Molecular Dynamics Studies on Peroxidases: A Structural Model for Horseradish Peroxidase and a Substrate Adduct

Lucia Banci,* Paolo Carloni, and Giovanni Gori Savellini

Department of Chemistry, University of Florence, Via G. Capponi 7, 50121 Florence, Italy

Received June 13, 1994®

ABSTRACT: Molecular dynamics (MD) calculations are performed on cytochrome *c* peroxidase (CcP) and on horseradish peroxidase, isoenzyme C (HRP), and its substrate adduct with *p*-cresol. For CcP, a refinement in solution of the X-ray structure is obtained which indicates that in solution the protein structure is very similar to that in the crystal. For HRP, the X-ray structure is not available. We have generated a model of this protein based on the recently reported structure of the similar lignin peroxidase (LiP) protein. This model involves the entire system as all the amino acid residues match the sequence. This HRP model was refined through energy minimization and MD calculations. A refined structural model for HRP, for the first time involving the entire protein, is therefore now available. The tertiary structure of HRP is close to that of LiP, and also the active site in the two proteins has significantly similar structures. The well-ordered water molecules and the extensive H-bond network present in the X-ray structure of CcP is maintained in the dynamics without any constraints, indicating that the active site residues produce a field strong enough to make all these interactions quite stable. Interestingly, also in HRP a network of ordered water molecules and H-bonds is present, again without constraints. This is consistent with the similarities of the active sites in the two proteins. Finally, we have calculated the MD structure of the adduct of HRP and a substrate molecule, *p*-cresol. This structural model is compared with the NMR data, which are in fairly good agreement. The binding site and the protein–substrate interactions are discussed.

The detailed knowledge of the structural and dynamical properties of proteins is a fundamental and key step in the comprehension of their biological behavior. Consequently, all the techniques aiming at this purpose have received many efforts in their development and in the application to biological systems. Among these techniques, molecular dynamics (MD) calculations can provide structural models for those systems for which the X-ray characterization is not available. These calculations allow also the study of the mobility of residues in the protein frame and are well suited for the structural analysis of the protein–substrate adducts, which in many cases cannot be crystallized and therefore cannot be studied through X-ray characterization (McCammon & Harvey, 1987; Kaptein et al., 1988; Brooks et al., 1988; Banci et al., 1992b).

Peroxidases represent an interesting class of proteins for the application of MD calculations. They are a class of heme enzymes which react with hydrogen peroxide to form a derivative (called compound I) which is two electrons higher in oxidation state with respect to the resting state (Dunford, 1982, 1991; Kirk & Farrell, 1987; Dawson, 1988; Tien, 1987; Tien & Kirk, 1983; Bosshard et al., 1991). In the latter state, peroxidases contain a five-coordinate high-spin iron(III) ion which is bound to the heme pyrroles and to an axial histidine called proximal histidine.

The X-ray structure is known for two of these proteins, cytochrome *c* peroxidase (CcP) (Finzel et al., 1984; Wang et al., 1990) and lignin peroxidase (LiP), recently reported (Edwards et al., 1993; Poulos et al., 1993; Piontek et al., 1993). In the case of horseradish peroxidase (HRP) and manganese peroxidase (MnP), the X-ray structure is not yet available, but extensive NMR characterization on the protein residues around the heme moiety has allowed the structural comparison among the various isoenzymes (Thanabal et al., 1987b; Banci et al., 1992a; La Mar et al., 1980, 1989). However, the characterization of HRP and MnP still suffers the lack of X-ray structure.

The distal site, where the iron ion interacts with hydrogen peroxide and compound I is formed, is characterized by two residues which are invariant in all the peroxidases sequenced up to now (Fee & DiCorleto, 1973; Jones, 1992; Welinder, 1992): the distal histidine and the distal arginine. They are proposed to assist the binding of H_2O_2 and the consequent formation of compound I.

On the proximal side, an invariant Asp residue is H-bonded to the δ_1 nitrogen of the proximal histidine, thus imparting a larger anionic character to the axial ligand in peroxidases with respect to globins (Poulos & Kraut, 1980a). This strong H-bond is believed to be the major factor responsible of the lower redox potentials of peroxidases compared to globins (Finzel et al., 1984; Poulos et al., 1993; Poulos & Kraut, 1980b; Banci et al., 1991a,b; Smulevich, 1993; von Bobman et al., 1986).

Compound I is able to perform two one-electron oxidations of a large variety of substrates. The various peroxidases characterized up to now show high selectivity on the nature of the catalyzed substrate and on the mode of interaction with the substrate itself. HRP oxidizes small aromatic molecules which can approach the porphyrin ring. On the other hand, CcP and LiP catalyze the oxidation of large molecules, i.e., a protein, cytochrome *c* (cyt *c*) in the case of CcP and lignin, an aromatic polymer, for LiP. These substrates can obviously only interact with the surface of the protein and can be oxidized only via a long-range electron transfer.

Furthermore, while the X-ray structure is known for the CcP–cyt *c* complex (Pelletier & Kraut, 1992), no structural data are yet available for the protein–substrate adducts of

* Address correspondence to this author; phone: +39-55-2757550; fax: +39-55-2757555; e-mail address: lucia@risc1.lrm.fi.cnr.it.

® Abstract published in *Advance ACS Abstracts,* September 1, 1994.

0006-2960/94/0433-12356$04.50/0

HRP, LiP, and MnP. For these proteins, the binding of substrates has been characterized through NMR spectroscopy (Veitch et al., 1992; Banci et al., 1992a, 1993; Morishima & Ogawa, 1979; La Mar & de Ropp, 1979; Casella et al., 1991; Thanabel et al., 1987a), through resonance Raman spectroscopy (Rogers et al., 1988; Kuriyan et al., 1986; Mabbutt & Wright, 1985), and through kinetic measurements on the complexes with suicide inhibitors (Ortiz de Montellano et al., 1988; Ator et al., 1987; Ator & Ortiz de Montellano, 1987). It has been learned that, in HRP, aromatic molecules bind in a cavity close to the heme edge bearing the 8-CH_3 group.

As the X-ray structure of HRP is still lacking, we thought it would have been useful to build a structural model for this protein and to refine it through energy minimization and MD calculations. A model for HRP was already made available by simple residue substitution (Welinder, 1992), but it involves only a part of the protein, so that all the proximal side and the substrate binding site are missing.

With this aim, we have addressed the problem starting first with the simulations of CcP for which the X-ray structure is known at a high level of refinement. The calculations on this protein have allowed us to test the validity of our force field parameters for the heme atoms, the iron ion, and the histidine axial ligand as well as the validity of the overall MD procedure. In addition, these calculations have allowed us to refine the structure of CcP in solution and to learn about its dynamic behavior.

MD simulations on the resting state and on compound I of CcP are already available (Collins et al., 1992), but the focus of that paper was mainly the comprehension of properties of compound I. Therefore, the calculations were performed for a much shorter time.

During the present work, the coordinates of LiP became available [entry 1LGA of the Protein Data Bank (Bernstein et al., 1977)]. The amino acid sequence of LiP on the active site is much closer to HRP than CcP. Indeed, LiP and HRP have an identical active site with distal and proximal Phe residues which, on the contrary, are both Trp in CcP. Both LiP and HRP produce a compound I derivative that is characterized by a heme delocalized cation radical. Finally, both LiP and HRP have a larger molecular weight than CcP and are glycosylated on the surface.

We therefore chose to build the starting structural model for MD simulations from the X-ray structure of LiP, which in the meantime became available. We also performed some test calculations on LiP to check if our force field parameters and MD procedure were capable of reproducing their structures and therefore were of general use for peroxidases.

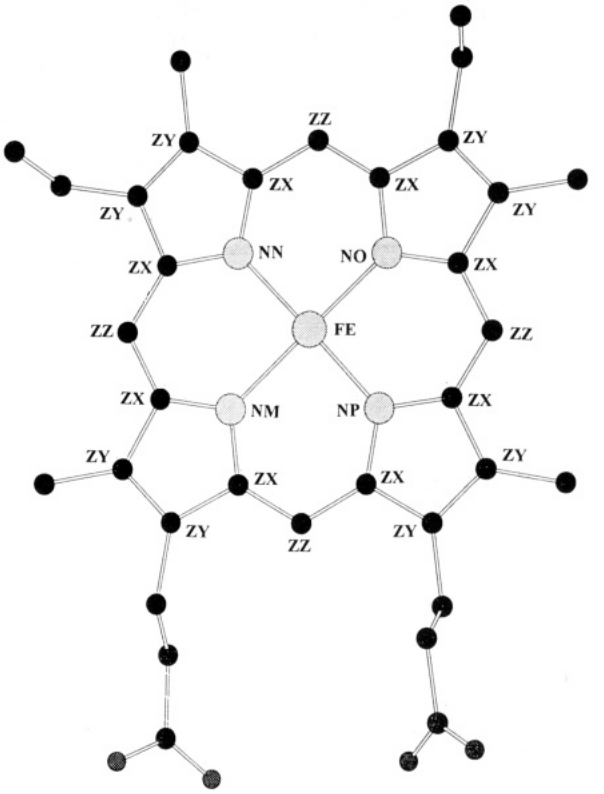
Once the validity of our approach was checked, we performed MD calculations on the structural model of HRP (isoenzyme C) we built from the X-ray structure of LiP. This MD structure would help in obtaining a better comprehension of the large amount of data collected on this protein.

Finally, the interaction of HRP with the substrate was studied by performing MD calculations on an adduct in which a substrate molecule was docked in the cavity of HRP. The results of these calculations were compared with the data obtained through NMR.

EXPERIMENTAL SECTION

The simulations were carried out using the AMBER 4.0 (Pearlman et al., 1991) program on IBM RISC/6000 550 and 530 workstations. Part of the calculations were carried out with the parallel version of AMBER 4.0 (Chin, 1992) running on a IBM/SP1 parallel machine.

Chart 1: Atom Type of Atoms in Porphyrin Group[a]



[a] The corresponding force field parameters are listed in Table 1.

Parametrization of Metal Active Site Parameters. (A) Bond Parameters. The stretching force constants were taken from the data file of AMBER 4.0. During all the simulations, however, the SHAKE algorithm (van Gunsteren & Berendsen, 1977) has been used for all the bonds of the protein and of the heme, thus constraining bond lengths to their equilibrium values. The values of the bending and torsion angle constants missing in the AMBER data file were derived from data already reported in the literature and obtained on some porphyrin model complexes (Argos & Mathews, 1975; Angelucci et al., 1993; Ruf et al., 1979; Dawson et al., 1982). Similar values have been also used for the bending constants in simulations on myoglobin (Case & Karplus, 1978). These parameters are able to maintain the iron ion slightly outside the heme plane as observed in the X-ray structure of CcP (Finzel et al., 1984). The iron atom was taken as five coordinate in all these proteins as it results from the X-ray and spectroscopic characterizations. Therefore, neither a bond nor a constraint was placed between the iron and the well-ordered water molecule present close to it.

(B) Nonbonded Interactions. Electrostatic interactions provide the most relevant contribution to the nonbonded interactions which are key factors for determining the tertiary structure. It is now well-established that the use of formal charges is unrealistic and could lead to wrong results. The parameters for heme moiety, the iron ion, and the axial histidine ligand, including their point charges, were mainly taken from the AMBER 4.0 database. The charge on iron(III) was set to +1.25. This value is in agreement with a recent paper which reports a MD characterization of CcP (Collins et al., 1992) and where INDO charges are used. The force field parameters for the heme atoms, whose type is defined in Chart 1, are reported in Table 1.

Table 1: Force Field Parameters for Atoms in Porphyrin Moiety[a]

bond	K_r[b]	r_0[c]	bond	K_r[b]	r_0[c]
Fe-NO	40.0	1.97	Fe-NM	40.0	2.04
Fe-NP	40.0	1.99	Fe-NB	100.0	1.95
Fe-NN	40.0	2.00			

angle	K_θ[d]	Θ_0[e]	angle	K_θ[d]	Θ_0[e]
Fe-NP-ZX	50.0	127.3	NP-Fe-NM	0.0	90
Fe-NO-ZX	50.0	127.3	NO-Fe-NN	0.0	90
Fe-NN-ZX	50.0	127.3	NN-Fe-NM	0.0	90
Fe-NM-ZX	50.0	127.3	NP-Fe-NB	50.0	98.1
NP-Fe-NN	0.0	179.9	NO-Fe-NB	50.0	98.8
NO-Fe-NM	0.0	179.9	NM-Fe-NB	50.0	92.0
NP-Fe-NO	0.0	90	NN-Fe-NB	50.0	89.1

dihedral	V_n[d]	Θ[e]	γ
X-Fe-NB-X	0.0	180.0	2
X-Fe-NP-X	0.0	180.0	2
X-Fe-NO-X	0.0	180.0	2
X-Fe-NN-X	0.0	180.0	2
X-Fe-NM-X	0.0	180.0	2
Fe-NP-ZX-ZY	15.0	180.0	2
Fe-NO-ZX-ZY	15.0	180.0	2
Fe-NN-ZX-ZY	15.0	180.0	2
Fe-NM-ZX-ZY	15.0	180.0	2
Fe-NO-ZX-ZZ	15.0	180.0	2
Fe-NP-ZX-ZZ	15.0	180.0	2
NM-Fe-NB-CV[f]	40.0	0.0	2
NM-Fe-NB-CR[f]	40.0	0.0	2
NN-Fe-NB-CV[f]	40.0	0.0	2
NN-Fe-NB-CR[f]	40.0	0.0	2
NO-Fe-NB-CV[f]	40.0	0.0	2
NO-Fe-NB-CR[f]	40.0	0.0	2
NP-Fe-NB-CV[f]	40.0	0.0	2
NP-Fe-NB-CR[f]	40.0	0.0	2

[a] The atom type definition is reported in Chart 1. NB indicates the nitrogen of the axial histidine bound to the iron. [b] In kcal mol^{-1} Å^{-2}. [c] In Å. [d] In kcal mol^{-1} rad^{-2}. [e] In degrees. [f] These parameters are applied only to HRP.

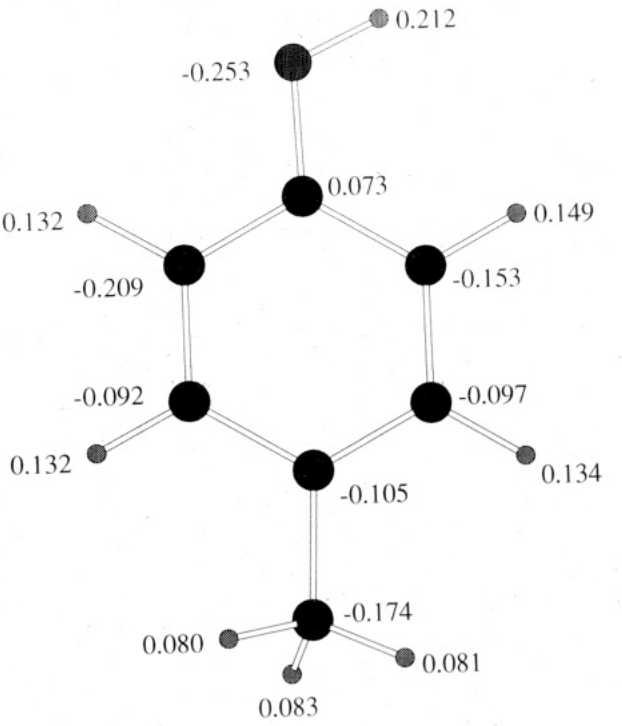
Parametrization of the Rest of the Protein. For the residues within a 15-Å sphere centered on the iron ion, we have used the all-atom AMBER force field parameters (Weiner et al., 1986), while for the rest of the protein the united-atom force field parameters (Weiner et al., 1984) have been used. We are using this mixed set of parameters, which is a widely used and well-accepted procedure (Dagget & Kollman, 1990; Ferguson et al., 1991; Merz et al., 1991), because it allows a detailed description of the protein around the active site without a very high computational time requirement.

Parametrization of the p-Cresol Molecule. The force field parameters of *p*-cresol were taken equal to those of similar moieties (i.e., the side chain of tyrosine) present in the AMBER 4.0 database. The atomic point charges, on the other hand, were obtained from semiempirical calculations using MOPAC 5.0 program (F. J. Seiler Research Laboratory, U.S. Air Force Academy, Colorado Springs, CO). The AM1 Hamiltonian was used for the SCF calculation. The charges are reported in Chart 2.

Protein Models. The calculations are run at neutral pH so that the histidines present in the molecule were taken as neutral and protonated on Nϵ1. The total protein charge of −14 for CcP and −2 for HRP was neutralized by the addition of 14 and 2 sodium ions, respectively, far from the active site and in positions so as not to interfere with salt bridges. For HRP, the sodium ions were added only for the runs in water (see later).

(A) CcP Model. The X-ray structure of CcP (Finzel et al., 1984) was used as a starting structure for our MD calculations. The 263 crystallographic water molecules found in the CcP X-ray structure were explicitly taken into account.

Chart 2: Atomic Point Charges of *p*-Cresol (see Experimental Section).



(B) HRP Model. The construction of a starting structural model for HRP is nontrivial, because this protein presents several insertions and deletions with respect to the two proteins for which the X-ray structure is available. As a starting structure, we used the crystal structure of LiP (Poulos et al., 1993). As mentioned in the introduction, we chose these coordinates because HRP is much closer, on many respects, to LiP than to CcP. Indeed, both LiP and HRP have in the active site a distal and a proximal Phe, which are Trp in CcP, and have a larger molecular weight than CcP. According to the sequence alignment reported by Welinder (Figure 2a in Welinder, 1992), we replaced the amino acids of LiP with those of HRP using the molecular graphics program Sybyl (1989). The positions of the Cβ atoms in HRP were kept as close as possible to those of LiP. Insertions and deletions of the backbone were needed to construct the model. We tried to maintain the structure of the backbone as much as possible near that of LiP. In particular, all the α helices of LiP were maintained in our HRP model, on the basis of the experimental observations that the helical content is 30–40% in HRP [as estimated from circular dichroism measurements (Strickland et al., 1968)] and 32% in LiP [from the crystal structure (Poulos et al., 1993)]. The inserted backbone atoms were added in a random conformation; however, the values of the dihedral angles ϕ and φ have been carefully checked and eventually modified when they were inducing a backbone conformation high in energy. This ended up with a model which shows a good Ramachandran plot (see Figure 1 of the supplementary material) with only a few residues in outlying regions. A similar plot was found in the X-ray structure of LiP (Poulos et al., 1993). The four S–S bridges experimentally found in HRP were explicitly included (see later). We did not include in the model the sugar chains which are on the surface of HRP. This approximation is not affecting our calculations because the sugar chains have been shown experimentally not to affect the overall structure of the protein and particularly the active site (Veitch et al., 1992). We have also not included in our calculations the calcium ions.

Table 2: Minimization Procedure for HRP *in Vacuo*[a]

SD steps	CG steps	k_α[b]	k_{Fe}[b]	k_S[b]
2400	0	–[c]	20.0	0.1
2400	0	20.0	20.0	10.0
2400	1500	3.0	20.0	50.0
2400	1500	0.1	0.1	160.0
500	0	3.0	0.1	0.0[d]

[a] SD = steepest descent method; CG = conjugate gradient method; k_α = harmonic constant for the α helix backbone atoms; k_{Fe} = harmonic constant for the atoms belonging to residues within a iron-centered 15-Å sphere; k_S = harmonic constant between bridging cysteine sulfur atoms. [b] Expressed in kcal mol^{-1} Å^{-2}. [c] α helix backbone atoms not moved. [d] Cysteine S–S covalent bond is imposed.

Table 3: MD Simulation Procedures for HRP *in Vacuo*[a]

simulation time[c]	k_α[b]	k_{Fe}[b]
0.0–9.0	3.0	0.1
9.0–13.5	2.5	0.05
13.5–18	2.0	0.025
18.0–22.5	1.5	0.012
22.5–27.0	1.0	0.0070
27.0–31.5	0.5	0.0020
31.5–36	0.1	0.0005
36–130	0.0	0.0

[a] The symbols k_α and k_{Fe} are as in Table 2. [b] Expressed in kcal mol^{-1} Å^{-2}. [c] Expressed in ps.

(C) Active Site in HPR Model. NMR studies (Thanabal et al., 1987a, 1988; Banci et al., 1991b; de Ropp et al., 1991a) have shown that the position of the distal and axial histidines is very similar in HRP and LiP, i.e., they are almost perpendicular to the heme plane in both proteins. Test calculations both *in vacuo* and in solution showed drifts of the position of the two residues; hence, constraints of their position during the equilibration and the production steps are needed. The conformation of His 170, bound to iron, was maintained by constraining the N(heme)-Fe-Nϵ2(His62)-Cϵ(His62) and N(heme)-Fe-Nϵ2(His62)-Cδ(His62) dihedral angles around their starting position (k = 50 kcal mol^{-1} rad^{-2}); His 42 was constrained with harmonic distance constraints between the nitrogen atoms of the heme pyrroles and Hϵ1 and Hδ2 atoms of the histidine (k = 10 kcal mol^{-1} Å^{-2}).

(D) Relaxation of HRP Model in Vacuo. Since the starting model is considerably far from equilibrium, large structural rearrangements are expected during the first steps of the calculations (equilibration phase), particularly on the arbitrarily added random coil insertions. Furthermore, the presence of S–S bridges in regions of HRP different with respect to LiP may induce significant changes in the backbone conformation. Hence, a long MD trajectory is needed to obtained a really equilibrated structure.

Unfortunately, for quite large systems like the present protein (about 4300 atoms of the proteins + 9000 of the solvent), the equilibrium phase would take very large computer resources. One way to overcome the problem with less computational effort is to perform first preliminary energy minimization and MD simulations *in vacuo*, the presence of the solvent being simulated by a distance-dependent dielectric constant. In this way, the largest rearrangements of the protein will be accomplished with relative less computer power. Then, the rearrangements due to the presence of the solvent, which are expected to be minor, can be obtained later with a simulation in water. Indeed, simulations *in vacuo*, particularly in the presence of constraints, are largely used for the refinement of structures in solutions obtained from NMR experiments (Scheek et al., 1989).

The Cys residues of the four cysteine pairs, which are known to be bound through a disulfide bridge (Welinder, 1985), turned to be quite far each other of the pair (from 10 to 17 Å) in the starting molecule.

Imposing a covalent bond (with the application of an harmonic potential) between the sulfur atoms would result in an enormous force acting on the cysteines and neighborhood residues and, consequently, in bad contacts and/or distortions in various regions of the protein. We decided therefore to apply forces with increasing strengths between the sulfur atoms forming the cysteine bridges. This procedure progressively reduced the S–S distance to the equilibrium value (2 Å) without causing unphysical distortions in the protein. Table 2 summarizes the minimization procedures we have used. In the first step, the atoms belonging to the backbone of the α helices were kept frozen and a 20 kcal mol^{-1} Å^{-2} harmonic constraint was set to their starting position for the residues within a 15-Å sphere centered on the iron atom (Table 2). In the second step, the α helix backbone atoms were allowed to move. The constraints applied to the α helix backbone atoms were progressively decreased during the minimization to a final value of 0.1 kcal mol^{-1} Å^{-2} (Table 2). The distance between the sulfur atoms of the bridged cysteines after this preliminary minimization was between 2 and 3 Å. In the last step of the minimization *in vacuo*, a covalent bond between the sulfur atoms of the cysteines was imposed. Few steps were needed to reach the convergence (gradient of the energy less than 0.1 kcal mol^{-1} Å^{-2}).

After the structure was minimized, a 130-ps MD simulation was carried out on the entire protein. Also in this case, at the beginning of the simulation, constraints on the α helix backbone atoms and the 15-Å Fe-centered sphere residues were applied and progressively decreased to zero (Table 3). During the MD simulation, the SHAKE algorithm (van Gunsteren & Berendsen, 1977) was used to constrain all bond lengths. The time step used was 1.5 fs. A residue-based cutoff value of 10 Å has been set for the nonbonded interaction and a distance-dependent dielectric constant was used.

The system was gradually warmed using the following procedure: 0–100 K, 3 ps; 100–200 K, 3 ps; and 200–300 K, 3 ps. The system was coupled with a thermal bath (Berendsen et al., 1984) at 300 K for the remaining 121 ps with a coupling constant of 0.1 ps. The average value of the temperature during this period was 299.5 K with a standard deviation of 4.0 K. During the last phase of the MD simulation, the system *in vacuo* is well equilibrated: indeed, the value of the total (as well as the potential energy) reaches a stable value for the last 55 ps. Also the plot of the rmsd of the backbone atoms vs time on the same time range confirms that the system has reached the equilibrium (Figure 2 of the supplementary material). The average rmsd of the last 55 ps with respect to the starting structural model is 3.3 Å (standard deviation 0.1 Å). The Ramachandran plot of the HRP structure after 126 ps shows also in this case only few ϕ–φ pairs in the highest energy regions (Figure 3 of the supplementary material).

Before performing the simulation in water, the crystal water molecules present in the active site of LiP (Poulos et al., 1993) were added to the structure of the final step of the simulation *in vacuo*.

(E) HRP–p-Cresol Adduct. The HRP–substrate (*p*-cresol) model was generated by docking the *p*-cresol in the structure of HRP obtained by the MD average structure in solution after full minimization. The *p*-cresol molecule was found to be, from NMR data (Banci et al., 1993), close to the heme edge bearing the 8-CH_3 group and close to a Phe residue.

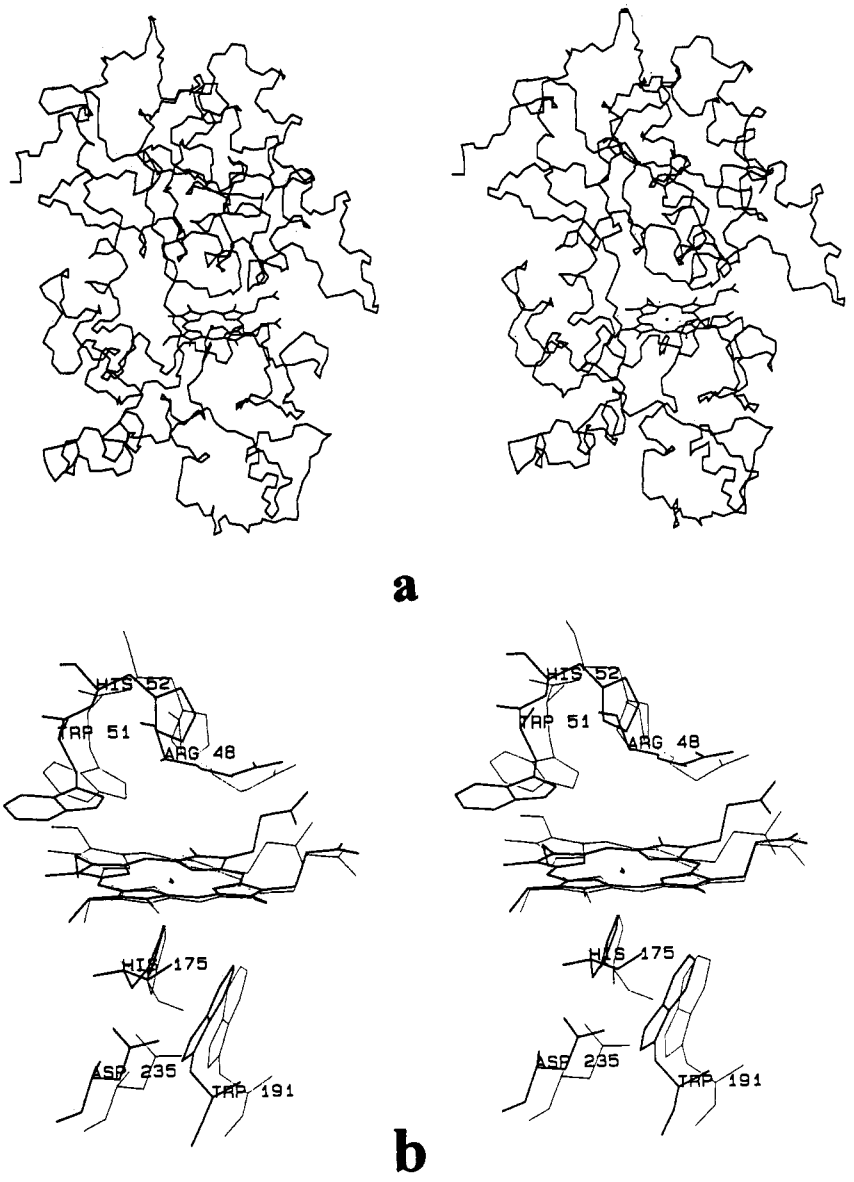


FIGURE 1: (a) Stereo view of the backbone atoms of the MD average structure of CcP. The carbonyl oxygen atoms are omitted for clarity. The heme moiety is also shown. (b) Comparison of the active site of CcP in the X-ray (thin) and MD (bold) structures.

NOEs are detected between 8-CH_3 and Hϵ's of an unassigned Phe residue and the 3,5-hydrogen atoms of *p*-cresol.

In the HRP MD average model, two Phe residues are at a position to satisfy this condition. *p*-Cresol was placed close to each of these two Phe residues, and MD simulations were performed for both models.

(F) Solvation of Models. The three systems (CcP, HRP, and HRP–*p*-cresol adduct) were solvated by a sphere of Montecarlo TIP3P water molecules with the BLOB-SOL option of the module EDIT of the AMBER program (Pearlman et al., 1991) A BLOB thickness of 10 Å was used. The total water molecules was about 3000 for HRP and the HRP–*p*-cresol adduct and 3500 for CcP. The final systems were constituted by about 15 000 atoms for CcP and 13 800 atoms for HRP and the HRP–*p*-cresol adduct.

Molecular Dynamics Calculations. For the simulations in water, we have used similar procedures for all the systems investigated here. A 10-Å residue-based cutoff value for the evaluation of nonbonded interactions was used. The number of pair interactions was about 3.7×10^6 for HRP and its adduct and 4.0×10^6 for CcP. The pair list was updated every 10 steps, and the values of the energies and the coordinates were saved every 100 steps. The time step of the dynamics was 1.5 fs, and a dielectric constant of 1 was chosen. All the water molecules and the sodium ions were equilibrated by minimizing the rms energy gradient to 0.10 kcal mol^{-1} $Å^{-2}$ followed by 9 ps of MD. Then the energy minimization of the whole protein molecule was performed until a rms energy gradient lower than 0.1 kcal mol^{-1} $Å^{-2}$ was obtained. Finally, MD calculations were performed. For CcP, they were done for a total of 142 ps. In the first 9 ps, the system was gradually heated by performing MD runs of 3 ps at the following temperatures: 100, 200, and 300 K. These runs as well as the coordinates of the following 18 ps of dynamics, performed at 300 K, were discarded. The final 88 ps of dynamics were used for structural analysis. In the case of HRP, the total simulation was performed for a much longer time with respect to CcP: 229.5 ps of trajectory was calculated, the last 120 ps of which was used for calculating the MD average structures.

For the HRP–*p*-cresol adduct, 30 ps of MD was calculated, the last 15 ps of which was used for generating the MD average structure.

Data Analysis. The averaged structures, generated with the CARNAL program by Wilson Ross (Ross, 1994), were

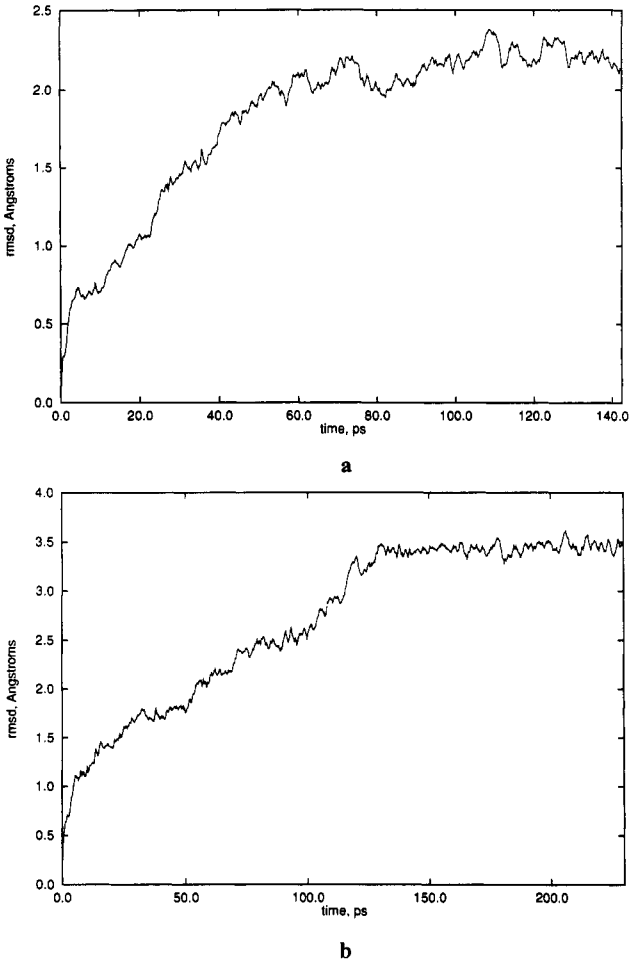


FIGURE 2: rmsd of the (a) CcP and (b) HRP backbone atoms with respect the minimized structure as a function of simulation time.

fully energy minimized. These structures were taken as models of the proteins in solution.

RESULTS

CcP. Figure 1 shows the MD average structure of CcP; Figure 1a reports the backbone of the MD structure while in Figure 1b the active sites of the X-ray and MD structures are compared. It is evident that the overall structure of the backbone is maintained, although the side chains of some residues on the protein surface experience large displacements during the dynamics with respect to the position in the starting structure. Also, the secondary structure features observed in the X-ray structure are maintained in the MD average structure. The presence of secondary motives in the MD average structure of CcP is further confirmed by its Ramachadran plot (Figure 4 of the supplementary material).

The overall protein structure experiences a relatively small deviation with respect to the crystallographic one. The rms deviation of the backbone is 2.1 and 2.2 Å for the heavy atoms. The deviation along the simulation, with respect to the minimized structure, is reported in Figure 2a. It can be noted that during the final 88 ps, used for the generation of the MD structure, the system is stable. The rms fluctuation is 1.1 and 1.2 Å for the backbone and the heavy atoms, respectively. These values indicate that the MD structure does not deviate significantly from the crystallographic one and that the protein has a reasonably small mobility. This is consistent with the presence of an extensive secondary structure with several α helices and β sheets. Indeed, the α helices constitute 50% of the entire protein, and the β sheets are present for 12% (Finzel et al., 1984). These motives are completely maintained in the MD structure and make the protein rigid as evidenced by the small fluctuations during the dynamics.

In the active site, the key residues essentially maintain the conformation they have in the X-ray structure. The proximal histidine (His 175) has a structural arrangement very close to that found from crystallographic studies. The hydrogen bond between this histidine and Asp 235 is maintained in the MD structure, and it stays quite stable along all the trajectory, with an average Oδ1 (Asp 235)–Hδ1 (His 175) distance of 2.0 Å. Trp 191 also has the same conformation in the X-ray and the MD structures, with the same relative orientation with respect to the proximal histidine and the heme ring. It experiences only a small translation from the proximal histidine, but the H bond with Asp 235 is maintained, with an average Oδ2 (Asp 235)–Hε1 (Trp 191) distance of 1.9 Å.

In the distal side, the distal histidine stays in a position close to that in the crystal structure. An extensive hydrogen bond network (Figure 3) includes WAT 595, WAT 596, WAT 648, WAT 348, and WAT 535 (X-ray structure numbering). This H-bond network links residues of the distal side, such as Trp 51 and Arg 48, with the heme propionate chains. WAT 348 is H-bonded to both Arg 48 and two oxygen atoms, one of the 6-propionate and one of the 7-propionate. It constitutes a link between the distal and the proximal sides. This link involves also the H-bond between the 7-propionate chain and the NH proton of His 181, which has its ring roughly coplanar with that of the heme and close to the heme edge bearing the propionate chains. Also this interaction is maintained throughout all the simulation. This extensive H-bond network is present in the crystal and can be catalytically relevant (Finzel et al., 1984). It is believed to be relevant in the determination of the coordination number of the iron and of its spin state, as evidenced by the characterization of several active site mutants (Smulevich, 1993). Interestingly, it is strong enough to be maintained also along all the MD simulation without any constraint on the water molecules. In particular, the water molecule close to the iron is quite rigid despite there being no bond and no constraint placed between these two groups. Due to this extensive H-bond network, the distal side is quite stable and quite rigid. The side chain of Arg 48 in the MD average and the X-ray structures is almost superimposed. Trp 51 also does not change significantly its conformation, even if it experiences a minor translation.

HRP. The structural model, built accordingly to the procedure discussed in the Experimental Section, has been first minimized *in vacuo.* This step produces the removal of some of the bad contacts and of the noncorrect bond lengths present in our starting model. However, the overall structure is not significantly affected. We then slowly warmed up and equilibrated the system *in vacuo* according to the procedure reported in the Experimental Section. These calculations allowed us to obtain a low-energy structure with relatively low computational cost. Then the system, solvated with a large sphere of water molecules, was treated with long MD simulations in order to have an equilibrated structure in solution.

The evolution of the overall HRP structure in water was followed by monitoring the rms deviations as a function of time. In Figure 2b, the rmsd of the backbone atoms is reported. It can be noted that in the last 120 ps the rms deviation is essentially constant within its fluctuations: the value of the rmsd of the backbone atoms with respect to the starting structure in solution is 3.4 Å (standard deviation, 0.1 Å).

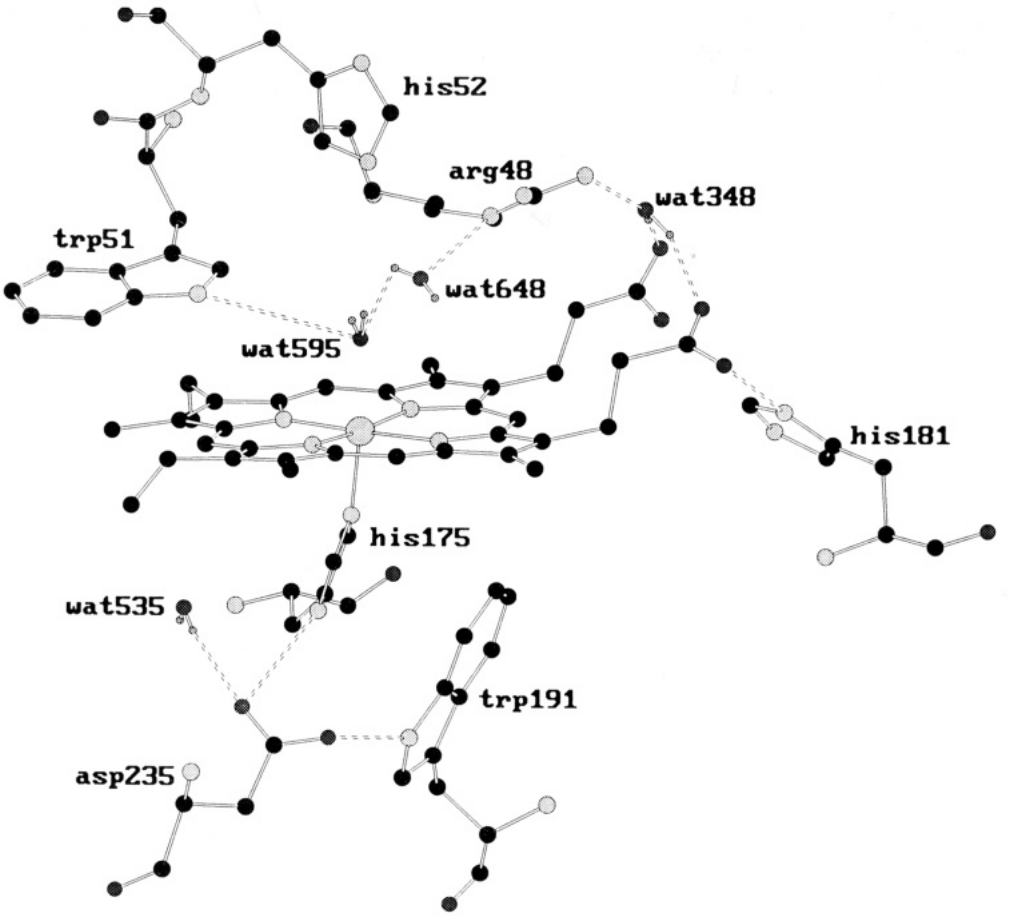


FIGURE 3: Active site of CcP MD average structure. Well-ordered water molecules are included. H-bonds are indicated by dashed lines. WAT 596 is not shown for clarity.

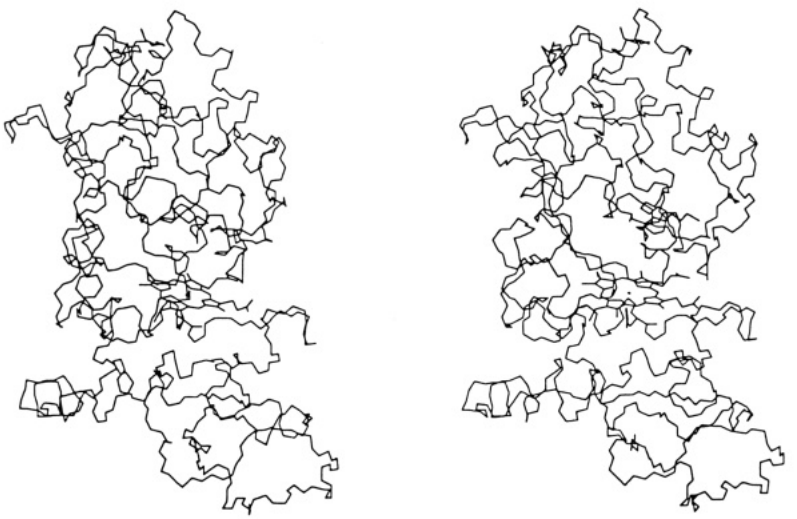


FIGURE 4: Stereo view of the backbone atoms of the MD average structure of HRP. The carbonyl oxygen atoms are omitted for clarity. The heme moiety is also shown.

Therefore, we generated an average MD structure from the last 120 ps of trajectory. This structure, as mentioned above, has been then fully energy minimized. Its Ramachandran plot shows that only few φ–ψ pairs are high in energy (Figure 5 of the supplementary material).

Figure 4 shows the backbone of the protein. As expected, the loops connecting the α helices are substantially changed during the dynamics. However, the secondary motives present in the LiP are almost completely maintained in our model of HRP: nine of the eleven α helices present in LiP are maintained in HRP (residues 9–27, 32–46, 75–87, 98–112, 144–156, 159–172, 244–253, 259–268, and 269–287), whereas the other two helices (residues 58–71 and 195–204) are partially or completely disrupted. This similarity among the structures of peroxidases have already been pointed out (Poulos et al., 1993). The structure comparison suggests that the overall fold will be the same for all peroxidases.

Figure 5 compares the active sites of the MD average structures of HRP with that of the MD average structure of CcP (a) and that of the X-ray structure of LiP (b). Beside the two histidines which are constrained, we can see that also most of the residues have a position similar to what they have

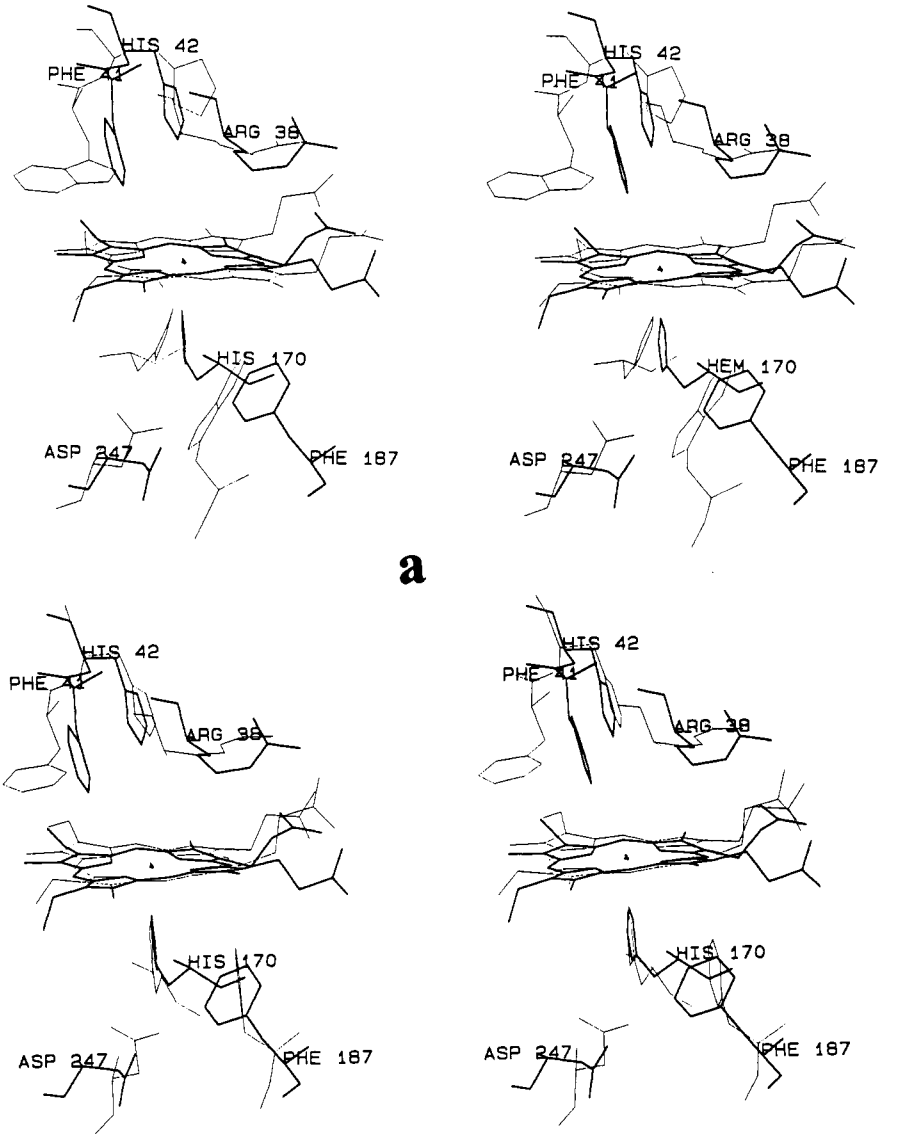


FIGURE 5: Comparison of the active site of HRP MD average structure (bold) and those of CcP MD average structure (a) and LiP X-ray structure (b) (thin).

in the other two peroxidases. The H-bond between the proximal histidine and the proximal Asp 230 residue is maintained, and it is stable along the trajectory, even if Asp 230 changes its conformation. Phe 187, which in HRP replaces Trp191 of CcP, experiences a rotation of the ring that is no more parallel to the ring of the proximal histidine (but see Added in Proof below). In the distal side, the long side chain of the Arg residue is in a position almost identical to that in CcP and LiP while the plane of Phe 41 undergoes a large movement.

The water molecule found between His 42 and the iron atom is well ordered and maintains its position during all trajectory despite the lack of any constraint (Figure 6). The distal cavity experiences other well-ordered water molecules and H-bond interactions, as observed in the active site of LiP and CcP. These water molecules form, through H-bonds, a bridge between the guanidinum group of Arg 38 and the two propionate residues. Other water molecules, which experience large movements during the MD simulation, are also present in the heme pocket. In the final step structure, 12 water molecules are present in a 9-Å sphere centered on the iron. The carboxylate of the 7-propionate, which is H-bonded to a His residue in CcP and to an Asp residue in LiP, is interacting in this HRP model with the side chain amide nitrogen of an Asn residue (Asn 175).

The structural features of the present MD structure of HRP are in agreement with the detailed NMR characterization performed on this protein (Thanabal et al., 1987a,b, 1988). As an example, we can consider the NOE connectivity detected between 8-CH_3 of the heme and a Ile residue that is still unidentified (de Ropp et al., 1991b). The signals of the latter residue were earlier assigned as due to a Leu residue (Thanabal et al., 1987a), based on the CcP structure and the comparison of the sequences. From inspection of the present HRP model, we find that the side chain of Ile 180 is pointing toward the 8-CH_3 and is located at an interacting distance (but see Added in Proof below). The above Leu results very far from the active site. This is just an example of the usefulness of the present structural model for the interpretation of experimental data. The coordinate file of the present structural model of HRP is available from the corresponding author.

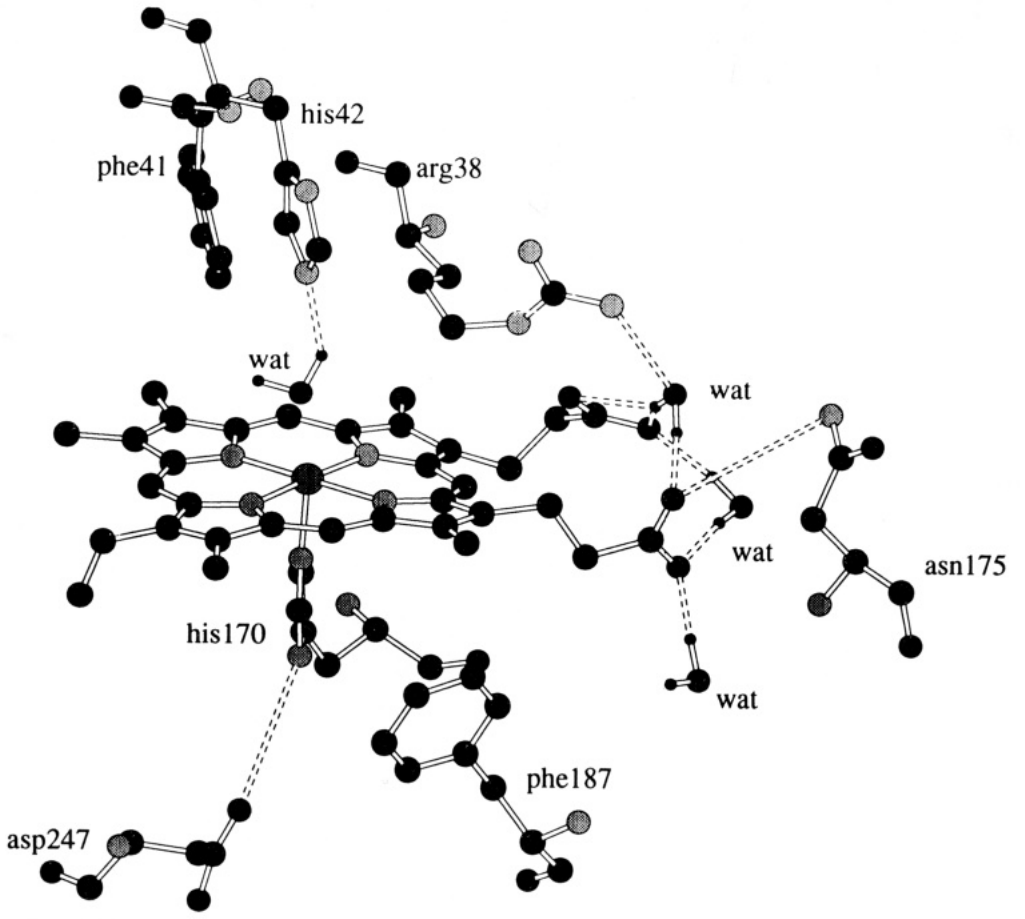


FIGURE 6: Active sites of HRP MD average structure. Well-ordered water molecules are included. H-bonds are indicated by dashed lines.

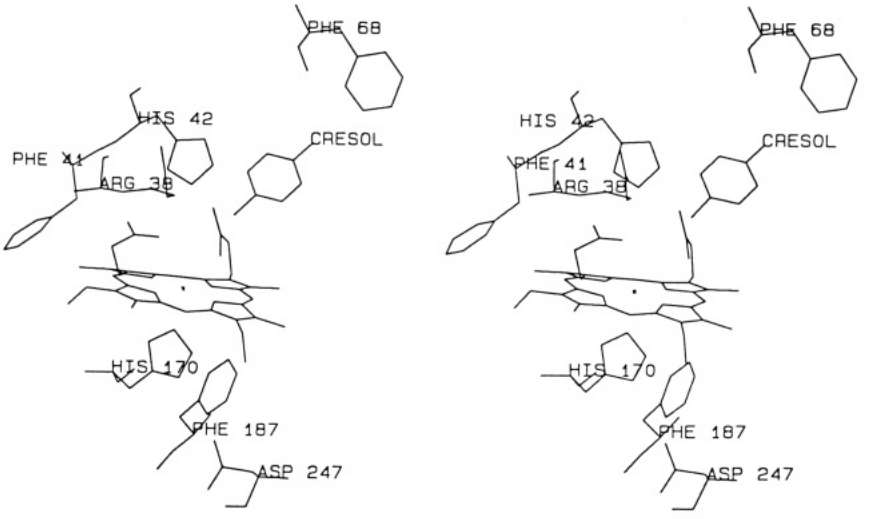


FIGURE 7: Stereo view of the active site of the HRP–*p*-cresol adduct MD average structure.

HRP–p-Cresol Substrate. As mentioned in the Experimental Section, *p*-cresol can be docked in two possible positions to satisfy the NMR data, close to Phe 68 or close to Phe 187 (but see Added in Proof below). However, in the latter model, after only 4 ps of dynamics, the substrate experiences large movements (about 8 Å) and migrates close to Phe 68. For the rest of the trajectory, it maintains this position. This clearly indicates that only the position close to Phe 68 is a stable location for the *p*-cresol molecule. As expected, when MD calculations are performed on the other model, i.e., *p*-cresol close to Phe 68, no major rearrangements of the substrate occur.

It should be pointed out that the movements of molecules detected through MD calculations are usually over smaller distances than those observed in this system. Thus, this finding could be an indication of how different the energy of the system is when the aromatic molecule is placed in the two different positions.

The overall backbone structure of the *p*-cresol adduct is not affected by the substrate binding. The active site substantially maintains the original structure even if, however, some rearrangements occur (Figure 7). The proximal and distal histidines stay in their original positions. In the distal side, some residue rearrangements, even if not very large, occur in

order to accommodate the aromatic molecule. The distal Arg has it side chain in a position very close to the starting one, while Phe 41 experiences a small movement of the ring. In the proximal site, Asp 230 still forms a hydrogen bond between its carboxylate and the Nδ1 of the proximal His residue.

The *p*-cresol molecule is located between the heme side with pyrrole IV and Phe 68. It interacts with the 8-CH_3 group and the ϵ protons of the Phe 68 ring, consistent with the NMR results and the NOE data (Banci et al., 1993; Veitch et al., 1992). The aromatic ring of *p*-cresol is almost perpendicular to the ring of Phe 68 and maintains this orientation along the trajectory, in a well-oriented conformation.

The extensive H-bond network among the well-ordered water molecules and protein residues in the distal cavity, observed in HRP, is not broken, even though some of the ordered water molecules experience some displacements.

DISCUSSION

The present MD calculations have, first of all, allowed the characterization of the solution structure of CcP and the study of its dynamical properties. The structure of this protein is fairly stable also in solution and shows a very low fluxionality. A relevant result is also that the extensive H-bond network among water molecules and protein residues around the heme active site is stable along all the simulation. This suggests that these interactions are relatively strong and are not broken even in solution.

The success in the application of MD simulations to CcP has made us confident in further applying these calculations to other peroxidases for which the X-ray structure is not yet available. Particularly, we have produced a structural model for HRP on which we performed energy minimization and then molecular dynamics calculations for a time long enough to reach the equilibration of the system. We are conscious that we are dealing with a "model" of the protein and not with its real structure and, therefore, that our model could deviate even sizably from a possible future X-ray structure. However, we would like to stress that (1) the present model is in agreement with *all* the experimental data available from a large variety of measurements (mainly NMR), with particular emphasis on the properties of the protein region around the heme where the catalytic reaction occurs; (2) a reasonable model can be quite useful for the interpretation of several experimental data.

For our calculations, we have taken advantage of a good starting model and a correct amino acid sequence alignment with a protein with known structure. Furthermore, in order to construct a correct HRP model, we have included in our computer modeling procedure information of valuable importance on several structural properties of the protein obtained through a variety of experimental evidences, such as circular dichroism and S–S bridge analysis.

As an example of the utility of such a structural model, we have reported here the simulation of a substrate–adduct of HRP. The knowledge of the binding mode of substrates in peroxidases is of great relevance for the comprehension of the overall enzymatic process in this class of enzymes. In the case of HRP, the aromatic molecules, which are substrates for the enzyme, were found through the present calculations to be located close to a heme edge and to interact with Phe 68. This binding site represents a local minimum in energy, as the aromatic molecules reaches this position even if it is initially located very far and then maintains its position for the rest of the MD simulation. This position is in agreement with the NMR data (Banci et al., 1993; Veitch et al., 1992), which show dipolar interactions between *p*-cresol protons and protons of the heme as well as of a Phe residue. The present MD calculations suggest that the hydrophobic interactions between the substrate and the protein residues in the binding site are strong enough to pull the aromatic molecule away from other possible positions and maintain it in a fixed position.

The MD calculations performed during the present study have allowed a deep structural characterization of CcP and particularly of HRP and its adducts. The latter results are quite meaningful due to the lack of X-ray structure of HRP and turn out to be relevant for the comprehension of the biological function of these systems.

ADDED IN PROOF

After a long discussion with Prof. Karen Welinder, who is warmly acknowledged, it appears that, if we use the alignment of Figure 2b in Welinder (1992), the side chain of Ile 244 instead of Ile 180 could be pointing toward 8-CH_3, and Phe 187 could be replaced by Phe 221. The arrangement of the residues in the proximal site would be, however, quite similar. The distal site would be not affected.

ACKNOWLEDGMENT

We are grateful to Prof. Ivano Bertini for the continuous encouragement and precious suggestions. We thank Prof. Peter Kollman and Prof. Giulietta Smulevich for helpful comments and discussion. We thank IBM Semea for the computer time given to us on the IBM/SP1 machine at IBM ECSEC; Rome, Italy. Dr. Orio Tomagnini is acknowledged for his technical assistance in running the parallel version of AMBER at IBM ECSEC.

SUPPLEMENTARY MATERIAL AVAILABLE

Five figures showing Ramachandran plots of HRP and rms deviations (6 pages). Ordering information given on any current masthead page.

REFERENCES

Angelucci, L., De Gioia, L., & Fantucci, P. (1993) *Gazz. Chim. Ital. 123*, 111.

Argos, P., & Mathews, F. S. (1975) *J. Biol. Chem. 250*, 747.

Ator, M. A., & Ortiz de Montellano, P. R. (1987) *J. Biol. Chem. 262*, 1542.

Ator, M. A., David, S. K., & Ortiz de Montellano, P. R. (1987) *J. Biol. Chem. 262*, 14954.

Banci, L., Bertini, L., Turano, P., Ferrer, J. C., & Mauk, A. G. (1991a) *Inorg. Chem. 30*, 4510.

Banci, L., Bertini, I., Turano, P., Tien, M., & Kirk, T. K. (1991b) *Proc. Natl. Acad. Sci. U.S.A. 88*, 6956.

Banci, L., Bertini, I., Pease, E., Tien, M., & Turano, P. (1992a) *Biochemistry 31*, 10009.

Banci, L., Schroeder, S., & Kollman, P. A. (1992b) *Proteins: Struct. Funct. Genet. 13*, 288.

Banci, L., Bertini, I., Bini, T., Tien, M., & Turano, P. (1993) *Biochemistry 32*, 5825.

Berendsen, H. J. C., Postma, J. P. M., van Gunsteren, W. F., DiNola, A., & Haak, J. R. (1984) *J. Chem. Phys. 81*, 3684.

Bernstein, F. C., Koetzle, T. F., Williams, G. J. B., Meyer, E. F., Jr., Rodgers, J. R., Kennard, O., Shimanouchi, T., & Tasumi, M. (1977) *J. Mol. Biol. 112*, 535.

Bosshard, H. R., Anni, H., & Yonetani, T. (1991) in *Peroxidases in Chemistry and Biology* (Everse, J., Everse, K. E., & Grisham, M. B., Eds.) pp 51–85. CRC Press, Boca Raton, FL.

Brooks, C. L., III, Karplus, M., & Petitt, B. M. (1988) *Proteins: a perspective of dynamics, structure and thermodynamics*, John Wiley and Sons, New York.
Case, D. A., & Karplus, M. (1978) *J. Mol. Biol. 132*, 343.
Casella, L., Gullotti, M., Poli, S., Bonfá, M., Ferrari, R. P., & Marchesini, A. (1991) *Biochem. J. 279*, 245.
Chin, S. (1992) *AMBER 4.0*, Parallel version, IBM Corporation, Kingston, NY.
Collins, J. R., Du, P., & Loew, G. H. (1992) *Biochemistry 31*, 11166.
Dagget, V., & Kollman, P. A. (1990) *Protein Eng. 3*, 677.
Dawson, J. H. (1988) *Science 240*, 433.
Dawson, J. H., Andersson, L. A., & Sono, M. (1982) *J. Biol. Chem. 257*, 3606.
de Ropp, J. S., La Mar, G. N., Wariishi, H., & Gold, M. H. (1991a) *J. Biol. Chem. 266*, 15001.
de Ropp, J. S., Yu, L. P., & La Mar, G. N. (1991b) *J. Biomol. NMR 1*, 175.
Dunford, H. B. (1982) *Adv. Inorg. Chem. 4*, 41.
Dunford, H. B. (1991) in *Peroxidases in Chemistry and Biology* (Everse, J., Everse, K. E., & Grisham, M. B., Eds.) pp 1–23. CRC Press, Boca Raton, FL.
Edwards, S. L., Raag, R., Wariishi, H., Gold, M. H., & Poulos, T. L. (1993) *Proc. Natl. Acad. Sci. U.S.A. 90*, 750.
Fee, J. A., & DiCorleto, P. E. (1973) *Biochemistry 12*, 4893.
Ferguson, D. M., Radmer, R. J., & Kollman, P. A. (1991) *J. Med. Chem. 34*, 2654.
Finzel, B. C., Poulos, T. L., & Kraut, J. (1984) *J. Biol. Chem. 259*, 13027.
Jones, T. A. (1992) in *Computational Crystallography* (Sayre, D., Ed.) pp 303–317, Claredon Press, Oxford.
Kaptein, R., Boelens, R., Scheek, R. M., & van Gunsteren, W. F. (1988) *Biochemistry 27*, 5389.
Kirk, T. K., & Farrell, R. L. (1987) *Annu. Rev. Microbiol. 41*, 465.
Kuriyan, J., Wiltz, S., Karplus, M., & Petsko, G. A. (1986) *J. Mol. Biol. 192*, 133.
La Mar, G. N., & de Ropp, J. S. (1979) *Biochem. Biophys. Res. Commun. 90*, 36.
La Mar, G. N., de Ropp, J. S., Smith, K. M., & Langry, K. C. (1980) *J. Biol. Chem. 255*, 6646.
La Mar, G. N., Thanabal, V., Johnson, R. D., Smith, K. M., & Parish, D. W. (1989) *J. Biol. Chem. 264*, 5428.
Mabbutt, B. C., & Wright, P. E. (1985) *Biochim. Biophys. Acta 832*, 175.
McCammon, J. A., & Harvey, S. (1987) *Dynamics of proteins and nucleic acids*, Cambridge University Press, Cambridge.
Merz, K. M. J., Murcko, M. A., & Kollman, P. A. (1991) *J. Am. Chem. Soc. 113*, 4484.
Morishima, I., & Ogawa, S. (1979) *J. Biol. Chem. 254*, 2814.
Ortiz de Montellano, P. R., David, S. K., Ator, M. A., & Tew, D. (1988) *Biochemistry 27*, 5470.
Pearlman, D. A., Case, D. A., Caldwell, G. C., Siebel, G. L., Singh, U. C., Weiner, P., & Kollman, P. A. (1991) *AMBER 4.0*, University of California, San Francisco, CA.
Pelletier, H., & Kraut, J. (1992) *Science 258*, 1748.
Piontek, K., Glumoff, T., & Winterhalter, K. (1993) *FEBS Lett. 315*, 119.
Poulos, T. L., & Kraut, J. (1980a) *J. Biol. Chem. 255*, 8199.
Poulos, T. L., & Kraut, J. (1980b) *J. Biol. Chem. 255*, 10322.
Poulos, T. L., Edwards, S. L., Wariishi, H., & Gold, M. H. (1993) *J. Biol. Chem. 268*, 4429.
Rogers, K. K., Pochapsky, T. C., & Sligar, S. G. (1988) *Science 240*, 1657.
Ross, W. S. (1994) *CARNAL*, Department of Pharmaceutical Chemistry, University of California, San Francisco, CA.
Ruf, H. H., Wende, P., & Ullrich, V. (1979) *J. Inorg. Biochem. 11*, 189.
Scheek, R. M., van Gunsteren, W. F., & Kaptein, R. (1989) *Methods Enzymol. 177*, 204.
Smulevich, G. (1993) in *Biomolecular Spectroscopy, Part A, Vol. 20, Advances in Spectroscopy* (Clark, R. J. H., & Hester, R. E., Eds.) pp 163–193, John Wiley and Sons, London.
SYBYL: Molecular Modeling Software, Version 5.2 (1989) Tripos Associates, Inc., St. Louis, MO.
Strickland, E. H., Kay, E., Shannon, M. L., & Horwitz, J. (1968) *J. Biol. Chem. 243*, 3560.
Thanabal, V., de Ropp, J. S., & La Mar, G. N. (1987a) *J. Am. Chem. Soc. 109*, 7516.
Thanabal, V., de Ropp, J. S., & La Mar, G. N. (1987b) *J. Am. Chem. Soc. 109*, 265.
Thanabal, V., de Ropp, J. S., & La Mar, G. N. (1988) *J. Am. Chem. Soc. 110*, 3027.
Tien, M. (1987) *CRC Crit. Rev. Biochem. 15*, 141.
Tien, M., & Kirk, T. K. (1983) *Science 221*, 661.
van Gunsteren, W. F., & Berendsen, H. J. C. (1977) *Mol. Phys. 34*, 1311.
Veitch, N. C., Williams, R. J. P., Smith, A. T., Sanders, R. N. F., Thornley, R. N. F., Bray, R. C., & Burke, J. F. (1992) *Biochem. Soc. Trans. 20*, 114S.
von Bobman, S. B., Schulder, M. A., Jollie, D. R., & Sligar, S. G. (1986) *Proc. Natl. Acad. Sci. U.S.A. 83*, 9443.
Wang, J., Mauro, J. M., Edwards, S. L., Oatley, S. J., Fishel, L. A., Ashford, V. A., Xuong, N., & Kraut, J. (1990) *Biochemistry 29*, 7160.
Weiner, S. J., Kollman, P. A., Case, D. A., Singh, U. C., Ghio, C., Alagona, G., & Profeta, S., Jr. (1984) *J. Am. Chem. Soc. 106*, 765.
Weiner, S. J., Kollman, P. A., Nguyen, D. T., & Case, D. A. (1986) *J. Comput. Chem. 7*, 287.
Welinder, K. G. (1985) *Eur. J. Biochem. 151*, 497.
Welinder, K. G. (1992) *Curr. Opin. Struct. Biol. 2*, 388.