# The distal residue–CO interaction in carbonmonoxy myoglobins: A molecular dynamics study of two distal histidine tautomers

**ARTICLE** *in* BIOPHYSICAL JOURNAL · JANUARY 1995

Impact Factor: 3.97 · DOI: 10.1016/S0006-3495(94)80708-8 · Source: PubMed



**2 AUTHORS**, INCLUDING:

Philip Jewsbury
AstraZeneca
**40** PUBLICATIONS **1,640** CITATIONS

SEE PROFILE

Biophysical Journal Volume 67 December 1994 2236–2250

# The Distal Residue-CO Interaction in Carbonmonoxy Myoglobins: A Molecular Dynamics Study of Two Distal Histidine Tautomers

Philip Jewsbury and Teizo Kitagawa
Institute for Molecular Science, Okazaki 444, Japan

ABSTRACT Four independent 90 ps molecular dynamics simulations of sperm-whale wild-type carbonmonoxy myoglobin (MbCO) have been calculated using a new AMBER force field for the haem prosthetic group. Two trajectories have the distal $64N_\delta$ nitrogen protonated, and two have the $64N_\epsilon$ nitrogen protonated; all water molecules within 16 Å of the carbonyl O are included. In three trajectories, the distal residue remains part of the haem pocket, with the protonated distal nitrogen pointing *into* the active site. This is in contrast with the neutron diffraction crystal structure, but is consistent with the solution phase CO stretching frequencies ($\nu_{CO}$) of MbCO and various of its mutants. There are significant differences in the "closed" pocket structures found for each tautomer: the $64N_\epsilon H$ trajectories both show stable distal-CO interactions, whereas the $64N_\delta H$ trajectory has a weaker interaction resulting in a more mobile distal side chain. One trajectory (a $64N_\delta H$ tautomer) has the distal histidine moving out into the "solvent," leaving the pocket in an "open" structure, with a large unhindered entrance to the active site. These trajectories suggest that the three $\nu_{CO}$ frequencies observed for wild-type MbCO in solution, rather than representing significantly different Fe-C-O geometries as such, arise from three different haem pocket structures, each with different electric fields at the ligand. Each pocket structure corresponds to a different distal histidine conformer: the $A_3$ band to the $64N_\epsilon H$ tautomer, the $A_{1,2}$ band to the $64N_\delta H$ tautomer, and the $A_0$ band to the absence of any significant interaction with the distal side chain.

## INTRODUCTION

Myoglobin (Mb) has been the subject of a large number of experimental studies in the years since its x-ray structure was first determined (Kendrew et al., 1960). Presently, over 20 x-ray structures are available in the Protein Data Bank (PDB) (Bernstein et al., 1977; Abola et al., 1987) for variously liganded and/or mutated Mb proteins, and a large number of resonance Raman (RR) (see for example Yu and Kerr (1988) and references therein) and infra-red (IR) (see, for example, Caughey et al. (1981) and references therein; Mourant et al. (1993) and references therein) studies of the ligand bound forms have also been published. Much effort has centred on carbonmonoxy myoglobin (MbCO) because its RR and IR spectra can be readily recorded and samples can be handled relatively easily. Although one of the most studied and best understood proteins, the nature of the structure-function relationship of MbCO is still in debate (see Springer et al. (1994) for a recent review). Only through x-ray crystallography can the Fe-ligand geometry be definitively established; nevertheless the three crystal structures available (PDB files: 1MBC, Kuriyan et al., 1986; 2MB5, Cheng and Schoenborn, 1991; 2MGK, Quillin et al., 1993) show a range of quite different Fe-C-O conformations, as seen in Table 1. To complicate the issue further, the crystalized carbonmonoxy myoglobins have different IR spectra than the solution species (Caughey et al., 1981; Makinen et al., 1979; Mourant et al., 1993); thus, the crystal structure may not represent the physiologically active species. The Fe-ligand properties have also been found to be significantly dependent on pH and temperature. This complexity has made a consistent interpretation of the experimental information difficult to achieve.

Since the haem pocket controls the reactivity of the active site (illustrated in Fig. 1), attention has focused on which residues in particular may be responsible. The prime candidates are the two amino acid residues closest to the Fe binding site: the proximal and distal histidines. The proximal histidine is the complementary axial ligand, and the distal histidine lies on the same side as the binding site, but translated relative to the axis through the haem plane normal at the Fe atom (see Fig. 1). The usual textbook explanation for the physiological role of these residues (for example, Stryer, 1988) is that the proximal residue controls the binding energy of the active site and communicates its status to the protein itself, whereas the distal residue controls the accessibility of the Fe atom and the conformational freedom of the bound ligand by sterically crowding the region above the Fe atom. Furthermore, the distal residue is usually considered to discourage the binding of the poisoning CO ligand by sterically forcing its binding conformation away from its preferred position normal to the haem (this does not effect the physiological ligand, $O_2$, which prefers a bent conformation), while stabilizing a bound oxygen ligand through a hydrogen bond. Neutron diffraction studies of MbCO (Norvell et al., 1975; Cheng and Schoenborn, 1991) have shown that the distal histidine is protonated at the $N_\delta$ side-chain nitrogen in the crystal, the lone pair of the $N_\epsilon$ nitrogen pointing into the haem pocket, as seen in Fig. 1 *a*. This is in contrast to the $MbO_2$ structure in which the $N_\epsilon$ nitrogen is protonated and hydrogen-bonded to the terminal ligand oxygen (Phillips and Schoenborn, 1981).

*Received for publication 26 May 1994 and in final form 1 September 1994.*

Address reprint requests to Dr. Philip Jewsbury, Protein Engineering Research Inst., 6-2-3 Furuedai, Suita, Osaka 565, Japan. Tel.: +81-6-872-8200; Fax: +81-6-872-8210; E-mail: philip@ala.peri.co.jp.

Dr. Jewsbury's present address: Protein Engineering Research Institute, 6-2-3 Furuedai, Suita, Osaka 565, Japan.



0006-3495/94/12/2236/15 $2.00

**TABLE 1 Geometry of the Fe-C-O unit in the Protein databank x-ray and neutron diffraction structures of MbCO**

| PDB file | | RFe-C (Å) | RC-O (Å) | Tilt* (degrees) | FeCO (degrees) | Φ(NFeCO)‡ (degrees) | CO orientation§ |
|---|---|---|---|---|---|---|---|
| 1MBC¶ | C | 1.92 | 1.17 | 87.3 | 141.4 | 60.3 | 40.4 |
| | D | 1.92 | 1.20 | 87.3 | 119.9 | −62.3 | 60.9 |
| 2MB5‖ | A | 2.17 | 1.21 | 76.8 | 145.8 | −36.2 | 47.1 |
| | B | 2.12 | 1.195 | 76.8 | 135.2 | 67.0 | 43.6 |
| 2MGK | | 1.90 | 1.12 | 81.9 | 168.7 | 38.2 | 19.4 |

See text for references.

* Orientation of the FeC bond relative to the mean plane through the 24 porphyrin heavy atoms (in degrees). An angle of 90° lies on the normal.

‡ With reference to the porphyrin $N_c$ nitrogen.

§ Orientation of the CO bond relative to the normal to the mean plane through the 24 porphyrin heavy atoms (in degrees). An angle of 0° lies on the normal.

¶ Two carbonyl O positions were resolved: C with occupancy 78% and D with occupancy 22%.

‖ Two carbonyl O positions were resolved: A with occupancy 58.2% and B with occupancy 41.8%.

Wild-type MbCO has three broad CO stretching peaks ($\nu_{CO}$) in its IR spectrum, with differing intensities depending on the conditions (Caughey et al. (1981) and references therein; Mourant et al. (1993) and references therein). Each has different average orientations of the CO dipole relative to the haem normal as established at 10 K by IR photoselection studies (Ormos et al., 1988): $A_0$ at 1966 $cm^{-1}$ (15 ± 10°), $A_{1,2}$ at 1948 $cm^{-1}$ (28 ± 2°) and $A_3$ at 1932 $cm^{-1}$ (33 ± 4°); whether the ligand is bent and/or tilted cannot be ascertained using this method. Each A state also shows different rebinding rates on photodissociation (Mourant et al., 1993).

There have been various structural interpretations of these A states. The $A_0$ state, being the predominant peak at low pH when the distal histidine is known to leave the haem pocket (Quillin et al., 1992), has been assigned to an "open" structure, i.e., free of distal interaction. Under physiological conditions, however, wild-type MbCO shows only two $\nu_{CO}$ frequencies, the $A_{1,2}$ and $A_3$ states. An early interpretation of these frequency shifts suggested they were caused by two different Fe-C-O conformations in the haem pocket (Moore et al., 1988). This "static" picture of the Fe-C-O geometry was largely based on the early crystal structures which resolved two sites for the carbonyl O (see Table 1); however, Henry (1993) has shown that over a 2 ns vacuum MbCO trajectory, the CO ligand has an essentially even distribution of conformations around an axis close to the haem normal. Consequently, there was no theoretical indication that the frequency shifts are attributable to significantly different Fe-C-O geometries in each A state. Recent IR (Li et al., 1994),

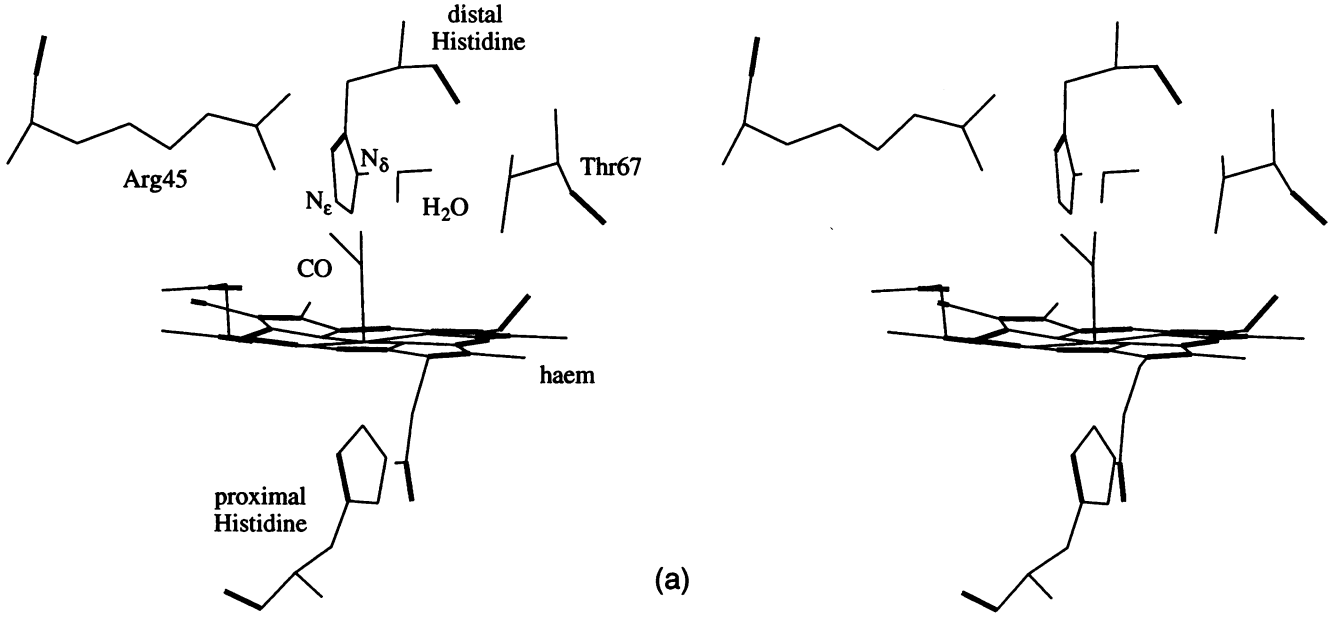


FIGURE 1 Stereo representations of certain key residues in the myoglobin haem pocket. (*a*) The "closed" 64$N_\delta$H haem pocket in the neutron diffraction structure (PDB file 2MB5); two carbonyl O positions were resolved. (*b*) The "closed" haem pocket of the "solution" phase 64$N_\delta$H simulation A. (*c*) The "closed" haem pocket of the "solution" phase 64$N_\varepsilon$H simulation B. (*d*) The "open" haem pocket in the x-ray structure of Mb bound with a bulky organic ligand, N-butyl isocyanide (PDB file 2MYA: Johnson et al., 1989). (*e*) The "open" haem pocket of the "solution" phase 64$N_\delta$H simulation B. The haem pocket geometries in *b*, *c*, and *e* were derived from the MD trajectories by averaging all coordinate sets corresponding to the appropriate "closed" or "open" conformations (see text).

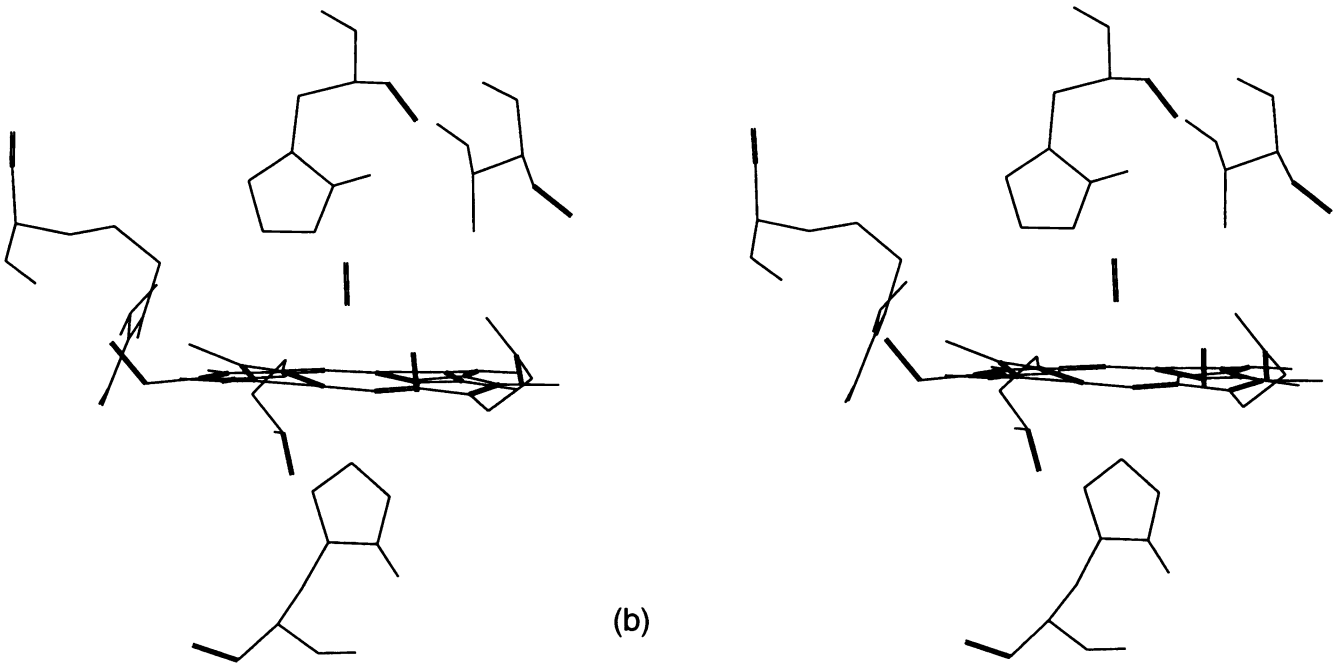


FIGURE 1 Continued

RR (Sakan et al., 1993), and x-ray (Quillin et al., 1993) studies have shown that there is no significant difference in the vibrational frequencies or the conformations of the Fe-C-O unit in a wide range of MbCOs with mutant distal residues of varying steric bulk. The vibrational frequencies, however, are found to vary with increasing polarity of the distal side chain (Sakan et al., 1993; Li et al., 1994). Consequently, Li et al. (1994) have proposed that the distal residue-CO interaction in wild-type myoglobins is largely electrostatic, *not steric*, in origin. This point of view has been independently supported by recent experimental studies of capped porphyrins (Ray et al., 1994), and ab initio calculations of a model haem prosthetic group (Jewsbury et al., 1994, in press). However, to account for the observed $\nu_{CO}$ frequencies, this interpretation requires the protonated nitrogen of the distal histidine side chain to point into the haem pocket, contradicting the neutron diffraction structure, the only direct experimental measure of the distal residue-CO conformation. In this interpretation, the proximity of the protonated distal nitrogen polarises the C═O bond to differing extents in each A state;

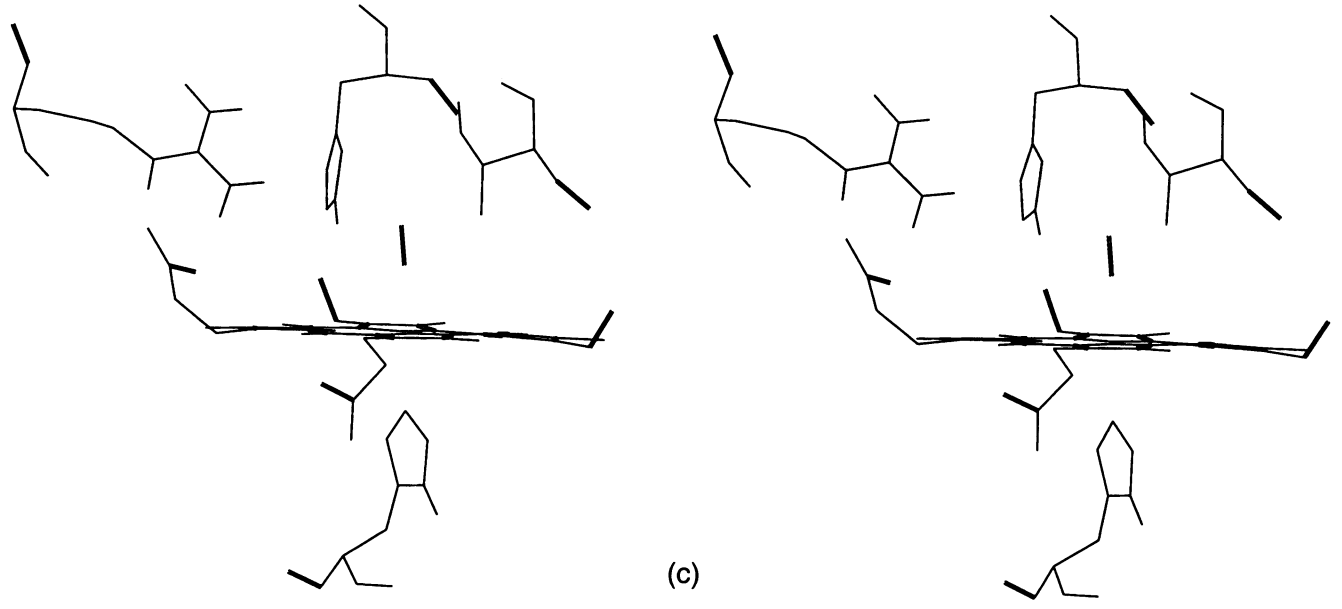


FIGURE 1 Continued

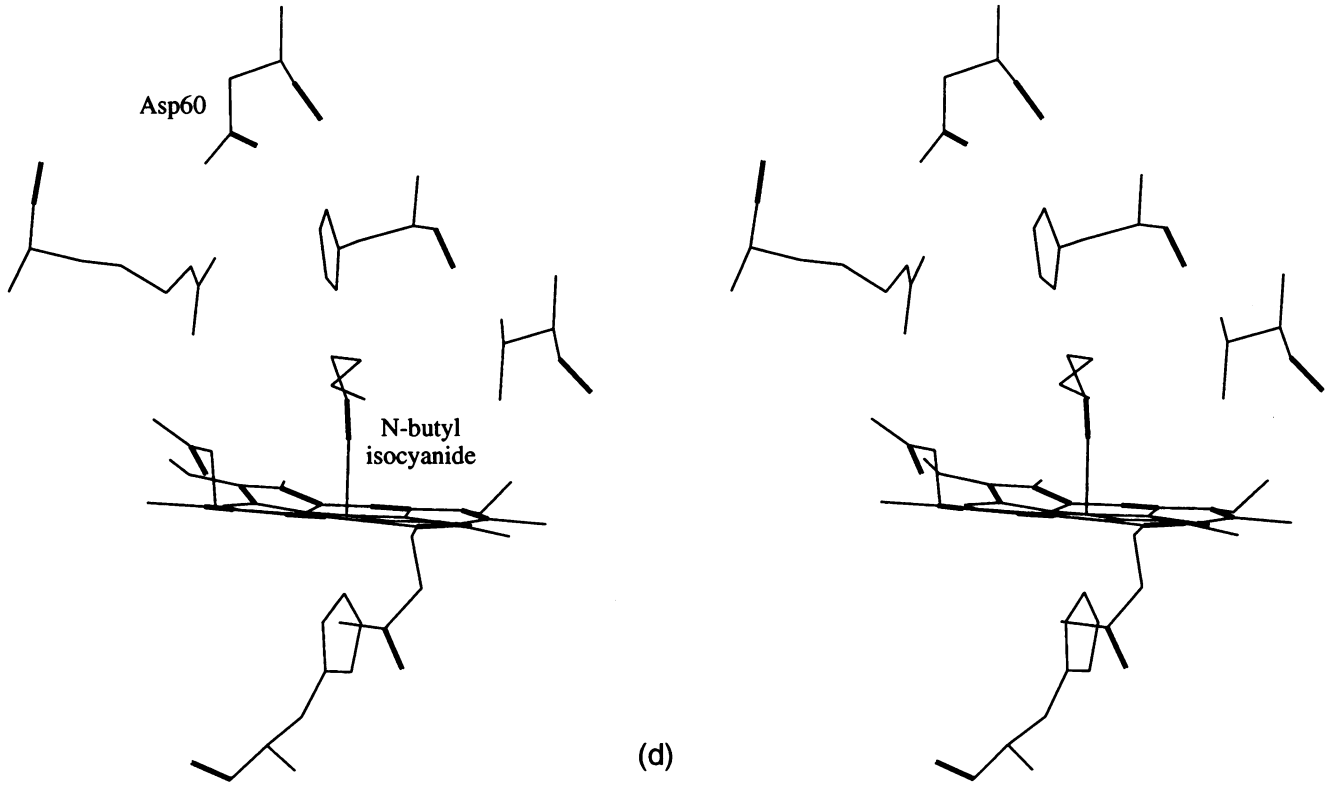


FIGURE 1 Continued

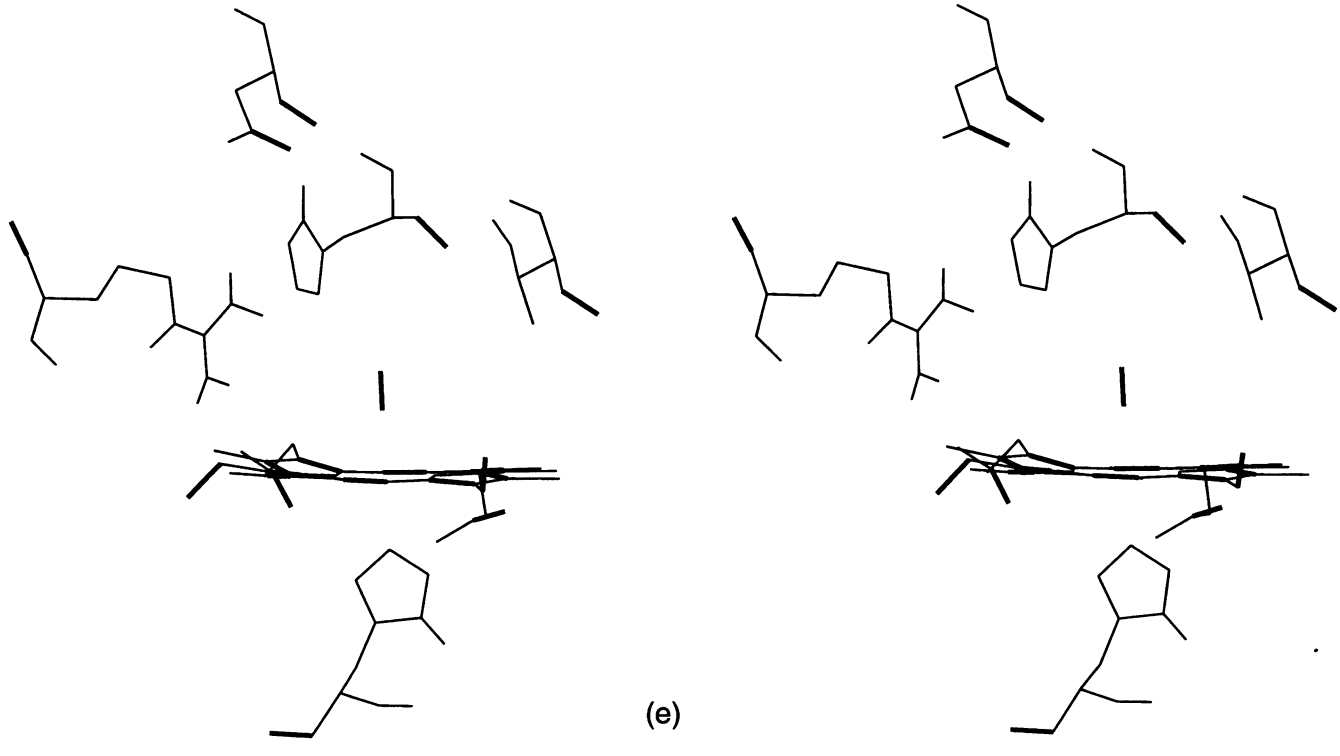


FIGURE 1 Continued

thus, the $A_{1,2}$ and $A_3$ states arise from two different haem pocket structures rather than from two different Fe-C-O geometries. Alternatively, Ray et al. (1994) have suggested that the $A_{1,2}$ band represents such a polarisation, but that the $A_3$ band represents a lone pair donation from the nonprotonated distal nitrogen to the carbonyl C. Oldfield et al. (1991) have proposed that the frequency shifts result from differing orientations of the distal side chain's dipole moment (depending on the orientation and protonation site of the imidazole ring) relative to the ligand. This paper addresses the structural interpretation of these $\nu_{CO}$ frequencies by considering the effect of the distal histidine tautomerism on the haem pocket structure.

Molecular dynamics is the theoretical method of choice to study the conformational dependence of the haem pocket-CO interaction because the interplay between the relative conformations of the CO ligand and the distal residue should be evident from the range of geometries explored during the trajectory. Only recently has a detailed force field been developed for the haem prosthetic group of the globin proteins (Kuczera et al., 1990), and a number of calculations have been published concerning the photodissociation dynamics of a CO ligand (considering the 64$N_\delta$H tautomer and the 64His$^+$ distal residues) (Straub and Karplus, 1991; Petrich et al., 1991), the vacuum dynamics of 64$N_\delta$H MbCO (Kuczera et al., 1990; Henry, 1993), and the diffusion of ligands through the protein (Elber and Karplus, 1990). The free energy difference between the two distal histidine tautomers of neutral pH MbCO has also been recently considered (Lopez and Kollman, 1993). In this study, a total of four 90 ps trajectories have been calculated, differing in primary structure only in the protonation site of the distal histidine side chain, which is shown in Fig. 2: two trajectories have the 64$N_\delta$ nitrogen protonated (which we will call simulations 64$N_\delta$H A and 64$N_\delta$H B), and two have the $N_\epsilon$ nitrogen protonated (simulations 64$N_\epsilon$H A and 64$N_\epsilon$H B). The starting geometries for each trajectory were established independently after a series of minimization, heating, and equilibration steps with a "computer fully-solvated" x-ray structure as the starting point. The final trajectories were run with only those water molecules within 16 Å of the carbonyl oxygen included (about 135 molecules), although the haem pocket can be considered to be effectively solvated. During the data collection, protein residues more than 16 Å from the carbonyl oxygen were weakly restrained to their starting positions to maintain the structural integrity of the solvated protein tertiary structure.

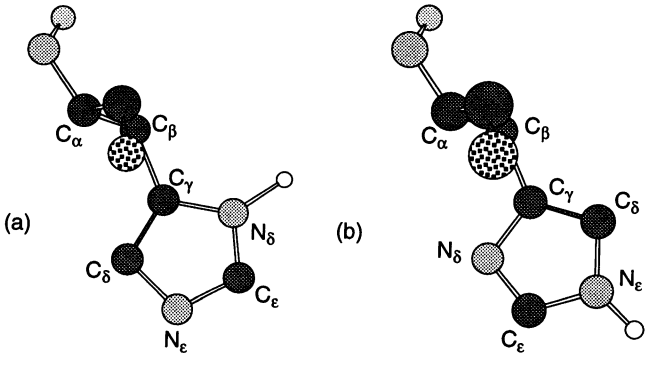


FIGURE 2 The $N_\delta$H (*a*) and $N_\epsilon$H (*b*) tautomers of the histidine side chain.

## MATERIALS AND METHODS

### Preparation of initial coordinates

The molecular dynamics trajectories were performed on a Sun workstation using AMBER4.0 (Pearlman et al., 1991). Initial geometries for each trajectory were developed independently using the following protocol. The initial protein coordinates were taken from the PDB file 1MBC (Kuriyan et al., 1986), and the protein (including crystal waters) was immersed in a "blob" of TI3P (Jorgensen et al., 1983) water molecules 8.5 Å thick. Standard AMBER4.0 topologies were used for establishing the positions of polar hydrogens. Table 2 gives the protonation sites of the histidine residues: these were assigned from the MbCO neutron diffraction structure (PDB structure 2MB5 (Cheng and Schoenborn, 1991)) wherever a significant preference was found; other histidine residues were protonated as described in Table 2. Standard AMBER4.0-united atom force fields were used for the protein residues (except the proximal histidine, see Appendix) and a new AMBER united atom force field was developed for the haem prosthetic group (see Appendix). The fully solvated protein was then energy-minimized for 100 steps with the SHAKE option (Ryckaert et al., 1977) switched off, followed by a further 400 steps with the SHAKE option on for all bonds to a hydrogen. This partially minimized geometry was then heated to 300 K in 50 steps of 6 K over 10 ps: each 0.2 ps trajectory was started at the final geometry of the previous trajectory, the time step in the integration was 0.002 ps, the nonbonded pair list was regenerated every 12 steps, the nonbonded cutoff was 8.0 Å, and the system loosely coupled, 0.2 ps, to a solvent bath at the relevant temperature (Berendsen et al., 1984). All noncrystal waters were restrained to their initial positions in that trajectory by a harmonic potential of 5.0 kcal/mol Å$^2$, and the initial velocities at the relevant temperature were randomly assigned at the start of each 0.2 ps step. The SHAKE algorithm was on for bonds to a hydrogen. The resulting geometry was then equilibrated at 280 K over two runs of 5 ps under the same conditions as above (although the restraining potential for solvent molecules was relaxed to 2.0 kcal/mol Å$^2$ in the second 5 ps trajectory) with the restraining coordinates and initial velocities reassigned between trajectories.

This structure was then taken as the starting structure for the data generation, although only water molecules within 16 Å of the carbonyl O (about 135 water molecules) were included to make the calculations tenable in real

**TABLE 2 Protonation site of the Mb histidine residues in this work**

| Residue no. | 64 $N_\delta$H | 64 $N_\epsilon$H | Neutron structure |
|---|---|---|---|
| 12 | x | | x |
| 24 | x | | |
| 36 | x | | x |
| 48 | | x | |
| 64 | n/a | n/a | |
| 81 | | x | x |
| 82 | | x | x |
| 93 | x | | x |
| 97 | | x | x |
| 113 | x | | x |
| 116 | x | | x |
| 119 | | x | |

An entry under "Neutron Structure" indicates the tautomer was assigned according to a high population density in the neutron study of Cheng and Schoenborn (1991) (PDB file 2MB5).

time. Protein residues further than 16 Å from the carbonyl O (residues 1–20, 50–57, 78–84, and 116–153) were restrained to their initial positions by a 2.0 kcal/mol Å$^2$ potential to maintain the structural integrity of the "solvated" protein during the simulation, and water molecules initially between 12 and 16 Å of the carbonyl O (about 100 of the water molecules) were restrained by 0.5 kcal/mol Å$^2$ to their initial position to prevent boiling off of the solvent. A distance-dependent dielectric function was used because the protein is not considered fully solvated. Four "solvated ions" (Singh et al., 1985), two positive and two negative, were also included in their energy-minimized positions near the exposed side chains of charged residues that were not involved in salt bridge interactions in the initial geometry. These "ions" were also restrained by 2.0 kcal/mol Å$^2$ harmonic potentials. The number of unrestrained centres was 823, and the number restrained was 712. 100 ps trajectories were calculated (integration step 0.001 ps, nonbonded pair list regenerated every 25 steps, nonbonded cutoff 10 Å, SHAKE algorithm on for bonds to hydrogen, system loosely coupled, 0.2 ps, to a solvent bath at 280 K) with data collected every 0.1 ps over the final 90 ps.

## Protein solvation

Early molecular dynamics simulations were performed with the protein in a vacuum, or with an "averaged" solvent implicitly included using stochastic dynamics. This was necessary because of the considerable memory and cpu demand of explicitly including a solvent. More recently, solvent molecules have been included explicitly giving much more realistic simulations of the protein in its physiological environment, although the demand on the computing resource is considerable. The inclusion of water is desirable in this study, however, because the haem pocket, and in particular the distal residue, is exposed to the bulk solvent.

In this work, we concentrate on the structure of the haem pocket and are not concerned with residues far from the active site. Consequently, we have chosen to include explicitly only those water molecules that lie close to the active site of a solvated, equilibrated MbCO molecule. The distal histidine lies around 4.5 Å from the carbonyl C in the x-ray structure, with the outward-facing edge of the imidazole ring in contact with the solvent (see Fig. 1). Since all water molecules within 16 Å of the carbonyl O have been included in these trajectories, and the nonbonded cutoff is 10 Å, we can consider the distal residue to be at least partially solvated. Full solvation would require that the distal histidine's solvent shell also be embedded in 10 Å of water and also include sufficient water molecules to allow for any movement of the distal histidine out of the haem pocket. Such a treatment, however, was beyond the computer resources available. More than 100 of the water molecules included in these trajectories lie over the haem pocket, and those molecules closest to the pocket can (and do) migrate freely in and out of the active site. The most distant water molecules have been restrained, however, to create a "skin," in particular over the haem pocket, preventing significant "boiling" away of the mobile solvent molecules. The distant residues have also been restrained to their solvated, equilibrated positions to maintain the solvent structure of those residues exposed to vacuum during the data collection stage of the calculation.

Since significant "boil off" of solvent molecules was not observed, the integrity of the "solution" phase was maintained throughout each of the trajectories.

## Stability of tertiary structure during trajectory

The average rms deviation of the protein backbone structure from that of the x-ray structure (PDB file 1MBC) of each trajectory was 1.123 ± 0.047 and 0.908 ± 0.032 Å for the two 64N$_\delta$H trajectories, and 1.080 ± 0.027 and 0.910 ± 0.033 Å for the two 64N$_\epsilon$H trajectories (where the uncertainty given is 1 SD). The rms deviations did not show large changes in value during the data collection stages of the simulations.

It is usually found that the backbone structure of the heated and equilibrated geometries used to initiate MD simulations deviate from the x-ray crystal structure by around 1 Å (see, for example, Kuczera et al., 1990; van Gunsteren and Mark, 1992; Henry, 1993). The backbone structure then continues to drift from the x-ray structure for up to 500 ps in a series of rapid rearrangements superimposed on largely stable segments of the trajectory, before stabilizing around rms deviations of 2 Å (van Gunsteren and Mark, 1992; Henry, 1993; Saito, 1994). Even during "stable" segments of the trajectory, the backbone structure still undergoes significant motion without drifting further from the x-ray structure (Henry, 1993). It is expected that the elevated temperature and different environment experienced by the protein in the calculation will lead to simulations deviating from the x-ray crystal structure, especially for surface residues of proteins in vacuo; however, artifacts in the calculation method (e.g., the finite cutoff used for the nonbonded interactions) may also, in some cases, cause a gradual drift from the initial structure over long time periods (see, for example, Saito, 1994. In general, it is not trivial to establish that a protein simulation has achieved equilibration (van Gunsteren and Mark, 1992): the length of the simulation cannot be increased indefinitely unless the protein is completely solvated, and all interactions included. This was impractical in this study. Since we are only concerned with the localized motions within the haem pocket, we have chosen to calculate a number of shorter simulations, independently initiated, and to draw conclusions from the similarities, or otherwise, in their behavior. Because no large jump in the total energy of the system, or systematic drift in the protein backbone structure, was observed over these relatively short trajectories, they can be considered realistic simulations of the protein dynamics within a limited region of the protein's conformational space.

# RESULTS AND DISCUSSION

## General description of the trajectories

In three of the simulations (64N$_\delta$H simulation A, and 64N$_\epsilon$H simulations A and B), the distal side chain stayed mainly inside the haem pocket creating "closed" pocket structures, but for most of 64N$_\delta$H simulation B the distal side chain lay outside the haem pocket leaving an "open" pocket. Stereo representations of the "closed" pocket structures of simulations 64N$_\delta$H A and 64N$_\epsilon$H B are shown in Fig. 1, *b* and *c*; the "open" pocket structure of simulation 64N$_\delta$H B is shown in Fig. 1 *e*. Fig. 3 shows the distance between the center of the distal imidazole ring and the carbonyl C as a function of time in each of the four trajectories; the separation in the starting x-ray structure, 1MBC, where the distal side chain forms part of the haem pocket, is 4.22 Å. If a separation of less than 5.5 Å is taken as indicative of a "closed" pocket structure, and a separation greater than 6 Å as an "open" pocket structure, then simulations 64N$_\epsilon$H A and B and 64N$_\delta$H A adopt "closed" pockets over 96, 96, and 70% of their trajectories, respectively, whereas 64N$_\delta$H B is "open" for over 74% of the trajectory. In the "open" conformations, the distal side chain still remains inside the unrestrained region of the solvent (up to 12 Å from the carbonyl O); however, the more restrained motion of the water molecules in the solvent "skin" (12–16 Å from the carbonyl O) may artificially prolong the lifetime of this conformation. When inside the haem pocket, the distal side chain is for the most part orientated with the protonated N inside the pocket.

The distal-CO interaction differs depending on the tautomeric state of the distal side chain. Fig. 4 shows the separation between the distal side chain polar hydrogen and the carbonyl O during the four trajectories. The close interaction between the distal polar hydrogen and the carbonyl O in the

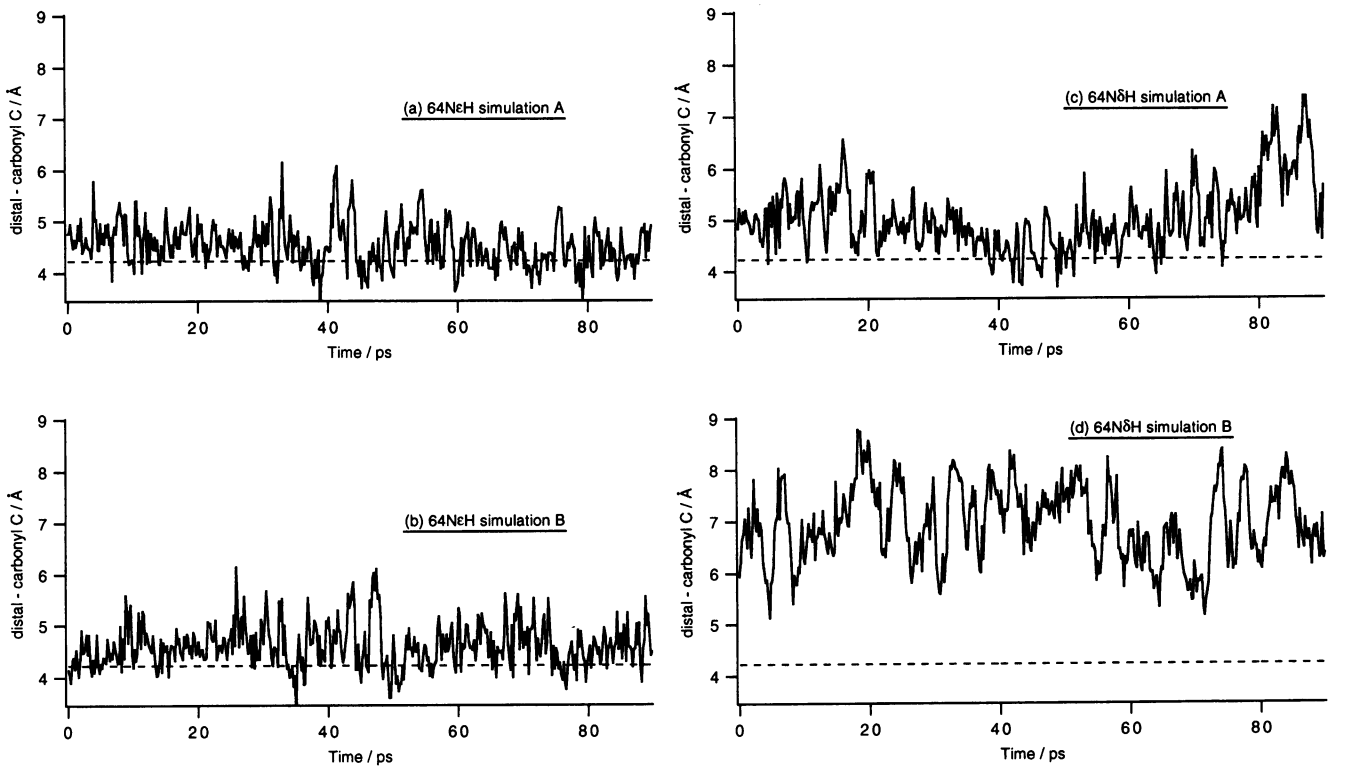


FIGURE 3 Separation between the centre of the distal side-chain ring and the carbonyl C during the four trajectories. The dotted line shows the separation in the x-ray structure (1MBC), 4.22 Å.

"closed" pocket structures of the 64$N_\epsilon$H tautomer is stable in the time scale of these simulations (Fig. 4, *a* and *b*), whereas the distal side chain of the 64$N_\delta$H tautomer appears to be more mobile (Fig. 4 *c*). The average position of the CO unit in the coordinate sets corresponding to "closed" (or "open") conformations of each trajectory makes angles with the haem normal of 3.1° and 6.6° for simulations 64$N_\epsilon$H A and B, and 2.2° and 2.8° for simulations 64$N_\delta$H A and B, respectively[1], where the normal has been calculated for the average plane through the 24 heavy atoms of the porphyrin macrocycle. Although the differences are small, the "closed" pocket structures of the 64$N_\epsilon$H tautomer are more distorted toward the distal side chain than those of the 64$N_\delta$H tautomer, reflecting the stronger interactions implied in Fig. 4, *a* and *b*. Larger distortions would be expected for the "closed" pocket structures if polarization functions were included in the active site force field. The CO ligand is known to be influenced by the Val-68 residue (Egeberg et al., 1990; Li et al., 1994) which lies, close enough to the haem to repel sterically the ligand in the direction of the distal residue (P. Jewsbury, unpublished results). The absence of counteracting interactions with the distal residue in the "open" conformations of simulation 64$N_\delta$H B account for the small distortions from the perpendicular found even in the "open" state.

[1] Note that this differs from the average orientation of the CO dipole (see later).

The differences in the distal-CO interactions of each tautomer are caused by the differing interactions available to the polar hydrogen depending on its position in the distal side chain. Protonation at the $N_\epsilon$ nitrogen allows for an electrostatic interaction with the carbonyl oxygen, stabilizing the "closed" pocket structure, but allows little interaction with nearby side chains (see Fig. 1 *c*). The 64$N_\epsilon$H-O interaction is therefore long-lived; the 64$N_\epsilon$H-O separation is, however, too large, and the 64$N_\epsilon$-H-O angle too far from linear, for this to represent a hydrogen bond, and no deuteration dependence is observed experimentally for the $\nu_{CO}$ stretching vibration. Protonation at the $N_\delta$ nitrogen has a reduced interaction with the ligand, because of an unfavorable geometry, but more favorable interactions with nearby residues (see Fig. 1 *b*). In the "open" structure, the 64$N_\delta$H hydrogen forms a long lived hydrogen bond with the main chain carbonyl O of Asp-60, as can be seen in Fig. 1 *e* and the plot of their separation in Fig. 5 (a separation of less than 2 Å is indicative of hydrogen bonding). In both tautomers, the unprotonated distal nitrogen interacts with the flexible Arg-45 side chain or solvent molecules, and so is not as important in determining the relative mobility of the distal side chain. The 64$N_\delta$H tautomer is therefore more likely to move between pocket structures than the 64$N_\epsilon$H tautomer.

Hydrogen bonding between the non-protonated distal N and the Arg-45 side chain is important in determining the width of the "entrance" to the active site. When hydrogen-bound to the distal side chain in a "closed" conformation, the

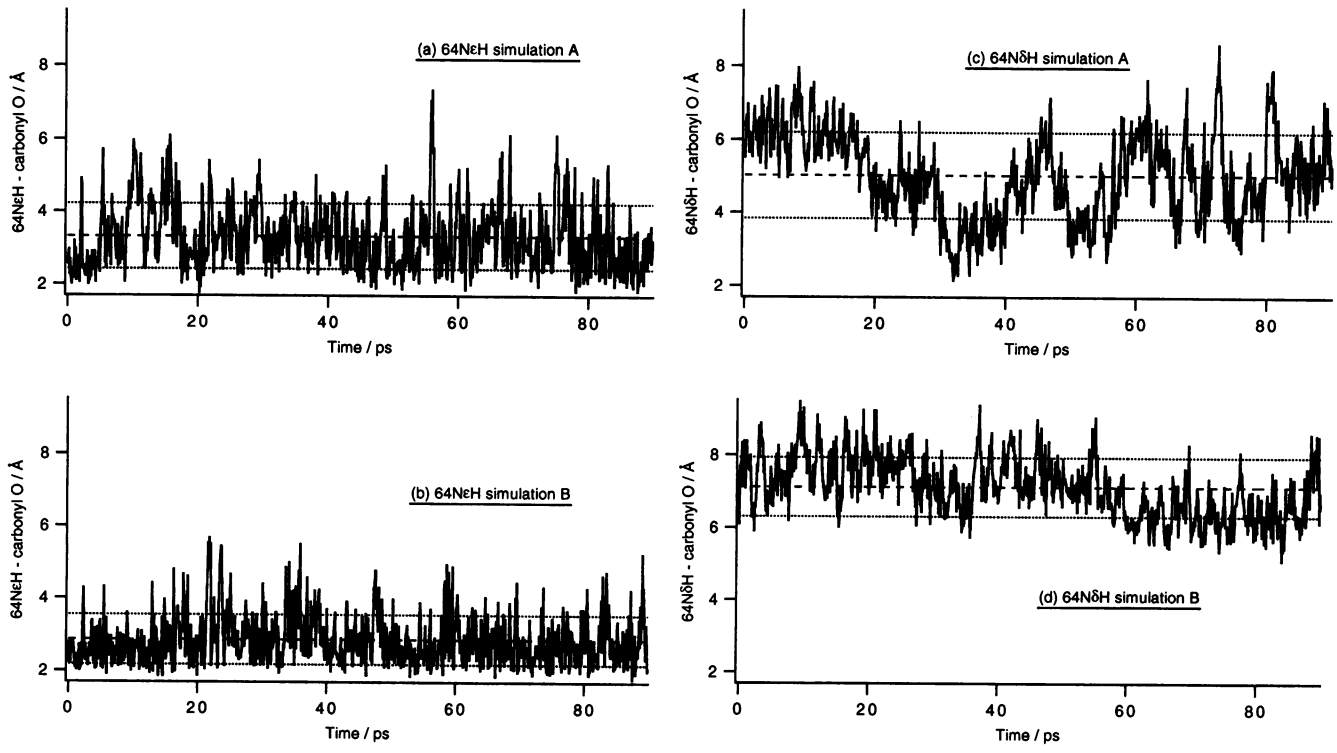


FIGURE 4 Separation between the distal side-chain polar hydrogen and the carbonyl O during the four trajectories. The dashed lines represent the average separations (3.32 and 2.86 Å for simulations 64$N_\epsilon$H A and B, and 5.02 and 7.12 Å for simulations 64$N_\delta$H A and B, respectively) over the whole trajectory, and the dotted lines 1 SD.

long Arg-45 side chain is held close to the haem pocket leaving a crowded entrance to the active site. Typical separations between Arg-45 $C_\zeta$ and Thr-67 Oγ are 6–8 Å. A significantly wider "entrance" to the active site is found if the hydrogen bond is broken, in which case the Arg-45 side chain can move out into the solution, or if the distal side chain is in an "open" conformation. In such conformations typical separations between the Arg-45 $C_\zeta$ and Thr-67 Oγ atoms are 10–12 Å.

## Comparison with crystal structures

The average positions of the CO ligand over the whole of each trajectory make angles relative to the haem normal of 3.0° and 6.5° for simulations 64$N_\epsilon$H A and B, and 2.1° and 2.9° for simulations 64$N_\delta$H A and B, respectively. On average, the simulations do not show the large displacements that are found in the x-ray structures (Table 1), nor does any one trajectory show a preference for two distinct Fe-C-O conformations. Henry (1993) has also found that, over a 2 ns vacuum MD simulation of MbCO, the CO ligand geometries are distributed isotropically around an axis close to the haem normal, although the large steric repulsion predicted for the x-ray structure would be unlikely near the surface of a protein in vacuo.

There is some uncertainty in resolving the Fe-C-O geometry in the x-ray structures: see Table 1, and Appendix 2 of Ray et al. (1994). Li et al. (1994) have recently concluded

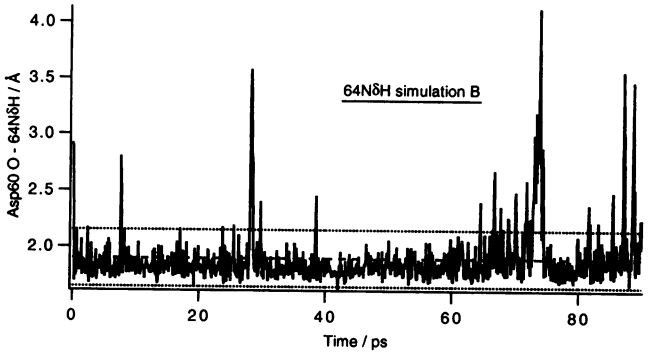


FIGURE 5 Separation between the distal side-chain polar hydrogen and the main chain oxygen of Asp-60 during 64$N_\delta$H simulation B. The dashed line represents the average separation (1.90 Å) over the whole trajectory, and the dotted lines 1 SD.

that there is no significant difference between the Fe-C-O conformations observed in a number of mutant carbonmonoxy myoglobins, despite recording bend angles in the range 156° to 171° and tilt angles (relative to the $N_A$ porphyrin nitrogen) in the range 84° to 96°. Furthermore, it is difficult to reconcile the two CO configurations in the neutron diffraction crystal structure with the tautomeric state of the distal side chain assigned by Cheng and Schoenborn (1991): one conformation has the carbonyl O repelled from the distal $N_\epsilon$ lone pair, as would be expected; however, the other conformation has the O moved *closer* than the perpendicular (see

Fig. 1 *a*). This suggests an attraction to the $64N_\epsilon$ nitrogen, i.e., that the $64N_\epsilon$ nitrogen is protonated in at least some of the MbCO molecules, contrary to Cheng and Schoenborn's assignment. The presence of the $64N_\epsilon H$ tautomer in the crystal would, however, explain the crystalline IR spectrum (Ray et al., 1994).

In an ab initio study of a model haem prosthetic unit based on the 1MBC geometry, however, Jewsbury et al. (1994, in press) have recently reproduced the distortions of the Fe-C-O unit reported in the x-ray structure. The distortions were not caused by the distal residue, but rather by the nonequilibrium orientation of the proximal residue through its competitive bonding to the Fe d orbitals. If such a mechanism is primarily determining the CO distortion in the protein, then current MD simulations will not correctly reproduce the distortions found in the x-ray structures since such a direct coupling is not allowed for in the AMBER or CHARMM force fields; the significance of the disparity between the x-ray and MD FeCO conformations, therefore, is not clear.

The dominant "closed" pocket distal side-chain orientations found in these simulations are quite different from that of the neutron diffraction crystal structure: compare Fig. 1 *a* with Fig. 1, *b* and *c*. In each simulation, although initiated from the x-ray coordinates and generated entirely independently of each other, the distal side chain has rotated to bring the protonated N into the haem pocket (or in one case the whole ring was outside the pocket). These orientations remained the dominant conformations throughout the trajectories: a conformation with the protonated $64N_\delta H$ pointing out of the haem pocket, similar to the neutron diffraction structure, was found to be only short-lived. The difference in the apparent distal conformations in the crystal and "solution" phases probably results from differences in the solvent shell local to the distal side chain. Cheng and Schoenborn (1991) resolved a well structured water network interacting with the Arg-45 and distal side chains, and anchored by the porphyrin IX propionic side chains, in their $P2_1$ crystal. One water molecule, shown in Fig. 1 *a*, forms a hydrogen-bonding bridge between the $64N_\delta H$ and the Arg-45 side chain and has a crystallographic B-factor of 31 Å$^2$; another, with a B-factor of 27 Å$^2$, forms part of a hydrogen bonding network to an $(NH_4)^+$ ion, which has a B-factor of 25 Å$^2$ and is held between the solvent exposed propionic side chains of the porphyrin IX macrocycle, around 6 Å from the distal side chain. A similar hydrogen-bonding network of four water molecules, but no counterion, was resolved with B-factors from 19 to 31 Å$^2$ in the P6 crystal structure 2MGK (Quillin et al., 1993), but not in the $P2_1$ structure 1MBC (Kuriyan et al., 1986). In comparison, the B-factors for the distal side chain are around 15 Å$^2$ in these crystals. The solvent molecules near the distal side chain show considerably more mobility in the simulations and have rms fluctuations equivalent to "B-factors" of ≈150 Å$^2$. It is likely that the increased mobility of the solvent in these "solution" phase simulations disrupts the structured water network, and the distal side chain forms longer lived interactions with the less mobile protein residues. Brown et al. (1983) have shown that the IR spectra of MbCO films are sensitive to the extent of hydration: removal of water causes a dramatic, but reversible, conversion in favor of the $A_0$ conformer. Below we interpret the three MbCO A states as representing different orientations of the distal side chain relative to the CO group; thus, Brown's result suggests that the extent of hydration has a significant influence on the distal side chain conformation. It is likely that this sensitivity accounts for the different distal conformations observed in the neutron diffraction structure and in these "solvent" phase simulations.

Soon after the wild-type crystal structure was determined, it was realized that motion of ligands in and out of the haem pocket was severely restricted by the distal pocket structure: the dynamics of the distal and nearby residues are important in reducing the barrier to ligand motion to and from the active site (Case and Karplus, 1978; Kuriyan et al., 1986). Kuriyan et al. (1986) proposed a reaction coordinate for ligand entry involving a combination of Arg-45 and His-64 side-chain motions, creating an "open" pocket structure. The importance of the combined motions of the Arg-45 and His-64 side chains in determining the width of the "entrance" to the active site as revealed in these simulations has been noted above. The likely structure of this "open" haem pocket was determined experimentally by resolving the crystal structures of myoglobin with bulky organic ligands bound at the Fe atom (Ringe et al., 1984; Johnson et al., 1989). These crystal structures all show the distal histidine side chain forced out of the haem pocket and hydrogen bonding to the Asp60 main chain oxygen, as shown in Fig. 1 *d*; simulation $64N_\delta H$ B shows a very similar pocket structure, Fig. 1 *e*, confirming that such an "open" conformation is possible for the wild-type protein, even with a small ligand bound at the active site.

## Comparison with solution spectroscopy: the electric field at the CO ligand

As discussed in the introduction, MbCO shows three characteristic IR $\nu_{CO}$ bands, seen with varying intensity depending on the conditions. The two peaks seen in the room temperature, solution phase, spectra, $A_{1,2}$ at 1944 cm$^{-1}$ (20 ± 8°), and $A_3$ at 1933 cm$^{-1}$ (35 ± 11°) have been variously interpreted to represent either varying degrees of electrostatic polarization by the distal histidine (Li et al., 1994), that and a lone pair bonding interaction with the distal side chain (Ray et al., 1994), or the different possible orientations of the imidazole dipole moment (Oldfield et al., 1991).

Ray et al. recently proposed (following the original suggestion of Maxwell and Caughey, 1976) that the $A_3$ peak corresponds to a weak dative bond between the non-protonated distal N and the carbonyl carbon. Despite the favorable electrostatic interaction possible in the force field used here, such a geometrical configuration is not approached in any of these "solution" phase trajectories. Their suggestion was based on the crystal MbCO IR spectrum, which has an enhanced $A_3$ intensity, and for which the distal

$N_\delta$ nitrogen is known to be protonated and the distal $N_\epsilon$ nitrogen lone pair orientated inside the haem pocket. They proposed that with such a dative bond to the carbon, the CO bonding will gain $sp^2$ character, leading to a bent Fe-C-O conformation and reduced CO stretching frequency (if the CO bonding is pure $sp^2$, then the Fe-C-O bend angle is expected to be 120°: the angle observed in the first x-ray structure, 1MBC). However, to attain such a dative bond, the distal side chain would need to move closer to the binding site by more than 1 Å, significantly disrupting the protein backbone structure. It seems improbable that such a dative bonding is determining the $A_3$ frequency.

The other two interpretations are essentially two aspects of the same model: that the CO bond is polarized by a positive electric field caused by the distal histidine, with a consequent decrease in the $\nu_{CO}$ frequency caused by an increase in the Fe-CO backbonding. They differ slightly in their details, however. Oldfield et al.'s interpretation considers the orientation of the imidazole dipole relative to the CO bond, and proposes that the $A_3$ state (the lowest frequency) has the imidazole dipole anti-parallel to the CO bond, the $A_0$ state (the highest frequency) with the imidazole dipole parallel, while the two intermediate frequencies (the $A_1$ and $A_2$ states that are not resolved separately at room temperature) have the dipole orientated near perpendicular to the CO bond. This interpretation has been challenged because it suggests that the unpolarized stretching frequency should be those of the $A_{1,2}$ states (i.e. around 1945 $cm^{-1}$), whereas the low pH spectrum, where interactions with the distal side chain are thought to be absent, has the $A_0$ state as its major peak.

Li et al.'s interpretation is slightly different from Oldfield's in that it considers that the polarization of the CO bond arises from the proximity of the protonated distal side chain NH causing a positive electric field at the carbonyl O, leading to increased backbonding electron donation from the Fe centre and, thus, a lower C═O stretching frequency. The two peaks observed in the solution spectrum then correspond to two dominant haem pocket conformations with differing extents of polarization of the C═O bond (i.e., differing proximity of the distal polar H). Both Li's and Oldfield's interpretations are consistent with the IR results; however, they require that the distal side chain be orientated with the protonated N inside the haem pocket, contrary to the neutron structure.

The results of these simulations are consistent with Li et al.'s interpretation. The two conformations they propose would correspond to the "closed" pocket structures of the two distal histidine tautomeric states at neutral pH. In both cases the preferred orientation of the distal side chain is with the protonated nitrogen inside the haem pocket, however the extent of distal-CO interaction differs in each tautomeric state. For the $64N_\epsilon H$ tautomer the position of protonation gives a favorable geometry for a stable interaction with the carbonyl ligand O; in the case of the $64N_\delta H$ tautomer, the site of protonation of the imidazole ring does not allow such a strong interaction with the ligand, but does retain the side chain inside the haem pocket for long periods, allowing for a weaker influence over the CO polarization. The third frequency, the $A_0$ state, would correspond to the "open" pocket structure seen in simulation $64N_\delta H$ B. The proximity of the distal NH group will cause changes in the electrostatic field at the binding site, polarizing the CO ligand and leading to shifts in the CO stretching frequency. Straub and Karplus (1991) have calculated that the CO stretching frequency of a hydrogen bonded CO-imidazole dimer is 17 $cm^{-1}$ lower than that of free CO. Although the interaction of Fe bound CO with the protein is expected to be weaker because of a less favorable geometry, the Fe-C-O unit itself is more polarizable; it is therefore plausible that the presence, or otherwise, of the distal side chain can account for the observed shifts in the $\nu_{CO}$ frequency of the A states. Lopez and Kollman (1993) have recently found a 0.74 kcal/mol free energy difference between the two distal histidine tautomers of MbCO, suggesting, as found here, that there are significant differences in the distal-CO interaction between the two tautomers. Their calculation was performed, however, in vacuo, and their result is subject to a large error (±0.78 kcal/mol).

Assigning the A states to different haem pocket structures is only applicable in broad terms, because the two "closed" pocket structures, in particular, are sufficiently similar to explore overlapping regions of conformational space in their dynamics. The local fluctuations in the orientation of the distal side chain would, as Oldfield et al. have suggested, lead to a range of electrostatic fields experienced by the ligand in the active site. This is reflected in the room temperature IR spectrum by the broad, overlapping nature of the $\nu_{CO}$ bands (see, for example, Li et al., 1994). Similarly, although the "open" structure represents a weak distal-CO interaction, its "open" nature will allow solvent molecules to migrate into the haem pocket and interact with the CO ligand, resulting in a polarized CO stretching frequency similar to the $A_{1,2}$ and $A_3$ states. Such transient hydrogen bonding accounts for the broad peak stretching from around 1935 $cm^{-1}$ to 1980 $cm^{-1}$ in the IR spectrum of the His-64Gly mutant (Li et al., 1994; Jewsbury and Kitagawa, 1994, unpublished data).

Li et al. (1994) measured the $\nu_{CO}$ frequencies of 41 carbonmonoxy myoglobins with mutations at various sites in the haem pocket, in particular at the distal residue. The x-ray structures of some of these mutants have also been solved (Carver et al., 1992; Quillin et *al*, 1993). Of the 11 distal mutants, only His-64Met, His-64Gln, and His-64Trp showed any peak other than at the $A_0$ frequency. MD simulations of the His-64Gln, His-64Leu, and His-64Gly mutants demonstrate the influence of a polar distal side chain, a hydrophobic distal side chain, and the presence of solvent waters in the haem pocket (Jewsbury and Kitagawa, 1994, unpublished data) and are consistent with the structural interpretation of the wild-type IR spectrum presented here. These trajectories support Li et al.'s picture that steric interactions with the distal side chain are in general not important in determining the C═O stretching frequency: only those distal mutants

with a side-chain structure that orientates a polar residue inside the haem pocket should result in shifts from the $A_0$ frequency.

Although the extension to other MbCO mutants is not certain until further studies are performed, structural interpretations of some of the key IR spectra reported by Li et al. can be suggested. The Asp-45 residue is known to interact with the distal side chain, both in the x-ray structures and the MD trajectories reported here; thus, it is expected that mutations of this residue will have a significant effect on the distal side chain's mobility. Li et al. have recorded the IR spectrum of the Arg-45Glu mutant at neutral and physiological pH. The glutamate side chain is significantly shorter than that of asparagine, making the bonding to the distal side chain that stabilises the "closed" pocket structures in the wild-type less likely in this mutant. Consequently, the "closed" pocket structures of this mutant are expected to be less stable than those found in the wild type, reducing the populations of the $A_{1,2}$ and $A_3$ conformers relative to the $A_0$ conformer. The proximity of the negatively charged glutamate side chain will, however, raise the $pK_a$ of the distal side chain, as Li et al. have suggested, so that the onset of significant populations of the $64His^+$ tautomer, and the increase in the $A_0$ intensity that this would correspond to, is expected at higher pH values than for the wild type. Li et al. report only a small redistribution of the A state intensities for this mutant relative to the wild-type spectra at pH 7.0: $A_0$ 19% (wild type: 0%), $A_{1,2}$ 59% (70%) and $A_3$ 22% (30%); however, at pH 5.5 the $A_0$ band has increased to about 50% of the total intensity in the mutant spectrum, whereas there is little change observed for the wild type. The Leu-29Phe mutation is also observed to have a considerable influence on the IR spectrum: the Leu-29Phe mutant has only one peak, at the $A_3$ position. The x-ray structure of this mutant is almost isomorphous with that of the wild type, with only a local disruption of the tertiary structure in the vicinity of the mutation (Leu-29Phe MbCO PDB structure 2SPL: Carver et al., 1992). The preliminary results of MD simulations of this mutant suggest that the bulky Leu-29Phe side chain separates the B and E helices to which it and the distal side chain are attached, disrupting the favorable interaction reported for the wild-type $64N_\epsilon H$ tautomer here. Both distal tautomers in this mutant show mobility characteristic of the $64N_\delta H$ wild-type tautomer (i.e., the $A_{1,2}$ state): Li et al. have suggested that the shift to $A_3$ frequencies may result from additional polarization of the CO group by the Leu-29Phe side chain that lies 3.0 Å from the carbonyl O in the x-ray structure, although it is not clear why the Leu-29Trp or Val-68Phe mutants do not show similar effects. An ab initio study of the electrostatic fields around these side chains and their likely effect on a carbonyl ligand in MbCO is in progress. Li et al. report that decreasing the bulk of the Leu-29 side chain to Ala, Val, or Ile increases the $A_3$ population at the expense of the $A_{1,2}$ population. It is possible that reduced bulk at the 29th residue allows the E helix, and therefore the distal side chain, to approach the binding site more closely; electrostatic interactions with the haem prosthetic group may then influence the relative $pK_a$s of the distal tautomers in favor of $64N_\epsilon H$, because it can approach the closer.

## Comparison with solution spectroscopy: orientation of the CO dipole

Ormos et al. (1988) have measured the angle made by the C═O dipoles relative to the haem plane normal for the spectroscopic states observed for a 10 mM solution of wild-type MbCO at 10K: $A_0$ (15 ± 10°), $A_{1,2}$ (28 ± 2°), and $A_3$ (33 ± 4°). Moore et al. (1988) have also measured these angles for the states found at room temperature in solution: the $A_{1,2}$ (20 ± 8°)and $A_3$ (35 ± 11°) states. Moore et al. suggested that the two room temperature results correspond to the two conformations observed in the x-ray structure, even though the orientations of the CO group in the x-ray structures are quite different (see Table 1), assuming that the differences were down to uncertainties in the experimental data. Henry has shown, however, that the dynamics of the prosthetic group contribute significantly to the angle measured in the photoselection experiments, and so the x-ray and IR results are not directly comparable (Henry, 1993).

The average angle made by the CO dipole relative to the haem plane normal can also be calculated for these simulations: they are 19.4 ± 10.7° and 19.1 ± 10.9° for the "closed" pocket conformations of simulations $64N_\epsilon H$ A and B, and 19.2 ± 10.7° and 20.2 ± 11.6° for the "closed," and "open," pocket conformations of simulations $64N_\delta H$ A and B. The average positions of the CO atoms derived from the same data sets, as would be found in an x-ray structural analysis are, however, much closer to the haem normal (see above). As Henry has shown, the apparent discrepancy between the average CO geometries and dipole orientations can be explained by considering a bent Fe-C-O unit rotating around the Fe-C bond: the CO bond describes a cone with the carbonyl C at its apex and the Fe-C bond as its axis. The orientation of the CO bond relative to the Fe-C axis is the same in all positions, i.e., the bend of the Fe-C-O unit; thus, the average orientation of the CO dipole is non-zero. This would be equivalent to the photoselection experimental result. The average CO position, however, is along the Fe-C bond; thus, the orientation of the CO bond with the CO atoms at their average positions is zero. This is similar to the geometry that would be resolved in an x-ray structure. That is, the x-ray data gives the orientation of the CO group in the average coordinate set, whereas the photoselection experiment gives the average orientation over all the data sets. The angle recorded in the IR photoselection experiment therefore also reflects the dynamics of the prosthetic group, and the discrepancy between the IR and x-ray Fe-C-O geometries is, amongst other things, an indication of the ligand dynamics. The IR data for the average angle subtended by the CO dipole in protohaem is around 20° depending on the solvent and method of analysis (Hansen et al., 1989), whereas this unhindered CO is expected to lie on the haem normal on average. Smaller discrepancies between the x-ray and IR results

have been found for carbonmonoxy haemoglobin (IR data: 17.5 ± 2.5° (Locke et al., 1991); orientation of the CO bond relative to the haem normal in x-ray PDB file 2HCO: α subunit 12.1°, β subunit 11.6° (Baldwin, 1980)), and for MbCO (IR data: 15 ± 10°, 28 ± 2°, 33 ± 4°; x-ray data PDB file 2MGK: 19.4°), although the wide range of CO orientations found in the crystal structures (as noted in Table 1 for MbCO) makes such an assertion difficult. A closer agreement between the low temperature IR results for MbCO and its x-ray structure would have been expected; however, it appears that even at low temperatures, the photoselection data represents a distribution of Fe-C-O conformations. Ormos et al. (1988) have noted that, even at 10 K, there is a systematic wavenumber dependence of the CO angle within a given A state, of the order of 4.5° and 8° for the $A_{1,2}$ and $A_0$ states, respectively, of a frozen 10 mM solution of MbCO. Such a distribution of Fe-C-O geometries is not resolved in the x-ray structures. This larger volume explored by the CO ligand in the $A_0$ state, as demonstrated by Ormos et al., is consistent with the view that the $A_0$ state is largely free of distal interaction; the $A_{1,2}$ and $A_3$ states have increasingly restricted motions, around axes increasingly distorted from that of the $A_0$ state, because of their increasingly strong attractive interaction with the distal side chain.

Although the average angles subtended by the CO dipole in these trajectories are of the same order as the experimental result, it is clear that the simulations have failed to accurately reproduce the details of the CO dynamics: there is no significant difference in the CO orientation in any of the trajectories recorded. The average deviations from the haem normal are significantly larger, in the simulations, than would be expected from the angle bending functions in the haem force field. Such large distortions arise from the strong electrostatic interactions between the CO ligand and the porphyrin macrocycle. These classical simulations using fixed atomic charges do not reproduce the conformation-dependent changes in the electronic structure of the haem prosthetic group; thus, any small difference in the CO orientation that may be due to interactions, or otherwise, with the distal side chain cannot be "resolved."

## CONCLUSIONS

These simulations suggest that in "closed" haem pocket structures, the solution phase orientation of the distal histidine side chain has the protonated nitrogen pointing into the haem pocket, unlike the neutron diffraction structure. This is consistent with recent interpretations of solution phase C=O stretching frequencies of wild-type MbCO and some of its distal mutants, where it is suggested that the distal protonated NH polarizes the CO ligand in the wild-type species (Li et al., 1994). The details of the "closed" haem pocket structure differ depending on the tautomeric state of the distal side chain: the $64N_\epsilon H$ tautomer adopts a stable conformation with the polar hydrogen close to the carbonyl oxygen, the $64N_\delta H$ tautomer is more mobile and moves between "open" and "closed" pocket structures. The differences in these conformations result simply from the differing geometries of the two tautomers: in each case, the polar hydrogen is restricted by the imidazole ring structure to explore different areas of the haem pocket. The lone pair of the nonprotonated N, which points out into the solvent, is either stabilized by hydrogen bonding to the flexible Arg-45 side chain or to solvent molecules. The differences in the haem pocket structures of these two distal tautomers will result in different electric fields at the active site, likely to correspond in broad terms to the decreases in the C=O stretching frequency observed experimentally for the $A_{1,2}$ and $A_3$ states. In all simulations, the time-averaged position is for the CO lying close to the haem normal. Thus, there is no significant overall steric distortion of the FeCO unit. The averaged orientations of the CO dipole, however, are at angles to the haem normal similar to those observed in photoselection experiments, confirming that this technique, rather than measuring a "static" geometry similar to an x-ray structure, also reflects differences in the dynamics of the ligand in each A state. Both the "closed" pocket structures differ from the "open" pocket structure explored in one of the trajectories. This structure has a large, unhindered entrance to the active site, with the distal side chain hydrogen-bonded to the main chain Asp-60 oxygen, as found in the x-ray structures of myoglobin with bulky organic ligands (Ringe et al., 1984; Johnson et al., 1989).

**TABLE 3 a Force field parameters for the haem prosthetic group**

| Bond stretching | $k_b$ | $r_{eq}$ | | $k_b$ | $r_{eq}$ |
|---|---|---|---|---|---|
| ZC-ZO | 1115.00 | 1.128 | PB-PR | 441.30 | 1.380 |
| FE-ZC | 258.00 | 1.900 | PB-P2 | 441.30 | 1.520 |
| *FE-PN* | *216.17* | *1.960* | PB-P3 | 441.30 | 1.520 |
| *PN-PA* | *276.64* | *1.380* | PR-PV | 512.00 | 1.375 |
| *PA-PB* | *257.40* | *1.440* | P2-P2 | 260.00 | 1.526 |
| *PA-PM* | *374.38* | *1.370* | P2-C | 317.00 | 1.522 |
| *PM-PH* | *360.94* | *1.090* | FE-NB | 65.00 | 2.200 |
| *PB-PB* | *429.36* | *1.350* | | | |
| **Angle bending** | $k_\theta$ | $\theta_{eq}$ | | $k_\theta$ | $\theta_{eq}$ |
| FE-ZC-ZO | 35.00 | 180.00 | PB-PB-P2 | 65.00 | 126.75 |
| PN-FE-ZC | 50.00 | 90.00 | PB-PB-P3 | 65.00 | 126.75 |
| *PA-PN-FE* | *62.45* | *128.05* | PA-PB-P2 | 65.00 | 126.75 |
| *PN-PA-PB* | *80.23* | *111.54* | PA-PB-P3 | 65.00 | 126.75 |
| *PN-PA-PM* | *19.13* | *124.39* | PB-PR-PV | 70.00 | 121.50 |
| *PA-PN-PA* | *137.36* | *103.90* | PB-P2-P2 | 70.00 | 113.00 |
| *PA-PB-PB* | *124.49* | *106.51* | P2-P2-C | 63.00 | 112.40 |
| *PA-PM-PA* | *169.42* | *125.12* | P2-C-O2 | 70.00 | 117.00 |
| *PA-PM-PH* | *30.19* | *117.44* | NB-FE-ZC | 50.00 | 180.00 |
| *PB-PA-PM* | *40.38* | *124.07* | NB-FE-PN | 50.00 | 90.00 |
| PA-PB-PR | 70.00 | 126.75 | FE-NB-CF | 30.00 | 122.35 |
| PB-PB-PR | 70.00 | 126.75 | FE-NB-CP | 30.00 | 132.35 |

All parameters defined according to Weiner et al. (1984). Atomic labels: ZC and ZO, carbonyl carbon and oxygen; PN, PA, PB, PM, and PH, porphyrin nitrogen, α carbon, β carbon, meso carbon, and meso hydrogen, respectively; PR, PV, P2, and P3, porphyrin substituent vinyl CH, vinyl $CH_2$, aliphatic $CH_2$ and aliphatic $CH_3$, respectively; NB, NA, CP, CF, and CG, proximal $N_\delta$ nitrogen, $N_\epsilon$ nitrogen, $C_\delta$ carbon, $C_\epsilon$ carbon, and $C_\gamma$ carbon, respectively. $k_r$ is the bond-stretching constant in kcal/mol Å$^2$; $r_{eq}$ is the equilibrium bond length in Å; $k_\theta$ is the angle bending constant in kcal/mol rad$^2$; $\theta_{eq}$ is the equilibrium angle in degrees. Entries in italics are from this work; zero energy entries not listed.

**TABLE 3 b Force field parameters for the haem prosthetic group, cont.**

| Proper torsions | $n$ | $V_n$ | $\gamma$ | | $n$ | $V_n$ | $\gamma$ |
|---|---|---|---|---|---|---|---|
| X-FE-ZC-X | 4 | 0.05 | 180.00 | *PN-PA-PM-PA* | *1* | *3.73* | *180.00* |
| X-PA-PB-X | 2 | 14.30 | 180.00 | *PN-PA-PM-PH* | *1* | *2.66* | *180.00* |
| X-PB-PB-X | 2 | 14.30 | 180.00 | *PA-PB-PB-PA* | *1* | *16.71* | *180.00* |
| X-PB-PR-X | 2 | 1.00 | 0.00 | *PB-PA-PN-PA* | *1* | *1.95* | *180.00* |
| X-FE-NB-X | 4 | 0.05 | 180.00 | *PB-PA-PM-PH* | *1* | *2.18* | *180.00* |
| | | | | *PB-PB-PA-PM* | *1* | *2.76* | *180.00* |
| *FE-PN-PA-PM* | *1* | *1.07* | *180.00* | PB-P2-P2-C | 1 | 2.00 | 0.00 |
| *PN-FE-PN-PA* | *1* | *2.71* | *180.00* | NA-CP-NB-FE | 1 | 12.50 | 180.00 |
| *PN-PA-PB-PB* | *1* | *1.44* | *180.00* | CC-CF-NB-FE | 1 | 12.50 | 180.00 |
| **Improper torsions** | $n$ | $V_n$ | $\gamma$ | | $n$ | $V_n$ | $\gamma$ |
| X-O2-C-O2 | 2 | 10.50 | 180.00 | X-X-NB-FE | 2 | 5.0 | 180.00 |

$n$ is the multiplicity of the cosine function; $V_n$ is the dihedral force constant in kcal/mol; $\gamma$ is the phase in degrees.

**TABLE 3 c Force field parameters for the haem prosthetic group, cont.**

| Van der Waals potential | $R^*$ | $\epsilon$ | | $R^*$ | $\epsilon$ |
|---|---|---|---|---|---|
| FE | 0.650 | 0.0000 | PM | 2.100 | 0.1200 |
| ZC | 1.850 | 0.1200 | PH | 1.540 | 0.0100 |
| ZO | 1.600 | 0.2000 | PR | 1.850 | 0.1200 |
| PN | 1.600 | 0.2384 | PV | 1.850 | 0.1200 |
| PA | 2.100 | 0.1200 | P2 | 1.925 | 0.1200 |
| PB | 2.100 | 0.1200 | P3 | 2.000 | 0.1500 |

$\epsilon$ and $R^*$ the van der Waals parameters inkcal/mol and Å, respectively.

Since the two distal tautomers show significantly different interactions with the CO ligand, but there is no large difference in the Fe-C-O geometries, it is suggested that the CO stretching frequencies of the different A states observed experimentally for MbCO in the solution phase do not result from significantly different average Fe-C-O geometries, but rather result from different haem pocket structures, with their different electrostatic influence on the Fe-bound CO ligand. The two "closed" structures correspond in broad terms to the $A_3$ ($64N_\epsilon H$ tautomer) and $A_{1,2}$ ($64N_\delta H$ tautomer) states, and the "open" structure corresponds to the $A_0$ state.

The authors thank Dr. M. Saito of the Protein Engineering Research Institute, Osaka, for stimulating discussions, and Drs. J. S. Olson and G. N. Phillips, Jr. of Rice University, Houston, for communicating results before publication. P. Jewsbury would like to thank the Japan Society for the Promotion of Science for a postdoctoral Fellowship through the grant 04NP0301 to T. Kitagawa from the Japanese Ministry of Education, Science and Culture for the new research program "Intelligent molecular systems with controlled functionality." These calculations were run at the Computer Centre of the Institute for Molecular Science.

## APPENDIX

### The haem group's force field

A force field for the haem prosthetic group was developed for AMBER similarly to that of Kuczera et al. (1990) for the CHARMM package. Lopez and Kollman (1989) have recently developed an AMBER haem force field based on the static x-ray structures of model haem compounds; however, this is not expected to reflect the dynamics of the porphyrin macrocycle well. Here we develop the first AMBER force field for the porphyrin macrocycle based on its observed vibrational frequencies.

The initial geometry of the porphyrin macrocycle was taken from the x-ray structure for Fe(TPP)(pyridine)(CO) (Peng and Ibers, 1976), slightly altered to give $D_{4h}$ symmetry. All atoms (i.e., including H atoms) were included in the fitting of the macrocycle's vibrational frequencies, as the H atoms have been shown to be important determinants of the macrocycle's normal modes (Li et al., 1989, 1990a, b). The equilibrium distances and angles were the same as those of Kuczera et al. (1990) and were held fixed during the fitting. The bend and stretch force constants were found by a least squares best fit of the calculated and observed (Li et al., 1990a) in-plane vibrational frequencies of Ni porphyrin (including the fully deuterated porphyrin isotopomer), using a modified quadratically convergent method of Powell (Press et al., 1986: p.298). The vibrations were separated into their symmetry groups and, after some initial constrained fitting, all 18 variables (7 stretch and 11 bend) were allowed to optimize freely. The macrocycle was constrained to $D_{4h}$ symmetry, and the fits to observed vibrational frequencies were equally weighted, except for those with large metal motions which were given only 50% weight. Frequencies not observed in the resonance Raman spectrum were assigned from Li et al.'s force field and given the lower weighting.

The dihedral force constants were optimized in a similar manner by fitting to the out-of-plane vibrational frequencies calculated by Warshel and Lappicirella (1981), and the $A_{2u}$ modes assigned from experiment for NiP by Li et al. (1989; their Table VIII). Experimentally determined frequencies were given a 30% higher weighting, and apparent errors in Warshel and Lappicirella's Table XI corrected.[2] The bend and stretch force constants were held fixed to their in-plane optimized values, and the 16 dihedral variables were allowed to optimize freely. Both the variable sets in these optimizations are overdetermined, so 1 bend and 3 dihedral parameters optimized to zero.

All parameters for the haem side chains were taken from the standard AMBER4.0 united atom parameter set for such groups, as were the nonbonded parameters. Force field parameters for the axial ligands were adapted from Kuczera et al. (1990). Charges for the porphyrin macrocycle were adapted from the extended Huckel calculations of Zerner et al. (1966), and the parameters supplied with AMBER4.0; those for the metal centre and axial ligands (including the proximal ligand) were adapted from those used by Lopez and Kollman (1989) for free energy calculations of model porphyrin compounds, giving the final force field described in Table 3 and Fig. 6.

The accuracy of the force field developed here is limited by a number of assumptions (for example, that the different substituents present in porphyrin IX do not significantly alter the force field from that of the macrocycle), and by the limited accuracy of the Huckel method in deriving the porphyrin atomic charges; however, the detailed motion of the macrocycle itself is expected to be of only secondary importance to the haem pocket

[2] Warshel and Lappicirella's Table XI has one too many $A_{2u}$ frequencies and one too few $A_{1u}$ frequencies. It also became evident during the optimization that all the frequencies could not be fitted satisfactorily as assigned in their Table XI, even if the extra high frequency $A_{2u}$ vibration was assumed to be $A_{1u}$ in symmetry. However, assigning the extra frequency to the $B_{2u}$ symmetry group and the highest $B_{2u}$ frequency to the $A_{1u}$ group gave a satisfactory fit, so this assignment was adopted in the final optimization.

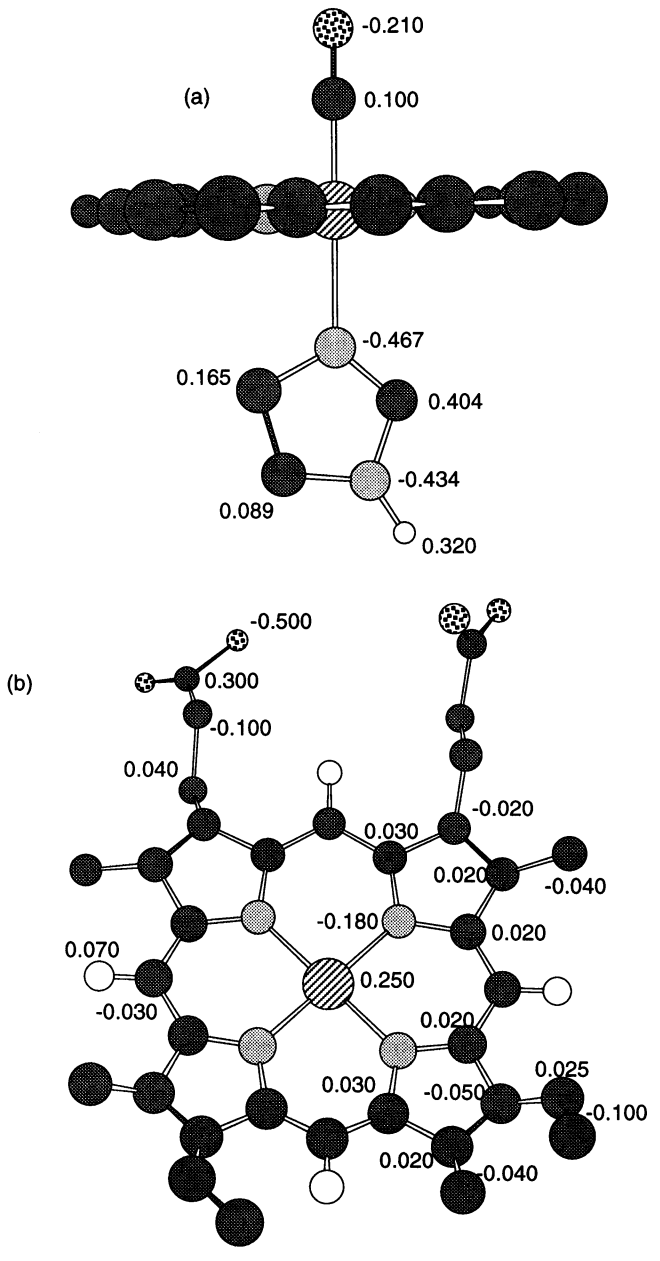


FIGURE 6 The electrostatic charges (in a.u.) used in the force field of the haem prosthetic group, (*a*) the carbonyl ligand and the proximal imidazole ring, and (*b*) the porphyrin IX macrocycle.

structure provided its average geometry and the force fields of its substituents are reasonable. A more detailed force field for the haem prosthetic group awaits the application of high-level theoretical techniques that can reproduce the observed frequencies. Such a calculation is not currently possible.

## REFERENCES


Abola, E. E., F. C. Bernstein, S. H. Bryant, T. F. Koetzle, and J. Weng. 1987. Protein Data Bank. *In* Crystallographic Databases–Information Content, Software Systems, Scientific Applications. F. H. Allen, G. Bergerhoff, and R. Sievers, editors. Data Commission of the International Union of Crystallography, Bonn/Cambridge/Chester. 107–132.

Baldwin, J. M. 1980. The structure of human carbonmonoxy haemoglobin at 2.7 Å resolution. *J. Mol. Biol.* 136:103–128.

Berendsen, H. J. C., J. P. M. Postma, W. F. van Gunsteren, A. DiNola, and J. R. Haak. 1984. Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* 81:3684–3690.

Bernstein, F. C., T. F. Koetzle, G. J. B. Williams, E. F. Meyer, Jr., M. D. Brice, J. R. Rodgers, O. Kennard, T. Shimanouchi, and M. Tasumi. 1977. The Protein Data Bank: a computer based archival file for macromolecular structures. *J. Mol. Biol.* 112:535–542.

Brown III, W. E., J. W. Sutcliffe, and P. D. Pulsinelli. 1983. Multiple internal reflectance infrared spectra of variably hydrated hemoglobin and myoglobin films: effects of globin hydration on ligand conformer dynamics and reactivity at the heme. *Biochemistry.* 22:2914–2923.

Carver, T. E., R. E. Brantley, Jr., E. W. Singleton, R. M. Arduini, M. L. Quillin, G. N. Phillips, and J. S. Olson. 1992. A novel site-directed mutant of myoglobin with an unusually high $O_2$ affinity and low autoxidation rate. *J. Biol. Chem.* 267:14443–14450.

Case, D. A., and M. Karplus. 1978. Stereochemistry of carbon monoxide binding to myoglobin and hemoglobin. *J. Mol. Biol.* 123:697–701.

Caughey, W. S., H. Shimada, G. C. Miles, and M. P. Tucker. 1981. Dynamic protein structures: infra red evidence for four discrete rapidly interconverting conformers at the carbonmonoxide binding site of bovine heart myoglobin. *Proc. Natl. Acad. Sci. USA.* 78:2903–2907.

Cheng, X., and B. P. Schoenborn. 1991. Neutron diffraction study of carbonmonoxy myoglobin. *J. Mol. Biol.* 220:381–399.

Egeberg, K. D., B. A. Springer, S. G. Silgar, T. E. Carver, R. J Rohlfs, and J. S. Olson. 1990. The role of Val68 in ligand binding to sperm whale myoglobin. *J. Biol. Chem.* 265:11788–11795.

Elber, R., and M. Karplus. 1990. Enhanced sampling in molecular dynamics: use of the time-dependent Hartree approximation for a simulation of carbon monoxide diffusion through myoglobin. *J. Am. Chem. Soc.* 112: 9161–9175.

Hansen, P. A., J. N. Moore, and R. M. Hochstrasser. 1989. Determination of the iron-carbonyl bond geometry of carboxyprotoheme in solution using picosecond infrared-optical photoselection. *Chem. Phys.* 131: 49–62.

Henry, E. R. 1993. Molecular dynamics simulations of heme reorientational motions in myoglobin. *Biophys. J.* 64:869–885.

Jewsbury, P., S. Yamamoto, T. Minato, M. Saito, and T. Kitagawa. 1995. The proximal residue largely determines the CO distortion in carbonmonoxy globin proteins. An *ab initio* study of a haem prosthetic unit. *J. Am. Chem. Soc.* In press.

Jorgensen, W. L., J. Chandrasekhar, J. D. Madura, R. W. Impey, and M. L. Klein. 1983. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* 79:926–935.

Johnson, K. A., J. S. Olson, and G. N. Phillips, Jr. 1989. Structure of myoglobin-ethyl isocyanide: histidine as a swinging door for ligand entry. *J. Mol. Biol.* 207:459–463.

Kendrew, J. C., R. E. Dickerson, B. E. Strandberg, R. G. Hart, D. R. Davies, D. C. Phillips, and V. C. Shore. 1960. Structure of myoglobin: a three dimensional fourier synthesis at 2 Å resolution. *Nature.* 185: 422–427.

Kuriyan, J., S. Wilz, M. Karplus, and G. A. Petsko. 1986. X-ray structure and refinement of carbon-monoxy (Fe II)-myoglobin at 1.5 Å resolution. *J. Mol. Biol.* 192:133–154.

Kuczera, K., J. Kuriyan, and M. Karplus. 1990. Temperature dependence of the structure and dynamics of myoglobin. *J. Mol. Biol.* 213: 351–373.

Li, T., M. L. Quillin, G. N. Phillips, Jr., and J. S. Olson. 1994. Structural determinants of the stretching frequency of CO bound to myoglobin. *Biochemistry.* 33:1433–1446.

Li, X.-Y., R. S. Czernuszewicz, J. R. Kincaid, Y. O. Su, and T. G. Spiro. 1989. Consistent porphyrin force field. 3. Out of plane modes in the resonance Raman spectra of planar and ruffled nickel octaethyl porphyrin. *J. Am. Chem. Soc.* 111:7012–7023.

Li, X.-Y., R. S. Czernuszewicz, J. R. Kincaid, Y. O. Su, and T. G. Spiro. 1990a. Consistent porphyrin force field. 1. Normal mode analysis for nickel porphine and nickel tetraphenylporphine from resonance Raman and infra red spectra and isotope shifts. *J. Phys. Chem.* 94:31–47.

Li, X.-Y., R. S. Czernuszewicz, J. R. Kincaid, P. Stein, and T. G. Spiro. 1990b. Consistent porphyrin force field. 2. Nickel octaethylporphyrin skeletal and substituent mode assignments form nitrogen-15, meso-$d_4$ and methylene-$d_{16}$ Raman and infra red isotope shifts. *J. Phys. Chem.* 94:47–61.

Locke, B., T. Lian, and R. M. Hochstrasser. 1991. Determination of Fe-CO geometry and heme rigidity in carbonmonoxyhemoglobin using femtosecond IR spectroscopy. *Chem. Phys.* 158:409–419.

Lopez, M. A., and P. A. Kollman. 1989. Application of molecular dynamics and free energy perturbation methods to metalloporphyrin ligand systems. 1. CO and dioxygen binding to four heme systems. *J. Am. Chem. Soc.* 111:6212–6222.

Lopez, M. A., and P. A. Kollman. 1993. Application of molecular dynamics and free energy perturbation methods to metalloporphyrin ligand systems. II. CO and dioxygen binding to myoglobin. *Protein Sci.* 2:1975–1986.

Makinen, M. W., R. A. Houtchens, and W. S. Caughey. 1979. Structure of carbonmonoxy myoglobin in crystals and in solution. *Proc. Natl. Acad. Sci. USA.* 76:6042–6046.

Maxwell, J. C., and W. S. Caughey. 1976. An infra red study of NO bonding to heme B and hemoglobin A. Evidence for inositol hexaphosphate induced cleavage of proximal histidine to iron bonds. *Biochemistry.* 15: 388–395.

Moore, J. H., P. A. Hansen, and R. M. Hochstrasser. 1988. Iron-carbonyl bond geometries of carboxymyoglobin and carboxyhemoglobin in solution determined by picosecond time-resolved infra red spectroscopy. *Proc. Natl. Acad. Sci. USA.* 85:5062–5066.

Mourant, J. R., D. P. Braunstein, K. Chu., H. Frauenfelder, G. U. Nienhaus, P. Ormos, and R. D. Young. 1993. Ligand binding to heme proteins. II. Transitions in the heme pocket of myoglobin. *Biophys. J.* 65:1496–1507.

Norvell, J. C., A. C. Nunes, and B. P. Schoenborn. 1975. Neutron diffraction analysis of myoglobin: structure of the carbon monoxide derivative. *Science.* 190:568–570.

Oldfield, E., K. Guo, J. D. Augspurger, and C. E. Dykstra. 1991. A molecular model for the major conformational substates in heme proteins. *J. Am. Chem. Soc.* 113:7537–7541.

Ormos, P., D. Braunstein, H. Fraunfelder, M. K. Hong, S. Lin, T. B. Sauke, and R. D. Young. 1988. Orientation of carbon monoxide and structure-function relationship in carbonmonoxy myoglobin. *Proc. Natl. Acad. Sci. USA.* 85:8492–8496.

Pearlman, D. A., D. A. Case, J. C. Caldwell, G. L. Siebel, U. C. Singh, P. Weiner, and P. A. Kollman. 1991. *AMBER 4.0.* University of California, San Francisco.

Peng, S.-M., and J. A. Ibers. 1976. Stereochemistry of carbonylmetalloporphyrins. The structure of (pyridine)(carbonyl)(5, 10, 15, 20-tetraphenylporphinato) iron(II). *J. Am. Chem. Soc.* 98:8032–8036.

Petrich, J. W., J.-C. Lambry, K. Kuczera, M. Karplus, M. Poyart, and J.-L. Martin. 1991. Ligand binding and protein relaxation in heme proteins: a room temperature analysis of NO geminate recombination. *Biochemistry.* 30:3975–3987.

Phillips, S. E. V., and B. P. Schoenborn. 1981. Neutron diffraction reveals oxygen-histidine hydrogen bond in oxymyoglobin. *Nature.* 292:81–82.

Press, W. H., B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling. 1986. Numerical Recipes: the art of scientific computing. Cambridge University Press, Cambridge.

Quillin, M. L., R. M. Arduini, J. S. Olson, and G. N. Phillips, Jr. 1993. High resolution crystal structures of distal histidine mutants of sperm whale myoglobin. *J. Mol. Biol.* 234:140–155.

Quillin, M. L., R. E. Brantley, Jr., K. A. Johnson, J. S. Olson, and G. N. Phillips, Jr. 1992. Kinetic and structural analysis of the effects of pH on carbon monoxide association in myoglobin. *Biophys. J.* 61:466*a*. (Abstr.)

Ray, G. B., X.-Y. Li, J. A. Ibers, J. L. Sessler, and T. G. Spiro. 1994. How far can proteins bend the FeCO unit? Distal polar effects in heme proteins and models. *J. Am. Chem. Soc.* 116:162–176.

Ringe, D., G. A. Petsko, D. Kerr, and P. R. Ortiz de Montellano. 1984. Reaction of myoglobin with phenylhydrazine: a molecular doorstop. *Biochemistry.* 23:2–4.

Ryckaert, J.-P., G. Cicotti, and H. J. C. Berendsen. 1977. Numerical integration of the cartesian coordinates of motion of a system with constraints: molecular dynamics of n-alkanes. *J. Comput. Phys.* 23:327–341.

Sakan, Y., T. Ogura, T. Kitagawa, F. A. Fraunfelter, R. Mattera, and M. Ikeda-Saito. 1993. Time-resolved resonance Raman study of the binding of carbon monoxide to recombinant human myoglobin and its distal mutants. *Biochemistry.* 32:5815–5824.

Saito, M. MD simulations of proteins in solution: artifacts caused by the cut-off approximation. *J. Chem. Phys.* 101:4055–4061.

Singh, U. C., S. J. Weiner, and P. A. Kollman. 1985. Molecular dynamics simulations of d(C-G-C-G-A).d(T-C-G-C-G) with and without "hydrated" counterions. *Proc. Natl. Acad. Sci. USA.* 82:755–759.

Springer, B. A., S. G. Silgar, J. S. Olson, and G. N. Phillips, Jr. 1994. Mechanisms of ligand recognition in myoglobin. *Chem. Rev.* 94:699–714.

Straub, J. E., and M. Karplus. 1991. Molecular dynamics study of the photodissociation of carbon monoxide from myoglobin: ligand dynamics in the first 10ps. *Chem. Phys.* 158:221–248.

Stryer, L. 1988. Biochemistry, 3rd edition. W. H. Freeman and Company, New York.

van Gunsteren, W. G., and A. E. Mark. 1992. On the interpretation of biochemical data by molecular dynamics computer simulation. *Eur. J. Biochem.* 204:947–961.

Warshel, A., and A. Lappicirella. 1981. Calculation of ground- and excited state potential surfaces for conjugated heteroatomic molecules. *J. Am. Chem. Soc.* 103:4664–4673.

Weiner, S. J., P. A. Kollman, D. A. Case, U. C. Singh, C. Ghio, G. Alagona, S. Profeta, Jr., and P. Weiner. 1984. A new force field for molecular mechanical simulation of nucleic acids and proteins. *J. Am. Chem. Soc.* 106:765–784.

Yu, N.-T., and E. A. Kerr. 1988. Vibrational modes of coordinated CO, $CN^-$, $O_2$ and NO. *In* Biological Applications of Raman Spectroscopy, Vol. 3: Resonance Raman Spectra of Heme and Metalloproteins. T. G. Spiro, editor. Wiley-Interscience, New York. 39–95.

Zerner, M., M. Gouterman, and H. Kobayashi. 1966. Porphyrins VIII. Extended Huckel calculations on iron porphyrins. *Theor. Chim. Acta.* 6:363–400.