered that the interaction of complex **1** with myoglobin, probably binding by an N-donor atom of a His residue, causes the removal of the azaindolate ligand initially bound to the Pt(II) through one of its N atoms. On the other hand, it can be assumed that when **1** interacts with MT1, it binds to the Cys amino acids through their S atoms, so that the elimination of the dmso is compulsory to allow this interaction. Hence, it appears that the kind of interaction of complex **1** with proteins is directly related to the type of amino acid side chains available as ligand groups on the protein surface, *i.e.* N-donor or S-donor residues. This assumption correlates well with the observed reactivity of **1** with cytochrome C (3 His and 2 Cys over 105 total aa) and transferrin (22 His, 30 Cys, 622 total amino acid content), where the release of the azaindolate and the dmso ligands at the same time was observed, this indicating the participation of both types of residues of the proteins as ligands.

Another interesting conclusion can be drawn when analyzing the number of Pt atoms that result bound to each protein at the different molar ratios assayed. When comparing the species formed at the 1:10 molar ratio, Fig. 2, it becomes evident that: (a) transferrin can bind up to 6 Pt atoms; (b) myoglobin can bind up to 4 Pt, although the major species detected binds only 1 Pt complex; (c) cytochrome C shows binding of up to 4 Pt, but the main species contain only 1 Pt; and (d) albumin showed the interaction with up to 2 Pt complexes. If assuming that under our working conditions (close to those of physiological environments), all the proteins maintain its functional folding, the low number of Pt bound observed (in comparison with the high number of putative coordination sites available in the proteins) suggests that the groups interacting with the Pt centre need to be easily accessible, presumably on the surface of the protein, as already mentioned for myoglobin. The 3D structures available in the respective *3rgk* Protein Data Bank entry fully support this hypothesis. Hence, the myoglobin fold places only 3 His, among 11, close to the protein surface, while cytochrome C shows all their His and Cys as solvent-exposed residues. In the case of transferrin, its considerable size and its high number of putative ligands complicate the assignment of availability of its amino acid side chains. The case of albumin must be considered apart from the rest of proteins. Albumin is a huge protein, which also contains an elevated number of His and Cys residues, and which is devoted to the systemic transport of several types of metabolites (small molecules, metal ions, peptides, etc.). In fact, the mass peaks registered when **1** reacts with albumin can be related to the formation of adducts with the intact Pt complex, this suggesting that the inclusion of complex **1** in any of the existing cavities of the albumin globular structure is more likely than the formation of covalent bonds, contrarily to what probably happens with the rest of the assayed proteins.

Altogether, we can consider that the interaction of the Pt complex **1** with the proteins, except for MT1, may not alter significantly their structure if the Pt moiety ends bound to their surface or enclosed into the special cavities available for transportation, and consequently, proteins will keep all or part of their functionality. However, the interaction of **1** with MT1 deserves a special consideration, as it has provided significant unique results. The formation of a single species containing 5 Pt complexes bound per MT (where the dmso ligands have been released from the complex) suggests an especial type of interaction, probably due to a strong interaction of the Pt centre with the Cys residues of MT1. In order to better understand the interaction of **1** with the functional form $Zn_7$-MT1, we have monitored their reaction by optical spectroscopy (UV–vis absorption and circular dichroism, CD). Thus, to confirm that the single species observed, $[Pt(dmba)(aza\text{-}N1)]_5$-MT1, corresponds to an especially favoured aggregate only depending on the amount of **1** added to the protein, and to discard the possibility of an artefact formed when adding large quantities of **1**, two further experiments have been carried out. In the first one, aliquots of the solution of **1** were consecutively added to a 10 μM solution of $Zn_7$-MT1 in such a way that 1, 3

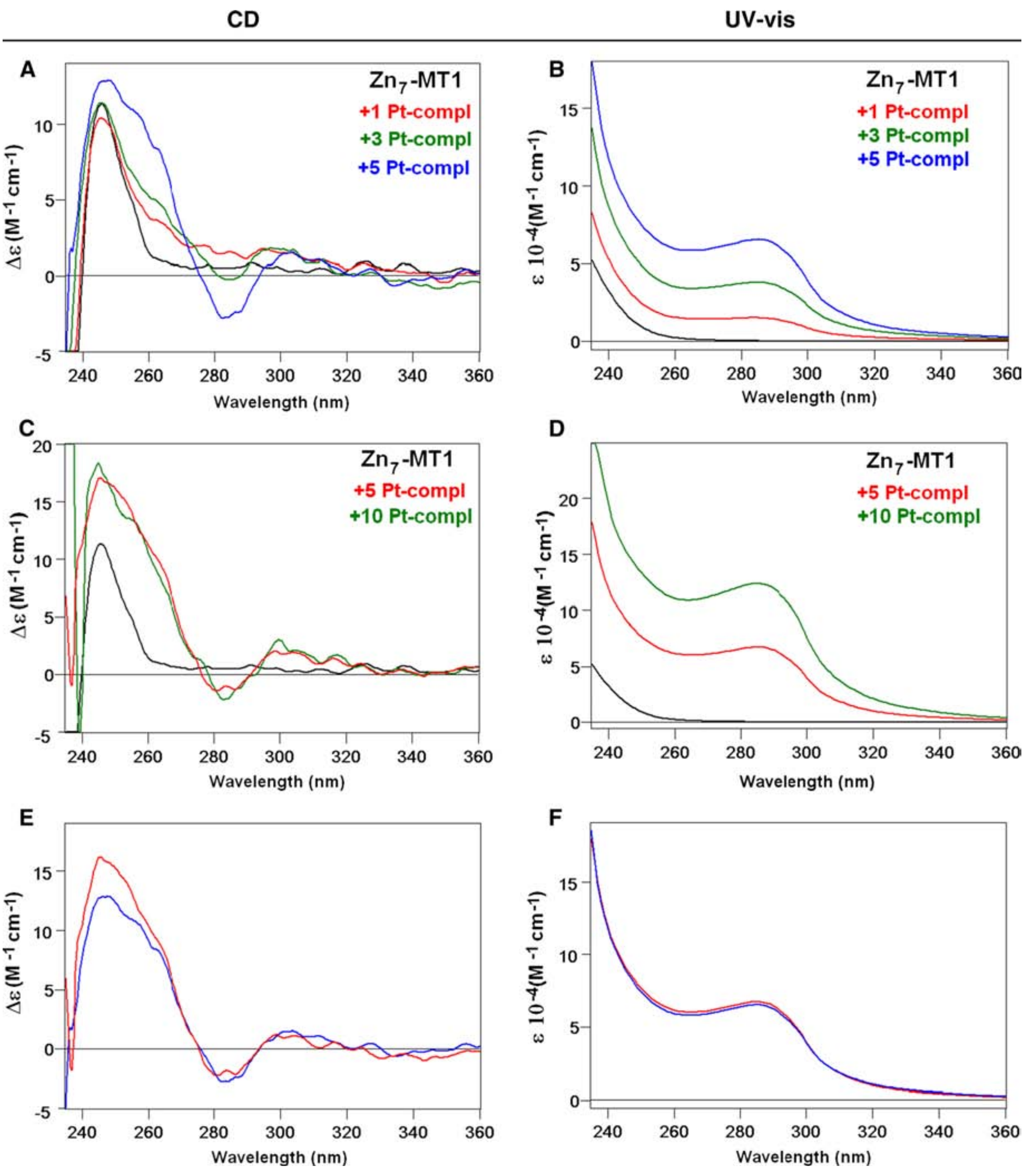


**Fig. 3** Optical spectra recorded after the reaction of a 10 µM solution of Zn$_7$-MT1 with **a**, **b** 1, 3 and 5, and **c**, **d** 5 and 10 molar equivalents of **1**. Comparison of **e** the CD and **f** UV–vis spectra corresponding to the addition of 5 molar equivalents of **1** in both experiments (in *blue* as a result of the successive additions of 1, 3 and 5 molar equivalents of **1** and in red after adding 5 molar equivalents of **1** at once). All spectra were recorded after the incubation (1 h at 40 °C) of the samples

and 5 molar equivalents of **1** were coexisting with the protein at each step of the assay. In the second experiment, a similar approach was undertaken to have 5 and 10 molar equivalents of **1** added to a 10 µM solution of Zn$_7$-MT1. After each addition of the complex, the samples were incubated (1 h at 40 °C) before registering the optical and the mass spectra of the resulting solution.

Addition of complex **1** to the $Zn_7$-MT1 solution implies the formation of new absorptions at ca. 290 nm (precisely an exciton coupling band in the CD spectra and a wide absorption band in the UV–vis spectra, Fig. 3) that can be associated to the binding of Pt to the protein. Interestingly, the spectra recorded after the addition of 5 molar equivalents of **1** in the two experiments, render practically identical spectra (Fig. 3e, f), which suggests that the species formed in both assays lead to a similar folding of the protein about the metal centre. The mass spectra recorded at each stage of each experiment (data not shown) yields identical information to that previously obtained (Fig. 2e and Table S2) and mass data collected after adding 5 molar equivalents of 1 in both assays are also practically coincident. Concluding, all our data reinforces the hypothesis that the species containing 5 Pt complexes bound to MT1, which is formed after releasing the initially-coordinated $Zn^{2+}$ ions, results especially favoured in structure and energy terms.

When searching analogous complexes in the literature, it is interesting to highlight that **1** is the unique monomeric Pt-azaindolate complex reported. This group recently reported the reactivity of a similar $Pt^{II}$-dmba complex (Samper et al. 2012), where triphenylphosphine and aminoacridine were additional ligands. While the results here presented show that complex **1** interacts with almost all the assayed proteins, it should be taken into account that the previously reported $Pt^{II}$-dmba complex (Samper et al. 2012) exhibited a poor interaction with the same proteins, and that the reactivity of **1** is significantly lower than that reported for other Pt-based complexes (Casini et al. 2007), –and especially than those containing labile chloride ligands (Esteban-Fernández et al. 2010)- hence highlighting the importance of the whole set of ligands in the reactivity of these platinum complexes.

## Theoretical calculations

The Pt complex, **1**, used in this work was specially designed with a dmso molecule bound to the platinum centre with the aim of facilitating its interaction with other molecules by releasing this ligand from the complex. Interestingly, experimental evidences show that **1** interacts with proteins in a different way than expected. Molecular Modeling was carried out to shed light on the origin of such behaviour focusing, in a first hypothesis, on pure bond energies.

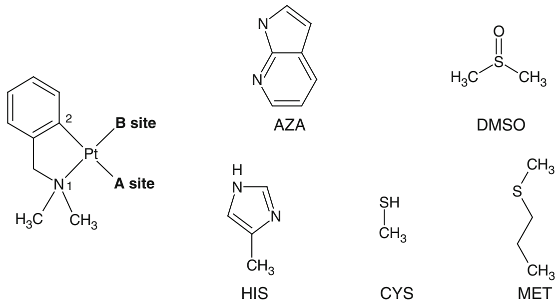


**Fig. 4** Two dimensional representations of the fragments used to model the binding energies at each binding site. The two fixed binding atoms of the chelate ligand are denoted as N1 and C2. The original ligands present in complex **1** are the azaindolate (AZA) and the dimethylsulfoxide (DMSO) molecules. The models employed to simulate the coordinating amino acids are: 1-ethylimidazole as histidine (HIS); methylmercaptan as cysteine (CYS); and propylmethylthiol as methionine (MET)

DFT calculations with the B3LYP functional were undertaken on compound **1** as well as on all the complexes resulting from the ligand substitution at the azaindolate site (site A) or the dmso site (site B) by models of protein amino acids (Fig. 4). Amino acids considered in this study are the most commonly accepted platinum binding residues and correspond to histidine (modelled by 1-ethylimidazole), cysteine (modelled by a methylmercaptan) and methionine (modelled by a propylmethylthiol). For His and Cys, coordination has been considered for a neutral and negatively charged residue. Calculations were also undertaken with platinum bound throughout $N_\varepsilon$ or $N_\delta$ atoms of the imidazole ring. Each system has been optimized with the Gaussian 09 package of program (Gaussian et al. 2009) with an extended basis set (see Experimental Section). In order to discuss the relative binding energies of the different residues to compound **1**, Energy Decomposition Analysis (EDA) as implemented in the ADF package (te Velde et al. 2001; Guerra et al. 1998; ADF 2010) has been carried out.

The optimized geometries of the different systems present square planar configurations about the metal centre, with little deviations from the ideality (Table S3 in Supplementary material). For sulfur containing residues, a slight distortion is observed with a dihedral angle, which is defined by the four coordinating atoms ($B_{site}$-$A_{site}$-$N_1$-$C_2$) directly connected to the platinum centre, of about 10°. Independently of the charge of the ligand, Pt–S bonds exhibit the longest length observed (about 2.4 Å for methionine and cysteine, against the 2.2 Å for other coordinating groups, including dmso), thus suggesting weaker interactions of such species with the platinum moiety. Predicted binding energies of the dmso and the azaindolate ligands are estimated to −41.74 and −130.72 kcal mol$^{-1}$, respectively. Energetic breakdowns show that these differences mainly arise from electrostatic and steric contributions; something consistent with the anionic character and larger size displayed by the latter. However, this also suggests a weaker release of the azaindolate with respect to dmso; a result apparently inconsistent with the experimental evidences described above.

Detailed analysis of possible ligand exchange between amino acids and the compound **1** (Table 2) have been further analyzed by comparing their interaction energy with the metallic system. Substitutions at the dmso site (B) by residue fragments lead to stronger binding energies than for dmso and range from approximately 6 (for neutral residues) to 35 kcal mol$^{-1}$ (for negatively charged residues). Interestingly, independently of the charge of the system, cysteine is always predicted as the best ligand for dmso substitution with about 1 to 6 kcal mol$^{-1}$ higher predicted affinity than histidine. Moreover, intermediate situations with a water exchange prior to residue binding are unlikely because of the weak binding of water at the B site (−23.52 kcal mol$^{-1}$). Substitutions at the azaindolate site (A) lead to a wider spectrum of binding energies. For neutral residues, sulfur containing chelates bind at about −27 kcal mol$^{-1}$; a result in the same range of values than for water (−21 kcal mol$^{-1}$) while histidine has a slightly better interaction energy of approximately −40 kcal mol$^{-1}$. None of them would be predicted to exchange with the original ligand in its anionic nature because of the large difference in energy to overcome (about 100 kcal mol$^{-1}$). For anionic species, predicted binding energies are in the same range, or even higher, than for the azaindolate. In particular, coordination of the cysteinate is close to 144 kcal mol$^{-1}$ (14 kcal mol$^{-1}$ stronger than the azaindolate) and coordination of the histidinate through the δ nitrogen is about −195 kcal mol$^{-1}$ (65 kcal mol$^{-1}$ less than the azaindolate). This shows that substitution at the A site could take place with the standard platinum binding residues with a preference, again, for histidine.

**Table 2** Main energetic contributions, steric and orbital, as well as total interaction energy, of different ligands at the A and B sites of complex **1**

| Ligand | Interactions (kcal mol$^{-1}$) | | Total bonding energy (kcal mol$^{-1}$) |
|---|---|---|---|
| | Steric | Orbital | |
| Site A | | | |
| AZA | −58.39 | −72.32 | −130.72 |
| CYS | 15.37 | −41.01 | −25.64 |
| HIS ($N_\varepsilon$) | 4.61 | −45.76 | −41.15 |
| HIS ($N_\delta$) | 7.34 | −45.77 | −38.42 |
| MET | 13.89 | −42.07 | −28.17 |
| $H_2O$ | 9.52 | −30.89 | −21.37 |
| AZA + H | 13.04 | −36.31 | −23.27 |
| CYS$^-$ | −60.14 | −84.29 | −144.43 |
| HIS$^-$ ($N_\varepsilon$) | −53.36 | −74.88 | −128.24 |
| HIS$^-$ ($N_\delta$) | 19.79 | −215.1 | −195.31 |
| Site B | | | |
| dmso | 31.68 | −73.42 | −41.74 |
| CYS | 23.31 | −71.38 | −48.07 |
| HIS ($N_\varepsilon$) | 17.82 | −64.91 | −47.09 |
| HIS ($N_\delta$) | 19.37 | −67.12 | −47.75 |
| MET | 23.03 | −69.06 | −46.03 |
| $H_2O$ | 13.42 | −36.95 | −23.52 |
| CYS$^-$ | 15.76 | −99.58 | −83.81 |
| HIS$^-$ ($N_\varepsilon$) | 12.76 | −87.97 | −75.21 |
| HIS$^-$ ($N_\delta$) | 13.76 | −88.07 | −74.66 |

At this point, the rationalization of the substitutions experimentally observed at site A can only be obtained by considering a negatively charged His$^-$ ($N_\delta$) exchange with the azaindolate ligand. This suggests that the protonation state of the azaindolate could have a weight into the binding process. Therefore, an additional calculation has been performed for which the azaindolate has been protonated, a situation possible considering the pKa values of the ligand. Such modification of the nature of the ligand leads to an energy of interaction of −23.27 kcal mol$^{-1}$. In this case, site A becomes far more labile than site B as the resulting binding energy shows that substitution by any residue can occur. Interestingly, such substitution should take place primarily with histidine. Although the dynamical considerations and the full mechanism of exchange between histidine and azaindolate are not taken into account at this point, our calculations clearly point at the importance of changes in the nature of the protonation state of the ligands to understand the experimental observations. However, more calculations are needed to be carried out to confirm the proposed mechanism.

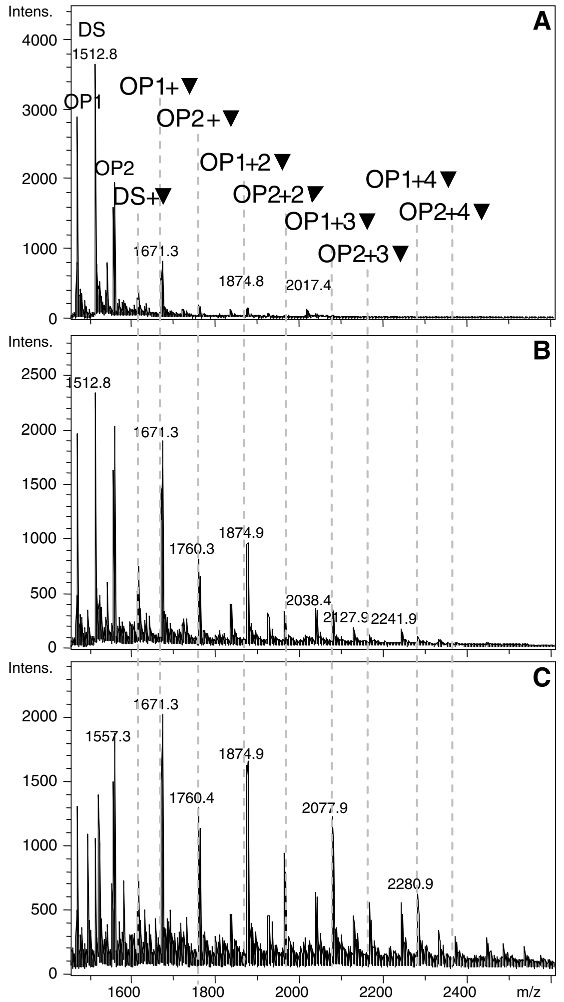


**Fig. 5** Negative ESI-TOF MS spectra obtained after incubation (24 h at 37 °C) of mixtures of the double-stranded chain (DS) with increasing amounts of complex **1** at **a** 1:1, **b** 1:5 and **c** 1:10 DS:Pt molar ratios. The notation "*n*▼" indicates a mass increase corresponding to the addition of "*n*" molecules of complex **1** after the elimination of its initial azaindolate ligand ($C_7H_5N_2$). (117.1 mass units at the shown charge state, which is m/z = 58.6). Other minor non-labelled peaks correlate with species formed between 1 and the oligonucleotides after elimination of both ligands, the azaindolate and the dmso

## Interaction of the platinum(II) complex with oligonucleotides

The oligonucleotides used in this work were designed as single stranded (SS) and double stranded (DS) Pt-binding probes. After preparation of the DS by mixing the OP1 and OP2 SS oligonucleotides under optimal experimental conditions (see *Experimental section*), this solution was incubated at 37 °C with **1** at different molar ratios (DS:Pt ratios of 1:1, 1:5, and 1:10) during ca. 24 h and was analyzed by ESI–MS spectrometry.

The results obtained (Fig. 5) show the presence of several MS peaks mainly attributable to the binding of **1** (after release of the azaindolate ligand) to the single stranded oligonucleotides. The number of peaks, and thus of different species, significantly increases with the increase of the Pt:oligonucleotide ratio, while concomitantly decreases the intensity of the DS peak. Even though the stoichiometry of the species formed (*i. e.* the number of Pt atoms bound to each oligonucleotide) directly depends on the amount of complex **1** added, at the 1:5 and 1:10 molar ratios the major peaks observed correspond to the binding of only one Pt complex to each SS oligonucleotide (OP1 + ▼; OP2 + ▼).

Other minor peaks were detected in the mass spectra: on the one hand, a single peak associated to the direct binding of **1** to the DS after the release of the azaindolate ligand (denoted as DS + ▼, with m/z = 1614.3); and on the other hand, small peaks that correlate with binding of the Pt complex **1** to the single stranded oligonucleotides after elimination of both the azaindolate and the dmso ligands, specially observed at the highest Pt:protein ratio assayed. Taking into consideration the previously studied (vide supra) reactivity of **1** with proteins, these results suggest that $Pt^{II}$ preferentially binds to an N-donor ligand present in the oligonucleotides, presumably their nitrogenous bases. Several attempts to determine the binding sites of the Pt-moiety were carried on by MS/MS experiments at different Pt:DS molar ratios but no results suggesting a preferential binding of **1** to any specific single strand or nucleotide were obtained (data not shown).

The results here presented also indicate that most of the observed peaks correlate with the binding of **1** to a single oligonucleotide (OP1 or OP2) rather than to DS as only one DS + ▼ peak was identified. This can be explained through two different hypotheses: (i) one possibility is that the final scenario corresponds to the interaction of complex **1** with the single stranded molecules that remain in the solution after a partial annealing process, and (ii) another possibility relies in considering that the dramatic loss of intensity of the original DS peak when increasing the Pt:oligonucleotide ratio, but not of those of OP1 and OP2, may indicate that the interaction of **1** with the double-

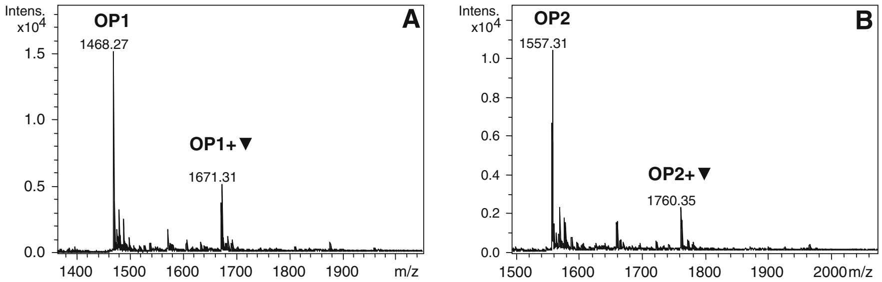


**Fig. 6** Negative ESI-TOF MS spectra recorded after incubation (24 h at 37 °C) of complex **1** with the individual oligonucleotides **a** OP1 and **b** OP2, at a 1:10 oligonucleotide:Pt ratio, showing a charge state of −2 for all the peaks. The notation "▼" indicates a mass increase corresponding to the addition of one molecule of **1** after the elimination of its initial azaindolate ligand ($C_7H_5N_2$). (117.1 mass units at the corresponding charge state)

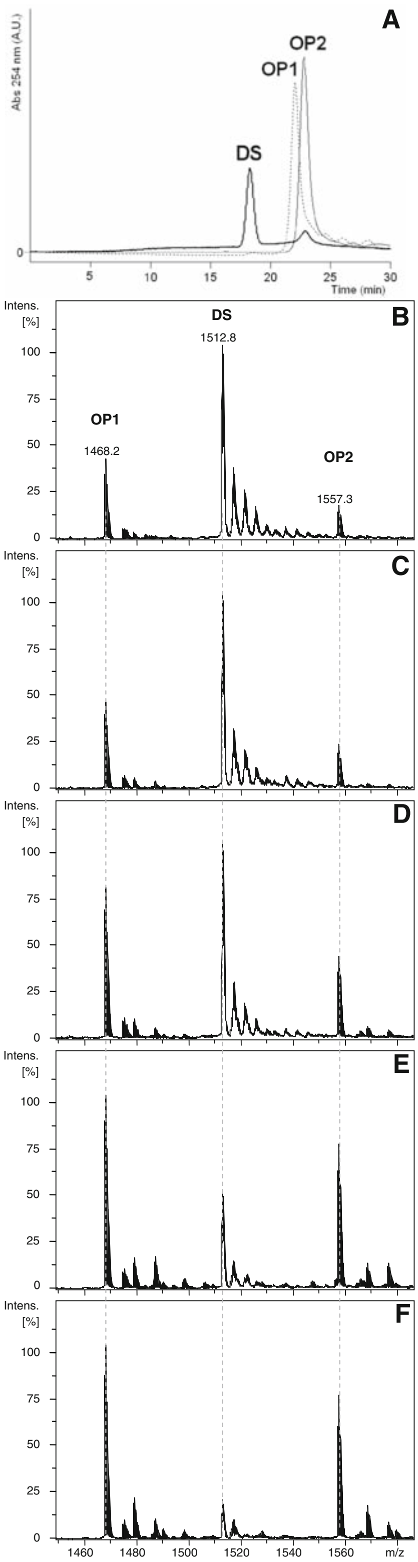
stranded molecule promotes its melting into the single-stranded oligonucleotides. In order to test these two possibilities, two further experiments were designed. First, each of the single stranded oligonucleotides, OP1 and OP2, was incubated at a 1:10 oligonucleotide:Pt ratio. As a result, only minor peaks suggestive of Pt-binding were observed in both cases (Fig. 6), a situation quite far away from the more intense peaks of the same species observed in the previous experiment at the same Pt:oligonucleotide ratio (Fig. 5c).

Secondly, the annealed DS form was separated by size exclusion-HPLC of the remaining SS oligonucleotides before incubation of the former with complex 1. The purification of DS was successfully achieved due to the different retention times of the single- and double-stranded oligonucleotides (Fig. 7a). The DS thus purified (Fig. 7b) was incubated with 1 at different molar ratios (Fig. 7c–f) and the mixtures analyzed by ESI–MS. The recorded mass spectra show a drastic decrease of the relative intensity of the peak corresponding to the free DS form with the increasing amounts of **1**, while the relative intensity of the peaks of OP1 and OP2 slightly increased. The absence of MS peaks at m/z higher than 1,600, *i.e.* those corresponding to the Pt-derivatives already observed in the previous experiments, can be explained if we consider that the concentration of the eluate resulting from chromatography was estimated to be lower than 5 μM, which clearly impairs the observation by mass spectrometry of several coexisting species. Interestingly, the same experiment carried out in the absence of **1**, *i. e.* incubation of the DS and further chromatographic purification, show that the intensities of DS and SS remained unaltered, recording the same mass spectra as that in Fig. 7b.

The results of these two last experiments (Figs. 6, 7) confirm the second initial hypothesis: the Pt complex 1 reacts almost exclusively with the DS and, due to this interaction, the double helix melts into the SS oligonucleotides. This could be probably explained taking into account the small size of the double chain (only 10 bp), which is probably destabilized after Pt binding.

In order to confirm the interaction of **1** with DS, it was incubated (12 h at 37 °C) with **1** at the 1:1 and 1:5 DS:Pt ratios, and the corresponding CD spectra registered (Fig. 8). These show that the presence of complex **1** provokes a clean decrease of the intensity of the initial spectropolarimetric fingerprint of DS in the 290 nm region at the 1:5 DS:Pt ratio from the early stages of the reaction (1 min), and that was completed after 12 h. Interestingly, no modification of the secondary structure of DS was observed at the 1:1 DS:Pt ratio, which is indicative of the stability of the 10-bp DS fragment under the assayed conditions. These results suggests that the complex can cause modifications in the secondary structure of DNA at the appropriate DNA:Pt ratios.

The reactivity here observed for complex **1** with DS oligonucleotides clearly differs from that we recently reported for a similar Pt$^{II}$-dmba complex (Samper et al. 2012). The strong covalent interaction of **1** after release of the azaindolate ligand contrasts with the π interaction reported for the latter, which is due to the presence of an aminoacridine ligand. This confirms that in spite of the stabilizing role of the dmba chelating ligand in the structure of the complex, its reactivity can be tuned through the modulation of the nature of the two other ligands bound to the Pt$^{II}$ moiety.



◀**Fig. 7** **a** Size exclusion HPLC chromatograms corresponding to the individual SS oligonucleotides, OP1 and OP2, and the profile obtained after their incubation (2 h at 70 °C) to form the double-stranded form (DS). Negative ESI-TOF MS spectra recorded **b** after separation of the DS fraction form the SS oligonucleotides, and after incubation of the DS fraction with **1** at the Pt:DS **c** 1:1, **d** 5:1, **e** 10:1, and **f** 20:1 molar ratios

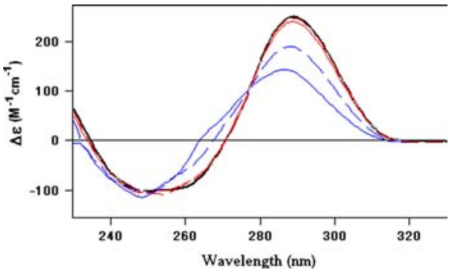


**Fig. 8** CD spectra recorded after the incubation of a 5 μM solution of DS (*black line*) with 1 (*red*) and 5 (*blue*) molar equivalents of **1** registered after 1 min (*dashed*) and after 12 h (*solid*) of incubation at 37 °C

### Competitive binding experiments

In order to further investigate the interaction between complex **1** and DNA, fluorescence competition experiments with ethidium bromide (EB) and Hoechst 33258 were carried out. EB is a planar cationic dye well-known to intercalate into the DNA double helix (Bresloff and Crothers 1975; Le Pecq 1971). While EB is only weakly fluorescent, the EB–DNA adduct is a strong emitter (near 620 nm) when excited near 520 nm. Quenching of the fluorescence may be used to determine the extent of the binding between the quencher **1** and commercial calf thymus DNA (ct-DNA). As seen in Fig. 9a, complex **1** can compete with EB for the DNA binding sites as there is a decrease in the fluorescence (by 28 % of the initial) at 602 nm with the increase of the amount of **1** added to the EB-DNA mixture. This suggests the idea that **1** could also interact with DNA by the intercalative mode, although the reason for the quenching of the EB-DNA adducts could also be the reduction in the number of available binding sites on DNA, presumably due to the competition with the complex, which is non-emissive under the experimental conditions (Beckford et al. 2011a, b; Ruiz et al. 2013).

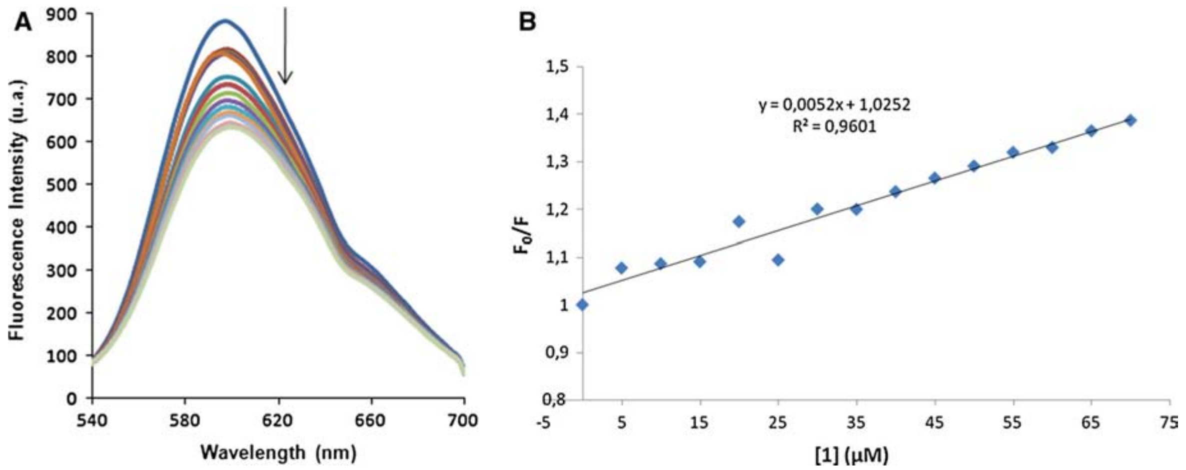


**Fig. 9** **a** Fluorescence spectra of EB bound to ct-DNA (*solid blue line*) in aqueous buffer solution in the presence of increasing amounts of **1**, at 298 K. $\lambda_{ex} = 520$ nm, [EB] = 0.33 μM, [DNA] = 10 μM, [complex **1**] (μM): 0–70 in 5 μM increments. **b** Stern–Volmer plot with the results obtained for the titration of EB bound to ct-DNA with **1**

In order to quantitatively assess the magnitude of the interaction between complex **1** and ct–DNA, the Stern–Volmer equation is used: $F_0/F = 1 + K_{SV}[Q]$ where F and $F_0$ respectively are the fluorescence intensities of the DNA solution in the presence and absence of the complex, $K_{SV}$ is the Stern–Volmer quenching constant and [Q] is the concentration of **1**. The good linearity of the Stern–Volmer plot (Fig. 9b) suggests a singular mode of quenching. The value of $K_{SV}$ was $5.2 \times 10^3\ M^{-1}$, which depicts complex **1** as a weak intercalator. The apparent binding constant ($K_{app}$) for the complex was $3.11 \times 10^3\ M^{-1}$, calculated using the equation: $K_{app} = K_{EB}[EB]/[Q]_{50}$, where $K_{EB} = 1.2 \times 10^6\ M^{-1}$ (Peberdy et al. 2007). The $K_{EB}$ value is the binding constant of EB to DNA and $[Q]_{50}$ is the concentration of **1** at 50 % of the initial fluorescence. One reason for the quenching of the EB-DNA adducts can be the already mentioned reduction in the number of available binding sites on DNA by competition with the complex.

Furthermore, we carried out another competition experiment using the Hoechst 33258 stain. This fluorescent dye binds to DNA and when this happens its fluorescence yield increases significantly (Weisblum and Haenssler 1974). Displacement by a competitor of the bound dye from its binding site leads to a decrease in the fluorescence intensity. It is well known that Hoechst 33258 binds to DNA in two concentration dependent ways, the first type of binding occurring in the minor groove at low dye-to-DNA ratios, which are the conditions we have assayed (Pjura et al. 1987; Guan et al. 2007). When complex **1** was added to Hoechst-ct-DNA solution a decrease (~33 %) in the fluorescence and the appearance of one peak at about 396 nm was observed (Fig. 10). The first effect suggests that complex **1** is able to weakly bind ct-DNA

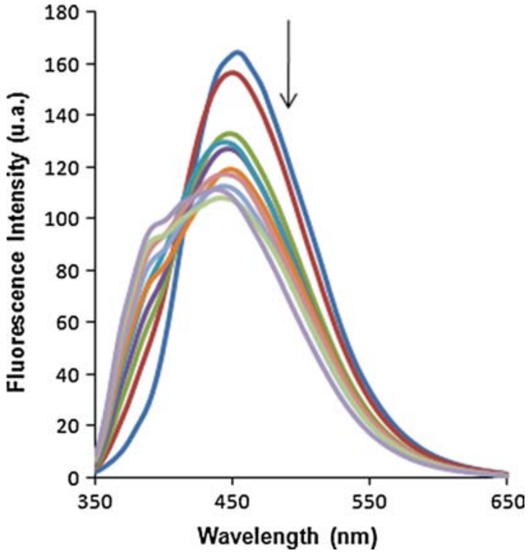


**Fig. 10** Fluorescence spectra of the Hoechst 33258-bound ct-DNA in aqueous buffer solution, in the absence (*solid blue line*) and presence of increasing amounts of complex **1**, recorded at 298 K. $\lambda_{ex} = 338$ nm, [Hoechst 33258] = 2 μM, [ct-DNA] = 20 μM, [complex **1**] (μM): 0–15 in 2.5 μM increments and 15–30 in 5 μM increments

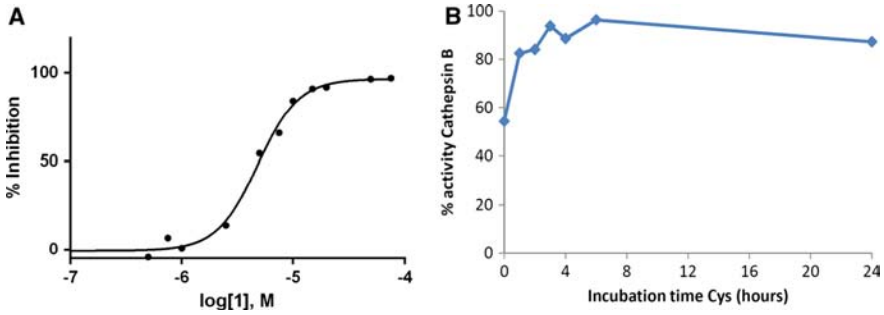


**Fig. 11** **a** Cat B activity inhibition curve for complex **1**. **b** Cysteine reactivation of cat B inhibited by $2 \times IC_{50}$ μM of **1**, in the presence of 1 mM cysteine. The first measurement was taken 5 min after the addition of cysteine

at the minor grove and the second was attributed to the fluorescence of the Pt(II) complex.

### In vitro evaluation of the bovine cat B inhibitory activity of complex 1

Cathepsin B (cat B) was proposed as a therapeutic target for the control of tumor progression (Strojnik et al. 1999) because the use of cat B inhibitors reduces both in vitro tumor cell mobility and invasiveness (Podgorski and Sloane 2003). Some metal complexes have been shown to be effective inhibitors of cat B (Casini et al. 2008). Hence, complex **1** was evaluated for activity against bovine cat B with an in vitro $IC_{50}$ of 4.86 ± 0.02 μM (obtained from Fig. 11a, which illustrates the inhibition of enzyme activity provoked by the presence of **1**). These results indicate that **1** is a very good cathepsin B inhibitor.

Additionally, the reactivation of cat B (after the previous inhibition) by the presence of cysteine (Fig. 11b) was evaluated in order to characterize the reversibility of the inhibition promoted by **1**. It was found that the addition of 1 mM cysteine to the 0.1 μM cat B inhibited solution results in almost full recovery of activity within a few hours (more than 80 % of initial enzyme activity was already restored after 3 h). Overall, these data support the hypothesis that loss of enzyme activity is mainly due to the specific interaction of **1** with the cat B active site instead of an enzyme denaturing process induced by the presence of the platinum complex.

## Conclusions

In this work, we propose that the reactivity of the (7-azaindolato-κN1)(*N*,*N*-dimethylbenzylamine-κN,κC)-(dimethylsulfoxide-κS)platinum(II) complex, **1**, with several proteins takes place, in most of the cases, by binding of compound **1** to the more accessible coordinating amino acids of the surface of the proteins. Consequently, this interaction would not lead to the subsequent protein denaturation or degradation. The specific reactivity of **1** with the chosen proteins allows to conclude that the Pt-ligand displaced in each case (dmso or azaindolate) is indicative of the nature of the protein ligand bound to the $Pt^{II}$ centre. Thus, when the $Pt^{II}$ moiety binds a S-atom of a Cys residue, the dmso is released from the complex, while the azaindolate is displaced when it binds to a N-atom of a His residue. The theoretical calculations here performed suggest that the release of the azaindolate ligand is promoted by a proton transfer to the non-coordinating N of the azaindolate ligand. Molecular modeling analysis of the interaction energies between common Pt-coordinating amino acids and complex **1** suggests that substitution at the initial azaindolate or dmso binding sites with histidines are a likely event. The experimental observation of a stronger interaction of the azaindolate than the dmso site with myoglobin is rationalized by assuming a change of the protonation state of the ligand at site A. This could result from a protonation of the azaindolate in solution or a deprotonation of the histidines of the protein. Although not reported yet for amino acids, this observation is consistent with the already reported huge shifts of pKa of biological building blocks in presence of metals (Lippert 2008; Roitzsch et al. 2005). Moreover, in both cases, such results consistently account for the importance of the *trans* effect in the substitution of cis-platinum derivatives (Montero et al. 2010; Manalastas et al. 2009). Further QM and QM/MM calculations on the entire mechanism of ligand substitutions and possible proton transfer occurring during such process would shed light on this aspect. The interaction of the Pt complex with the mammalian $Zn_7$-MT1 complex indicates the formation of a single, unexpected species containing 5 Pt-complex units without the initial dmso ligand, with displacement of all the initial $Zn^{2+}$ ions bound to the peptide. Furthermore, the formation of this particular species appears to depend only on the protein:complex ratio assayed.

The interaction of complex **1** with the OP1 and OP2 oligonucleotides show a strong covalent binding of the $Pt^{II}$ centre to both of them, in a similar way -it is without a special preference for one or another-, and in all cases with release of the initial azaindolate ligand. Interestingly, this interaction seems to occur only when both oligonucleotides are annealed forming a double strand. Circular dichroism data corroborate the modification of the secondary structure of DS by the presence of complex **1**. Furthermore, competing experiments with ethidium bromide and Hoechst 33258 displacement show that **1** is able to weakly bind to ct-DNA, probably in the minor groove. Additionally, complex **1** has been shown to be a good cathepsin B inhibitor, most probably interacting with the active site of the enzyme.

Concluding, compound **1** should be taken into consideration as a putative anticancer drug due to its strong interaction with oligonucleotides and its effective inhibition of cathepsin B, although it is definitely able to bind proteins that can hamper its arrival to the nuclear target. Significantly, the specific reactivity of **1** when interacting with S- or N-donor ligands, owing to the presence of the azaindolate ligand, opens a new way to modulate and direct $Pt^{II}$ binding to specific targets.

**Acknowledgments** This work was supported by the Spanish *Ministerio de Ciencia e Innovación* and FEDER through the following projects: SAF2011-26611 to J. Ruiz, BIO2012-39682-C02-01 to S. Atrian, and BIO2012-39682-C02-02 to M. Capdevila; CTQ2008-06866-C02-01 and consolider-ingenio 2010 to J.-D. Marechal. J. Ruiz also acknowledges the financial support received from *Fundación Seneca-CARM* (Project 08666/PI/08). The authors from UAB and UB are members of the *Grup de Recerca de la Generalitat de Catalunya* refs. 2009SGR-1457 and 2009SGR-68. S. Artime, at S. Atrian's lab, was responsible of the recombinant synthesis of $Zn_7$-MT1.

## References

ADF (2010) SCM, Theoretical Chemostry, Vrije Universiteit, Amsterdam, The Netherlands, http://www.scm.com

Barnes KR, Lippard SJ (2004) Cisplatin and related anticancer drugs: recent advances and insights. Met Ions Biol Syst 42:143–147

Barton JK, Goldberg JM, Kumar CV, Turro NJ (1986) Binding modes and base specificity of tris(phenanthroline)ruthenium(II) enantiomers with nucleic acids: tuning the stereoselectivity. J Am Chem Soc 108(8):2081–2088

Becke AD (1988) Density-functional exchange-energy approximation with correct asymptotic behaviour. Phys Rev A 38(6):3098–3100

Becke AD (1993) Density-functional thermochemistry. III. The role of exact exchange. J Chem Phys 98:5648–5652

Beckford F, Dourth D, Shaloski M Jr, Didion J, Thessing J, Woods J, Crowell V, Gerasimchuck N, Gonzalez-Sarrías A, Seeram NP (2011a) Half-sandwich ruthenium-arene complexes with thiosemicarbazones: synthesis and biological evaluation of [(η6-p-cymene)Ru(piperonal thiosemicarbazones)Cl]Cl complexes. J Inorg Biochem 105(8):1019–1029

Beckford F, Thessing J, Woods J, Didion J, Gerasimchuck N, Gonzalez-Sarrías A, Seeram NP (2011b) Synthesis and structure of [(η6-p-cymene)Ru(2-anthracen-9-ylmethylene-N-ethylhydrazinecarbothioamide)Cl]Cl; biological evaluation, topoisomerase II inhibition and reaction with DNA and human serum albumin. Metallomics 3(5): 491–502

Bresloff JL, Crothers DM (1975) DNA-ethidium reaction kinetics. Demonstration of direct ligand transfer between DNA binding sites. J Mol Biol 95(1):103–123

Casini A, Gabbiani C, Mastrobuoni G, Pellicani RZ, Intini FP, Arnesano F, Natile G, Moneti G, Francese S, Messori L (2007) Insights into the molecular mechanisms of protein platination from a case study: the reaction of anticancer platinum(II) iminoethers with horse heart cytochrome C. Biochemistry 46(43):12220–12230

Casini A, Gabbiani C, Sorrentino F, Rigobello MP, Geldbach ABTJ, Marrone A, Re N, Hartinger CG, Dyson PJ, Messori L (2008) Emerging protein targets for anticancer metallodrugs: inhibition of thioredoxin reductase and cathepsin B by antitumor ruthenium(II)-arene compounds. J Med Chem 51(21):6773–6781

Castillo-Busto ME, Meija J, Montes-Bayón M, Sanz-Medel A (2009) Diophantine analysis complements electrospray-Q-TOF data for structure elucidation of transferrin glycoforms used for clinical diagnosis in human serum and cerebrospinal fluid. Proteomics 9(4):1109–1113

Clark T, Chandrasekhar J, Spitznagel GW, Schleyer PR (1983) Efficient diffuse function-augmented basis sets for anion calculations. III. The 3–21 + G basis set for first-row elements, lithium to fluorine. J Comput Chem 4(3): 294–301

Cols N, Romero-Isart N, Capdevila M, Oliva B, Gonzàlez-Duarte P, Gonzàlez-Duarte R, Atrian S (1997) Binding of excess cadmium(II) to Cd7-metallothionein from recombinant mouse Zn7-metallothionein 1. UV-VIS absorption and circular dichroism studies and theoretical location approach by surface accessibility analysis. J Inorg Biochem 68(3):157–166

Egger AE, Hartinger CG, Hamidane HB, Tsybin YO, Keppler BK, Dyson PJ (2008) High resolution mass spectrometry for studying the interactions of cisplatin with oligonucleotides. Inorg Chem 47(22):10626–10633

Esteban-Fernández D, Moreno-Gordaliza E, Cañas B, Palacios MA, Gómez-Gómez MM (2010) Analytical methodologies for metallomics studies of antitumor Pt-containing drugs. Metallomics 2(1):19–38

Fernandez P, Farre X, Nadal A, Fernandez E, Peiro N, Sloane BF, Sih G, Chapman HA, Campo E, Cardesa A (2001) Expression of cathepsins B and S in the progression of prostate carcinoma. Int J Cancer 95(1):51–55

Gaussian 09, Frisch MJ, Trucks GW, Schlegel HB, Scuseria GE, Robb MA, Cheeseman JR, Scalmani G, Barone V, Mennucci B, Petersson GA, Nakatsuji H, Caricato M, Li X, Hratchian HP, Izmaylov AF, Bloino J, Zheng G, Sonnenberg JL, Hada M, Ehara M, Toyota K, Fukuda R, Hasegawa J, Ishida M, Nakajima T, Honda Y, Kitao O, Nakai H, Vreven T, Montgomery JA, Peralta Jr.JE, Ogliaro F, Bearpark M, Heyd JJ, Brothers E, Kudin KN, Staroverov VN, Kobayashi R, Normand J, Raghavachari K, Rendell A, Burant JC, Iyengar SS, Tomasi J, Cossi M, Rega N, Millam JM, Klene M, Knox JE, Cross JB, Bakken V, Adamo C, Jaramillo J, Gomperts R, Stratmann RE, Yazyev O, Austin AJ, Cammi R, Pomelli C, Ochterski JW, Martin RL, Morokuma K, Zakrzewski VG, Voth GA, Salvador P, Dannenberg JJ, Dapprich S, Daniels AD, Farkas Ö, Foresman JB, Ortiz JV, Cioslowski J, Fox DJ (2009) Gaussian, Inc., Wallingford CT

Guan Y, Shi R, Li X, Zhao M, Li Y (2007) Multiple binding modes for dicationic Hoechst 33258 to DNA. J Phys Chem B 111(25):7336–7344

Guerra CF, Snijders JG, te Velde G, Baerends EJ (1998) Towards an order-N DFT method. Theor Chem Acc 99(6):391–403

Hariharan PC, Pople JA (1973) Influence of polarization functions on MO hydrogenation energies. Theor Chim Acta 28(3): 213–222

Harper BW, Krause-Heuer AM, Grant MP, Manohar M, Garbutcheon-Singh KB, Aldrich-Wright JR (2010) Advances in Platinum Chemotherapeutics. Chem Eur J 16(24): 7064–7077

Hay J, Wadt WR (1985) Ab initio effective core potentials for molecular calculations. Potentials for potassium to gold including the outermost core orbitals. J Chem Phys 82(1):299–310

Hehre WJ, Dirchfield R, Pople JA (1972) Self-consistent molecular orbital methods. XII. Further extensions of gaussian-type basis sets for use in molecular orbital studies of organic molecules. J Chem Phys 56:2257–2261

Ivanov AI, Christodoulou J, Parkinson JA, Barnham KJ, Tucker A, Woodrow J, Sadler PJ (1998) Cisplatin binding sites on human albumin. J Biol Chem 273(24):14721–14730

Jakupec MA, Galanski M, Arion VB, Hartinger CG, Keppler BK (2008) Antitumor metal compounds: more than theme and variations. Dalton Trans 2:183–194

Jamieson ER, Lippard SJ (1999) Structure, recognition, and processing of cisplatin-DNA adducts. Chem Rev 99(9): 2467–2498

Karotki AV, Vasak M (2008) Interaction of Metallothionein-2 with Platinum-Modified 5′-Guanosine Monophosphate and DNA. Biochemistry 47(41):10961–10969

Knipp M, Karotki AV, Chesnov S, Natile G, Sadler PJ, Brabec V, Vasak M (2007) Reaction of Zn7Metallothionein with cis- and trans-[Pt(N-donor)2Cl2] anticancer complexes: trans-PtII complexes retain their N-donor ligands. J Med Chem 50(17):4075–4086

Le Pecq JB (1971) Use of ethidium bromide for separation and determination of nucleic acids of various conformational forms and measurement of their associated enzymes. Methods Biochem Anal 20:41–86

Lee C, Yang W, Parr RG (1988) Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density. Phys Rev B 37(2):785–789

Lippert B (1999) Cisplatin: chemistry and biochemistry of a leading anticancer drug, 1st edn. Helvetia Chimica Acta/ Wiley-VCH, Zurich/Weinheim, p 576

Lippert B (2008) Ligand-pKa shifts through metals: potential relevance to ribozyme chemistry. Chem Biodivers 5(8):1455–1474

Manalastas WW Jr, Dy ES, Quevada NP, Kasai H (2009) Trans-influence of nitrogen- and sulfur-containing ligands in trans-platinum complexes: a density functional theory study. J Phys Matter 21:064210/1–064210/6

Montero EI, Benedetti BT, Mangrum JB, Oehlsen MJ, Qu Y, Farrell NP (2007) Pre-association of polynuclear platinum anticancer agents on a protein, human serum albumin. Implications for drug design. Dalton Trans 43:4938–4942

Montero EI, Zhang J, Moniodis JJ, Berners-Price SJ, Farrel P (2010) The trans influence in the modulation of platinum anticancer agent biology: the effect of nitrite leaving group on aquation, reactions with S-nucleophiles and DNA binding of dinuclear and trinuclear compounds. Chem Eur J 16(30):9175–9185

Paolicchi A, Lorenzini E, Perego P, Supino R, Zunino F, Compòrti M, Pompella A (2002) Extra-cellular thiol metabolism in clones of human metastatic melanoma with different gamma-glutamyl transpeptidase expression: implications for cell response to platinum-based drugs. Int J Cancer 97(6):740–745

Peberdy JP, Malina J, Khalid S, Haman MJ, Rodger A (2007) Influence of surface shape on DNA binding of bimetallo helicates. J Inorg Biochem 101:1937–1945

Perdew JP, Chevary JA, Vosko SH, Jackson KA, Pederson MR, Sing DJ, Fiolhais C (1992) Atoms, molecules, solids, and surfaces: applications of the generalized gradient approximation for exchange and correlation. Phys Rev B 46(11):6671–6687

Pjura PE, Grzeskowiak K, Dickerson RE (1987) Binding of Hoechst 33258 to the minor groove of B-DNA. J Mol Biol 197(2):257–271

Podgorski I, Sloane BF (2003) Cathepsin B and its role(s) in cancer progression. Biochem Soc Symp 70:263–276

Roitzsch M, Añorbe MG, Miguel PJS, Müller B, Lippert B (2005) The role of intramolecular hydrogen bonding on nucleobase acidification following metal coordination: possible implications of an "indirect" role of metals in acid-base catalysis of nucleic acids. J Biol Inorg Chem 10(7):800–812

Ruiz J, Rodriguez V, de Haro C, Espinosa A, Perez J, Janiak C (2010) New 7-azaindole palladium and platinum complexes: crystal structures and theoretical calculations.

In vitro anticancer activity of the platinum compounds. Dalton Trans 39(13):3290–3301

Ruiz J, Vicente C, de Haro C, Bautista D (2013) Novel bis-C, N-cyclometalated iridium(III) thiosemicarbazide antitumor complexes: interactions with human serum albumin and DNA, and Inhibition of cathepsin B. Inorg Chem 52(2):974–982

Samper KG, Vicente C, Rodríguez V, Atrian S, Cutillas N, Capdevila M, Ruiz J, Palacios Ò (2012) Studying the interactions of a platinum(II) 9-aminoacridine complex with proteins and oligonucleotides by ESI-TOF MS. Dalton Trans 41(1):300–306

Spitznagel GW, Clark T, Chandrasekhar J, Schleyer PR (1982) Stabilization of methyl anions by first-row substituents. The superiority of diffuse function-augmented basis sets for anion calculations. J Comput Chem 3(3):3633–3640

Strojnik T, Kos J, Zidanik B, Golouh R, Lah T (1999) Cathepsin B immunohistochemical staining in tumor and endothelial cells is a new prognostic factor for survival in patients with brain tumors. Clin Cancer Res 5(3):559–567

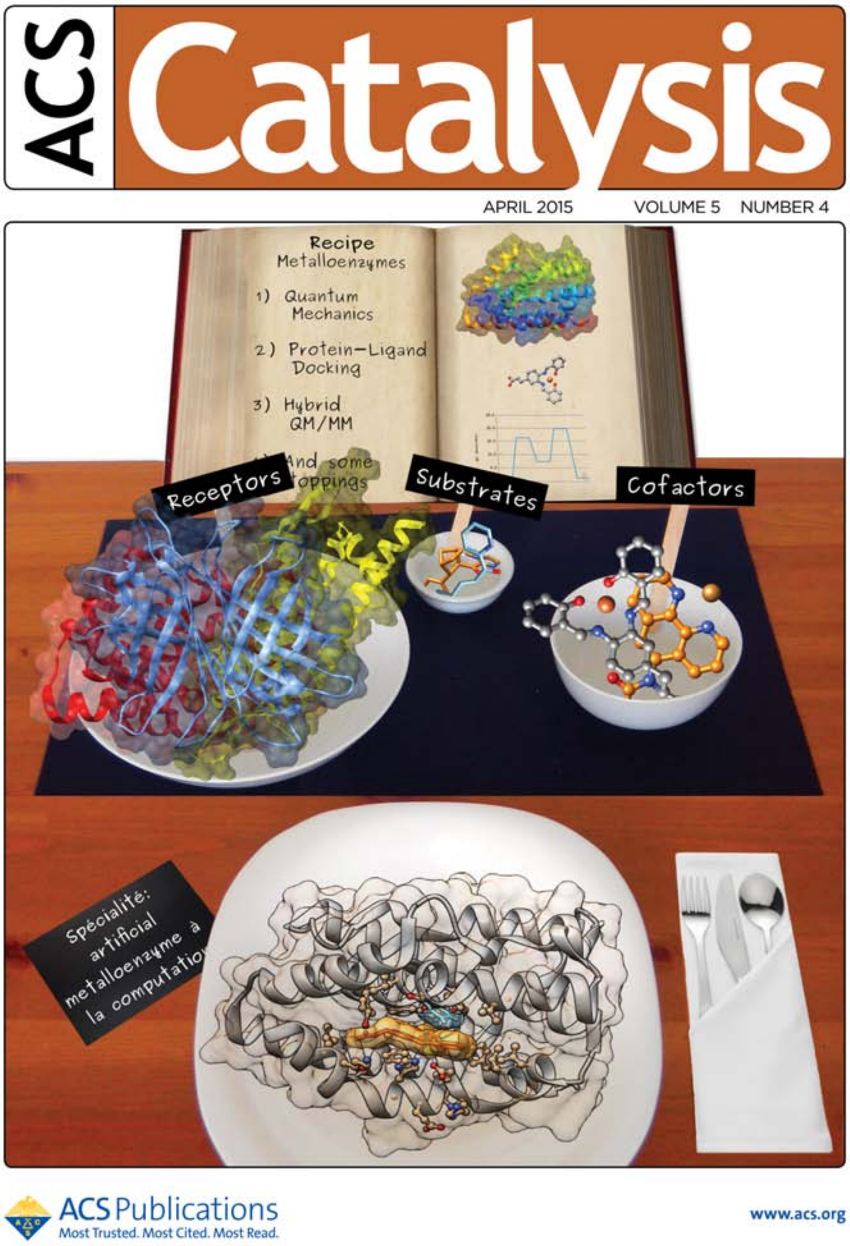
te Velde G, Bickelhaupt FM, van Gisbergen SJA, Guerra CF, Baerends EJ, Snijders JG, Ziegler T (2001) Chemistry with ADF. J. Comp. Chem. 22(9):931–967

van Lenthe E, Ehlers AE, Baerends EJ (1999) Geometry optimizations in the zero order regular approximation for relativistic effects. J Chem Phys 110(18):8943–8953

Weisblum B, Haenssler E (1974) Fluorometric properties of the bibenzimidazole specific for AT [adenine-thymine] concentration in chromosomal DNA. Chromosoma 46(3): 255–260

# 2 Toward the Computational Design of Artificial Metalloenzymes: From Protein-Ligand Docking to Multiscale Approaches

Robles, V. M.; Ortega-Carrasco, E.; Alonso-Cotchico, L.; Rodriguez-Guerra, J.; Lledós, A., Maréchal, J. D. *ACS Catal.* **2015**, *5*, 2469-2480.



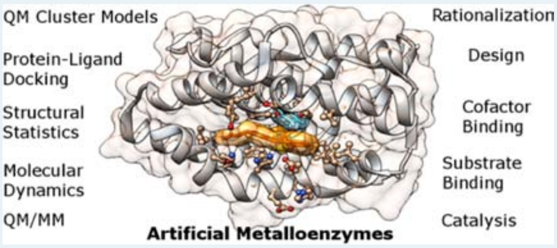
ACS Catalysis
Perspective
pubs.acs.org/acscatalysis

# Toward the Computational Design of Artificial Metalloenzymes: From Protein−Ligand Docking to Multiscale Approaches

Victor Muñoz Robles, Elisabeth Ortega-Carrasco, Lur Alonso-Cotchico, Jaime Rodriguez-Guerra, Agustí Lledós, and Jean-Didier Maréchal*

Departament de Química, Universitat Autònoma de Barcelona, 08193 Cerdanyola del Vallès, Barcelona, Spain

**ABSTRACT:** The development of artificial enzymes aims at expanding the scope of biocatalysis. Over recent years, artificial metalloenzymes based on the insertion of homogeneous catalysts in biomolecules have received an increasing amount of attention. Rational or pseudorational design of these composites is a challenging task because of the complexity of the identification of efficient complementarities among the cofactor, the substrate, and the biological partner. Molecular modeling represents an interesting alternative to help in this task. However, little attention has been paid to this field so far. In this manuscript, we aim at reviewing our efforts in developing strategies efficient to computationally drive the design of artificial metalloenzymes. From protein−ligand dockings to multiscale approaches, we intend to demonstrate that modeling could be useful at the different steps of the design. This Perspective ultimately aims at providing computational chemists with illustration of the applications of their tools for artificial metalloenzymes and convincing enzyme designers of the capabilities, qualitative and quantitative, of computational methodologies.



**KEYWORDS:** *artificial metalloenzymes, biocatalysis, molecular modeling, multiscale approaches, protein−ligand dockings*

## INTRODUCTION

Biocatalysis consists of the industrial application of enzymes for the manufacturing of chemical compounds. It is one of the cornerstones for green and sustainable chemistry because enzymes are by nature biodegradable, biocompatible, and easily renewable.[1] Despite being widespread in current industries, most biocatalysts are based on naturally occurring enzymes that, despite their variety, cover only a narrow spectrum of the needs of chemical industries.

During the past century, homogeneous catalysis has been the most prolific chemical field in discovering new chemical reactivities. The award of two recent Nobel Prizes of Chemistry (Chauvin, Grubbs, and Schrock in 2005; Heck, Negishi, and Suzuki in 2010) appears particularly illustrative. However, the transition metal complexes that sustain homogeneous catalysis are in their majority functional under nonenvironmentally friendly conditions, which include apolar solvents and low or high temperatures, among others. Moreover, control over substrate and regio- and enantiospecificities is generally challenging in these complexes; conversely, they are properties inherent to enzymatic activities.

With one-third of naturally occurring biocatalysts containing metal ions, metalloenzymes have been the focus of attention of enzyme designers. One possible framework consists of mutating residues that coordinate the metal in the native biomolecule or simply switch the metal by another. Such approaches have led to interesting outcomes in recent years, although modulating the activity of these scaffolds resides mainly in the biochemical space afforded by the 20 amino acids available in Nature.[2−5]

Another framework consists of physically merging homogeneous catalysts within a biomolecular host. Conceptually mimicking natural hemoenzymes, this strategy is increasingly applied to the development of biocatalysts absent from the biological realm.[6] In the resulting hybrids, also called artificial metalloenzymes, the cofactor (synthetic in this case) provides most of the catalytic specificity of the system. The protein environment protects the homogeneous catalyst from the solvent and generates an asymmetric second coordination sphere that dictates substrate, regio- and enantioselectivities, and specificities (Figure 1). Today, numerous systems developed using this concept have already been reported and include reactivities such as hydration of ketone,[7] transfer hydrogenation,[8] and sulfoxidation.[9] Strategies used to incorporate the cofactor inside the protein include pure host−guest interactions, "Trojan horse" insertion in which the cofactor is covalently bound to the natural ligand of a protein, or covalent anchoring in which peripheral substitutents of the organometallic catalyst chemically bind to the host.[10] A nonexhaustive list of artificial metalloenzymes with their catalytic activities can be found in Table 1.

The successful development of artificial metalloenzymes stands on the quality of the molecular partnership between

**Received:** January 4, 2015
**Revised:** February 19, 2015
**Published:** February 23, 2015

DOI: 10.1021/acscatal.5b00010
*ACS Catal.* 2015, 5, 2469−2480

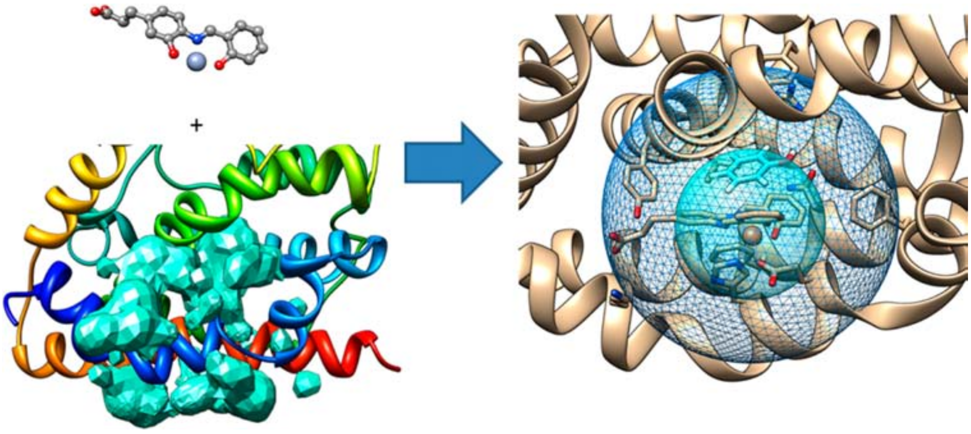


**Figure 1.** Schematic representation of the process of designing an artificial metalloenzyme. Homogenous catalysts (top left) and a protein host (bottom left) with sufficient vacant sites (solid blue blobs) are merged to provide artificial metalloenzymes (right onside). Their activity is driven for the first coordination sphere of the metal (blue sphere) and substrate (blue stick atoms) binding and orientation defined by the second coordination sphere environment (dark blue mesh sphere).

**Table 1. A List of Artificial Metalloenzymes and Their Catalytic Activities**

| biomolecular scaffold | transition metal | organic cofactor | catalytic activity | substrate | enantioselectivity, % | ee | ref |
|---|---|---|---|---|---|---|---|
| NikA transport protein | iron(III) | organic ligand L1 | oxidation | sulfides | 10 | | 13 |
| LmrR | copper(II) | phenanthroline | syn hydration | ketones | 84 | ? | 14 |
| | | phenanthroline | Diels–Alder reaction | azachalcone | >97 | + | 15 |
| bovine *β*-lactoglobulin | rhodium(III) | fatty acid derivatives | hydrogenation | trifluoroacetophenone | 26 | R | 16 |
| *β*-helical bionanotube | scandium(III) | bipyridine, Ser, Thr | epoxide ring-opening reaction | *cis*-stilbene oxide | 17 | R | 17 |
| streptavidin mutants | osmium(VIII) | quinidine or quinine derivatives | asymmetric dihydroxylation | olefins | 95 | R | 18 |
| | iridium(III) | biotinilated complex | hydrogenation | cyclic imines | 96, 78 | RS | 8 |
| DNA | copper(II) | phenanthroline | syn hydration | enones | 72 | R | 19 |
| DNA | copper(II) | DNA intercalating moiety | Diels–Alder reaction | dienophiles | 90 | exo | 20 |

substrate, organometallic and biological partners in terms of binding and catalytic control.[11] The actual strategies for their design are time- and resource-consuming and consist mostly of trial-and-error procedures. In general, the main steps involved in the process consist of (1) the identification (by biochemical intuition of the researchers) of one possible protein scaffold able to bind a given homogeneous catalyst, (2) experimental binding assays of the artificial cofactor in this particular host, (3) testing of the catalytic activity for a prototypical substrate, and (4) optimization of the initial hit toward catalytic preferences.[7,12−20]

Molecular information is fundamental in all aspects of the process. However, structural knowledge provided by approaches such as X-ray or NMR are rather scarce for artificial metalloenzymes. These techniques generally fail because the complementarity between subsystems is not optimal especially when dealing with the first candidates of these systems. The interaction among the three partners has not suffered evolutionary pressures, affinity constants are generally low, and substantial protein engineering is needed to stabilize the structure of the hybrid system. Molecular modeling offers an interesting alternative to reach atomic details on the mechanism of artificial metalloenzymes and help in their design; however, the development of synthetic enzymes through in silico approaches is still in its infancy, and only a few attempts have been performed on the particular case of artificial metalloenzymes.

This manuscript aims to underline the particularities of artificial metalloenzymes in the area of in-silico-based enzyme design; give an overview of the strategies we have been empowering in the recent years to establish an efficient framework in this field as well as their consequent results; and finally, to focus on what we believed should be the future of modeling-based artificial metalloenzymes. It is a Perspective that intends to motivate computational chemists to consider artificial metalloenzymes as an interesting (but challenging) target as well as present to experimentalists how the variety of computational tools could be relevant for their designs.

**A Brief Overview on Molecular Modeling Tools.** Molecular modeling is now widespread at the interface between chemistry and biology, with models increasingly accurate, but molecular modeling is also a general term for defining a series of computational methods based on physical models with different degrees of accuracy and computational needs. Methods based on force field approaches, also called molecular mechanics (MM), allow vast geometrical samplings because of the relatively low ratio between the number of atoms and the computational cost. MM approaches are used mainly to study systems of large dimensionality (i.e., an entire protein in a solvated medium) and allow the exploration of large conformational spaces. Generally combined with deterministic (i.e., molecular dynamics) or stochastic (i.e., Monte Carlo) search algorithms, MM allows the study of changes in the shape of the molecule associated with their motions, extraction of statistical thermodynamics values, or handling of the prediction of the structure of large databases of the compounds, among others. When dealing with the interaction between partners, the additional degrees of freedom associated with translation and rotation increase substantially the geometrical space to explore. In these cases, a common solution consists of using simplified force fields centered on noncovalent terms (called scoring functions) and reducing the number of degrees explored during the conformational search (i.e., all the degrees of freedom of the ligand are considered, but a reduced number of amino acids or, eventually, collective motions are allowed to move during the docking process). Protein−ligand dockings, which aim at predicting the structure of the complexes formed between small molecules and proteins, are based on these premises. In any case, only very specific and nonstandardized MM approaches are able to predict fine electronic effects.[21,22]

Computational methods based on quantum mechanics (QM) accurately reproduce the nature of the electronic properties of the molecules and allow simulating changes in their chemical state. QM approaches are used for very different molecular problems, including spectroscopic and photoelectronic processes and any system in which its coordination or the covalent linkages change during a chemical process. A vast ensemble of QM methods is accessible nowadays. It is likely that those with the wider number of applications are based on the density functional theory (DFT). These methods allow the insertion of fine electronic effects (correlation) for a relatively low additional cost over the typical Hartree−Fock calculations and are particularly relevant in fields such as organometallics. Despite their success, DFT techniques are based on a series of approximations that could substantially limit the reliability of their results (i.e., dealing with changes in spin states of a transition metal is still a challenging task).[23,24] The quest for the best DFT method is a vivid field of research, and still today, DFT capabilities seem system-dependent.[25] Whatever their ground, though, QM approaches are counterbalanced by expansive computational costs that do not allow sampling of large-dimensional problems.

Approaches that combine several methodologies together are increasingly applied in molecular sciences to overcome the limitations of individual methodologies. Generally referred to as multiscale, integrative, or hierarchical methods, their potential has already been widely recognized, including by the Nobel Prize in chemistry awarded to Karplus, Warshel, and Levitt in 2013 "for the development of multiscale models for complex chemical systems". Prototypical multiscale approaches are the hybrid quantum mechanics/molecular mechanics (QM/MM) methods, which considers part of the molecule under a quantum mechanical framework and the remaining part under a molecular mechanics approximation. QM/MM methods are now legion and key in the simulation of biomolecular systems. They differ in the algorithms used in each subset of atoms and how the information is transferred from one to another.[26] Major breakthroughs in metalloenzymes have been reached with these methodologies in decoding enzymatic mechanisms.[27−31]

Other combinations are frequent in biosimulation either under successive steps of different methods or integrated under a unique protocol. Focusing on those related to the study of enzymatic systems, some bridge molecular dynamics and QM/MM calculations. This combination is particularly interesting when an enzyme−substrate complex is relatively well-defined (for example, from an X-ray structure obtained with a substrate analogue) and is aimed to improve the quality of the catalytic path explored under the QM/MM energetic landscape.[32] However, when more complex binding processes need to be modeled, protocols integrating protein−ligand dockings are necessary. Although such combinations are less frequent in the modeling of enzymatic reactions, their use has been particularly relevant in several recent studies, including those related to the study of the reactivity of cytochromes P450s 3A4,[33] the change of specificity of cytochrome P450 2D6,[34] the elucidation of the catalytic mechanism of *Trametopsis cervina* lignin peroxidase,[35] or the promiscuous activity of human carbonic anhydrase against cyanic acid.[36] In all of them, the dockings are used, alone or in combination with MD runs, to provide physically sound complexes between the substrate and the enzyme prior to catalysis.

**In-Silico-Based Enzyme Design.** Different ways of developing new enzymatic activities are under the scrutiny of designers. Their differences arise from the degree of molecular diversity involved in the biomolecular scaffolds and include the engineering of a few amino acids in the active site, development of catalytically active peptides, or the redesign of a pre-existing scaffold.[37] Computation has been increasingly involved in several of these strategies. We here focus on computer-based designs in which conceptual frameworks best overlap with those that could lead to artificial metalloenzymes.

De novo design of artificial enzymes consists of identifying a protein scaffold and its consequent mutations to catalyze a nonnatural reaction on a given substrate. The combinatorial space to reach an active scaffold is tremendous and not yet achievable by experimental means. Part of de novo enzymes are based on relatively small peptides that could self-assemble. Systems with those dimensions confine the search for activity into a sequential space easier (but still challenging) to handle with respect to large folded proteins.[38−40] For designs considering larger folds, computation is more frequently required.

The most established procedures for computer aided de novo design of artificial enzymes stand on hypothetical transition state structures of a nonnatural reaction that could be embedded in a protein medium. The identification of such geometry is often performed by quantum mechanical calculations on a minimalist active site, which includes the substrate and a series of functional groups representing side chain of amino acids that could stabilize its orientation and participate in the reaction. These cluster models, also referred to as *theozyme*s under the definition of Houk and co-workers,[41] are used as starting points for posterior search algorithms under an explicit protein environment.

Mayo et al. were among the first to generate a novel proteic scaffold from in silico approaches and reached a synthetic $\beta\beta\alpha$ motif designed by screening $1.9 \times 10^{27}$ possible amino acid sequences.[42] Pursuing their efforts, they computationally identified mutations in the 108-residue *Escherichia coli* thioredoxin, leading to a "protozyme" able to catalyze the histidine-mediated nucleophilic hydrolysis of *p*-nitrophenyl acetate into *p*-nitrophenol and acetate.[43] Subsequently, Mayo and co-workers implemented a new method to place the substrate within the active site of the protein while the designing algorithm is exploring the conformational and chemical space.[44] In their more recent successes, they iteratively

DOI: 10.1021/acscatal.5b00010
*ACS Catal.* 2015, 5, 2469−2480

execute computational simulations (best designing algorithm and MD simulations) with X-ray crystallography to obtain one of the best Kemp eliminases so far reported ($K_{cat}/K_m$ of 430 $M^{-1}$ $s^{-1}$ rate after three iterations).[45]

The Baker laboratory is another clear example of success in de novo design of artificial enzymes by in silico approaches using a procedure that combines rational design and directed evolution.[46,47] Briefly, their methodology accounts for an extensive search of pre-existing high-resolution protein structures that could accommodate the transition state structure using the RosettaMatch algorithm.[48] A scaffold is considered a match if it satisfies that all the amino acid side chains of the theozyme that can be placed on the protein scaffold. Each match is then optimized using the RosettaDesign methodology for proteins and small molecules.[49,50] Except for catalytic residues of the theozyme, all the remaining amino acids in the vicinity are redesigned so that the final cavity has the maximum shape complementarity with the modeled transition state. All the resulting structures are then screened for compatibility with substrate/product binding and ranked according to the catalytic geometry and the computed transition state binding energy. This way, a handful of different putative new enzymes are selected for experimental characterization. Those that present the final activity will undergo a series of directed evolution steps for further optimization.

Computational involvements in manipulating enzymes are limited not only to de novo designs but also to bioengineering processes, some of them relevant for the present work. For enzymes with proven reactivity, computation can be used to rationalize their mechanism, improve it, or ultimately reorient their activity. An increasing number of studies with this objective have appeared in the literature over the past decade. Both pure quantum mechanical on large models of the active site[51] and hybrid QM/MM calculations[30,35,52,53] are used to this end.

Artificial metalloenzymes constructed by the insertion of homogeneous catalysts into protein have thus far received very little attention from computation.[52,54−56] Conceptually, their design stands on the same premises as pure organic systems: the modeling should identify transition state structures stabilized under a biocompatible host. However, these composites work because of a complementarity of the three different molecular entities and not only on protein−substrate recognition and activation. In this case, molecular modeling needs to handle the cofactor−host−substrate triad as best it can. The need in dealing with metal-mediated recognition processes, the effects they could induce on the structure of the host, and the reactivity of the final composite provide extra complexity for molecular modeling.

**Our Computational Framework for the Modeling of Artificial Metalloenzymes.** In the field of artificial metalloenzymes, molecular modeling needs to address processes involving large conformational sampling on one side and fine electronic effects on the other. The former are related to the binding of the artificial cofactor in the host and the orientation of the substrate in an efficient manner for the reaction to proceed. In principle, those steps can be achieved by protein−ligand dockings. The latter consists of events related to the identification of stable geometries of the isolated cofactor, the changes in its coordination sphere upon binding, and the characterization of low-energy reactive paths with the emphasis on identifying transition states structures. Pure QM and hybrid QM/MM calculations represent the best candidates for leading this part of the modeling.

In recent years, the objective of our group has been to generate computational protocols efficient for the design of artificial metalloenzymes. As a framework, we decided to use standardized (or lowly tuned) computational chemistry methods as well as multiscale approaches as a function of the problem presented. Regarding protein−ligand dockings, we decided to use the commercial software GOLD, which is one of the few that contain metal parameters, although at the beginning of our work, none were designed for metal-containing ligands.[57,58] It also affords flexible schemes for both receptors and ligands, which we applied in most of our calculations. Quantum simulations are performed using the Gaussian package (Gaussian09)[59] for both pure DFT calculations and QM/MM calculations. The latter are performed using the ONIOM approach using mechanical and electronic embedding.[60] Finally, structural modeling and statistics are performed in the UCSF Chimera[61] platform and include, nonexclusively, the exploration of rotameric conformation of amino acids[62] or the clustering of large sets of geometries using the NMRClust approach.[63] To ease our development of integrated approaches, we also developed a series of interfaces written in Python into the UCSF Chimera environment, which allows rapid input/output exchanges between the different methodologies we use.

## 1. BINDING OF ORGANOMETALLIC COMPOUNDS TO PROTEIN: THE QUEST FOR RESTING STATE MODELS

The design of artificial metalloenzymes relies, on a first instance, on the identification of structural matches between a biomolecule and a homogeneous catalyst. Only efficient complementarities should lead to a precatalytic state. The availability of 3D models of protein−artificial cofactor complexes is therefore fundamental at this stage. Although protein−ligand dockings represent one of the cornerstones in medicinal chemistry and drug design projects, little attention has been paid to the interaction of organometallic compounds with proteins. Indeed, only a small amount of drug candidates contain transition metal ions. However, metals are considered in several of these techniques for their presence in the active site of metaloproteins and how they influence the binding of organic drugs. In this case, different strategies are available to introduce metal−ligand interactions in the calculation of the energy, ranging from simple electrostatics (hydrogen-bond-donor-like function)[64] to coordination rules.[58,65,66]

For the interaction of organometallic entities with proteins, an accurate computational prediction has not yet been standardized. On the basis of bioinorganic considerations, efficient modeling should take into account (1) changes of the electronic state and geometry of the first coordination sphere of the metal upon binding, (2) geometrical changes on the entire cofactor, and (3) possible induced effects on the protein scaffold. Dealing with all these variables is beyond the scope of standard protein−ligand docking software, and different levels of approximation are mandatory.

A first case scenario consists of the situation in which no ligand exchanges occur on the metal when migrating from solution to its cavity in the host. These so-called "inert scaffold"[67] interactions imply that only subtle rearrangements of the first coordination sphere of the metal happen upon binding but that its overall geometry is little affected. From a

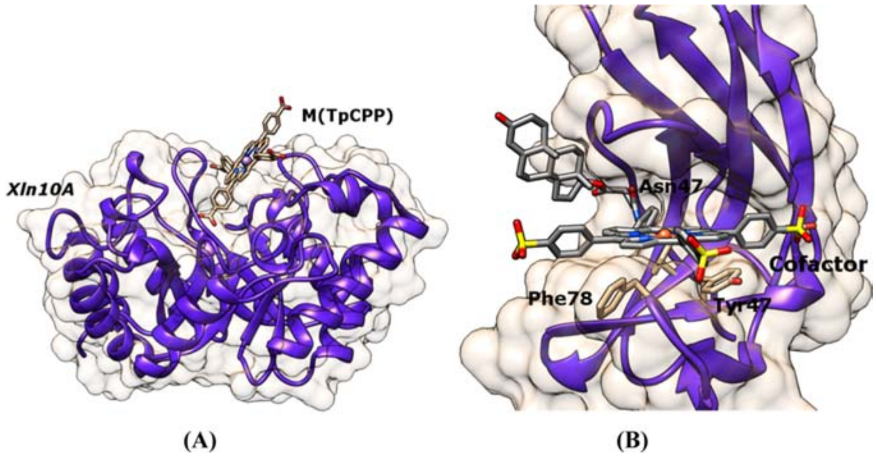


**Figure 2.** General view of different artificial metalloenzymes for which complementarity between host and cofactor were studied by protein ligand docking: (A) supramolecular interaction between iron porphyrin (concretely Fe(III)TCPP) and the xylanases 10A; (B) NCS bound with a Trojan horse testosterone−porphyrin derivative.

computational point of view, such systems could be simulated by dockings in which the close vicinity of the metal in the ligand remains rigid and the rest of the scaffold is optimized during the conformational search.

On the basis of this hypothesis, we recently showed that protein−ligand docking software behaves for inert scaffolds as well as it does for organic ligands. Limited by the reduced number of crystal structures of organometallics bound to protein available in the Protein Data Bank,[68] we performed a benchmark on structures corresponding to metal-containing inhibitors bound to their kinase targets and designed at Meggers' Laboratory.[67] Using GOLD as the method of choice, we tested different flexible schemes and scoring functions. The results were of very good quality. First, excellent structural matching between calculated low-energy structures and experimental complexes is observed. Between 75 and 94% of the theoretical complexes presented an RMSD lower than 2.5 Å from their experimental counterparts. The scores obtained were also of good quality, with correlations between experiment and theory reaching $R^2$ values up to 0.8 for those scoring functions that best behave. Of the scoring functions available in Gold, ChemScore[57] appears the most robust for both structural and energetic predictions. Outliers were encountered only when the geometry of the ligand bound to the metal differs substantially between the isolated conditions and the proteic complexes, something happening rarely in this set but that illustrates that the improvement of how dockings can explore the conformational changes related to the first coordination sphere of the metal or the coupling of dockings with accurate electronic methods such as QM/MM would represent a major step forward in those predictions.[69]

The prediction of the binding of synthetic cofactors to proteins under an inert interaction represents an initial step along the quest of 3D models of the resting state of artificial metalloenzymes. Despite an apparent simplicity, such approximation still provides crucial information on the most important features in defining protein−ligand binding: shape, hydrophobic and hydrogen bonding complementarity. Such information is extremely valuable when dealing with a first generation of an artificial enzyme for which structural information is missing. As such, dockings performed under this assumption have been one of the cornerstones in our collaboration with Mahy and co-workers, whose main objective is the development of artificial oxidases, more particularly, peroxidases and cytochromes P450.[70]

The first of our studies allows rationalization of the difference in activity of iron(III)-tetra-$\alpha_4$-*ortho*-carboxyphenylporphyrin (Fe(ToCPP)) and iron(III)-tetra-*para*-carboxyphenylporphyrin (Fe(TpCPP)) systems embedded into xylanase A (Xln10A) from *Streptomyces lividans*. Xln10A is a glycoside hydrolase that hydrolyzes $\beta$-1,4 bonds in the main chain of xylan and is available at low cost and in large quantities.[71] Of the most important results, the protein−ligand dockings showed that Fe(TpCPP) enters deeper into the large Xln10A cleft than its Fe(ToCpp) counterpart. This better complementarity is due to a major part of the porphyrin ring anchored into the binding site as well as a substantial hydrogen-bonding network between the peripheral carboxylates of consecutive aromatic substituents and two polar patches of the receptor (Figure 2A). Moreover, the calculated Fe(TpCPP)-Xln10A complex shows the cofactor with one of its faces slightly packed on the surface of the binding side protein and the other accessible to the solvent. This orientation is in agreement with the experimental observation that only one imidazole could coordinate the iron of the porphyrin. Similar approaches on the same target also concluded that a sulfoxinated tetra-*para*-phenylporphyrin shows different binding modes with regard to its carboxylic counterparts with wider variability in interacting with polar patches of the receptor.[72] A final study with Xln10A as a receptor for porphyrin complexes showed that metallic Schiff base cofactors displayed very limited complementarities to the Xln10A binding site but that Mn(TpCPP) afforded a cavity vast enough to accommodate a substrate for the epoxidation of the series of aromatic styrenes. Interestingly, one of the residues of the receptor (Arg139) is identified to control the access of the substrates.[73]

More recently, we focused on another receptor: an engineered mutant of neocarzinostatin (NCS). NCS is a 113 amino acid chromoprotein secreted by *Streptomyces* that binds a nine-membered enediyne "chromophore" responsible for the cytotoxic and antibiotic activities of the protein−ligand complex.[74] The NCS 3.24 mutant allows the binding of two

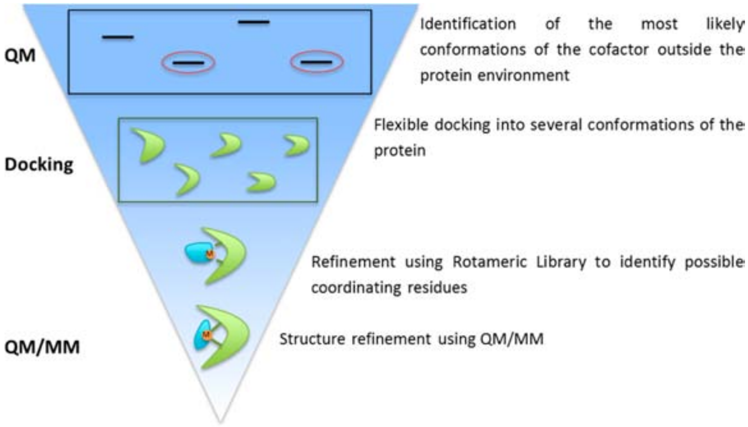


**Figure 3.** Steps of our integrative procedure combining docking, structural statistics and quantum mechanical based calculations.

testosterone molecules in its hydrophobic binding site in place of the natural chromophore. Using a so-called "Trojan horse" strategy, Mahy and co-workers synthesized an iron(III)−porphyrin−testosterone derivative able to bind to NCS-3.24. The resulting hybrid is able to catalyze the chemoselective and slightly enantioselective (ee = 13%) sulfoxidation of thioanisole by $H_2O_2$.

To increase the yield and the enantioselectivity of the construct, protein−ligand dockings were applied to look for improvement of the interaction between the cofactor and the protein. The molecular modeling showed that the porphyrin macrocycle fits perfectly into the protein binding site and is well sandwiched between the two subdomains of the protein (Figure 2B). However, the metal ion remains exposed to the solvent, which could explain the moderate enantioselectivity observed. The study also gave hints on possible improvements in the "Trojan horse" strategy because the artificial cofactor has filled up the two testosterone sites entirely and displaced its conjugated scavenger out to the solvent. Smaller cofactors are therefore expected to better fit inside the binding site of the enzyme and provide a wider asymmetric environment for enantioselective reactions.[75,76] A final scaffold we studied is a family of porphyrin-binding catalytic antibodies that are able to perform peroxidase activities. In conjunction with X-ray structures that were not conclusive on the geometry of the cofactor in the hapten recognition site, we could qualitatively rationalize both activity and binding.[77]

Although the binding of organometallics to their host in an inert fashion is frequent with drug compounds, for homogeneous catalysts, this hypothesis is valuable only as a "first shoot" for structural knowledge. In a wider context, results obtained under this assumption have to be nuanced. On one side, the absence of coordination changes during binding needs to be compared with experimental data, mainly spectroscopic, to validate such an approximation. Moreover, this approximation is interesting mainly for resting state structures. Indeed, either prior to or during the catalysis, one or several groups bound to the metal are likely to be displaced from its isolated situation in solvent to its binding to the protein cavity. When disposing of the clear idea on which groups could leave the cofactor (i.e., labile water on the top of the iron in a heme like complex), a possible strategy consists of mixing dockings and quantum-based approaches to identify correct resting states of the artificial metalloenzymes.

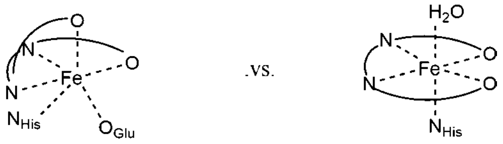
To this end, we developed an integrative procedure that combines docking, structural statistics, and quantum mechanical-based calculations (Figure 3). In this process, stable structures of the isolated cofactor obtained either from pure quantum mechanical calculations or from a database of small molecules (considering, if necessary, spin and oxidation states) are initially docked into the receptor cavity.

During the docking, we simulate the formation of possible coordination bonds between the metal and atoms of the protein by removing the most likely leaving group from its first coordination sphere and using a pseudometal atom type. In Gold, our program of choice, a hydrogen-like function is located at the vacant coordination site with directions that respect the coordination rules of the metal (i.e., octahedral, square planar, etc.) and can interact with Lewis basis atoms. The resulting binding modes are further analyzed to identify additional residues that could reach the metal ion. On the basis of statistics of metaloprotein three-dimensional structures, any residue with the $C_\alpha$ under the cut-off of 9 Å from the metal could display one of its rotameric states coordinating the ion. Once those amino acids are identified, the final step along the process consists of generating the different coordination modes of a given docking solution by rotameric refinement and pursuing with QM/MM calculations of the resulting complex. QM/MM calculations are generated with an initial minimization constraining the coordination bond to a reasonable distance and subsequently releasing the constraint to avoid artifacts along the optimization. The potential energies of the final models are compared together, and those with the lowest energy are compared and discussed and could eventually be used for further designs.

We tested this approach for the first time in 2010 on the structure of an artificial metalloenzyme obtained by the substitution of the heme by a Fe(Schiff base) salophen in *Corynebacterium diphtheria* heme oxygenase (*cd*HO).[78] *cd*HO is a small all-$\alpha$ enzyme that performs the first step of the oxidation of the heme.[54] The Fe(Schiff base)−*cd*HO resulted as a superoxidase able to work thanks to successive reductions performed by the electron partner of the natural enzyme. Importantly, the crystal structure of Fe(Schiff base)-*cd*HO shows major differences from other salophen and heme-bound enzymes (Scheme 1 and Figure 4A).

First, the iron displays an octahedral configuration with a distorted cofactor and diverges from the planar geometry

**Scheme 1. General Geometries of X-ray Geometry of the Fe(III)Schiff Base·*cd*HO System (left) and Resting States of Naturally Occurring Hemoenzymes (right)**



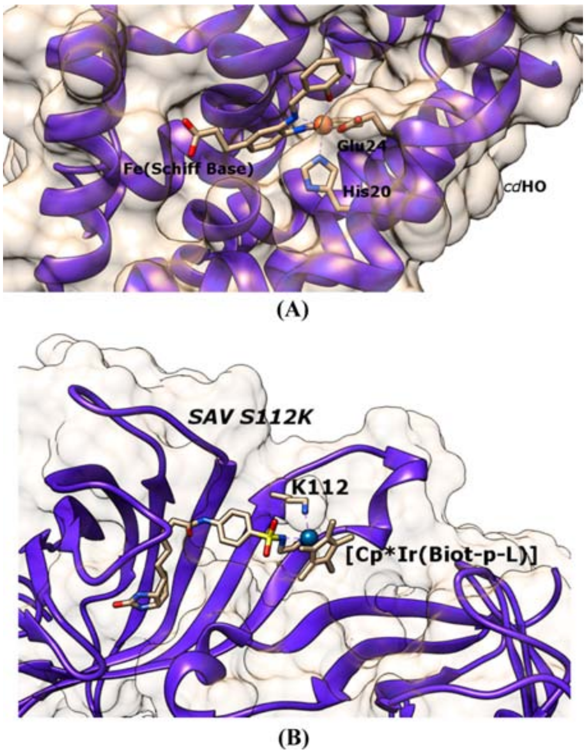


**Figure 4.** Examples of artificial metalloenzymes in which the binding of the cofactor occurs under an active coordination sphere: (A) predicted geometries of the Fe(III)(Schiff Base)-*cd*HO resting state and (B) structure of the binding site of the [Cp*Ir(Biot-p-L)Cl] ⊂ S112K Sav system with a coordinated lysine 112 to the iridium complex, as predicted by the docking procedure.

observed with salophen and porphyrin systems (Figure 4A). In this geometry, one of the oxygen atoms of the catalyst migrated from the equatorial position to the axial position. Second, two residues coordinate the iron, the His20 that occupies one of the axial positions, and the glutamate 24 that fulfills the fourth position of the equatorial plane. This geometry does not follow the general trend of heme-bound enzymes, including the natural substrate-bound heme oxygenases. In those systems, the iron coordinates the four atoms of the macrocycle in equatorial position and binds His20 as a unique protein ligand at the proximal axial site. Moreover, the resting state of hemoenzymes remains either square planar or octahedral by the binding of an additional labile ligand to the metal, generally a water molecule. Finally, the *cd*HO structure also presents a displacement of the helix A, to which both His20 and Glu24 are bound, that has been unreported before in X-ray structures of heme oxygenases. As a whole, the geometry presents all the molecular features that challenge computational prediction of the binding of organometallics to protein.

Our study consisted of a blank experiment in which we started with a set of structures of the heme oxygenase (none of them corresponding to the crystal structure of Fe(Schiff base)-*cd*HO) and the chemical structure of the cofactor. Applying the protocol described previously, we obtained two low-energy structures: one with excellent structural similarity to the crystal structure of the artificial metalloenzyme, the second with a square pyramidal geometry reminiscent of the heme enzymes and in which glutamate 24 was eliminated from the first coordination of the metal.[54]

The first structure clearly illustrates that the combination of protein−ligand dockings and QM/MM approaches would lead to excellent predictions for active binding of homogeneous catalysts to a protein host, even with changes of the coordination sphere of the metal. The second apparently suggests a failure of the simulation in discriminating between different binding modes but also that a possible equilibrium between hexacoordinated and pentacoordinated geometries exists in solution. The crystal structure snapshot could have somehow trapped an intermediate out of the catalytic path of the enzyme.

We further studied this aspect by investigating the transition between both structures considering all the spin and oxidation states conceivable in the initiation step of the catalysis by QM/MM calculations. We showed that the X-ray structure corresponds to the real resting state of the enzyme in Fe(III) state, and the square pyramidal one corresponds to the reduced Fe(II) form of the enzyme. Energy decomposition using different QM/MM partitions allowed identifying that the first coordination sphere of the metal is the most important factor in dictating the geometry of the final complex. In addition, this study clearly demonstrated that the transition between both structures is energetically feasible only when the reduction has occurred, hence providing additional evidence of the divergence in the mechanism of action of artificial and natural heme-like enzymes. The transition state vector also shows that the reorganization of the cofactor, the displacement of the glutamate out of the first coordination sphere of the metal and the entire rearrangement of the helix A, are intrinsically related.[79]

More recently, we applied the same procedure to artificial imine reductases designed by Ward and co-workers and resulting from the incorporation of a biotinylated Cp*Ir Noyori's-like catalyst (Cp* = $C_5Me_5-$) within different mutants of the homotetrameric streptavidin (Sav) (referred to as Cp*Ir(Biot-p-L)Cl] ⊂ Sav). Mutants at position S112 reveal major differences in both the Ir/streptavidin ratio and the enantioselectivity for the production of salsolidine. For [Cp*Ir(Biot-p-L)Cl] ⊂ S112A Sav, the reaction rate and the enantioselectivity (which reach up to 96% ee for (*R*)-salsolidine) decrease upon saturating all biotin binding sites, whereas for [Cp*Ir(Biot-p-L)Cl] ⊂ S112K Sav, the rate and the ee remain almost constant as a function of the ratio Ir/streptavidin (ee near 78% for (*S*)-salsolidine). Our docking complemented the X-ray structures that only partially resolve the location and the orientation of the cofactor into the cavity of the hosts. In collaboration with Ward's group, our calculations verify that the S112A and S112K Sav mutants prefer binding the $S_{Ir}$ and $R_{Ir}$ enantiomeric forms of the cofactor, respectively, a phenomenon not observed on natural enzymes binding organometallic cofactors. Moreover, it shows that the binding in the S112K mutant could be stabilized by a coordination of the metal with the $N_\varepsilon$ atom of lysine 112, forming a resting state structure stabilizing the orientation of the cofactor and differing from those of the S112A mutant in which no additional coordination is observed between the metal and the protein (Figure 4B).[80]

## 2. MODELING THE CATALYTIC ACTIVITY OF ARTIFICIAL METALLOENZYMES

The binding of the synthetic cofactor is part of the molecular events that condition the design of artificial metalozymes, but their activity can be understood only from the binding of the substrate and its activation. Protein−ligand dockings on its own can provide some relevant insights by allowing the identification of substrate binding modes that are catalytically consistent.[81−83]

The composite resulting from the insertion of Mn(III)-meso-tetrakis(*p*-carboxyphenyl)porphyrin (Mn(TpCPP)) into Xln10A displays epoxidative activity on a series of styrenes. Subtle enantioselective preferences toward *S* products are observed in most substrate but the most remarkable ee is observed with *p*-methoxystyrene, featuring a stereoselectivity of 80% in favor of the *R* isomer. The docking of the different substrates into the binding site of a model of the artificial metalloenzyme previously generated showed that the predicted orientations of the substrate in the active site of Mn(TpCPP)-Xln10A consistent with the formation of *S*-epoxide are slightly more stable than those for *R* epoxide ones. However, for *p*-methoxystyrene, the trend is inverted, with orientations consistent with *R* epoxide formation being more stable. Such inversion is associated with an additional H-bond between tyrosine 172 and the oxygen atom of the *p*-methoxy substituent (Figure 5).[73]

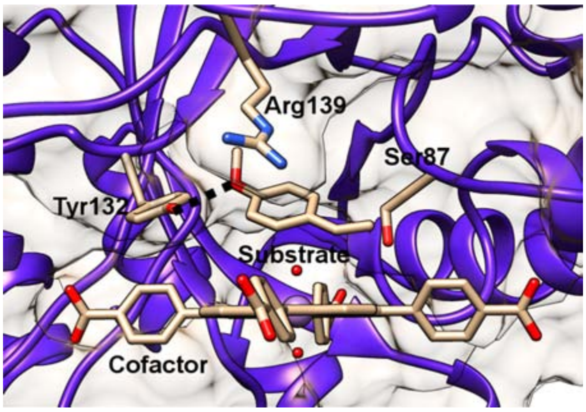


**Figure 5.** Predicted low-energy complexes of the *p*-methoxystyrene inside the cavity of the Mn(TpCPP))-Xyl10A model, corresponding to the orientation of the substrate consistent with the formation of the *R* epoxide product.

Protein−ligand dockings are not able to identify true transition state structures along an enzymatic reaction. They are limited to providing substrate binding modes that are to be contextualized in terms of prereactive orientations. The characterization of real, true transition state structures can be performed only by means of QM/MM calculations. However, for most artificial metalloenzymes, this task cannot be dissociated to a wide conformational sampling because little molecular information is available for the location and orientation of the cofactor and its interaction with the substrate.

We developed a protocol combining docking, QM, and QM/MM calculations in which both substrate and cofactor are taken into account during the geometrical search. This methodology, reminiscent of the work of Houk, Mayo, and Baker, consists of three successive steps: (1) Study of the catalytic mechanism in a cluster model of the enzyme by DFT calculations. The reduced model consists generally of the cofactor, the substrate, and amino acids likely to interact on the reaction center. (2) Docking of the geometries of the transition state structures obtained in step 1 into the binding site of the artificial metalloenzyme. Those pseudotransition states are generated while imposing few geometry variables extracted from the structure of the transition state models. (3) Refinement by QM/MM calculations of the pseudotransition state structures obtained in step 2 and identification of true transition state structures on the full potential energy surface. From this step, the lowest energy paths can be identified and compared with experiment. At each step along the process, the models with the substantially highest energies are neglected for the next step forward. A scheme of the protocol employed, showing the sequential steps and its application to an example, is depicted in Figure 6.

A first application of this methodology has been the study of the catalytic mechanism of the artificial [Cp*Ir(Biot-p-L)Cl] ⊂ S112A transfer hydrogenase mutant mentioned earlier.[8] Although the mechanism of reduction of ketones by Noyori's-like complexes is now widely accepted, the one leading to the reduction of imine has not yet reached a consensus. Although the metal center is well-known to provide the transfer of hydride, there is still discussion on the source of the proton. As a consequence, our model system considers several mechanistic hypotheses for this step, including the organometallic moiety itself, a hydronium from the medium, or a positively charged lysine that the active site could contain. Calculations were performed for processes leading to *R* and *S* chiral reduced imine.

The first step of our protocol allowed discarding mechanisms in which hydride and a proton are transferred from the homogeneous catalysts. All of those pathways are systematically 15 kcal $mol^{-1}$ higher than any other ones, a magnitude difficult to imagine that a protein scaffold could counterbalance. After docking the remaining pseudotransition state structures in the streptavidine vestibule, QM/MM refinements led to the identification of eight different reaction paths. Those involving proton transfer from the lysine residues located in the binding site are the less favored. The lowest-energy mechanism implies the transfer of the hydride on the substrate that was protonated in solution prior its access to the SAV112A site. The lowest-energy *R* and *S* paths clearly indicate preference toward the formation of the *R* product. The corresponding ee calculated on the difference in energy of the transition states reaches 80%, a magnitude in good agreement with the 98% reported experimentally. Interestingly, the geometries of the predicted transition state structures of *proR* and *proS* mechanisms clearly show a major drift of the cofactor and the substrate into the SAV cavity, something absent from natural hemoenzymes in which the location of the cofactor is well stabilized (Figure 7).

The relevance of a correct identification of the location of the cofactor is consistent with the first QM/MM study reported on an artificial metalloenzyme by Morokuma et al.[52] They analyzed the reaction mechanism of the polymerization of phenyl-

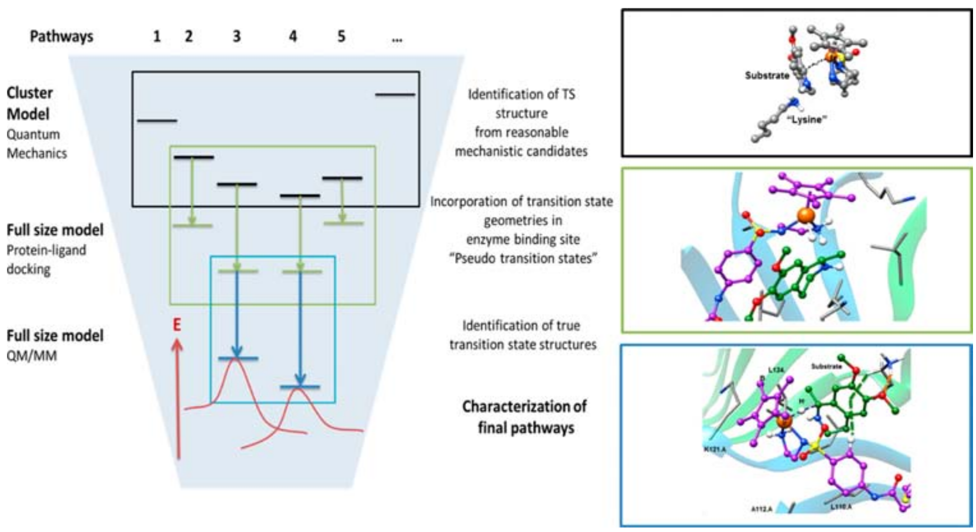


**Figure 6.** Schematic procedure for the identification of catalytic mechanism inside a proteic scaffold of artificial metalloenzymes considering the uncertainty of the location of the cofactor (left) and its application to mechanistic study of an artificial transfer hydrogenase (right).

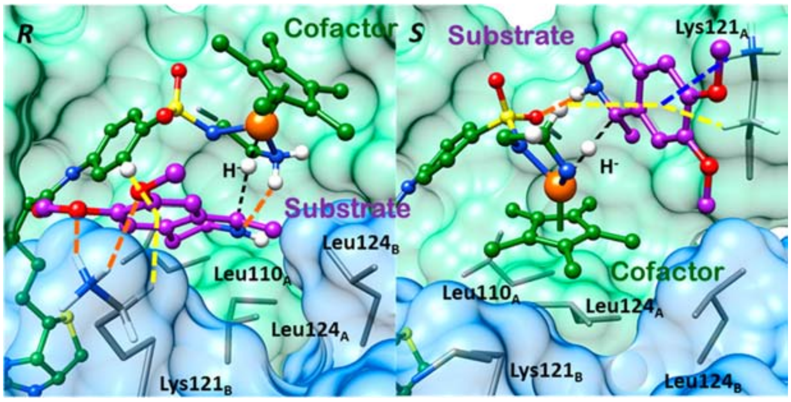


**Figure 7.** Calculated transition state structures of the lowest energy paths leading to the formation of *R* (left) and *S* (right) salsolidine.

acetylene by an artificial metalloenzyme developed by Ueno and co-workers. This biometallic hybrid was obtained through the insertion of a $[Rh(norbornadiene)Cl]_2$ catalyst inside a horse L-chain *apo*-ferritin. After investigating different reaction mechanisms, the authors were able to characterize the most likely cavity in the *apo*-ferritin structure that could shelter the inorganic cofactor and favor the polymerization reaction.

## ■ CONCLUSION AND PERSPECTIVE

The development of artificial metalloenzymes is becoming a major field of investigation. To date, designers have based most of their work on (bio)chemical intuitions, with limited structural information. Molecular modeling can be useful in this field, although dealing with all the energetic and structural aspects that need to be considered represents a real tour de force. Among them, metal-mediated recognition processes involved in cofactor binding are fundamental but out of the scope of current state-of-the-art protein−ligand docking methodologies. Another fundamental aspect is to determine three-dimensional models of catalytically active conformations of the cofactor−substrate−receptor triad.

A few years ago and based on our experience on organometallics, drug design, and bioinorganics, we embarked on establishing and benchmarking procedures convenient for artificial metalloenzyme design. The fruits of the first steps in this venture are summarized in this manuscript. Here, we try to show which approximation we had to contemplate, the evolution of our approaches, the most relevant elements of our achievements, and what accuracy is to be expected.

Regarding the binding of the homogeneous catalyst to a protein, we validated standard protein−ligand docking procedures in generating accurate 3D models if no chemical changes of the first coordination sphere of the metal occur upon binding. Although scoring functions and parameters to deal specifically with metal ions in ligands still leave room for improvement, calculations performed under this hypothesis are extremely instructive. Such dockings are particularly relevant for composites that represent the first line of candidates for artificial metalloenzymes. This benchmark also suggests that high-throughput virtual screening of large databases of organometallics and proteins to detect novel frameworks is achievable.

For the prediction of 3D models involving the modification of the first coordination of the metal upon binding, we thought that the best chances of success were with allying protein–ligand docking with QM and QM/MM approaches. This strategy allows reproducing complex processes and gives highlights on major induced effects, such as the distortion of the cofactor and large scale motions of the receptor, and eventually complement crystallographic observations. However, its success depends on hypothesizing which coordinated ligand(s) leave(s) the cofactor and is (are) replaced by protein residues. This is relatively straightforward for labile water molecules in heme-like systems but still challenging for organometallic compounds for which interaction with biological scaffolds are less documented. Generating multiple coordination sphere candidates during the docking run is therefore primordial for more advanced designs and a methodological challenge we are now exploring.

Finally, we showed that bridging pure quantum mechanics calculations on model systems, protein–ligand dockings, and QM/MM calculations allow identification of true transition state structures in artificial metalloenzymes. This approach is efficient enough to characterize reaction paths even when the location of the cofactor is uncertain. It also provides information on fine structural events, such as those that control the enantioselective profile of artificial metalloenzymes. Importantly, this procedure is computationally far less demanding than other deterministic protocols, such as those performed with stirred molecular dynamics or metadynamics.

Although most of the tools for computer-aided design of artificial metalloenzymes are now part of the toolbox of computational chemists, their success in this field will also be dependent on our ability to deal with the fine-tuning between simulation of binding processes and catalytic mechanisms. In particular, better sampling protein–ligand docking techniques are of the most important aspects to incorporate into an integrative framework. As such, methods than allow fast introduction of large-scale (collective) motions[84] and enhanced sampling in docking will be a major asset.[85]

With few years dedicated to this field, we believe that our experience illustrates the potential of molecular modeling tools for the rationalization of the reactivity of existing artificial metalloenzymes. Decoding their molecular mechanism at the atomic level first provides useful information for further optimization steps (i.e., control over regio- and enantioselectivities and specificities) and also affords conceptual knowledge on nonnatural bioinorganic interactions. With the lack of molecular information on these composites, we hope this could serve in the development of the entire field of artificial metalloenzymes.

The challenge in the years ahead consists of expanding our modeling framework so that computation could become an interesting tool for design purposes. To reach such in silico designs, focus should be given, among others, on strategies that allow the identification of protein scaffolds that could host the artificial cofactors and satisfy the chemical requirements for the reactivity to occur as well as predict suitable redesign of the protein–substrate–cofactor interface. To this end, combinations and adaptations of the approaches described in this manuscript with those already established in enzyme design are among the most interesting. Still, the main barrier to overcome consists of the simultaneous exploration of both the biological (i.e., mutations of the protein scaffold) and chemical (i.e., nature of the cofactor and its substituents) spaces so that calculations could ascertain the most promising complementarity of the protein–substrate–cofactor triad. In this, there is no doubt that the integration of different methodologies will be crucial to success.

## ■ AUTHOR INFORMATION

### Corresponding Author

*E-mail: JeanDidier.Marechal@uab.cat.

### Notes

The authors declare no competing financial interest.

## ■ ACKNOWLEDGMENTS

Financial support from the Spanish Ministerio de Economia y Competitividad (Project CTQ2014-54071-P) is acknowledged. V.M.R. is grateful to the Spanish MINECO for a FPI Fellowship. J.D.M. also deeply thanks Profs. Mahy and Ward and Dr. Ricoux for their introduction to the world of artificial metalloenzymes.

## ■ REFERENCES

(1) Golynskiy, M. V.; Seelig, B. *Trends Biotechnol.* **2010**, *28*, 340–345.

(2) Okrasa, K.; Kazlauskas, R. J. *Chem.—Eur. J.* **2006**, *12*, 1587–1596.

(3) Lin, Y.-W.; Yeung, N.; Gao, Y.-G.; Miner, K. D.; Lei, L.; Robinson, H.; Lu, Y. *J. Am. Chem. Soc.* **2010**, *132*, 9970–9972.

(4) Jing, Q.; Okrasa, K.; Kazlauskas, R. J. *Chem.—Eur. J.* **2009**, *15*, 1370–1376.

(5) Podtetenieff, J.; Taglieber, A.; Bill, E.; Reijerse, E. J.; Reetz, M. T. *Angew. Chem., Int. Ed.* **2010**, *122*, 5277–5281.

(6) Köhler, V.; Wilson, Y. M.; Lo, C.; Sardo, A.; Ward, T. R. *Curr. Opin. Biotechnol.* **2010**, *21*, 744–752.

(7) Rosati, F.; Roelfes, G. *ChemCatChem* **2010**, *2*, 916–927.

(8) Dürrenberger, M.; Heinisch, T.; Wilson, Y. M.; Rossel, T.; Nogueira, E.; Knörr, L.; Mutschler, A.; Kersten, K.; Zimbron, M. J.; Pierron, J.; Schirmer, T.; Ward, T. R. *Angew. Chem., Int. Ed.* **2011**, *123*, 3082–3085.

(9) Pordea, A.; Creus, M.; Panek, J.; Duboc, C.; Mathis, D.; Novic, M.; Ward, T. R. *J. Am. Chem. Soc.* **2008**, *130*, 8085–8088.

(10) Heinisch, T.; Ward, T. R. *Curr. Opin. Chem. Biol.* **2010**, *14*, 184–199.

(11) Williams, R. J. P. *Chem. Commun.* **2003**, 1109–1113.

(12) Ringenberg, M. R.; Ward, T. R. *Chem. Commun.* **2011**, *47*, 8470–8476.

(13) Esmieu, C.; Cherrier, M. V.; Amara, P.; Girgenti, E.; Marchi-Delapierre, C.; Oddon, F.; Iannello, M.; Jorge-Robin, A.; Cavazza, C.; Ménage, S. *Angew. Chem., Int. Ed.* **2013**, *52*, 3922–3925.

(14) Bos, J.; García-Herraiz, A.; Roelfes, G. *Chem. Sci.* **2013**, *4*, 3578–3582.

(15) Bos, J.; Fusetti, F.; Driessen, A. J. M.; Roelfes, G. *Angew. Chem., Int. Ed.* **2012**, *51*, 7472–7475.

(16) Chevalley, A.; Salmain, M. *Chem. Commun.* **2012**, *48*, 11984–11986.

(17) Inaba, H.; Kanamaru, S.; Arisaka, F.; Kitagawa, S.; Ueno, T. *Dalton Trans.* **2012**, *41*, 11424–11427.

(18) Köhler, V.; Mao, J.; Heinisch, T.; Pordea, A.; Sardo, A.; Wilson, Y. M.; Knörr, L.; Creus, M.; Prost, J.-C.; Schirmer, T.; Ward, T. R. *Angew. Chem., Int. Ed.* **2011**, *50*, 10863–10866.

(19) Boersma, A. J.; Coquière, D.; Geerdink, D.; Rosati, F.; Feringa, B. L.; Roelfes, G. *Nat. Chem.* **2010**, *2*, 991–995.

(20) Roelfes, G.; Feringa, B. L. *Angew. Chem., Int. Ed.* **2005**, *44*, 3230–3232.

(21) Gresh, N.; Cisneros, G. A.; Darden, T. A.; Piquemal, J.-P. *J. Chem. Theor* **2007**, *3*, 1960–1986.

(22) Deeth, R. J.; Anastasi, A.; Diedrich, C.; Randell, K. *Coord. Chem. Rev.* **2009**, *253*, 795–816.

(23) Harvey, J. N. *Struct. Bond.* **2004**, *112*, 151–184.

(24) Swart, M.; Groenhof, R.; Ehlers, A. W.; Lammertsma, K. *J. Phys. Chem. A* **2004**, *108*, 5479−5483.

(25) Zhao, Y.; Truhlar, D. G. *Acc. Chem. Res.* **2008**, *41*, 157−167.

(26) Senn, H. M.; Thiel, W. *Angew. Chem., Int. Ed.* **2009**, *48*, 1198−1229.

(27) Bathelt, C. M.; Mulholland, A. J.; Harvey, J. N. *J. Phys. Chem. A* **2008**, *112*, 13149−13156.

(28) Oláh, J.; Mulholland, A. J.; Harvey, J. N. *Proc. Natl. Acad. Sci. U.S.A.* **2011**, *108*, 6050−6055.

(29) Piazzetta, P.; Marino, T.; Russo, N. *Inorg. Chem.* **2014**, *53*, 3488−3493.

(30) Lin, H.; Truhlar, D. G. *Theor. Chem. Acc.* **2007**, *117*, 185−199.

(31) Bathelt, C. M.; Zurek, J.; Mulholland, A. J.; Harvey, J. N. *J. Am. Chem. Soc.* **2005**, *127*, 12900−12908.

(32) Martí, S.; Andrés, J.; Silla, E.; Moliner, V.; Tuñón, I.; Bertrán, J. *Angew. Chem., Int. Ed.* **2007**, *46*, 286−290.

(33) Lonsdale, R.; Rouse, S. L.; Sansom, M. S. P.; Mulholland, A. J. *PLOS Comput. Biol.* **2014**, *10*, e1003714.

(34) Shi, R.; Li, W.; Liu, G.; Tang, Y. *Chin. J. Chem.* **2013**, *31*, 1219−1227.

(35) Miki, Y.; Pogni, R.; Acebes, S.; Lucas, F.; Fernández-Fueyo, E.; Baratto, M. C.; Fernández, M. I.; de los Ríos, V.; Ruiz-Dueñas, F. J.; Sinicropi, A.; Basosi, R.; Hammel, K. E.; Guallar, V.; Martínez, A. T. *Biochem. J.* **2013**, *452*, 575−584.

(36) Piazzetta, P.; Marino, T.; Russo, N. *Phys. Chem. Chem. Phys.* **2014**, *16*, 16671−16676.

(37) Kiss, G.; Çelebi-Ölçüm, N.; Moretti, R.; Baker, D.; Houk, K. N. *Angew. Chem., Int. Ed.* **2013**, *52*, 5700−5725.

(38) Koder, R. L.; Anderson, J. L. R.; Solomon, L.; Reddy, K. S.; Moser, C. C.; Dutton, P. L. *Nature* **2009**, *458*, 305−9.

(39) Dieckmann, G. R.; McRorie, D. K.; Tierney, D. L.; Utschig, L. M.; Singer, C. P.; O'Halloran, T. V.; Penner-Hahn, J. E.; DeGrado, W. F.; Pecoraro, V. L. *J. Am. Chem. Soc.* **1997**, *119*, 6195−6196.

(40) Ghosh, D.; Pecoraro, V. L. *Curr. Opin. Chem. Biol.* **2005**, *9*, 97−103.

(41) Tantillo, D. J.; Houk, K. N. *Curr. Opin. Chem. Biol.* **1998**, *2*, 743−750.

(42) Dahiyat, B. I.; Mayo, S. L. *Science* **1997**, *278*, 82−87.

(43) Bolon, D. N.; Mayo, S. L. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 14274−14279.

(44) Lassila, J. K.; Privett, H. K.; Allen, B. D.; Mayo, S. L. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 16710−16715.

(45) Privett, H. K.; Kiss, G.; Lee, T. M.; Blomberg, R.; Chica, R.; Thomas, L. M.; Hilvert, D.; Houk, K. N.; Mayo, S. L. *Proc. Natl. Acad. Sci. U.S.A.* **2012**, *109*, 3790−3795.

(46) Röthlisberger, D.; Khersonsky, O.; Wollacott, A. M.; Jiang, L.; DeChancie, J.; Betker, J.; Gallaher, J. L.; Althoff, E. a; Zanghellini, A.; Dym, O.; Albeck, S.; Houk, K. N.; Tawfik, D. S.; Baker, D. *Nature* **2008**, *453*, 190−195.

(47) Siegel, J. B.; Zanghellini, A.; Lovick, H. M.; Kiss, G.; Lambert, A. R.; St. Clair, J. L.; Gallaher, J. L.; Hilvert, D.; Gelb, M. H.; Stoddard, B. L.; Houk, K. N.; Michael, F. E.; Baker, D. *Science* **2010**, *329*, 309−313.

(48) Zanghellini, A.; Jiang, L. I. N.; Wollacott, A. M.; Cheng, G.; Meiler, J.; Althoff, E. A.; Ro, D. *Protein Sci.* **2006**, *15*, 2785−2794.

(49) Kuhlman, B.; Dantas, G.; Ireton, G. C.; Varani, G.; Stoddard, B. L.; Baker, D. *Science* **2003**, *302*, 1364−1368.

(50) Meiler, J.; Baker, D. *Proteins* **2006**, *65*, 538−548.

(51) Lind, M. E. S.; Himo, F. *Angew. Chem., Int. Ed.* **2013**, *125*, 4661−4665.

(52) Ke, Z.; Abe, S.; Ueno, T.; Morokuma, K. *J. Am. Chem. Soc.* **2012**, *134*, 15418−15429.

(53) Frushicheva, M. P.; Warshel, A. *ChemBioChem* **2012**, *13*, 215−223.

(54) Robles, V. M.; Ortega-Carrasco, E.; Fuentes, E. G.; Lledós, A.; Maréchal, J.-D. *Faraday Discuss.* **2011**, *148*, 137−159.

(55) Muñoz Robles, V.; Vidossich, P.; Lledós, A.; Ward, T. R.; Maréchal, J.-D. *ACS Catal.* **2014**, *4*, 833−842.

(56) Allard, M.; Dupont, C.; Muñoz Robles, V.; Doucet, N.; Lledós, A.; Maréchal, J.-D.; Urvoas, A.; Mahy, J.-P.; Ricoux, R. *ChemBioChem* **2012**, *13*, 240−251.

(57) Verdonk, M. L.; Cole, J. C.; Hartshorn, M. J.; Murray, C. W.; Taylor, R. D. *Proteins* **2003**, *52*, 609−623.

(58) Kirton, S. B.; Murray, C. W.; Verdonk, M. L.; Taylor, R. D. *Proteins* **2005**, *58*, 836−844.

(59) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, J. A.; Vreven, T., Jr.; Kudin, K. N.; Burant, J. C.; Millam, J. M.; Iyengar, S. S.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; Petersson, G. A.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Klene, M.; Li, X.; Knox, J. E.; Hratchian, H. P.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J.; Ayala, P. Y.; Morokuma, K.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Zakrzewski, V. G.; Dapprich, S.; Daniels, A. D.; Strain, M. C.; Farkas, O.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Ortiz, J. V.; Cui, Q.; Baboul, A. G.; Clifford, S.; Cioslowski, J.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B. G.; Chen, W.; Wong, M. W.; C, G.; Pople, J. A. *Gaussian 09*; Revision D.01, Gaussian, Inc.: Wallingford, CT, 2009.

(60) Dapprich, S.; Komaromi, I.; Byun, S. K.; Morokuma, K. S.; Frisch, M. J. *J. Mol. Struct. THEOCHEM* **1999**, *461−462*, 1−21.

(61) 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.

(62) Chase, F.; Avenue, B. *Curr. Opin. Struct. Biol.* **2002**, *12*, 431−440.

(63) Kelley, L. A.; Gardner, S. P.; Sutcliffe, M. J. *Protein Eng.* **1996**, *9*, 1063−1065.

(64) Trott, O.; Olson, A. J. *J. Comput. Chem.* **2009**, *31*, 455−461.

(65) 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.; Shaw, D. E.; Francis, P.; Shenkin, P. S. *J. Med. Chem.* **2004**, *47*, 1739−1749.

(66) Seebeck, B.; Reulecke, I.; Kämper, A.; Rarey, M. *Proteins* **2008**, *71*, 1237−54.

(67) Meggers, E. *Chem. Commun.* **2009**, 1001−1010.

(68) Bernstein, F. C.; Koetzle, T. F.; Williams, G. J.; Meyer, E. E., Jr.; Brice, M. D.; Rodgers, J. R.; Kennard, O.; Shimanouchi, T.; Tasumi, M. *J. Mol. Biol.* **1977**, *112*, 535.

(69) Ortega-Carrasco, E.; Lledós, A.; Maréchal, J.-D. *J. Comput. Chem.* **2014**, *35*, 192−198.

(70) Mahy, J.-P.; Maréchal, J.-D.; Ricoux, R. *Chem. Commun.* **2015**, *51*, 2476−2494.

(71) Ricoux, R.; Dubuc, R.; Dupont, C.; Maréchal, J.-D.; Martin, A.; Sellier, M. *Bioconjugate Chem.* **2008**, *19*, 899−910.

(72) Ricoux, R.; Allard, M.; Dubuc, R.; Dupont, C.; Maréchal, J.-D.; Mahy, J. P. *Org. Biomol. Chem.* **2009**, *7*, 3208−3211.

(73) Allard, M.; Dupont, C.; Muñoz Robles, V.; Doucet, N.; Lledós, A.; Maréchal, J.-D.; Urvoas, A.; Mahy, J.-P.; Ricoux, R. *ChemBioChem* **2012**, *13*, 240−251.

(74) Drevelle, A.; Graille, M.; Heyd, B.; Sorel, I.; Ulryck, N.; Pecorari, F.; Desmadril, M.; van Tilbeurgh, H.; Minard, P. *J. Mol. Biol.* **2006**, *358*, 455−471.

(75) Sansiaume-Dagousset, E.; Urvoas, A.; Chelly, K.; Ghattas, W.; Maréchal, J.-D.; Mahy, J.-P.; Ricoux, R. *Dalt. Trans.* **2014**, *43*, 8344−8354.

(76) Urvoas, A.; Ghattas, W.; Maréchal, J.-D.; Avenier, F.; Bellande, F.; Mao, W.; Ricoux, R.; Mahy, J.-P. *Bioorgan. Med. Chem.* **2014**, *22*, 5678−5686.

(77) Muñoz Robles, V.; Maréchal, J.-D.; Bahloul, A.; Sari, M.-A.; Mahy, J.-P.; Golinelli-Pimpaneau, B. *PLoS One* **2012**, *7*, e51128.

(78) Ueno, T.; Yokoi, N.; Unno, M.; Matsui, T.; Tokita, Y.; Yamada, M.; Ikeda-Saito, M.; Nakajima, H.; Watanabe, Y. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 9416−9421.

(79) Ortega-Carrasco, E.; Lledós, A.; Maréchal, J. *J. R. Soc. Interface* **2014**, *11*, 20140090.

(80) Muñoz Robles, V.; Dürrenberger, M.; Heinisch, T.; Lledós, A.; Schirmeister, T.; Ward, T. R.; Maréchal, J.-D. *J. Am. Chem. Soc.* **2014**, *136*, 15676−15683.

(81) Toledo, L.; Masgrau, L.; Maréchal, J.-D.; Lluch, J. M.; González-Lafont, A. *J. Phys. Chem. B* **2010**, *114*, 7037−7046.

(82) Maréchal, J.; Yu, J.; Brown, S.; Kapelioukh, I.; Rankin, E. M.; Wolf, C. R.; Roberts, G. C. K.; Paine, M. J. I.; Sutcliffe, M. J. *Drug Metab. Dispos.* **2006**, *34*, 534−538.

(83) McLaughlin, L.; Paine, M. J. I.; Kemp, C.; Maréchal, J.-D.; Flanagan, J. U.; Ward, C. J.; Sutcliffe, M. J.; Roberts, G. C. K.; Wolf, C. R. *J. Biol. Chem.* **2005**, *280*, 38617−38624.

(84) Borrelli, K. W.; Vitalis, A.; Alcantara, R.; Guallar, V. *J. Chem. Theory Comput.* **2005**, *1*, 1304−1311.

(85) Amaro, R. E.; Baron, R.; McCammon, J. A. *J. Comput. Aided Mol. Des.* **2008**, *22*, 693−705.