Available online at www.sciencedirect.com
SCIENCE DIRECT
Chemical Physics 323 (2006) 129–137
Chemical Physics
www.elsevier.com/locate/chemphys

A first principles study of the binding of formic acid in catalase complementing high resolution X-ray structures

Carme Rovira [a,*], Mercedes Alfonso-Prieto [a], Xevi Biarnés [a], Xavi Carpena [b], Ignacio Fita [b], Peter C. Loewen [c]

[a] Centre especial de Recerca en Química Teòrica, Parc Científic de Barcelona, Josep Samitier 1-5, 08028 Barcelona, Spain
[b] Consejo Superior de Investigaciones Científicas y Parc Científic de Barcelona (CSIC-PCB), Josep Samitier 1-5, 08028 Barcelona, Spain
[c] Department of Microbiology, University of Manitoba, Winnipeg, Canada MB R3T 2N2

Received 12 May 2005; accepted 16 August 2005
Available online 25 October 2005

Abstract

Density functional molecular dynamics simulations using a QM/MM approach are used to get insight into the binding modes of formic acid in catalase. Two ligand binding sites are found, named **A** and **B**, in agreement with recent high resolution structures of catalase with bound formic acid. In addition, the calculations show that the His56 residue is protonated and the ligand is present as a formate anion. The lowest energy minimum structure (**A**) corresponds to the ligand interacting with both the heme iron and the catalytic residues (His56 and Asn129). The second minimum energy structure (**B**) corresponds to the situation in which the ligand interacts solely with the catalytic residues. A mechanism for the process of formic acid binding in catalase is suggested.

1. Introduction

Heme-based catalases are present in almost all aerobically respiring organism [1]. They play critical roles in protecting the cell against the toxic effects of hydrogen peroxide (H_2O_2) by degrading it to water and oxygen (reaction 1)

$$2H_2O_2 \rightarrow 2H_2O + O_2 \quad (1)$$

The ubiquity of the enzyme and the easy availability of the substrates (H_2O_2 and alkylperoxides) have made heme-catalases the focus of many biochemical and molecular biology studies over the years [2]. It is generally accepted that the "catalatic" reaction occurs in two main steps. In the first step the Fe(III)–heme reacts with H_2O_2 to form an oxoferryl intermediate (named compound I, Cpd I).

$$\text{Enz (Heme–Fe}^{\text{III}}\text{)} + H_2O_2 \rightarrow \text{Cpd I (Heme}^{+\cdot}\text{–Fe}^{\text{IV}}\text{=O)} + H_2O \quad (2)$$

In the second step, Cpd I reacts with another molecule of H_2O_2 and the enzyme returns to the resting state.

$$\text{Cpd I (Heme}^{+\cdot}\text{–Fe}^{\text{IV}}\text{=O)} + H_2O_2 \rightarrow \text{Enz (Heme–Fe}^{\text{III}}\text{)} + H_2O + O_2 \quad (3)$$

At low H_2O_2 concentrations and in the presence of hydrogen donors, compound I may undergo a one-electron reduction towards the so-called Cpd II intermediate, which can subsequently be converted to another inactive form named compound III.

$$\text{Cpd I (Heme}^{+\cdot}\text{–Fe}^{\text{IV}}\text{=O)} + \text{HA} \rightarrow \text{Cpd II (Heme –Fe}^{\text{IV}}\text{–OH)} + \text{A}^{\cdot} \quad (4)$$

$$\text{Cpd II (Heme–Fe}^{\text{IV}}\text{–OH}^{+}\text{)} + H_2O_2 \rightarrow \text{Cpd III (Heme–Fe}^{\text{II}}\text{–OOH}^{+}\text{)} + H_2O \quad (5)$$

Several inhibitors of the reaction are known, including azide, cyanide, acetate, aminotriazole and formate [3]. The binding of these molecules impedes the binding of hydrogen peroxide in the active center, blocking the enzymatic reaction. Formic acid is a particularly interesting

* Corresponding author. Tel.: +34 93 4037112; fax: +34 93 4037225.
E-mail address: crovira@pcb.ub.es (C. Rovira).

0301-0104/$ - see front matter © 2005 Elsevier B.V. All rights reserved.
doi:10.1016/j.chemphys.2005.08.049

inhibitor as it displays a complex reactivity. In this respect, it can react in at least three ways: as a ligand with the heme iron (i.e., blocking the enzymatic reaction), as an electron donor for Cpd I reduction (HA in reaction 4) or as a ligand in a reaction with Cpd II to form an unstable complex [1,4].

The binding mode of formic acid in heme proteins has recently been elucidated, as the structures of the catalase complex with formic acid were reported for two different species: *Heliobacter pylori* Catalase (HPC) at 1.6 A resolution [5] and *Proteus mirabillis* Catalase (PMC) at 2.3 Å resolution [6]. Both structures show a ligand molecule inside the heme pocket and close to the heme iron (at 2.66 Å for HPC and 2.8 Å distance for PMC) as well as close to the His56 and Asn129 catalytic residues. In addition, a recent study of the complex of Horse Radish Peroxidase (HRP) with formic acid shows a similar ligand orientation [7], with a Fe···O distance of 2.3 Å. This type of configuration, hereafter referred as **A**, is shown in Fig. 1A. One of the formate oxygen atoms (O_1) is close to the heme iron, while the other one (O_2) is close to both the N_ε of His56 and the NH_2 unit of Asn129.

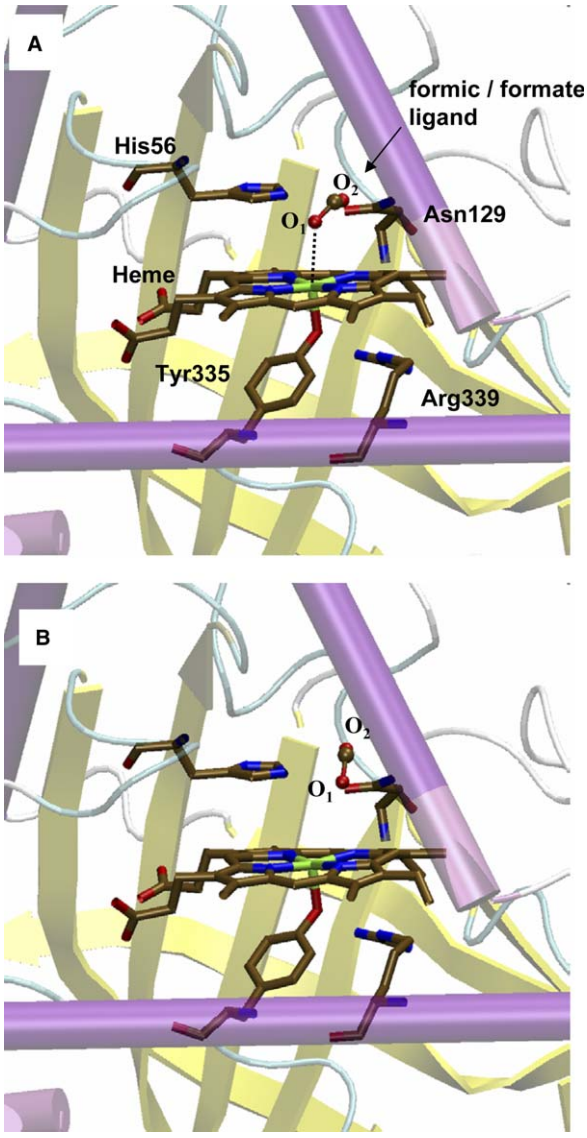

Fig. 1. Ligand binding sites observed in the X-ray structure of HPC with bound formate (PDB entry 1qwm).

In addition, the structure of HPC shows a second ligand binding site in which the ligand sits between the catalytic residues (**B** in Fig. 1). In this configuration, the O_2 atom of the ligand points towards the entry channel and the O_1 atom is between the catalytic residues. However, as the positions of hydrogen atoms are not available from the X-ray structure, it is not known whether formic acid is present as a neutral or anionic species. Taking into account the pH at which the experiments were done (5.6 for HPC and 6.5 for PMC) and the pKa for formic acid (3.75), it is expected that the anionic or formate form is favoured in solution. However, even though the ligand might reach the heme cavity as formate, its binding is accompanied by proton uptake [8], giving rise to the ambiguity. The asymmetry of the two C–O ligand distances could also help to elucidate the protonation form. If the ligand is present as formic acid, one C–O should be much larger than the other, while if it is present as formate, the asymmetry should be less pronounced. However, the crystal structure does not reflect such asymmetry, as both C–O distances have practically the same length (1.23 and 1.24 Å).

Theoretical studies could be very useful to decipher these issues and to determine the precise internal structure of the ligand. Among all quantum chemistry methods, Density Functional Theory (DFT) provides a good compromise between accuracy and computational requirements [9]. DFT has been applied with success in the study of heme-based systems relevant for the chemistry of catalases, peroxidases and other heme proteins [10]. In our previous work, we performed DFT-based molecular dynamics (Car-Parrinello) simulations in active site models [5]. We could show that the formic acid ligand only binds to the heme after having transferred its proton to the His56 residue. This provided an explanation for the binding mode **A**. The binding mode **B** was rationalized as the formic acid molecule interacting with the catalytic residues and sharing its proton with the His56 residue. Nevertheless, it was not possible to discern whether one or both oxygen atoms are involved in these interactions. In addition, the orientation of the ligand with respect to the catalytic residues turned out to be qualitatively different from the X-ray structure.

In this paper, we extend our previous analysis by including the protein environment by means of the QM/MM approach. We analyze the binding of formate/formic acid in the presence or not of the catalytic residues His56 and Asn129 and demonstrate that inclusion of the complete protein is necessary in order to describe the ligand orientation in configuration **B**.

2. Methods

2.1. Models

The calculations were performed using three different models, two small models taking into account the residues that directly interact with the ligand, and one large model including the protein environment within a QM/MM approach.

In order to describe the intrinsic properties of a formic acid/formate ligand bound to the heme, the first model was built of an iron–porphyrin and the ligand. Two protonation forms, formic acid and formate, were considered. The Tyr335 protein residue, which is directly bonded to the iron–porphyrin, and its hydrogen-bonded Arg339 residue were also included in the model (the tyrosine was replaced by a phenolate anion and the arginine was replaced by a methylguanidinium cation). Since it was previously found that this residue modulates the binding energy of the Fe–Tyr bond and governs the relative orientation of the Tyr residue with respect to the heme [11]. The system with formic acid was computed as a cation (due to the positive charge of Arg339), while the one with formate was computed as neutral. In both cases the total spin of the system was taken as $S = 5/2$ (high spin, the experimental ground state for the 5-coordinated heme in most catalases).

In order to describe the properties of the ligand interacting only with the catalytic residues, a second model was built that included the catalytic residues (His56 and Asn129, replaced by methylimidazole and acetamide, respectively) and the formic acid/formate ligand. In addition, the residues which are hydrogen bonded to His56 (Ser95 and Thr96) were also included in order to describe a putative proton transfer between the ligand and His56. The starting position of the formic/formate ligands was taken from the X-ray structure (position **B** in Fig. 1).

The third model used in our calculations consists of a large fragment of the protein surrounding the heme. All residues at less than 20 Å distance from the heme iron in the native HPC were included (4605 atoms). The formic acid ligand was inserted into the binding pocket, in a position corresponding with structures **A** or **B**, generating two starting structures. The heme, ligand, Tyr335, Arg339, His56, Asn129, Ser95 residues were treated quantum-mechanically, while the rest of the system was treated with the AMBER force field [12]. Only the quantum region was allowed to relax and temperature effects were not taken into account. The details of the QM/MM implementation are given in the next section.

2.2. Computational details

The calculations were performed using the Car-Parrinello molecular dynamics method [13], which is based on Density Functional Theory (DFT) [9]. Previous work has demonstrated the reliability of this method in the description of structural, energetic and dynamical properties of systems of biological interest [14]. The Kohn–Sham orbitals are expanded in a plane wave basis set with the kinetic energy cutoff of 70 Ry. We employed ab initio pseudopotentials, generated within the Troullier–Martins scheme [15], including the non-linear core-correction for the iron atom [16]. Our calculations were made using the generalized gradient-corrected approximation of the spin-dependent density functional theory (DFT-LSD), following the prescription of Becke and Perdew [17]. Structural optimizations were performed by means of molecular dynamics with annealing of the atomic velocities, using a time step of 0.12 fs and the fictitious mass of the electrons was set at 700 a.u. The systems are enclosed in an isolated supercell of sizes $18 \times 17 \times 12$ Å^3 (small models) and $15 \times 22 \times 20$ Å^3 (large models). The calculations for small models were performed using the CPMD program, written by Hutter et al. [18]. Structure analysis was performed with VMD [19].

The QM/MM calculations were performed using the approach developed by Laio, VandeVondele and Röthlisberger [20] which combines the first principles molecular dynamics method of Car and Parrinello (CPMD) [13] with a force-field molecular dynamics methodology. In this approach, the system is partitioned into a QM fragment and a MM fragment. The dynamics of the atoms on the QM fragment depends on the electronic density, $\rho(r)$, computed with DFT, while that of the atoms on the MM fragment is ruled by an empirical force field. The electrostatic interactions between the QM and MM regions are handled via a fully Hamiltonian coupling scheme [20] where the short range electrostatic interactions between the QM and the MM regions were explicitly taken into account for all atoms. An appropriately modified Coulomb potential was used to ensure that no unphysical escape of the electronic density from the QM to the MM region occurs. The electrostatic interactions with the more distant MM atoms are treated with a multipole expansion for the QM region. Bonded and van der Waals interactions between the QM and the MM part are treated with the standard AMBER force field [12]. Long-range electrostatic interactions between MM atoms have been described with P3M implementation [21]. The mesh used for P3M was $64 \times 64 \times 64$.

The QM/MM partition used is depicted in Fig. 2. All the protein residues up to 20 Å from the iron atom were considered in the calculation. This includes residues from the four protein subunits, as they are strongly interconnected in catalase. The following residues were taken quantum mechanically: the heme group (except the propionates), the phenolate group of Tyr335, the methylguanidinium fragment of Arg339, the imidazole of His56, the –CH_2–OH side chain of Ser95, the water molecule interacting with the OH of Ser95, the acetamide side chain of Asn129 and the formic acid ligand. The total size of the QM region is 123 atoms, over a total of 4605 atoms. The size of the QM region is large enough to capture the chemistry of the bound ligand, as all residues directly interacting with the ligand are included in the QM region. The rest of the

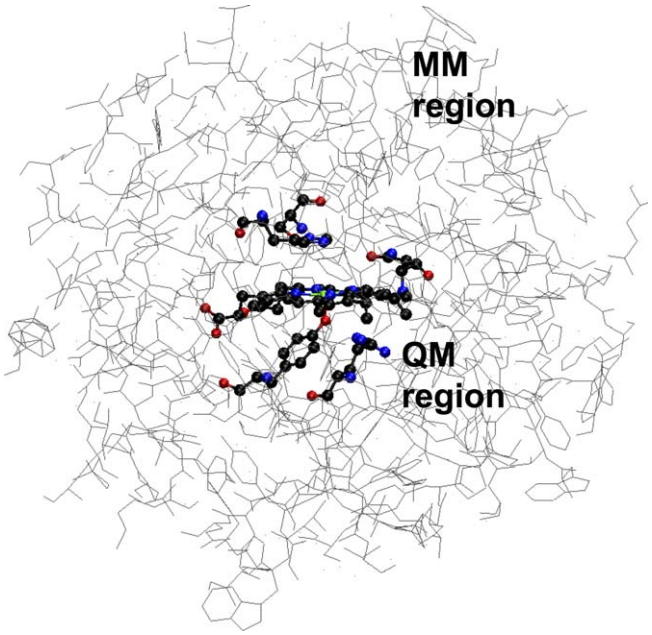

Fig. 2. QM–MM partition used in the CPMD QM/MM calculations. The atoms of the QM region are shown in ball and stick, while the protein environment considered is shown in solid lines.

system was treated with molecular mechanics using the AMBER force field [12].

3. Results and discussion

3.1. Ligand binding to the heme

In order to elucidate which protonation form of the ligand better binds to the heme, two calculations were performed starting with either formic acid or formate. The ligand was placed above the heme iron, with a Fe···O distance of 1.9 Å. During the course of the optimization, the formic acid ligand started to separate from the heme (Fig. 3(a)), until the Fe···O distance is 3.0 Å. Simultaneously, the iron atom moves out of the porphyrin plane in a structure very similar to the unbound five-coordinated heme. The formic acid molecule undergoes little change with respect of an isolated molecule (Table 1). In particular, the C=O bond distances remain unchanged. All this indicates that the formic acid molecule is very weakly bound (Fig. 3).

In contrast, when formic acid is replaced by formate the ligand remains bonded to the heme, as reflected by the short Fe–O distance (2.03 Å) and the fact that the iron atom remains on the porphyrin plane. Table 1 shows that the internal structure of the formate anion undergoes significant changes upon ligand binding. The C–O distance is larger for the carbonyl that interacts with the iron atom (1.31 Å) than that of the non-interacting carbonyl (1.23 Å).

The different binding behaviour between formic acid and formate can be rationalized in terms of the different donor character of the carbonyl oxygen. In order to make

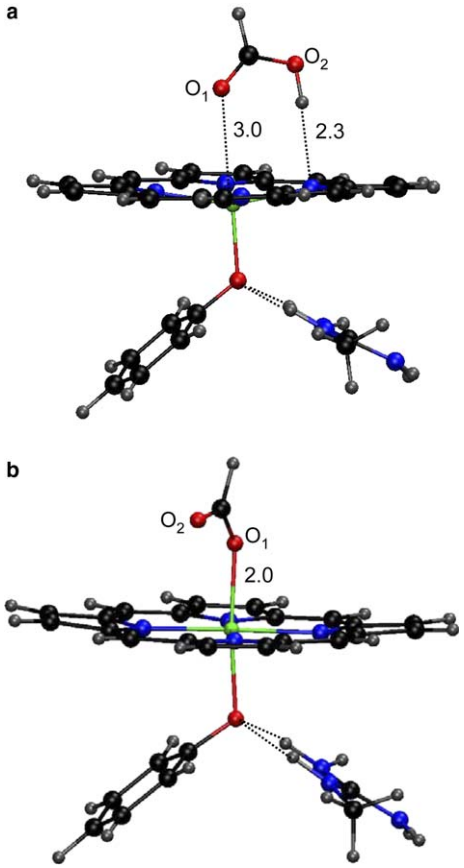

Fig. 3. (a) Optimized structure of FeP(Tyr + Arg)(HCOOH) complex. (b) Optimized structure of the FeP(Tyr + Arg) ($HCOO^-$) complex.

a bond to the iron, the sp^2 lone orbital of the carbonyl oxygen should interact with the antibonding Fe(dz^2) orbital (filled with one electron in high spin Fe^{III}). In the case of formate, having one electron more, the oxygen "sp^2" orbital lies high in energy and close to the iron dz^2 orbital, thus the interaction is maximal. Instead, the same orbital for formic acid lies much lower in energy and the interaction with the heme iron is weaker. It is apparent from Fig. 3 that the ligand prefers to interact with the heme through a OH···π interaction with the porphyrin π system than by forming a covalent Fe–O bond (the distance between the hydroxyl hydrogen and the closest porphyrin N is shorter than the Fe···O distance).

Therefore, for an isolated heme/ligand complex, only the formate anion shows a clear tendency to bind to the heme. These observations will be used later on to rationalize the ligand binding modes in the calculations using large models.

3.2. Ligand binding in the His/Asn region

As mentioned in the introduction, the structure of HPC shows a second ligand molecule sitting between the catalytic residues with the O_2 pointing towards the entry channel and O_1 located in between the catalytic residues (**B** in Fig. 1). The long distance between O_1 and

Table 1
Main parameters defining the optimized structure of the complexes of formate and formic acid with an iron porphyrin (Fig. 2)

Structural parameter	FeP(Tyr + Arg)($HCOO^-$)	$HCOO^-$	FeP(Tyr + Arg)(HCOOH)	HCOOH
Fe···O_1	2.03	–	3.01	–
C–O_1	1.31	1.26	1.22	1.22
C–O_2	1.23	1.26	1.35	1.37
O–H	–	–	1.00	0.99
∠O_1–C–O_2	127.9	127.0	125.0	122.2
Fe out-of-plane	−0.05	–	−0.4	–
Fe–N_p	2.07–2.10	–	2.05–2.09	–
Fe–O_{Tyr}	2.12	–	1.98	–
∠Fe–O_{Tyr}–C	122.1	–	124.8	–

The results of the optimized structures of an isolated ligand molecule are also reported for comparison. Distances are given in angstrom and angles in degrees.

the Fe atom (3.88 Å) suggests that there is little interaction between the ligand and the heme. If this is the case, a small model consisting of the catalytic residues and the ligand could be sufficient to have a first approximation to the binding mode. Moreover, the small size of this model (39 atoms) allows us to perform first principles MD simulations for a few picoseconds. To this aim, we constructed a model consisting of the His56 and Asn129 residues (modelled by methylimidazole and acetamide) and the formic acid ligand. The Ser95 and Thr96 residues, which are at hydrogen bond distance from the N_δ–H of His56, were also considered. The effect of these residues is crucial to describe a putative proton transfer between His56 and the formic acid ligand. Finally, a water molecule that interacts with the hydroxyl group of Ser95 was also included in the model (Fig. 4).

The starting position of the residues and the ligand was taken from the X-ray structure (**B** in Fig. 1). A structural optimization was performed, followed by a molecular dynamics simulation of 3 ps at 300 K. The orientation of the ligand found in the X-ray structure turned out not to be a minimum for this model, as it evolved towards a configuration in which the O–C–O plane is almost parallel with the plane of the imidazole ring of His56 (Fig. 4(a)). Molecular dynamics simulation shows that the carboxyl proton transfers back and forth between the ligand (His···HOOCH) and the imidazole ring (Fig. 4(b)), forming either a His–$H^{(+)}$···$^{(-)}$OOCH or a His···HCOOH type of interaction. Such transfers occurs three to five times per picosecond, with a corresponding oscillation of the COO plane with respect to the imidazole plane of 60°. The interchange of the COOH proton involves the same carboxylate oxygen within the time scale that was followed, and a longer simulation time would be required to exclude the possibility of proton exchange between carboxylate oxygen atoms via the N_ε of the imidazole.

Additional calculations starting with a formate anion and a neutral His56 result in the repulsive interaction between one formate oxygen and the N_ε of His56 forcing the ligand away from the imidazole ring, leaving it hydrogen bonded only with the NH_2 of Asn129. In addition, the molecular plane of the formate is almost perpendicular to

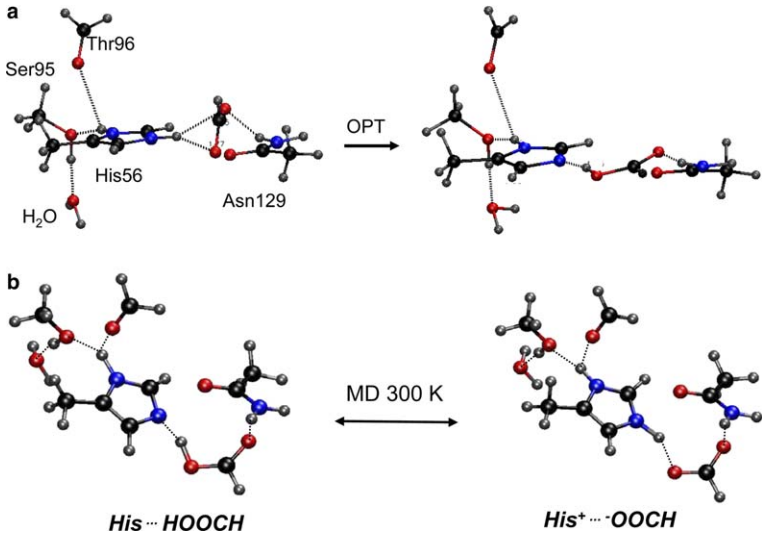

Fig. 4. (a) Evolution of the relative position of formic acid in the binding site **B**, upon structure optimization (see text). Only the catalytic residues and its close environment is included in the model. (b) Type of configurations obtained during the CPMD simulation at 300 K.

the NH_2 plane, and the ligand placement is inconsistent with the ligand locations in experiments.

Therefore, the calculations show that, unless His56 is protonated, the formate ion is unlikely to reach the region of the catalytic residues mainly because the combination of the formate anion and an unprotonated imidazole is unfavourable. Instead, the configuration with a formic acid molecule located in between the catalytic residues leads to a stable minimum, in which the proton is transferred between the ligand and the His56 residue. Although the main features remain unchanged, the inclusion of the protein environment in the calculation introduces quantitative changes to this picture.

3.3. Ligand binding in the presence of the protein environment

As a last step in our study, the two ligand positions found in the X-ray structure of HPC were analyzed by means of QM/MM calculations (see details in Section 2) in which the effect of the protein environment is taken into account by means of a classical force-field.

As a first step, the formic acid ligand was placed in the position **A** of the crystal structure (Fig. 1), with the unprotonated oxygen atom pointing towards the Fe atom and the protonated oxygen atom at hydrogen bond distance with the N_ε of His56. This configuration turned out to be unstable, as the formic acid proton transfers to the His in the course of the optimization. This is consistent with the fact that formate anion forms a stronger bond with the heme than formic acid (Section 3.1). Thus, when a good proton acceptor is present (His56) the ligand prefers to donate the OH proton and to bind to the heme as a formate anion. In the final optimized structure, one oxygen atom interacts with the heme iron ($O_1 \cdots Fe = 2.12$ Å) while the other interacts with both His56 (1.75 Å between the N_ε–H proton and O_2) and Asn129 (2.43 Å between the NH_2 proton and O_2). The asymmetry in the ligand C–O distances (1.29 and 1.26 Å in Table 2) reflects the strong coordination of the ligand with the heme iron. It is noteworthy that the elongation of the C–O_1 bond is less pronounced than it was in the absence of catalytic residues (Table 1) where the C–O_1 distance is 1.31 Å. This trend is expected as the Fe–O_1 distance is 0.12 Å larger than the optimum distance for a formate–heme complex (Fig. 3b). This indicates that the heme-ligand bond is not so strong as for an isolated heme, as the ligand is interacting simultaneously with the catalytic residues.

The computed structure reproduces well one of the ligand positions (**A**) found in the X-ray structure (Table 2). In particular, the distances between the ligand and the His56 and Asn129 residues are in very good agreement with experiments. The largest discrepancy is found for the Fe–O_1 distance, which is 0.5 Å shorter in the calculation. Part of this discrepancy might be due to the anharmonicity of the Fe–O bond which makes the average distance at 100 K larger than the distance at 0 K (our calculations refer to the last situation, as temperature effects were not taken into account). Unfortunately, a finite temperature MD simulation for this system is out of our computational capabilities. Another possible source of error is the method used (DFT, with the BP exchange correlation functional). Nevertheless, previous work on iron–porphyrin derivatives [10a] shows that the iron–ligand bond distances obtained using the same methodology reproduce fairly well the experimental bond distances. Another aspect that should be taken into account is the uncertainty of the experimental determination (the r.m.s.d. in bond lengths amounts to 0.02 Å and the estimated coordinate error is 0.16 Å [5]). The two C–O bond distances, for instance, are almost

Table 2
Main parameters defining the optimized structure of the QM region in the QM/MM calculations of the models depicted in Fig. 5

Structural parameter	Minimum A	Exp. (A)	Minimum B	Exp. (B)
Fe···O_1	2.12	2.66	3.94	3.88
C–O_1	1.29	1.24	1.28	1.22
C–O_2	1.26	1.23	1.26	1.23
∠O_1–C–O_2	125.5	124.4	125.2	123.7
N(His)–H···O_1	2.58	–	1.50	–
N(His)–H···O_2	1.75	–	2.33	–
N(Asn)–H···O_1	4.20	–	3.25	–
N(Asn)–H···O_2	2.43	–	1.81	–
N(His)···O_1	3.19	3.13	2.55	2.77
N(His)···O_2	2.81	2.63	3.02	2.81
N(Asn)···O_1	5.24	5.43	4.24	3.61
N(Asn)···O_2	3.44	3.37	2.80	2.58
N(His)–H	1.07	–	1.11	–
N(Asn)–H	1.03	–	1.04	–
Fe out-of-plane	+0.1	−0.2	−0.4	−0.2
Fe–N_p	2.04–2.11	2.03–2.09	2.02–2.08	2.04–2.07
Fe–O_{Tyr}	2.09	1.87	1.96	1.93
∠Fe–O_{Tyr}–C	122.9	126.29	121.05	124.47

Distances are given in angstrom and angles in degrees. The experimental data is taken from [5] (PDB entry 1qwm).

equivalent in the X-ray structure while they should have an asymmetry of 0.03 Å due to the different coordination of the two oxygen atoms. Nevertheless, even considering all these sources of small errors, both experiments and theory show that the ligand is inside the heme pocket and interacts simultaneously with the iron atom and the two catalytic residues. In addition, the calculations show that the ligand is present as formate and the His56 residue is protonated.

A second calculation was performed starting with the formic acid ligand in the alternative position found in the X-ray structure (**B** in Fig. 1). In this case, we also found that the OH proton transfers to the His56 residue. The final optimized structure corresponds to a formate anion in which one oxygen interacts with N_ε of His56 and the other oxygen interacts with the NH_2 group of Asn129, as shown in Fig. 5(B). Interestingly, the ligand keeps the orientation found in the crystal structure, unlike what was observed using a small model (Fig. 4(a)). Therefore, the protein environment is responsible for the orientation of the ligand relative to the His56 and Asn129 residues (i.e., at the entry gate).

Comparing the total energy of this minimum with that of structure **A** (Fig. 5(A)) gives an energy difference of 16 kcal/mol, favouring structure **A**. As the protein environment is fixed in the calculation, this difference can be fully attributed to differences in the structure of the quantum region, i.e., the different binding energy of the ligand. On another hand, it is expected that migration of the ligand from **B** to **A** involves a sizable barrier, as each of these structures alone is observed in the X-ray structure [5,6].

Finally, it should be pointed out that entropic contributions (not taken into account in our model) could play a role in this picture. The fact that the protein environment dictates the orientation of the ligand in **B** suggests that protein fluctuations are important for the binding in **B**, but not in **A**. In other words, entropic contributions could affect to the free energy difference between **A** and **B**, favoring conformation **B**. For this reason, the energy difference obtained in the calculation should be regarded as an upper bound to the real free energy difference.

In summary, two minimum energy structures are found. The lowest energy structure corresponds to formate anion interacting with the heme iron and the catalytic residues (His56-H^+ and Asn129), while the second corresponds to the formate anion trapped in the His56/Asn129 region and forming hydrogen bond interactions with the catalytic residues. The His56 residue is found to be protonated in both cases, although we cannot exclude that the proton is shared between His56 and the ligand in structure **B**, as MD simulations in the second model predict.

The binding modes of formic acid found in HPC and the fact that the proton easily transfers to His56 allows us to draw some suggestions on the process of formic acid binding. The most likely pathway for the ligand to enter the heme pocket is through the main channel located above the heme. Once the ligand reaches the region of the catalytic residues, it transfers the hydrogen to the His56 residue

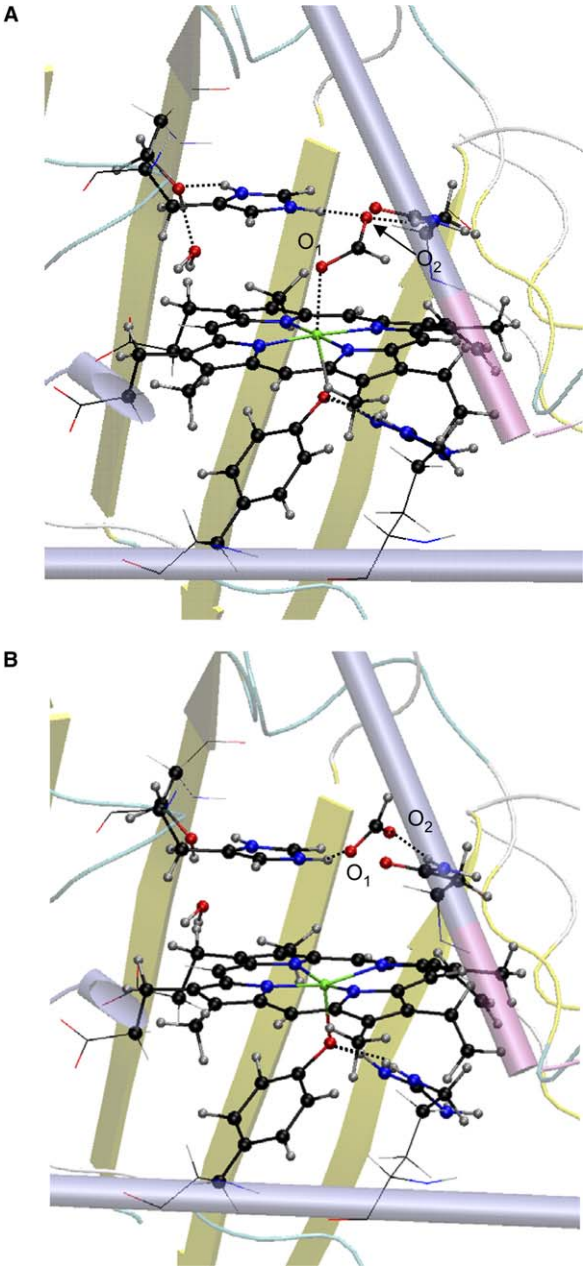

Fig. 5. Optimized structures obtained from the CPMD QM/MM simulation for binding sites **A** and **B**. The atoms of the QM region are shown in ball and stick.

(Fig. 6) and remains trapped in the region above the catalytic residues (conformation **B**), forming hydrogen bond interactions with His56 and Asn129 [(His)N_ε–$H^{(+)}\cdots{}^{(-)}$O=C(H)=O$\cdots H_2$N(Asn)]. The same final configuration would be obtained if the ligand enters the main channel as formate and binds to the protonated His56 residue. In the absence of the appropriate protein fluctuations stabilizing conformation **B**, the ligand penetrates into the binding pocket and interacts with the heme, leading to configuration **A**. Because of the limitations of our model and the short time scales sampled in first-principles calculations, we cannot elucidate the structural determinants for

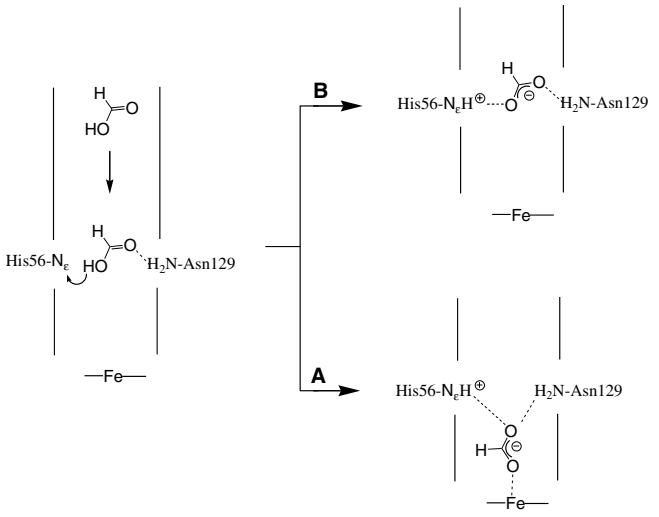

Fig. 6. Proposed scheme of the binding process of formic acid in catalase.

conformation **B**. Long time scale MD simulations without structural restraints, combined with QM/MM calculations, are needed to clarify the role of protein fluctuations in the ligand binding process.

4. Conclusions

In this work, the most stable forms of formic acid bound to catalase are analyzed in order to interpret the active site structure reported in the recent X-ray structures of HPC and PMC with a bound formic acid/formate. Two different binding sites are found in the X-ray structure of HPC (**A** and **B**). One of them (**A**, also found in PMC), corresponds to the ligand in the heme cavity, interacting with both the heme iron and the catalytic residues, His56 and Asn129. In the other binding site (**B**), the ligand is far from the heme and interacts only with the catalytic residues. The protonation state of the His56 residue and formate ligand are not available from experiments, therefore DFT calculations were performed to clarify this aspect, as well as to determine the internal structure of the formic acid/formate ligand.

Calculations using small models show that only the formate anion binds to the heme. However, only formic acid binds to the His/Asn region, unless His56 is protonated. The formate anion is unlikely to enter the heme cavity in the absence of an accompanying proton to protonate the imidazole, as the repulsive interaction between formate O^- and the N_ε of His56 forces the ligand away from the entry channel.

The above conclusions are further supported by the results of calculations using large models including the protein environment by means of a QM/MM approach. The calculations show that when formic acid reaches the catalytic residues, it transfers the carboxyl proton to the His56 residue and subsequently binds in the His/Asn region in an orientation that is very sensitive to changes in the protein environment. However, if the ligand is able to reach the heme, it forms a very stable structure in which one oxygen binds to the heme iron and the other one is involved in hydrogen bond interactions with the catalytic residues.

Finally, it is suggested that protein fluctuations control the access of the ligand to the heme iron, being key to trap the ligand in the His/Asn region. Long time scale MD simulations on the free protein, combined with QM/MM calculations, are needed to approach this problem.

Acknowledgements

This work was supported by Grants 2001SGR-00044 (CIRIT) and FIS2005-00655 (MEC). The computer resources were provided by the Barcelona Supercomputing Center (BSC). Financial support from the Ramon y Cajal program of the MCYT and the ICREA foundation was provided. Support was also provided by NSERC (069600) and Canada Research Chair Program (PCL).

References

[1] P. Nichols, I. Fita, P.C. Loewen, in: A.G. Sykes, G. Mauk (Eds.), Enzymology and Structure of Catalases, Advances in Inorganic Chemistry, vol. 51, Academic Press, 2001, p. 51.
[2] P. Chelikani, I. Fita, P.C. Loewen, Cell. Mol. Life Sci. 61 (2004) 192.
[3] M.J. Maté, G. Murshudov, J. Bravo, W. Melik-Adamyan, P.C. Loewen, I. Fita, in: A. Messerschmidt, R. Huber, T. Poulos (Eds.), The Handbook of Metalloproteins, Willey, Chichester, 2001, p. 486.
[4] P. Nichols, G.R. Schonbaum, in: Catalases, second ed., in: P. Boyer, H. Lardy, K. Myrback (Eds.), The Enzymes, vol. VIII, Academic Press, New York, 1963, p. 147.
[5] P.C. Loewen, X. Carpena, R. Perez-Luque, C. Rovira, R. Haas, S. Odenbreit, P. Nichols, I. Fita, Biochemistry 43 (2004) 3089.
[6] P. Andreoletti, A. Pernoud, G. Sainz, P. Gouet, H. Jouve, Acta Crystallogr. D59 (2003) 2163.
[7] G.H. Carlsson, P. Nicholls, D. Svistunenko, G.I. Berglung, J. Hadju, Biochemistry 44 (2005) 635.
[8] M. Maj, P. Loewen, P. Nicholls, Biochim. Biophys. Acta 1384 (1998) 209.
[9] W. Koch, M.C. Holthausen (Eds.), A Chemist's guide to Density Functional Theory, Wiley-VCH, Weinheim, Germany, 2000.
[10] (a) See for instance: C. Rovira, K. Kunc, J. Hutter, P. Ballone, M. Parrinello, J. Phys. Chem. A 101 (1997) 8914;
(b) C. Rovira, M. Parrinello, Biophys. J. 78 (2000) 93;
(c) C. Rovira, B. Schulze, M. Eichinger, J.D. Evanseck, M. Parrinello, Biophys. J. 81 (2001) 435;
(d) M.T. Green, J.H. Dawson, H. Gray, Science 304 (2004) 1653;
(e) B. Meunier, S. de Visser, S. Shaik, Chem. Rev. 104 (2004) 3947.
[11] C. Rovira, I. Fita, J. Phys. Chem. B 107 (2003) 5300.
[12] D.A. Pearlman, D.A. Case, J.W. Caldwell, W.S. Ross, T.E. Cheatham, S. Debolt, D. Ferguson, G. Seibel, P. Kollman, Comput. Phys. Commun. 91 (1995) 1.
[13] (a) R. Car, M. Parrinello, Phys. Rev. Lett. 55 (1985) 2471;
(b) D. Marx, J. Hutter, in: J. Grotendorst (Ed.), Ab Initio Molecular Dynamics: Theory and implementation, Modern methods and algorithms of Quantum Chemistry, John von Neumann Institute for Computing, Julich, Germany, 2000, p. 301.
[14] (a) P. Carloni, U. Röthlisberger, M. Parrinello, Acc. Chem. Res. 35 (2002) 455;
(b) P. Carloni, U. Röthlisberger, in: L. Eriksson (Ed.), Simulations of enzymatic systems: Perspectives from Car-Parrinello molecular

dynamics simulations, Theoretical Biochemistry-Processes and Properties of Biological Systems, Elsevier Science, Amsterdam, p. 215.
[15] M. Troullier, J.L. Martins, Phys. Rev. B 43 (1991) 1993.
[16] S.G. Louie, S. Froyen, M.L. Cohen, Phys. Rev. B 26 (1982) 1738.
[17] (a) A.D. Becke, J. Chem. Phys. 84 (1986) 4524;
(b) J.P. Perdew, Phys. Rev. B. 33 (1986) 8822.
[18] CPMD program, Copyright IBM Corp. 1990–2003, Copyright MPI für Festkörperforschung, Stuttgart, 1997–2001. URL: http://www.cpmd.org.
[19] W. Humphrey, A. Dalke, K. Schulten, J. Mol. Graph. 14 (1996) 33.
[20] A. Laio, J. vandeVondele, U. Röthlisberger, J. Chem. Phys. 116 (2002) 6941.
[21] P.J. Hünenberger, Chem. Phys. 113 (2000) 10464.