

# A Density Functional Theory Study of Conformers in the Ferrous CO Complex of Horseradish Peroxidase with Distinct Fe–C–O Configurations

Sergio D. Dalosto,\* Ninad V. Prabhu, Jane M. Vanderkooi, and Kim A. Sharp

Johnson Research Foundation, Department of Biochemistry & Biophysics, School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania 19104

Received: September 6, 2002; In Final Form: December 19, 2002

The reactive site structure and CO IR spectral signature of horseradish peroxidase complexed with CO were studied using a combination of density functional theory and molecular modeling. The quantum chemical model used included the proximal His170, the distal amino acids His42 and Arg38, and the crystallographic water molecule that interacts with CO, His42, and Arg38. The mutant H42L was simulated by removing distal histidine from the cluster and the mutant R38L by removing the arginine. In the models where the Arg38 is present a hydrogen bond interaction is found between the H<sub>ε</sub> of Arg and the oxygen of adduct –CO. When both His42 and Arg38 are present they interact with the CO moiety, but they also interact with each other so their effect is not simply additive. The effect of a change in pH was modeled by removing potentially titratable hydrogen atoms identified by molecular mechanics and Poisson–Boltzmann pK<sub>a</sub> calculations on the active site. Previously observed IR spectral features of CO as a function of pH and active site mutations could be interpreted primarily in terms of changes in CO and Fe–CO bond lengths produced by changes in active site composition and charge state. The effect of the charge relay due to the possible second deprotonation (assuming that the first deprotonation took place in the His42) from the Arg38 or H<sub>δ</sub> in the proximal His170 gives a decrease of the CO stretching frequency.

## Introduction

There is considerable interest in the use of the CO group as an IR probe to explore the environment and the structure of specific sites in proteins. Experimentally, the CO stretch frequency is found in an uncongested region of the spectrum and its oscillator strength is large. The prototype for using CO in this manner is the myoglobin–CO complex. The stretching frequency of the CO moiety when bound to heme iron in deoxymyoglobin (myoglobin–CO) has been extensively studied. The heme–CO bond in myoglobin–CO is quite robust, which is reflected in the strong affinity of CO for binding. At the same time, both the CO stretch frequency and the strength of CO···His64 interaction are very dependent on the environment, including details of the orientation and tautomerization state of the distal histidine residue His64.<sup>1,2</sup> Here we analyze the CO probe methodology applied to horseradish peroxidase type C (HRPC). HRPC is an enzyme that catalyzes the oxidation of a variety of organic compounds by hydrogen peroxidase and is currently a model protein for development of ideas on redox catalysis. The structure of this protein has been determined for the ferric molecule<sup>3</sup> (1ATJ in the Protein Data Bank), and the heme is significantly more buried than it is in myoglobin. The crystal structure shows that HRPC has two water molecules in the active site, which interact with the ferric Fe<sup>3+</sup>, Arg38, and His42 in the distal side. The water bound directly to the Fe<sup>3+</sup> occupies the CO binding site. Unlike myoglobin, HRPC has an arginine in the distal side. This Arg appears to serve to stabilize the peroxy complex.<sup>4</sup>

HRPC is an attractive model to study the effect of the protein on the regulation of heme functions since multiple types of

conformers in the ferrous CO complex of (HRPC) have been identified with the use of native and deuterio-heme-substituted enzymes and site-directed mutagenesis.<sup>5</sup> Depending on the pH, the populations of these conformers change. At alkaline pH (~9), one CO stretch band is observed at  $\nu_{\text{CO}} = 1934 \text{ cm}^{-1}$  and the Fe–CO stretching ( $\nu_{\text{Fe–CO}}$ ) Raman line is present at  $530 \text{ cm}^{-1}$ . At pH = 6.0 there are two bands, a major band at low frequency,  $\nu_{\text{CO}} = 1905 \text{ cm}^{-1}$  with  $\nu_{\text{Fe–CO}} = 539 \text{ cm}^{-1}$  and a second minor peak at  $\nu_{\text{CO}} = 1934 \text{ cm}^{-1}$ , with an associated Raman band at  $\nu_{\text{Fe–CO}} = 516 \text{ cm}^{-1}$ . At acid pH (~4) there are two species with bands at  $\nu_{\text{CO}} = 1905$  and  $1934 \text{ cm}^{-1}$ ; the corresponding Raman frequencies are  $\nu_{\text{Fe–CO}} = 539$  and  $524 \text{ cm}^{-1}$ , respectively. There is also a broad band at  $\nu_{\text{CO}} = 1967 \text{ cm}^{-1}$  with  $\nu_{\text{Fe–CO}} = 492 \text{ cm}^{-1}$ . The interconversion between conformers is accomplished with a change in the pH; the pK<sub>a</sub> is 8.7.<sup>6,7</sup>

The Raman spectra<sup>5,8</sup> of the CO complex of the site-mutated derivative of the recombinant HRPC His42Leu mutant have bands at  $\nu_{\text{CO}} = 1924 \text{ cm}^{-1}$  and  $\nu_{\text{Fe–CO}} = 525 \text{ cm}^{-1}$  between pH 3.3–10.0.<sup>7</sup> The Arg38Leu mutant has  $\nu_{\text{CO}}$  bands at a higher frequency than the wild-type enzyme; at pH 6 the CO protein has bands at  $\nu_{\text{CO}} = 1941.5 \text{ cm}^{-1}$  and  $\nu_{\text{Fe–CO}} = 515 \text{ cm}^{-1}$ ; at alkaline pH these values are  $\nu_{\text{CO}} = 1944 \text{ cm}^{-1}$  and  $\nu_{\text{Fe–CO}} = 496 \text{ cm}^{-1}$ . The band at  $1905 \text{ cm}^{-1}$  and its Raman associated band at  $539 \text{ cm}^{-1}$  disappear when Arg38 is substituted by leucine.

The polar and hydrogen bond interactions of the amino acids His42 and Arg38 with CO, the protonation of the heme propionic acid,<sup>5</sup> changes in the protonation of the proximal His170,<sup>6</sup> and a slow conformational transition of the heme<sup>9</sup> have all been suggested as possible explanations for the different observed IR and Raman bands. While these explanations are not mutually exclusive, a definitive explanation of major spectral

\* Corresponding author. Phone (215) 573-3507, fax (215) 898-4217, e-mail sdalosto@mail.med.upenn.edu

**TABLE 1: Chemical Models Used to Study the Stretching and Bending Frequencies of CO and F-CO<sup>a</sup>**

model <sup>b</sup>	name <sup>b</sup>	His42	Arg38	H <sub>2</sub> O	Q
1	P-CO-His170				0
2a*	His42N <sub>δ</sub> H <sub>δ</sub> N <sub>ε</sub> :	x			0
2b*	His42N <sub>δ</sub> H <sub>δ</sub> N <sub>ε</sub> :	x			0
2c	His42N <sub>δ</sub> H <sub>δ</sub> N <sub>ε</sub> H <sub>ε</sub>	x			+1
2d	His42N <sub>ε</sub> H <sub>δ</sub> N <sub>δ</sub> :	x			0
3	Arg		x		+1
4	Arg + H <sub>2</sub> O	x	x	x	+1
5a	Arg + His42N <sub>δ</sub> H <sub>δ</sub> N <sub>ε</sub> H <sub>ε</sub>	x	x		+2
5b	Arg + His42N <sub>δ</sub> H <sub>δ</sub> N <sub>ε</sub> :	x	x		+1
5c	Arg + His42N <sub>δ</sub> H <sub>δ</sub> N <sub>ε</sub> H <sub>ε</sub> + H <sub>2</sub> O	x	x	x	+2
5d	Arg + His42N <sub>δ</sub> H <sub>δ</sub> N <sub>ε</sub> : + H <sub>2</sub> O	x	x	x	+1
6	Arg <sup>neutral</sup> + His42N <sub>δ</sub> H <sub>δ</sub> N <sub>ε</sub> :	x	x		0
7	P-CO-His170N <sub>δ</sub> :				-1
8	Arg + His42N <sub>δ</sub> H <sub>δ</sub> N <sub>ε</sub> : -His170N <sub>δ</sub> :	x	x		0
9	HRPC-CO				0

<sup>a</sup> The basic components present in each model are the porphine (P) and the proximal histidine, His170. *Q* indicates the total charge of the corresponding model. <sup>b</sup> The symbol (:) indicates lone-pair orbital and the symbol \* indicates different constraints.

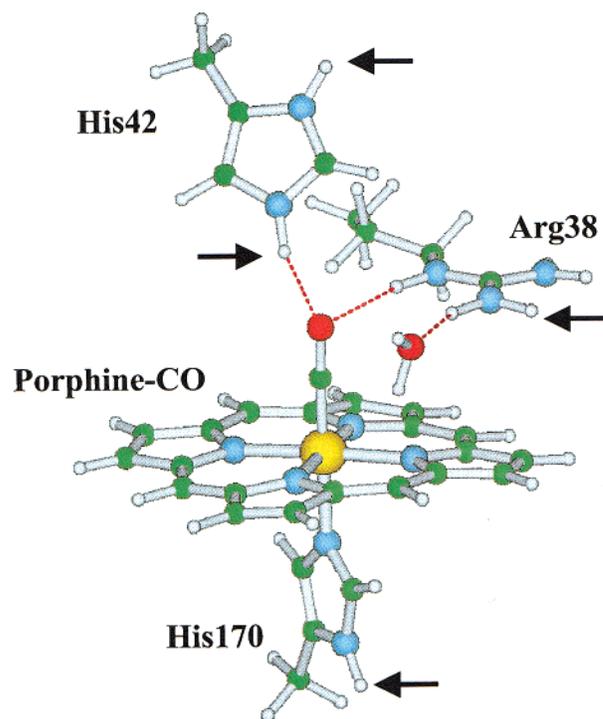
features is still lacking. This compromises the use of CO as a probe of this protein's active site environment. The purpose of this paper is to better understand the factors that affect the bound CO, in particular the effects of the presence of distal amino acids on CO bond lengths and interpretation of observed IR bands. To do this we have carried out a density functional theory (DFT) study of the bound CO in various environments by modeling changes in the protonation states of the His42, Arg38, and His170 with different clusters. We have also examined the nature of the interaction between His42, Arg38, and the CO adduct using molecular dynamics simulations of the entire protein in a continuum solvent environment. In addition we have computed the p*K*<sub>a</sub> of these residues using the Poisson-Boltzmann electrostatics model to evaluate candidates for pH-dependent deprotonation.

Our long-term goal is to evaluate the role of the protein forming the cavity surrounding the heme in HRP using methods that combine self-consistent quantum mechanics and molecular mechanics. A necessary preliminary step is the detailed quantum mechanical study of the active site described here. The DFT technique gives positions of atoms, and we emphasize that these positions change depending upon the conditions selected for the cluster.

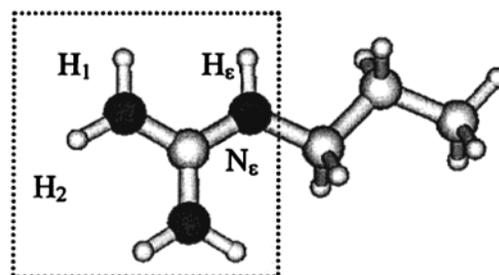
## Computational Methods

**Density Functional Theory Calculations.** DFT calculations using Becke's three-parameter hybrid functional with the correlation functional of Lee, Yang, and Parr (B3LYP)<sup>10</sup> were performed using the Gaussian 98 program package.<sup>11</sup> All structures were optimized with the 6-31G basis set. Due to computational limitations we have not used basis sets with polarization or diffuse functions during the relaxation. The geometries of all models studied were optimized until the total energy fell below a threshold of 0.0002 au. The frequencies were calculated at the same level of theory. Also, model 1 (models are described below) was optimized with TZV of Ahlrichs basis.<sup>12</sup> The oxidation state of HRP when -CO is coordinated to Fe is ferrous and the spin state is *S* = 0. The geometry of each model, Fe-CO, C-O stretching, and the Fe-C-O bending frequencies, were identified using the program MOLLEN.<sup>13</sup>

Table 1 summarizes the components and nomenclature of the different clusters and their total charge. The structures of these



**Figure 1.** Model showing the components of the various clusters used to model the active site of HRP-CO. The Figure shows in the distal side the His42, Arg38, and a water molecule and in the proximal side, the His170. The arrows indicate the positions that can undergo protonation/deprotonation. The figure is the relaxed structure of model 5c.



**Figure 2.** Model molecule for the arginine used in the DFT calculations. The guanidyl group is showed inside the dotted square, we named for convenience their hydrogens. H<sub>ε</sub> interacts with the carbonyl oxygen and the hydrogen H<sub>1</sub> with a water molecule. H<sub>2</sub> is a high probable exchangeable proton (see text for more details). In model 6, the hydrogen H<sub>2</sub> was removed giving a neutral arginine.

models are displayed in Figure 1. Arrows on the figure indicate the positions at which pH-dependent protonation/deprotonation could conceivably occur. To represent the different moieties that compose the active site clusters within the DFT calculations, truncated and capped analogues were used, as follows: The heme was modeled as Porphine, the distal and proximal histidine by Im-CH<sub>3</sub> (Im means Imidazole) and arginine by its side chain guanidinium group and the delta methyl groups (NH<sub>2</sub>)<sub>2</sub>C-NH-CH<sub>2</sub>-CH<sub>3</sub> (Figure 2 identifies the relevant atoms of arginine). Hereafter when we speak about histidine or arginine in regard to the DFT calculations we refer to their analogues.

Histidine possesses one (H<sub>δ</sub> in His170) or two (H<sub>ε</sub> and H<sub>δ</sub> in His42) labile ring protons, depending on the state of protonation, and thus His170 can have charge zero or -1 while His42 can have charge +1, 0, or -1. In most of the models where the distal histidine is present, the Fe-methyl terminal distance of His42 was kept fixed to the 1ATJ value 8.89 Å. The exceptions

are the models 2a, c, and d. In models 5c and d we added a restriction in the dihedral angle  $\text{Fe}^{\text{---methyl}}\text{C}^{\text{---his}}\text{C}^{\text{---N}}_{\delta} = 129.0^{\circ}$ .

Additionally, mutants such as His42Leu and Arg38Leu are approximated by deleting the mutated residue from the cluster. The zero total charge and nonpolar nature of leucine is expected to have a negligible effect on the CO orbitals and hence its frequency. Each model cluster is described in the results section.

**Molecular Dynamics.** Since it is currently impractical to do high quality DFT calculations on an entire protein, molecular dynamics (MD) simulations were used to guide the generation of the initial cluster geometry for the DFT calculations, to check the consistency of geometries produced by the DFT calculations with the structure of the complete protein present, and to include the effect of the dynamics of the protein atoms. These simulations were also used to generate energetically reasonable structures as input to the Poisson–Boltzmann  $pK_a$  calculations. The force field for the molecular dynamics simulations of HRPc was the CHARMM 22 all atom parameter set,<sup>14</sup> and simulations were run using the CHARMM 27 program suite,<sup>15</sup> with 1ATJ as the starting structure. To model a system of the size of HRPc in a box of explicit solvent with periodic boundary conditions is computationally demanding (the total system size would be greater than 40,000 atoms). Hence, the solvent was modeled as a continuum by incorporating the solvent contribution to the free energy and forces from a high-speed Poisson–Boltzmann solver<sup>16</sup> implemented as the ZAP C++ toolkit (OpenEye Software, Sante Fe). The ZAP toolkit was interfaced with the CHARMM MD code. For the continuum solvent model, a dielectric of 80 was used. The apolar contribution to energies and forces from solvent was obtained using the solvent accessible surface area model, with a hydrophobic coefficient of 6 cal/mol/Å<sup>2</sup>. Surface areas and derivatives are also obtained from ZAP. Continuum solvent forces were updated every 10 fs. All alpha carbons and the two Ca<sup>2+</sup> ions were restrained by a force of 24 kcal/mol/Å<sup>2</sup> to ensure the overall stability of the protein structure during the course of the simulations, and all bond lengths were constrained by SHAKE. Comparison of this protocol (with ZAP and similar constraints) with an explicit water MD simulation of a smaller protein calmodulin (with constraints) will be published elsewhere (Prabhu, N. V.; Sharp, K. A., in preparation).

Three simulations are presented here, and in each the water molecule bound to the heme Fe in 1ATJ is replaced by CO. One simulation has positively charged His42 with a hydrogen attached to the nitrogen N<sub>ε</sub>. The second simulation has His42 neutral, and the third has a neutral His42 with Arg38 replaced by Leu. All structures were first minimized and, after short heating and equilibration, were run for 500 ps. The purpose of these simulations, given the approximations employed, is limited to the examination of side chain motions in the interior of the protein that are not accompanied by large main-chain rearrangements. These still give interesting information about the dynamics of the protein that may be compared with experiment and quantum mechanical calculations.

**$pK_a$  Shifts.** Bulk solution properties such as pH cannot currently be incorporated directly into DFT calculations, which are restricted to a relatively small number of atoms. However, the effect of a change in pH can be incorporated by using the correct protonation state of each residue as input to the DFT calculations. This requires identification of all titratable protons, along with a consistent set of  $pK_a$ 's for them.  $pK_a$  calculations were performed using finite difference Poisson–Boltzmann (FDPB) calculations on the protein as described previously.<sup>17,18</sup> The methodology in principle allows the self-consistent calcula-

tion of the entire set of ionizable groups in the protein; however, for a large protein such as HRPc the calculation is intensive. For computational tractability, it was assumed that all residues outside the active site have unperturbed  $pK_a$ 's. Then the  $pK_a$  shifts of residues His42, Arg38, and His170 due to electrostatic effects were obtained from the FDPB calculations using

$$pK_a = pK_a^0 + \left( \frac{1}{2.303kT} \right) \left( \frac{1}{2} \sum_i (q_{i,p} \Delta\phi_{i,p}^{\text{rxn}} - q_{i,u} \Delta\phi_{i,u}^{\text{rxn}}) + \sum_i (q_{i,p} - q_{i,u}) \Delta\phi_i^{\text{int}} \right)$$

where  $pK_a^0$  is the reference state (isolated in water)  $pK_a$  for that residue, and the sum is over all the residue charges. The subscripts u and p refer to the unprotonated and protonated states of the residue, respectively, and  $\Delta\phi^{\text{rxn}}$  and  $\Delta\phi^{\text{int}}$  are the changes in the reaction potential and charge–charge interaction potential at the residue upon transferring it from the reference state to the protein.<sup>17,18</sup> In removing a proton of His and Arg, there is a choice of two and five protons, respectively. FDPB calculations were performed for each case, and the deprotonation that resulted in the lowest electrostatic energy was used.<sup>19</sup> The electrostatic reaction potential and charge–charge interaction potential were calculated as described previously<sup>19</sup> using the FDPB method implemented in the software package DelPhi.<sup>20–22</sup> Parameters used in the FDPB calculations were as follows: grid dimensions were 161 in each direction, with a scale of about 2 grids/Å. Solutions were obtained for the linear Poisson–Boltzmann equation with Debye–Huckel type boundary conditions<sup>20</sup> using the multigridding iteration to a final convergence value of  $1 \times 10^{-4}$  kT/e total residual error in the potential. The solvent and protein dielectric constants used were 80 and 4, respectively. The CHARMM param22 atomic charges were used for the protein, the heme, and carbon monoxide.<sup>23</sup>

## DFT Models and Structural Results

We now describe each of the models. The calculated C–O ( $d_{\text{CO}}$ ), Fe–C ( $d_{\text{Fe–CO}}$ ) and other distances are given in Table 2. In the Supporting Information, we present the final structure of some of the models studied.

**Porphine and the Effect of His42.** The first models studied, 1 and 2 a–d, serve as a comparison with previous work. Since models 2a–d lack the Arg38, they also correspond closely to the case where this residue is mutated to leucine. Model cluster 1 is the P–CO–His170 complex and includes Porphine–CO and His170. These two groups are present in all the models studied. The nitrogen N<sub>δ</sub> of His170 is protonated (as in all models except in models 7 and 8). Table 3 gives the calculated distances and frequencies using two basis sets. The distances  $d_{\text{CO}}$  and  $d_{\text{Fe–CO}}$  for the optimized model 1 with the basis TZV are 1.164 and 1.795 Å, respectively, 0.007 greater and 0.027 Å smaller than with the basis 6-31G. These small discrepancies between these two basis sets are present in all the models studied here. However, we expect that the overall trend in the distances seen with the smaller basis will be the same as with the basis TZV. Comparison of these results with other works is shown in Table 3, indicating good agreement.

**His42 Protonation.** In models 2a–d, His42 is included, and as mentioned, they resemble the Arg38Leu mutant. In model 2a and b, two geometries are compared arising from different constraints in the cases where the proton H<sub>ε</sub> was removed in the distal histidine, so that its total charge is 0. When there is no positional restraint in the methyl terminal of His42, one sees

TABLE 2: Selected Bond Lengths [Å] after Optimization<sup>a</sup>

model <sup>b</sup>	name <sup>b</sup>	Fe–C	C–O	Fe–N <sub>p</sub>	Fe–His170N <sub>ε</sub>	His42N <sub>ε</sub> –O or His42H <sub>ε</sub> –O or H <sub>2</sub> O–O	ArgH–O	His42N <sub>ε</sub> –ArgH <sub>1</sub>
1	P–CO–His170	1.768	1.171	2.024	2.052			
2a*	His42N <sub>δ</sub> H <sub>δ</sub> N <sub>ε</sub> :	1.780	1.167	2.024	2.054	3.007		
2b*	His42N <sub>δ</sub> H <sub>δ</sub> N <sub>ε</sub> :	1.774	1.170	2.024	2.047	3.982		
2c	His42N <sub>ε</sub> H <sub>ε</sub> N <sub>δ</sub> H <sub>δ</sub>	1.716	1.189	2.025	2.056	1.668		
2d	His42N <sub>ε</sub> H <sub>ε</sub> N <sub>δ</sub> :	1.749	1.176	2.024	2.054	1.997		
3	Arg	1.722	1.188	2.027	2.051		1.893	
4	Arg + H <sub>2</sub> O	1.716	1.194	2.025	2.051	1.790	2.165	
5a	Arg + His42N <sub>δ</sub> H <sub>δ</sub> N <sub>ε</sub> H <sub>ε</sub>	1.699	1.204	2.026	2.050	1.801	1.883	
5b	Arg + His42N <sub>δ</sub> H <sub>δ</sub> N <sub>ε</sub> :	1.741	1.182	2.022	2.050	4.410	1.978	1.754
5c	Arg + His42N <sub>δ</sub> H <sub>δ</sub> N <sub>ε</sub> H <sub>ε</sub> + H <sub>2</sub> O	1.708	1.199	2.028	2.052	1.903	2.097	
5d	Arg + His42N <sub>δ</sub> H <sub>δ</sub> N <sub>ε</sub> : + H <sub>2</sub> O	1.727	1.188	2.024	2.053	3.211	1.856	3.63
6	Arg <sup>neutral</sup> + His42N <sub>δ</sub> H <sub>δ</sub> N <sub>ε</sub> :	1.759	1.174	2.024	2.054	3.62	2.191	2.41
7	P–CO–His170N <sub>δ</sub> :	1.775	1.177	2.014	2.025			
8	P–CO–His170N <sub>δ</sub> : + Arg + HisN <sub>δ</sub> H <sub>δ</sub> N <sub>ε</sub> :	1.760	1.185	2.028	2.008	3.859	1.924	1.81
9	HRPC–CO	1.767	1.171					

<sup>a</sup> Fe–N<sub>p</sub> is the mean distance Fe–nitrogen atoms of porphine. Fe–N<sub>ε</sub> is the distance N<sub>ε</sub> of His170 and Fe. <sup>b</sup> The symbol (:) indicates lone-pair orbital, and the symbol \* indicates different constrains.

TABLE 3: Calculated Bond Lengths and Frequencies for Some Models

model	Fe–C	C–O	$\nu_{\text{Fe–CO}}$	$\nu_{\text{CO}}$	refs
P–Im–CO <sup>a</sup>	1.733	1.165	524	1985	45
P–Im–CO <sup>a</sup>	1.781	1.168	501	1991	33
P–Im–CO <sup>a</sup>	1.813	1.159	498	1995	32
P–Im–CO <sup>a</sup>	1.736	1.1699	570	1982	2
P–Im–CO···Im <sup>b</sup>	1.722	1.1746	581	1952	2
Mb–CO QM/MM N <sub>ε</sub> ···O <sup>c</sup>	1.74	1.16	−62 <sup>c</sup>	+14 <sup>c</sup>	1
Mb–CO QM/MM N <sub>ε</sub> H <sub>ε</sub> ···O <sup>c</sup>	1.72	1.17	+61 <sup>c</sup>	−23 <sup>c</sup>	1
P–CO–His170 (model 1) (B3LYP/6-31G)	1.768	1.171	505	1976	this work
P–CO–His170 (model 1) (B3LYP/TZV)	1.795	1.164		1938	this work
His42N <sub>ε</sub> H <sub>ε</sub> N <sub>δ</sub> : (model 2d)	1.749	1.176	537	1948	this work
Arg + His42N <sub>δ</sub> H <sub>δ</sub> N <sub>ε</sub> : (model 5b)	1.741	1.182	550	1901	this work
Arg + His42N <sub>δ</sub> H <sub>δ</sub> N <sub>ε</sub> : + H <sub>2</sub> O (model 5d)	1.727	1.188	562	1875	this work

<sup>a</sup> Porphine–CO with imidazole as proximal ligand. <sup>b</sup> Myoglobin–CO model; the effect of the distal imidazole with charge zero where the H<sub>δ</sub> was removed. <sup>c</sup> Hybrid QM/MM simulation of myoglobin–CO (Mb–CO). Shift of  $\nu_{\text{Fe–CO}}$  and  $\nu_{\text{CO}}$  stretch frequencies with respect to the isolated heme–CO model. N<sub>ε</sub>···O and N<sub>ε</sub>H<sub>ε</sub>···O means the interaction of His64 unprotonated and protonated with the CO moiety, respectively.

the imidazole moving away from CO, indicating a repulsive interaction between P–CO–His170 and neutral His42. To avoid this in model 2a the distance N<sub>ε</sub>···O<sup>oc</sup> (O<sup>oc</sup> is the oxygen atom of the carbon monoxide moiety) and the dihedral angle Fe–C–O–N<sub>ε</sub> were fixed to 3.01 Å and −53°, respectively, according with the crystal structure. This increases the interaction of the lone pair orbital in N<sub>ε</sub> of the His42 with the CO moiety. In model 2b the histidine is allowed to rotate around the terminal carbon of the methyl group. On relaxation, the His42 moves away from CO with the N<sub>ε</sub> lone-pair orbital being away from the O<sup>oc</sup> moiety at a distance His42N<sub>ε</sub>–O<sup>oc</sup> ~ 4 Å, see Table 2. C<sub>ε</sub> moves slightly closer but it is still far from the O<sup>oc</sup> moiety.

In model 2c, His42 has charge +1 and the hydrogens in N<sub>ε</sub> and N<sub>δ</sub> were retained. The relaxed structure shows a rotation in the histidine which increases its interaction with O<sup>oc</sup> moving to the plane defined by the imidazole, and the Fe–C–O angle is 179.2°. In comparison with model 1, there is a lengthening of the C–O bond.

In model 2d (P–CO–His170 and His42N<sub>ε</sub>H N<sub>δ</sub>:) the H<sub>ε</sub> is retained and instead the H<sub>δ</sub> of His42 is removed and there are

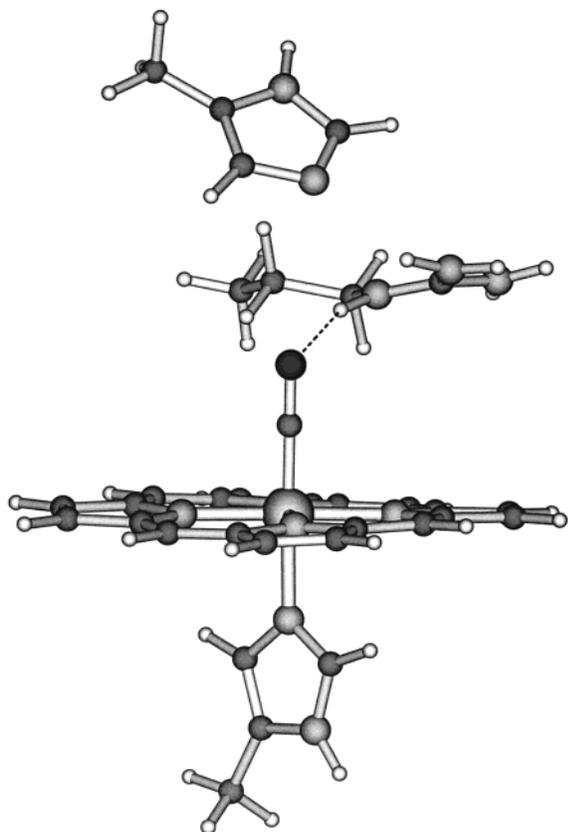
no positional constraints in the minimization. This is the only case where the H<sub>δ</sub> of His42 is removed. This configuration relaxed the histidine 0.35 Å farther away than in model 2c. The Fe–C–O angle is 179.9°, and the plane of imidazole of His42 has an angle of ~8° with respect to the normal of the porphine plane. The results of these changes produce a decrease in  $\nu_{\text{CO}}$ , relative to model 1, Table 3.

**Effect of Arg38 and Water.** In the rest of the model clusters we consider the effect of Arg38 on the CO stretching band. The arginine possesses five labile protons on the guanidyl group; its charge is +1, except for model 6 where we removed the hydrogen H<sub>2</sub> (Figure 2, proton HN22 in the file 1ATJ), making Arg38 neutral. The position of the terminal methyl was fixed with respect to the porphine plane according to the crystal structure.<sup>24</sup>

Model 3 (P–CO–His170 and Arg38) represents the mutant His42Leu.<sup>25</sup> A hydrogen bond (1.89 Å) between the hydrogen H<sub>ε</sub> of the Arg38 and the oxygen O<sup>oc</sup> is observed, and the angle ArgN<sub>ε</sub>–H<sub>ε</sub>···O<sup>oc</sup> is 168°. The bending angle Fe–C–O is 177°; there is no tilting.

Model 4 is same as model 3, but with one crystallographically observed water molecule added. This water molecule was suggested to be a participant in the active site in ref 5, based on the fact that two water molecules are retained and interact with adducted CO and/or with the distal amino acids in cytochrome *c* peroxidase–CO.<sup>26,27</sup> Thus, it is possible that at least one water molecule is interacting with the distal amino acids and CO in HRPc. The initial position of the water oxygen is taken from 1ATJ, and there is no positional restriction in the optimization of this model. The water molecule interacts with both the distal arginine (O<sup>water</sup>···ArgH<sub>1</sub>) and O<sup>oc</sup>, see Figure 1. Our relaxed structure shows an increase in the distance ArgH<sub>ε</sub>···O<sup>oc</sup> of ~0.27 Å with an angle ArgN<sub>ε</sub>H<sub>ε</sub>···O<sup>oc</sup> of 163° with respect to the model without H<sub>2</sub>O. The bending angle Fe–C–O of 176° is in the direction of the bond ArgN<sub>ε</sub>H<sub>ε</sub>···O<sup>oc</sup> and without tilting.

**Effect of Arg38 and His42.** Models 5a,b (P–CO–His170, Arg38 and His42) contain both the distal histidine and arginine and represent HRPc–CO at low (5a) and high (5b) pH (latter case has an unprotonated N<sub>ε</sub> at the distal histidine), respectively. In both models, the H<sub>δ</sub> proton is retained.<sup>28</sup> The relaxed structure of model 5a is similar to the structure showed in Figure 1 but without considering the water molecule seen in that figure. When both distal amino acids are present, the optimized positions are



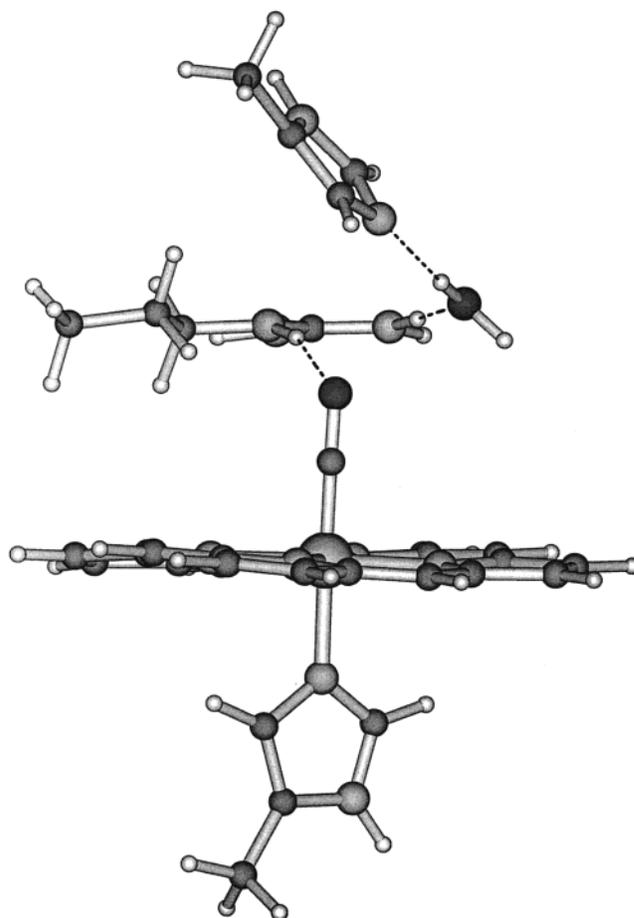
**Figure 3.** Relaxed structure of model 5b, (Arg + His42N $\delta$ H N $\epsilon$ ), representing the high pH situation. The His42 is unprotonated and interacts with the guanidyl group of Arg38, which is distorted. The dotted line shows the interaction of H $\epsilon$  of Arg38 with the oxygen O $^{\text{oc}}$  of the CO moiety.

different than when only one amino acid interacts with P-CO-His170, since both the protonated histidine and arginine compete for the CO moiety. The bending angle Fe-C-O is 176 $^{\circ}$  in a plane perpendicular to the  $^{\text{Arg}}\text{N}_{\epsilon}\text{H}_{\epsilon}\cdots\text{O}^{\text{oc}}$  and  $^{\text{His}}\text{N}_{\epsilon}\text{H}_{\epsilon}\cdots\text{O}^{\text{oc}}$  bonds.

Model 5b is shown in Figure 3; on relaxation its structure is drastically different than for model 5a. The histidine and the guanidyl group of arginine rotate, increasing their interaction.<sup>29</sup> Likewise, as in model 2b, where the histidine is unprotonated, the lone pair on  $^{\text{His}}\text{N}_{\epsilon}$  is oriented far from O $^{\text{oc}}$ . The guanidyl group of Arg38 also moves far from the porphyrin plane.

Models 5c,d are similar to a and b but also include the water molecule seen in Figure 1. Compared to model 5a, in model 5c the distance  $d_{\text{C-O}}$  decreases and  $d_{\text{Fe-CO}}$  increases. In contrast, compared to model 5b, in model 5d (Figure 4)  $d_{\text{C-O}}$  increases and  $d_{\text{Fe-CO}}$  decreases with respect to the models without water. In model 5c the water molecule is a hydrogen acceptor for the Arg38, and relaxed, the positions of the His42 and Arg38 are a bit farther away than in model 5a, which lacks water. In the presence of water, the lone pair orbital of unprotonated His42 (model 5d) interacts with the water molecule which restricts the motion of the His42 and of the guanidyl group of the Arg38. The water is a hydrogen bond donor for the His42 and an acceptor for the Arg38. The frequencies for  $\nu_{\text{CO}}$  for both model 5b and model 5d are lower than for models 1 and 2d (Table 3).

Model 6 contains the arginine with the guanidyl group without the proton H $_2$  (see Figure 2); the molecule is neutral. This model was based on our  $\text{p}K_{\text{a}}$  calculations, which indicate that the proton H $_2$  has the highest chance of being dissociated at high pH. In model 6, the neutral arginine relaxed to a position relatively



**Figure 4.** Model 5d (Arg + His42N $\delta$ H N $\epsilon$  + H $_2$ O), the water molecule interacts with the guanidyl group of Arg38 and with the His42. The dotted line shows the hydrogen bond interaction.

farther from the carbonyl oxygen than in the other models studied. The change in  $d_{\text{C-O}}$  and  $d_{\text{Fe-CO}}$  follows the same correlation as the other models with values close to the model 1, thus justifying our approximation of modeling the leucine mutants by deleting the mutated residues.

#### Effect of Proximal His (Comparison with Distal Effects).

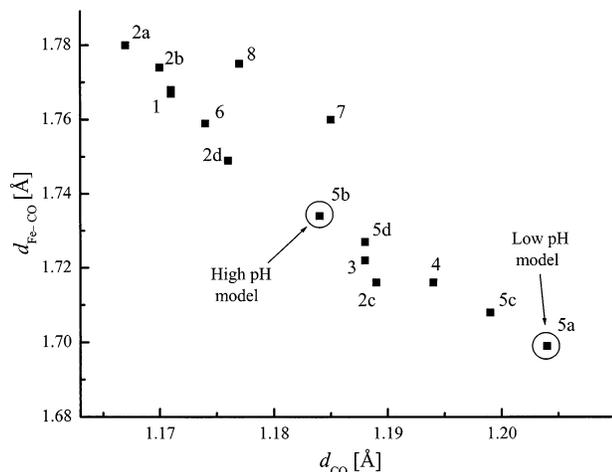
Models 7 and 8 show the effect of the trans ligand. The trans effect was studied recently in porphine-Im-CO in reference 30. In HRPC, the His170 has a hydrogen bond (His170N $\delta$ -H $\delta\cdots\text{Asp}^{\text{O}}\text{O}_{\delta}$ ) with the Asp247, which affects the character of the His170.<sup>3</sup> We relaxed the structure, model 7 P-His170N $\delta$ -CO, without the hydrogen H $\delta$  of His170 and also the distal amino acids (see Tables 1 and 2) to rationalize the distal influence. We started the relaxation of model 7 with the optimized structure of model 1. On comparison with model 1, where His170 is protonated, the distances  $d_{\text{CO}}$  and  $d_{\text{Fe-CO}}$  in model 7 increase. To study the distal contribution of His42 and Arg38 to  $d_{\text{CO}}$  and  $d_{\text{Fe-CO}}$  when the proximal His170 is unprotonated, we relaxed model 8, model 5b without H $\delta$  in the proximal His170. The distances  $d_{\text{CO}}$  and  $d_{\text{Fe-CO}}$  increase further, being larger than when H $\delta$  is present in the proximal histidine. However, in comparison to model 7  $d_{\text{Fe-CO}}$  is more sensitive to deprotonation of His170 when the distal amino acids are present than  $d_{\text{CO}}$  is. The distance  $d_{\text{Fe-N}_{\epsilon}}$  (here N $\epsilon$  is from His170) is the shortest of all models, see Table 2, indicating an increase in the interaction Fe-His170.

**Protoporphyrin IX-CO.** Model 9 is protoporphyrin IX-CO, the natural heme of HRPC, and allows us to study the effect of the peripheral atoms on the CO frequency. The initial coordinates were from the 1ATJ file. We added two hydrogen

**TABLE 4: Average Distances and Standard Deviation,  $\sigma$ , (in bracket) [Å] Observed during the Molecular Dynamics Study with CHARMM<sup>a</sup>**

model	ArgH <sub>ε</sub> ...O <sup>o</sup> c	LeuH <sub>γ</sub> ...O <sup>o</sup> c	HisH <sub>ε</sub> ...O <sup>o</sup> c	HisN <sub>ε</sub> ...O <sup>o</sup> c	HisH <sub>ε</sub> ...ArgH <sub>1</sub>	HisN <sub>ε</sub> ...ArgH <sub>1</sub>	HisH <sub>δ</sub> ...Asn70O
Arg + His42N <sub>δ</sub> H <sub>δ</sub> N <sub>ε</sub> H <sub>ε</sub>	2.41 (0.05)		2.67 (0.05)		3.82 (0.07)		1.77 (0.06)
Arg + His42N <sub>δ</sub> H <sub>δ</sub> N <sub>ε</sub> :	2.69 (0.09)			3.01 (0.04)		3.3 (0.3)	1.89 (0.04)
Leu + His42N <sub>δ</sub> H <sub>δ</sub> N <sub>ε</sub> :		2.70 (0.04)		3.14 (0.05)			1.92 (0.04)

<sup>a</sup> The model Leu + His42N<sub>δ</sub>H<sub>δ</sub>N<sub>ε</sub>H<sub>ε</sub> (Leu + His42N<sub>δ</sub>H<sub>δ</sub>N<sub>ε</sub>: ) corresponds to the mutation in which the Arg38 was replaced by leucine and the distal His42 is protonated (unprotonated).

**Figure 5.** Plot of calculated distance  $d_{\text{CO}}$  vs  $d_{\text{Fe-CO}}$  in Fe(II)CO model complexes of HRPc-CO. Numbers refer to Table 1.

atoms to the propionic acids to fill the valence shell. During the optimization the propionic acids move closer, showing a final structure with the propionic acid very close. To avoid the distortion we fixed the internal angles in the propionic acids. Interestingly, the final structure shows the same distance for CO and Fe-CO as observed for model 1 (Table 2). Since protoporphyrin IX has methyl, vinyl, and propionic side groups, this result suggests that its side groups have very little effect on the bound CO.

### Relationship between Bond Lengths

Figure 5 shows the correlation between  $d_{\text{CO}}$  and  $d_{\text{Fe-CO}}$  for the models studied. The trend seen here is similar to other studies and reflects the inverse dependence between  $\nu_{\text{CO}}$  and  $\nu_{\text{Fe-CO}}$ .<sup>2,31</sup> As expected, a positive electrostatic environment around CO increases the  $d_{\text{C-O}}$  bond distance and decreases  $d_{\text{Fe-CO}}$ . The points corresponding to models 1 (porphine-CO) and 9 (protoporphyrin IX-CO) in Figure 5 overlap each other, showing that the influence of the peripheral ligands is very small and the porphine-His170-CO is an adequate model.

### pK<sub>a</sub> of Active Site Residues

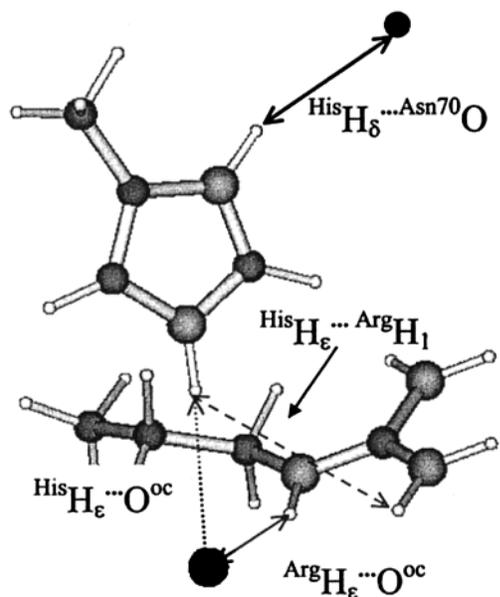
Poisson-Boltzmann calculations of pK<sub>a</sub> shifts are summarized in Table 5. Contributions are broken down into a desolvation contribution due to the reduced accessibility of the active residue to solvent, and interactions with other ionizable or partially charged (dipolar) groups. The desolvation term favors the neutral form of the residue and is always present. Some of the other charge-charge/dipole contributions depend on the charged state of other parts of the protein. Some, such as the ionized state of the heme propionates, are not known precisely, but the uncertainty in pK<sub>a</sub> from this source may be estimated from the separately tabulated contributions. For the proximal His170 liganded to the Fe, the reference state value for the removal of the second nitrogen proton of the now neutral side is not known exactly, but it is certainly much higher than the first proton

**TABLE 5: pK<sub>a</sub> Shifts of Active Site Residues**

residue/proton (ref pK <sub>a</sub> )	contribution	$\Delta\text{pK}_a$
His170/H <sub>δ</sub> (13.1) <sup>d</sup>	desolvation	9.5
	CO+HRPC+Heme <sup>b</sup>	11.7
His42/H <sub>ε</sub> (6.8)	desolvation	-9.8
	CO	2
	HRPC <sup>c</sup>	9.8
	calcium	-3.0
	heme <sup>d</sup>	0.9
	propionates	2.9
	water <sup>e</sup>	0
	Arg <sup>+</sup> 38	-5.2
	net <sup>f</sup>	4.4
	Arg38/H <sub>2</sub> (12)	desolvation
	CO	0.3
	HRPC <sup>c</sup>	2.8
	calcium	-1.7
	heme <sup>d</sup>	-1.2
	propionates	9.6
	water <sup>e</sup>	-0.2
	His <sup>+</sup> 42 <sup>g</sup>	-4.6
	net <sup>f</sup>	12.0 <sup>g</sup> (16.8) <sup>h</sup>

<sup>a</sup> The reference pK<sub>a</sub> for removal of the second proton of an Fe<sup>2+</sup>-liganded histidine is not known. This estimate taken from His liganded to an iron-sulfur cluster. <sup>b</sup> Assuming ionized propionates. <sup>c</sup> Excluding bound calcium ions, tabulated separately, and the other distal residue (Arg38 or His42). <sup>d</sup> Excluding ionized propionates, tabulated separately. <sup>e</sup> Crystallographic water residue #468 in active site which interacts with CO, Arg38, and His42. <sup>f</sup> Sum of contributions above plus reference value. <sup>g</sup> Arg pK<sub>a</sub> assuming His42 is protonated, i.e., the low pH (<4) value. Figure in brackets assuming His42 is unprotonated, i.e., the Arg pK<sub>a</sub> at mid to high pH (> 5).

removal pK<sub>a</sub> of 6.8. Since the desolvation and charge/dipolar terms both give positive pK<sub>a</sub> shifts, we can conclude with some certainty that the net pK<sub>a</sub> of this group is very high, and rule out deprotonation of this group as a source of observed spectral changes in the pH 4–10 range. The pK<sub>a</sub>'s of His42 and Arg38 may be estimated with considerably more certainty. From the data, the histidine residue will be the first to deprotonate, since its pK<sub>a</sub> is considerably lower than that of Arg38. The calculations suggest that the deprotonation occurs around pH 5, although it may occur as low as pH 3 if the propionates are not ionized. The Arg38 pK<sub>a</sub> is relatively unperturbed when the distal histidine is charged (low pH), but it increases to 16.8 upon raising the pH, due to deprotonation of His42. From the data it appears then that Arg38 is not the source of a second deprotonation within the normal pH range. However, this possibility cannot be ruled out, since there is a strong interaction between ionized heme propionates and Arg38, which provides a large positive pK<sub>a</sub> shift. If the propionates were not ionized, the pK<sub>a</sub> of Arg in the presence of neutral His42 would drop to 7.2, well within the experimental range of pH. Deprotonated Arg38 would then be expected to contribute to the observed pH-dependent spectral shifts. Model 6 considers the effect of deprotonation of Arg38. Values for pK<sub>a</sub> shifts of Arg38 were calculated for all five possible deprotonations of the guanidinium group. Table 5 presents results for the HN22 (H<sub>2</sub> in Figure 2) proton, which in the FDPB calculations is the easiest to remove (i.e., has the lowest pK<sub>a</sub>). The H<sub>ε</sub> proton was slightly harder to remove, but



**Figure 6.** Distances between the different atoms studied with molecular dynamics when the His42 is protonated ( $^{\text{Arg}}\text{H}_\epsilon \cdots \text{O}^{\text{oc}}$ ,  $^{\text{His}}\text{H}_\epsilon \cdots \text{O}^{\text{oc}}$ , and  $^{\text{His}}\text{H}_\epsilon \cdots \text{ArgH}_1$ ). When His42 is unprotonated the distances are  $^{\text{Arg}}\text{H}_\epsilon \cdots \text{O}^{\text{oc}}$ ,  $^{\text{His}}\text{N}_\epsilon \cdots \text{O}^{\text{oc}}$ , and  $^{\text{His}}\text{N}_\epsilon \cdots \text{ArgH}_1$ .

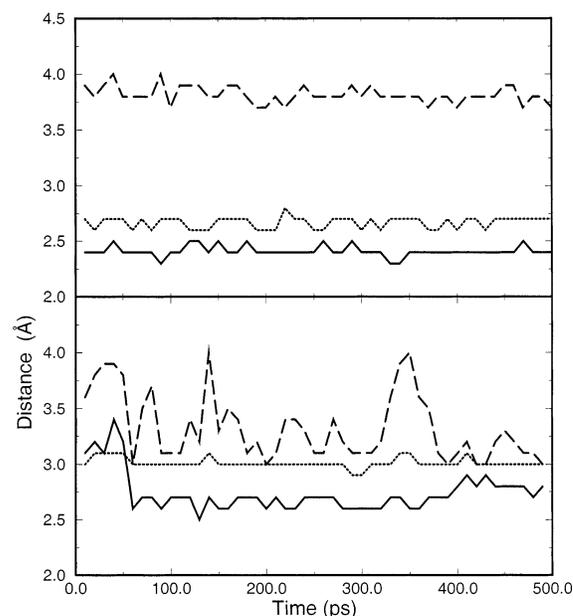
gives an almost identical  $\text{pK}_a$  value (not shown). The other three guanidinium protons, which point more to the outer part of the active site (i.e., away from the CO, toward the more solvent exposed heme edge), have significantly higher  $\text{pK}_a$ 's because of their small desolvation penalty, and greater interaction with the propionates. These were judged to be very unlikely candidates for deprotonation at any pH.

### Molecular Dynamics

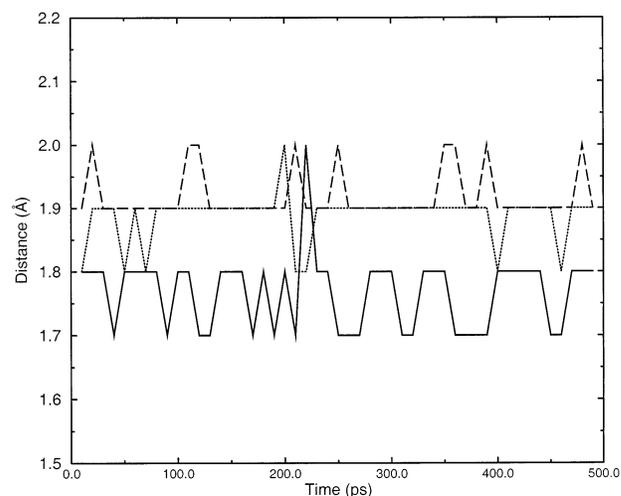
The DFT computations above have indicated that Arg38 and His42 have an important effect on the CO frequency. However, these were performed with truncated models and static structures, and due to these limitations the effects of dynamics of the structure and the steric and electrostatic effects of the rest of the protein environment could not be included.<sup>1</sup> The time histories of the positions and the fluctuations around the CO adducts of His42 (protonated and unprotonated) and Arg38 (Figure 6) and in the mutant Arg38Leu from the MD simulations are presented for comparison with DFT.

Figure 7a shows the time histories from the MD simulation with protonated His42 (which correlates with the DFT models 5a and 5c). The three distances  $^{\text{Arg}}\text{H}_\epsilon \cdots \text{O}^{\text{oc}}$ ,  $^{\text{His}}\text{H}_\epsilon \cdots \text{O}^{\text{oc}}$ , and  $^{\text{His}}\text{H}_\epsilon \cdots \text{ArgH}_1$  and their small standard deviations (Table 4) show that His42 and Arg38 are stiff. The average distances of Arg38 and His42 are larger in the MD simulation (Table 4) than seen from the DFT calculations on the corresponding models 5a and b but reflect the strong interaction. In analogy to the increased distances seen in model 5c as compared to 5a, on inclusion of the water molecule in the former cluster, the increased distances in the MD simulation may reflect the effects of the inclusion of the solvent environment in the interactions.

Figure 7b shows the time evolution of the protein when His42 is unprotonated (which correlates with the models 5b and 5d). In this case the distance Arg38 and  $\text{O}^{\text{oc}}$  increases and the distance  $^{\text{His}}\text{N}_\epsilon \cdots \text{ArgH}_1$  decreases but shows a strong fluctuation ( $\sigma \sim 0.3 \text{ \AA}$ ), compared to the protonated His42 MD simulation. The flexibility of the guanidyl group of the Arg38 and its interaction with His42 may be the reason for this large fluctuation.<sup>32</sup> Once again, the average His42–CO distance and



**Figure 7.** (a and b) Plot of the distances during the MD simulation. Panel 7a, the  $^{\text{Arg}}\text{H}_\epsilon \cdots \text{O}^{\text{oc}}$  (solid line),  $^{\text{His}}\text{H}_\epsilon \cdots \text{O}^{\text{oc}}$  (dotted line), and  $^{\text{His}}\text{H}_\epsilon \cdots \text{ArgH}_1$  with the distal His42 protonated (corresponding to model 5a in DFT calculations). Panel 7b shows the distances  $^{\text{Arg}}\text{H}_\epsilon \cdots \text{O}^{\text{oc}}$  (solid line),  $^{\text{His}}\text{N}_\epsilon \cdots \text{O}^{\text{oc}}$  (dotted line), and  $^{\text{His}}\text{N}_\epsilon \cdots \text{ArgH}_1$  (dashed line) with the distal His42 unprotonated (model 5b in DFT).



**Figure 8.** Distances between  $^{\text{His}}\text{H}_\delta \cdots \text{Asn70O}$  when the His42 is protonated (solid line), unprotonated (dotted line), and in the mutant Arg38Leu (dashed line).

His42–Arg38 distances are in much better agreement with the corresponding partially solvated DFT model 5d than with the unsolvated model 5b. However, MD overestimates the Arg–CO distances compared to DFT.

The MD simulation of the mutant Arg38 by leucine with the His42 unprotonated (which would correspond to the alkaline pH) shows an average distance  $^{\text{His}}\text{N}_\epsilon \cdots \text{O}^{\text{oc}}$ , slightly larger than when His42 is protonated. The distance  $\text{LeuC}_\gamma\text{H}_\gamma \cdots \text{O}^{\text{oc}}$  (where  $\text{H}_\gamma$  is the hydrogen of  $\text{C}_\gamma$  of leucine) has an average distance of  $3.14 \text{ \AA}$  ( $\sigma = 0.05$ ).

Figure 8 shows the time evolution of the distance  $\text{H}_\delta$  of the His42 and the oxygen,  $\text{O}_\delta$  of the Asn70, ( $^{\text{His}}\text{H}_\delta \cdots \text{Asn70O}_\delta$ ) when the His42 is protonated or unprotonated and in the mutant Arg38–leucine with the His42 unprotonated. When the His42 is protonated, the distance  $^{\text{His}}\text{H}_\delta \cdots \text{Asn70O}_\delta$  is shorter than when His42 is unprotonated. In the mutant, the distance  $^{\text{His}}\text{H}_\delta \cdots \text{Asn70O}_\delta$

is greater than in the other two cases, see Table 4. However, these distances vary only between 1.7 and 2.0 Å, and hence within our theoretical approach, the hydrogen bond between His42 and the Asp70 is strong and may constrain the position of His42 and its interaction with CO.<sup>33</sup> This hydrogen bond may be cause for the MD simulations having the His42 close to O<sup>oc</sup> regardless of the protonation state.

## Discussion

The DFT calculations show that the Fe–C and C–O bond lengths and thus the stretch frequency of CO bound to heme are especially sensitive to the nearby amino acids in the protein. Experimentally, the values for  $\nu_{\text{CO}}$  range from  $\sim 1900$  to  $1970 \text{ cm}^{-1}$  in different proteins and mutants.<sup>34</sup> As indicated in the Introduction, it was shown by mutagenesis<sup>35</sup> and Stark experiments that local fields generated by surrounding atoms play a role in shifting CO frequencies.<sup>36</sup> The importance of local electric fields in CO–HRPC has been invoked to explain features of the optical absorption spectra.<sup>37</sup> The Q(0,0) band of HRPC–CO at pH 6.0 shows a single peak that narrows as the temperature decreases. Increasing the pH to 9, causes a split in the band to  $200 \text{ cm}^{-1}$ . Interestingly, the addition of substrates such as benzhydroxamic acid (BHA) or naphthohydroxamic acid (NHA) also produces a splitting of  $\sim 320 \text{ cm}^{-1}$  at low temperature. Local electric fields as well as heme distortions are also contributors of changes in the heme optical spectra.<sup>38–41</sup>

The changes in bonding geometry and stretch frequency of Fe bound CO due to protein effects, as revealed by our DFT calculations and earlier experimental and theoretical work, are collected in Tables 2 and 3. The overall trends and general physical picture can be summarized and interpreted at different levels of physical description.<sup>42</sup> The most basic description is to consider the charged and polar groups surrounding the CO as contributing to the local field. Simple considerations show that, for example, a positive charge above the partially negative O of CO (or an electric field in the direction O→C) increases the population of the antibonding orbitals  $2\pi^*$ , which would lengthen the CO bond and decrease the CO stretch frequency, while a negative charge would have the opposite effect. This level of description is sufficient to understand at a qualitative level, the data in Table 2. A more sophisticated interpretation would include the effect of polarization, particularly from HRPC's distal groups, which are close to the bound ligand. Furthermore, when a change in charge state occurs (for example by (de)protonation of active site residues) there will also be changes in geometry and frequency, since there is a change in the local electric field. Protonation related changes in field have two components: first, the direct effect of the change in charge state; second, the indirect effect, due to relaxation of the protein, especially changes in position of groups close to CO, which in turn contribute to the local electric field. By DFT calculations combining geometry optimization, molecular modeling, and  $\text{pK}_a$  calculations, we have attempted to include all these effects in as self-consistent manner as possible.

Some generalities regarding the models can be made. The bend angle for all the models studied is less than  $\sim 4^\circ$  and the tilt is less than  $\sim 1^\circ$ . The linearity indicates that the overlap of both filled Fe  $d_\pi$  orbitals with the two empty  $\pi^*$  orbitals of CO is important. The distal effects increase the overlap optimizing the back-bonding. The energy required in the in-phase  $\tau$ – $\beta$  CO deformation is small as was shown by Ghosh et al.<sup>43</sup> in the porphine–Im–CO complex. The relaxed models studied show the structures in the minimum of the  $\tau$ – $\beta$  potential energy surface of Ghosh et al. In HRPC–CO the presence of Arg38

could change at least the energy landscape for motion in the direction of Arg38.

A main emphasis of our work is that Arg38 has an important contribution to  $d_{\text{CO}}$  and  $d_{\text{Fe–CO}}$  and consequently in their associated frequencies. The presence of Arg decreases the frequency. This is, in general, seen experimentally, for instance  $\nu_{\text{CO}}$  for myoglobin, which lacks Arg in the active site, have higher frequencies the CO–HRPC.

The involvement of Arg in HRPC is supported by the following evidence from our models:

(a) Whether histidine 42 is protonated or unprotonated, Arg38 interacts with the CO moiety. The low pH ( $\sim 6$ ) case is studied with models 5a and 5c. When His42 is protonated, His42 and Arg38 compete for the CO electron density and polarize its frontier orbitals. The presence of a water molecule, model 5c, in the distal site has a small effect in the position of both His42 and Arg38; however,  $d_{\text{CO}}$  and  $d_{\text{Fe–CO}}$  distances are sensitive to these small changes (Figure 5). The molecular dynamics shows a His42 and Arg38 rigid. The guanidyl group of Arg38 is very flexible but has a more planar conformation when the distal His42 is protonated (below) or a water molecule is included interacting with it.

The high pH ( $>9$ ) case was studied with models 5b and 5d, where His42 is unprotonated. In model 5b, the guanidyl group (hydrogen H<sub>1</sub>) of Arg38 interacts strongly with the lone-pair orbital of His42N<sub>ε</sub>, pulling the whole amino acid His42 far from the carbonyl oxygen. This is reflected in a long C–O distance (Figure 5). The presence of a water molecule, model 5d, relaxed the position of His42N<sub>δ</sub>N<sub>ε</sub> in the direction of the water molecule. The guanidyl group of Arg38 interacts with the water and less with the lone-pair orbital of the His42N<sub>ε</sub>. The MD study which simulated model 5a shows the guanidyl group to be very flexible (Figure 7).

(b) The mutants Arg38Leu (model 2 a–d) and His42Leu (models 3 and 4) show the reinforcement effect of His42 and Arg38 with the CO moiety. The mutant His42Leu shows experimentally a frequency between the frequencies found at high and low pH, which were modeled with models 5a and 5b, respectively. The predicted structure for models 3 and 4 are in the expected place. Experimentally, the mutant Arg38Leu has a  $\nu_{\text{CO}}$  frequency higher than the low pH frequency for the wild type. However, our predicted structure, model 2c, shows a point in an unexpected place. One would expect it to show  $d_{\text{Fe–CO}}$  and  $d_{\text{CO}}$  between model 1 and 5b. The likely reason of that disagreement is because model 2c does not contain leucine, and the histidine can move closer to CO moiety.<sup>25</sup> It is possible that in the mutant there are additional structural changes, and water molecules can reorient the histidine in the cavity, weakening its interaction with CO moiety.

(c) Effect of charge of Arg38. The deprotonation of Arg38 occurs at high pH according with our  $\text{pK}_a$  calculation (Table 5). However, as noted in the calculation of  $\text{pK}'$ s, changing their charge on Arg38 might be experimentally feasible. In addition, the neutral Arg38 model allows understanding of the effect of changes in the positive charge distribution of Arg38 and its consequence on the CO moiety. Model 6 (neutral Arg38 and neutral His42) shows a decrease in the  $d_{\text{CO}}$  distance followed by an increase in  $d_{\text{Fe–C}}$ , relaxing the structure close to the model 1 (without distal amino acids) and also in the  $d_{\text{CO}}$  vs  $d_{\text{Fe–CO}}$  correlation.

Finally, we comment on the role of the distal ligand. Models 7 and 8 show the effects of the trans (proximal) ligand on  $d_{\text{CO}}$  and  $d_{\text{Fe–CO}}$  in the absence of distal amino acids and in the presence of the unprotonated His42 and the Arg38, respectively.

$d_{\text{Fe-CO}}$  is more sensitive to deprotonation of His170 when the distal amino acids are present than  $d_{\text{CO}}$ . This is an interesting result which suggests the need for a complementary study. The  $pK_a$  calculation of the buried  $H_\delta$  of His170 is higher than the other cases studied (Table 5). We speculate that, within our theoretical approximations (clusters, basis sets, and functionals), the changes in  $d_{\text{CO}}$  and  $d_{\text{Fe-CO}}$  when the distal amino acids are present is a response to the protein-coupled “push-pull” mechanism.<sup>44,45</sup> Part of the electronic changes due to Imidazolate character of deprotonated His170 are reflected in the  $-\text{CO}$  adducts which seem to act as a link between the proximal and distal sides.

This study represents an advance in computation in that both distal and proximal effects of the protein were included in the calculation. The effects of one water, deprotonation of proximal His170 and distal His42 and Arg38 were systematically examined. The results emphasize the importance of the distal amino acids. Further work would involve mixing quantum mechanics and molecular mechanics to evaluate the role of the protein in the cavity surrounding the heme in HRPc.

**Acknowledgment.** The National Institute of Health grant PO1 GM48310 supported this work. We thank W. W. Wright for complementary experiments.

**Supporting Information Available:** Atomic positions in cartesian coordinates for the relaxed models 1 and 5c. This material is available free of charge via the Internet at <http://pub.acs.org>.

## References and Notes

- Rovira, C.; Shulze, B.; Eichinger, M.; Evanseck, J. D.; Parrinello M. *Biophys. J.* **2001**, *81*, 435–445.
- Sigfridsson, E.; Ryde U. *J. Biol. Inorg. Chem.* **1999**, *4*, 99–110.
- Gajhede, M.; Schuller, D. J.; Henriksen, A.; Smith, A. T.; Poulos, T. L. *Nature Struct. Biol.* **1997**, *4*, 1032–1038.
- Li, H.; Poulos, T. L. *Structure* **1994**, *2*, 461–464.
- Feis, A.; Rodriguez-Lopez, J. N.; Thorneley, R. N. F.; Smulevich, G. *Biochemistry* **1998**, *37*, 13575–13581.
- Holzbaur, I. E.; English, A. M.; Ismail, A. A. *J. Am. Chem. Soc.* **1996**, *118*, 3354–3359.
- Rodriguez-Lopez, J. N.; George, S. J.; Thorneley R. N. F. *J. Biol. Inorg. Chem.* **1998**, *3*, 44–52.
- Uno, T.; Nishimura, Y.; Tsuboi, M.; Makino, R.; Iizuka, T.; Ishimura, Y. *J. Biol. Chem.* **1987**, *262*, 4549–4556.
- Smulevich, G.; Paoli, M.; De Sanctis, G.; Martini, A. R.; Ascoli, F.; Coletta, M. *Biochemistry* **1997**, *36*, 640–649.
- Becke, A. D. *Phys. Rev. A* **1988**, *38*, 3098–3100. Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 1372–1377. Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 5648–5652.
- Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Zakrzewski, V. G.; Montgomery, J. A., Jr.; Stratmann, R. E.; Burant, J. C.; Dapprich, S.; Millam, J. M.; Daniels, A. D.; Kudin, K. N.; Strain, M. C.; Farkas, O.; Tomasi, J.; Barone, V.; Cossi, M.; Cammi, R.; Mennucci, B.; Pomelli, C.; Adamo, C.; Clifford, S.; Ochterski, J.; Petersson, G. A.; Ayala, P. Y.; Cui, Q.; Morokuma, K.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Cioslowski, J.; Ortiz, J. V.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Gomperts, R.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Gonzalez, C.; Challacombe, M.; Gill, P. M. W.; Johnson, B. G.; Chen, W.; Wong, M. W.; Andres, J. L.; Head-Gordon, M.; Replogle, E. S.; Pople, J. A. *Gaussian 98*; Gaussian, Inc.: Pittsburgh, PA, 1998.
- Shaefer, A.; Huber, C.; Ahlrichs, R. *J. Chem. Phys.* **1994**, *100*, 5829–5835.
- Schaftenaar, G.; Noordik, J. H. Molden: a pre- and post-processing program for molecular and electronic structures; *J. Comput.-Aided Mol. Des.* **2000**, *14*, 123–134.
- MacKerell, A. A., Jr.; Bashford, M.; Bellott, R. L.; Dunbrack, J. D., Jr.; Evanseck, J. J.; Field, S.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph-McCarthy, D.; Kuchnir, L.; Kuczera, K.; Lau, F. T. K.; Mattos, C.; Michnick, S.; Ngo, D. T.; Nguyen, D. B.; Prodhom, W. E.; Reiher, B., III; Roux, M.; Schlenkrich, J. C.; Smith, R.; Stote, J.; Straub, J.; Watanabe, J.; Wiorkiewicz-Kuczera, D.; Yin, D.; Karplus, M. *J. Phys. Chem. B* **1998**, *102*, 3586–3616.
- Brooks, B. R.; Brucoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. *J. Comput. Chem.* **1983**, *4*, 187–217.
- Grant, J. A.; Pickup, B.; Nicholls, A. *J. Comput. Chem.* **2001**, *22*, 608–640.
- Yang, A.; Gunner, M.; Sampogna, R.; Sharp, K.; Honig, B. *Proteins: Struct., Funct., Genet.* **1993**, *15*, 252–265.
- Langsetmo, K.; Fuchs, J. A.; Woodward, C.; Sharp, K. A. *Biochemistry* **1991**, *30*, 7609–7614.
- Sharp, K. A. *Biophys. Chem.* **1996**, *61*, 37–49.
- Gilson, M.; Sharp, K. A.; Honig, B. *J. Comput. Chem.* **1988**, *9*, 327–335.
- Nicholls, A.; Sharp, K. A.; Honig, B. *Proteins* **1991**, *11*, 281–296.
- Sharp, K.; Honig, B. *Annu. Rev. Biophys. Biophys. Chem.* **1990**, *19*, 301–332.
- MacKerell, A. D.; Brooks, B.; Brooks, C. L.; Nilsson, L.; Roux, B.; Won, Y.; Karplus, M. CHARMM: The Energy Function and Its Parametrization with an Overview of the Program. In *The Encyclopedia of Computational Chemistry*; John Wiley & Sons: Chichester, 1998.
- To test these restrictions we ran the same jobs without constraints and found that they do not affect the final structure.
- We consider that the effect of the leucine at  $\sim 4 \text{ \AA}$  over CO is very small. A partial relaxation of a model with the protonated His42 and leucine (here leucine is the small molecule  $(\text{CH}_3)_2\text{C}_\gamma\text{H}_\gamma\text{-CH}_3$ ) shows  $d_{\text{CO}} \sim 1.191 \text{ \AA}$  and  $d_{\text{Fe-CO}} \sim 1.714 \text{ \AA}$  and  $d_{\text{H}_\gamma \cdots \text{O}^{\text{O}^c}} \sim 3.01 \text{ \AA}$  very close to model 2c. The starting structure for leucine was from CHARMM minimization.
- Edwards, S. L.; Poulos, T. L. *J. Biol. Chem.* **1990**, *265*, 2588–2595.
- The crystal structure of cytochrome *c* peroxidase, ref 26, has three water molecules interacting with His52, Arg48, and Trp51. Interestingly, cytochrome *c* peroxidase in the presence of CO, CN, NO, and F retains two water molecules that interact with the distal amino acids and with the adducted ligand atoms.
- At neutral pH,  $N_\delta$  of His42 is H-bonded to the  $O_\delta$  of Asn70, which maintains the basicity of His 42. Tanaka, M.; Nagano, S.; Ishimori K.; Morishima I. *Biochem. J.* **1997**, *36*, 9791–9798.
- We started with a structure where the  $\text{His}^{\text{N}_\epsilon}$  is oriented to the  $\text{Arg}^{\text{N}_\epsilon}$  of arginine and the final structure was similar to the model 5b.
- Franzen, S. *J. Am. Chem. Soc.* **2001**, *123*, 12578–12589.
- Vogel, K. M.; Kozlowski, P. M.; Zgierski, M. Z.; Spiro, T. G. *Inorg. Chim. Acta* **2000**, *297*, 11–17.
- A similar result is observed in cytochrome *c* peroxidase, where the distal Arg48 moves in or out to interact with the heme-bound ligand, see ref 28.
- MD simulations and crystallographic studies of Mb-CO show that the distal histidine, His64, is more flexible than in HRPc-CO and is the origin of the A states in Mb. See references, Yang, F.; Phillips, G. N., Jr. *J. Mol. Biol.* **1996**, *256*, 762–774. Schulze, B.; Evanseck, J. D. *J. Am. Chem. Soc.* **1999**, *121*, 6444–6454.
- Spiro, T. G.; Zgierski, M. Z.; Kozlowski, P. M. *Coord. Chem. Rev.* **2001**, *219*–221, 923–936.
- Phillips, G. N.; Teodoro, M. L.; Li, T.; Smith, B.; Olson, J. S. *J. Phys. Chem. B*, **1999**, 8817–8829.
- Bublitz, G. U.; Boxer, S. G. *Annu. Rev. Phys. Chem.* **1997**, *48*, 213–242.
- Kaposi, A. D.; Wright, W. W.; Fidy, J.; Stavrov, S. S.; Vanderkooi, J. M.; Rasnik, I. *Biochemistry* **2001**, *40*, 3483–3491.
- Manas, E.; Vanderkooi, J. M.; Sharp, K. *J. Phys. Chem. B* **1999**, *103*, 6334–6348.
- Manas, E. S.; Wright, W. W.; Sharp, K.; Friedrich, J.; Vanderkooi, J. M. *J. Phys. Chem. B* **2000**, *104*, 6932–6841.
- Prabhu, N. V.; Dalosto, S. D.; Sharp, K. A.; Wright, W. W.; Vanderkooi, J. M. *J. Phys. Chem. B* **2002**, *106*, 5561–5571.
- Rasnik, I.; Sharp, K.; Fee, J. A.; Vanderkooi, J. M. *J. Phys. Chem. B* **2001**, *105*, 282–286.
- Kushkuley, B.; Stavrov, S. S. *Biophys. J.* **1996**, *70*, 1214–1229.
- Ghosh, A.; Bocian, D. F. *J. Phys. Chem.* **1996**, *100*, 6363–6367.
- Finzel, B. C.; Poulos, T. L.; Kraut, J. *J. Biol. Chem.* **1984**, *259*, 13027–13036.
- de Ropp, J. S.; Sham, S.; Asokan, A.; Newmyer, S.; Ortiz de Montellano, P. R.; La Mar G. N. *J. Am. Chem. Soc.* **2002**, *124*, 11029–11037.