# On the significance of hydrogen bonds for the discrimination between CO and $O_2$ by myoglobin

Sigfridsson, Emma; Ryde, Ulf

*Published in:*
Journal of Biological Inorganic Chemistry

*DOI:*
10.1007/s007750050293

1999

*Document Version:*
Peer reviewed version (aka post-print)

# On the significance of hydrogen bonds for the discrimination between CO and $O_2$ by myoglobin

**Emma Sigfridsson and Ulf Ryde**

Department of Theoretical Chemistry,

Lund University

Chemical Centre, P.O.B 124

S-221 00 Lund, Sweden

*Correspondence to U. Ryde*

Tel: 46-46-22245 02

Fax: 46-46-222 45 43

Email: Ulf.Ryde@teokem.lu.se

**Abstract**

Quantum chemical geometry optimisations have been performed on realistic models of the active site of myoglobin using density functional methods. The energy of the hydrogen bond between the distal histidine residue and CO or $O_2$ has been estimated to 8 and 32 kJ/mole, respectively. This 24 kJ/mole energy difference accounts for most of the discrimination between CO and $O_2$ by myoglobin (about 17 kJ/mole). Thus, steric effects seem to be of minor importance for this discrimination. The Fe–C and C–O vibrational frequencies of CO-myoglobin have also been studied and the results indicate that CO forms hydrogen bonds to either the distal histidine residue or a water molecule during normal conditions. We have made several attempts to optimise structures with the deprotonated nitrogen atom of histidine directed towards CO. However, all such structures lead unfavourable interactions between the histidine and CO, and $\nu_{CO}$ frequencies higher than those observed experimentally.

**Key words**: myoglobin, hydrogen bond, quantum chemical calculations, CO/$O_2$ discrimination, vibrational frequencies

## Introduction

Myoglobin is a haem protein in muscle that reversibly binds biologically relevant ligands, such as $O_2$ and CO. It has been known for a long time that the protein matrix affects the ligand-binding affinity of the haem group. For example, CO binds to free haem in solution around 20 000 times as strongly as $O_2$, but in myoglobin this factor is only 25 [1]. Thus, myoglobin seems to favour $O_2$ before CO by about 17 kJ/mole. This discrimination is essential for life, since CO is formed during the degradation of the haem group in the body, and at least 1% of the haem groups in non-smokers are poisoned by CO [2]. The mechanism of this discrimination is a central question in inorganic biochemistry, and although myoglobin is among the best studied proteins, the question remains controversial [3].

Twenty years ago, Collman et al. [4] suggested that the highly conserved distal histidine residue (His-64) sterically forces CO to bind in a bent manner to the haem iron ion. $O_2$ is not affected by this restriction, since it prefers to bind to iron with a Fe–O–O angle around 120°. This idea is supported by X-ray as well as neutron structures, showing Fe–C–O angles of 120–140° [5-9] and it has made its way into the textbooks [10]. However, the hypothesis has lately been questioned. Two recent crystal structures show a larger Fe–C–O angle (160-170°) [11-13] and experimental evidence from vibrational spectroscopy indicates that the Fe–C–O angle is nearly linear [4,10,14-18]. Moreover, recent theoretical calculations have indicated that energy of bending and tilting CO is small (<8 kJ/mole) for moderate angles (up to about 25°) [19,20].

An alternative hypothesis is that the protein stabilises $O_2$ by electrostatic interactions [21]. Neutron scattering structures [22] show that there is a hydrogen bond between $O_2$ and the distal histidine residue, whereas no such bond can be seen in the CO complex [9]. Moreover, studies by site-directed mutagenesis and synthetic haem models [1,16,23,24] have suggested that electrostatic interactions are of major importance for the discrimination between CO and $O_2$, whereas steric hindrance plays only a minor role.

In this paper we calculate the energies of hydrogen bonds between the distal histidine residue and CO or $O_2$ by quantum chemical methods. This provides a quantisation of the hydrogen-bond hypothesis and thereby complements the theoretical calculations of the bending energies [19,20]. In

addition, we calculate the vibrational frequencies of CO-myoglobin in order to interpret the experimental spectra and to gain information about possible hydrogen-bond interactions to CO.

## Methods

*Computational details*

Myoglobin was modelled by complexes of two different sizes. The largest models consisted of a porphine molecule (Por; the full porphyrin ring without any substituents) with a central Fe(II) ion and imidazole (Im) as a model of the proximal as well as the distal histidine residues (57 atoms in total). For test calculations, imidazole was replaced by $NH_3$ and porphine was replaced by two molecules of diformamidate (Dfa; $NHCHNH^-$), which has been suggested by Newton and Hall [25] to provide a reasonable compromise between accuracy and economy.

If not otherwise stated, geometry optimisations were performed with the density functional Becke–Perdew86 (BP) method, which consists of Becke's 1988 gradient-corrected exchange potential [26] together with Perdew's 1986 gradient correction [27] to the Vosko–Wilk–Nusair correlation functional (Ceperly–Adler solution) [28]. In addition, the Coulomb operators were treated with the RI (resolution of identity) approximation. In the optimisations, all atoms not involved in the hydrogen bond were treated by effective core potentials (ECP). For Fe we used the 10-electron ECP and double-ζ basis set (21/21/31) of Hay–Wadt [29], for C, N, and O, the 2-electron ECP and double-ζ basis set (31/31) of Stevens et al. [30] enhanced with a polarising *d* orbital with the exponent 0.8, and for H, the 6-31G basis set. For atoms involved in the hydrogen bond ($N_{donor}$, H, and CO or $O_2$), we used the 6-31+G** [31] basis set. Moreover, for the RI approximation we employed the default auxiliary basis sets in Turbomole (6*s*4*p*2*d*3*f*2*g* for Fe, 6*s*3*p*3*d*2*f* for C and N, 2*s*1*p*for H, but 7*s*3*p*3*d*2*f* for C and O, and 2*s*2*p* for H in the hydrogen bonds) [32,33]. The full geometry was optimised for all models until the change in energy were below 2.6 J/mole and the internal gradients were below 0.053 pm or 0.057°.

After the geometries were optimised, energies were calculated with the more accurate B3LYP method [34] as implemented in the Turbomole package [35,36]. These calculations employed the 6-31G* basis set [31], except for atoms in the hydrogen bond, for which the larger 6-31+G** basis set was used [31]. For iron, the double-ζ basis set (62111111/33111/311) of Schäfer et al. [37] was used, enhanced with diffuse *p*, *d* and *f* functions with exponents 0.134915, 0.041843 (two *p*

functions), 0.1244, and 1.339 (called DZpdf). In all calculations, only the pure five *d* and seven *f* functions were used.

Experiments have shown that the $Fe^{2+}$ ion in myoglobin without any ligands is in the high-spin state, whereas the complexes with $O_2$ or CO are paramagnetic [38]. Therefore, if not otherwise stated, a singlet state was considered for all models with a six-coordinate iron ion, and a quintet spin state was employed for the five-coordinate models. All models with Fe and $O_2$ were treated within the unrestricted-spin formalism since the lowest electronic state is an open-shell singlet [39].

Bond energies were corrected for the basis set superposition error using the standard counter-poise method [40]. However, when calculating the binding energy of CO and $O_2$ to the porphyrins with this method, the spin state is crucial. Therefore, we used a more sophisticated formula for these energies:

$$\Delta E(PO_2) = E_{opt}(^1PO_2) - E_{po}(^sP\bullet) - E_{po}(^3\bullet O_2) - E_{opt}(^5P) + E_{po}(^sP) - E_{opt}(^3O_2) + E_{po}(^3O_2) \qquad (1)$$

and similar for $^1CO$ (treated as a singlet). Here, P is the five-coordinate porphyrin model (ImPorFe), $E_{opt}$ and $E_{po}$ means the energy calculated at the optimum geometry and at the geometry of the $PO_2$ complex, respectively, and • means a calculation with the basis functions but no nuclear charge of that part of the complex (P or $O_2$). This formula is insensitive to the actual spin state of P and P• (*s*); for example, the binding energy of $O_2$ to ImPorFe changes by only 0.2 kJ/mole if *s* in Eqn. 1 is changed from 1 to 5. In the following we only present the results obtained with $s = 5$. It should be noted that the first three terms constitute the normal formula for the counter-poise corrected binding energy, whereas the last four terms correct this energy for the relaxation of the substructures to their optimal geometry. In fact, it turns out that for the rather soft porphyrin complexes, both terms are essential to obtain accurate binding energies. For the binding energy of CO or $O_2$ to ImPorFe, the counter-poise correction is 13–14 kJ/mole, whereas the relaxation correction is appreciably larger, 24–43 kJ/mole. For hydrogen bonds, the two terms are smaller and of similar magnitude, 1–5 kJ/mole. Therefore, Eqn. 1 was used also for the most important hydrogen-bond energies. If not otherwise stated, all calculations were performed using the Turbomole software [35] on IBM RS/6000 workstations.

*Frequency calculations*

For the large porphyrin models, it is too time consuming to calculate analytical frequencies. Therefore, we instead estimated the frequencies using a harmonic approximation. The force constant of a particular bond is given by

$$k = \frac{\partial^2 E}{\partial^2 x} \approx \frac{\left.\frac{\partial E}{\partial x}\right|_{x=x_0+\Delta x} - \left.\frac{\partial E}{\partial x}\right|_{x=x_0}}{\Delta x} \tag{2}$$

where $x$ is the bond length ($x_0$ at equilibrium) and all the other bonds are kept at their equilibrium values. The second derivative was calculated from the gradients rather than from the energy, since this gave more accurate results. Force constants were calculated with positive as well as negative displacements and the average value is reported. The frequency was calculated from the force constant using the standard equation for a harmonic oscillator:

$$\nu = \frac{1}{2\pi}\sqrt{\frac{k}{\mu}} \tag{3}$$

where ν is the frequency and μ is the reduced mass. For the C–O stretching frequency, μ was taken as the reduced mass of a free CO molecule. On the other hand, for the weaker Fe–C bond, the reduced mass was calculated by

$$\frac{1}{\mu} = \frac{1}{m_{\mathrm{Fe}}} + \frac{1}{m_{\mathrm{C}} + m_{\mathrm{O}}} \tag{4}$$

i.e. the molecule was approximated as CO bound to an iron ion.

After a thorough calibration, the displacement ($\Delta x$) was chosen to 0.053 pm (0.001 au). In Table 1, frequencies calculated with this method are compared to the exact (analytical) ones for two model systems: $NH_3$DfaFeCO and $NH_3$DfaFeCO···$NH_3$. The result shows that positive and negative displacements give the same results within 5 $cm^{-1}$. The Fe–C frequency is rather accurately reproduced (error 3–5 $cm^{-1}$), whereas for the C–O vibration, the error is larger, about 25

$cm^{-1}$. Since the error is almost constant for the two structures, it is probably caused mainly by the estimate of the reduced mass.

Frequencies calculated with this method turned out to be strongly correlated to the corresponding bond lengths ($\rho = 0.996$ for $\nu_{CO}$). This shows that the frequencies could as well be calculated directly from the C–O or Fe–C bond lengths. In fact, such frequencies gave more stable results than those obtained by Eqns. 2–4. Therefore, frequencies presented in Figure 5 were obtained from the bond lengths using the following linear relations (the regression curves for the bond lengths and the frequencies):

$$\nu_{CO} = 7957.2 - 51.13\, r_{CO} \quad (5)$$

$$\nu_{FeC} = 3100.9 - 14.619\, r_{FeC} \quad (6)$$

where $r_{CO}$ and $r_{FeC}$ are the C–O and Fe–C bond lengths in pm, respectively, and the frequencies are given in $cm^{-1}$.

*Calibration of the method*

In order to evaluate the performance of our theoretical method for the relevant hydrogen bonds, the geometry of CO···Im was optimised by a number of ab initio and density functional methods: local density functional approximation (S–VWN), the hybrid density functional B3LYP method, the Becke-Perdew86 method, with and without the RI approximation (BP and BP–RI), the ab initio Hartree-Fock (HF) method, and Møller-Plesset second-order perturbation theory (MP2). The results in Table 2 show that MP2, B3LYP, BP, and BP–RI give rather similar result both for the geometry and the energy of the hydrogen bond, whereas HF and local density theory (S–VWN) give much worse results. B3LYP gives an energy that is closer to MP2 (2 kJ/mole too low) than the two BP variants (1 kJ/mole lower), but the hydrogen bond distance and the C–O bond length of the BP method are closer to MP2. It is especially gratifying that the BP–RI method gives virtually the same results as the BP method (within 0.02 pm and 0.04 kJ/mole). Since this method is more than five times faster than the B3LYP and MP2 methods, we decided to use it for all geometry

optimisations, whereas we use the more accurate B3LYP method [34] for single-point energy calculations on the optimised structures. This gave virtually the same hydrogen–bond energy as the full B3LYP method, as can be seen in Table 2.

Second, we tested the performance of effective core potentials (ECPs) on two medium-size porphyrin model systems. Table 3 shows that the results obtained with the ECPs are close to those obtained with a full basis set; the geometry change by less than 1 pm (2.6 pm for Fe–O) and the hydrogen-bond energy by less than 1 kJ/mole.

Third, we compared the results obtained with the small and the large porphyrin models, Dfa and Por. From Tables 4–6, it can be seen that Dfa gives results that are reasonably close to those obtained with Por. The bond lengths to the iron ion are the same as for the Por models within 4 pm and the angles are even better reproduced. The hydrogen-bonds are also reasonably well reproduced if the donor is imidazole; $NH_3$DfaFe gives hydrogen bonds that are 2–4 kJ/mole stronger than ImPorFe with about 6–7 pm shorter N–O and H–O distances. On the other hand, if the hydrogen-bond donor is ammonia, appreciably weaker hydrogen bonds are obtained. This is partly due to that ammonia also interacts with the polar hydrogens of the Dfa group, which distorts the hydrogen-bond geometry. On the other hand, the results are quite insensitive to the model of the proximal histidine; imidazole and ammonia give identical results (within 2 kJ/mole and 2 pm). Since the differences between the results obtained with the Por and Dfa models seem to be reasonably constant, we have used the $NH_3$DfaFeCO/$O_2$···Im model for some investigative calculations. For quantitative calculations, however, the full ImProFeCO/$O_2$···Im model was used.

## Result and Discussion

### *Geometries*

First, we optimised the geometry of the five-coordinate ImPorFe complex. Although experiments have shown that the five-coordinate $Fe^{2+}$ ion in myoglobin is high-spin [23], earlier quantum chemical calculations have indicated that the isolated ImPorFe complex is most stable in the triplet spin state (by 27 kJ/mole) [39]. Therefore, we optimised all three possible spin states of

this complex. As can be seen in Table 4, we also get a triplet ground state, but the energy difference to the quintet state is very small, only 0.5 kJ/mole. The reason why we get a smaller energy difference than Parrinello et al. [39] is most likely that we have used a more accurate functional; this energy difference depends strongly on the method used. If the geometries of the three optimised complexes are compared with the structure of five-coordinate model compounds and crystal structures [23] (Table 4) it is clear that the quintet state is more similar to the experimental structures than are the other states. For the quintet state, both the Fe–N distances and the distance of the iron ion out of the porphyrin plane are within the range of the values observed experimentally, whereas the two other states give a too small out-of-plane distance, and the Fe–$N_{His}$ bond is too long (225 pm) in the triplet state and too short in the singlet state (189 pm compared to 200–222 pm in the crystal structures). Thus, it seems that our method slightly overestimates the stability of the triplet state. Alternatively, myoglobin might preferentially stabilise the quintet state of the five-coordinate porphyrin. For example, Parrinello have suggested that the proteins may stabilise the quintet state by favouring a large iron out-of-plane distance [39]. Most importantly, however, it should be noted that the energies presented in this paper are insensitive to the spin state used in the calculations.

Next, we optimised the geometry of ImPorFe$O_2$. As can be seen from the result in Figure 1, the iron ion is octahedrally coordinated to the six ligands, the porphyrin ring is almost perfectly planar, and the Fe–O–O angle is bent, 121°. In Table 5, we compare the iron geometry of the optimised structure with the one of oxy-myoglobin and oxygenated porphyrin model systems [15,23,41–45]. Most of the optimised parameters are close to the experimental ones, e.g. the Fe–O and Fe–N bond lengths. However, the optimised O–O bond length seems to be slightly too long (129 pm compared to 121-124 pm, respectively). Yet, very similar bond lengths have been obtained in quantum chemical calculations with other methods (128-130 pm) [39,49].

Finally, imidazole as a model of His-64 was added. Imidazole forms a strong hydrogen bond to $O_2$ by the protonated nitrogen atom. The H–O distance is 190 pm, which is close to the value observed in the neutron structure of myoglobin, 198 pm [43]. Likewise, the N–O distance in the optimised model, 293 pm, is similar to the distance in myoglobin (290 pm) and intermediate between the distances observed in the α and β subunits of haemoglobin (270 and 320–340 nm,

respectively). Thus, our geometry optimisation method gives a proper hydrogen-bond geometry. Interestingly, the strong hydrogen bond induces very small changes in the structure. Figure 1 is actually a superposition of the optimised structures of $ImPorFeO_2$ with and without the His-64 model, but very small differences between the two structures can be seen; the root-mean-squared deviation between the two structures is only 2.1 pm and the geometric parameters around the iron ion differ by less than 2 pm and 1° (see Table 5).

However, when the optimised structure of $ImPorFeO_2\cdots Im$ is compared to the crystal structure of oxy-myoglobin [41] (Figure 2), appreciable differences can be seen. For example, the relative orientation of $O_2$, the proximal imidazole ring, and the porphyrin ring is different. Yet, the orientation of these three fragments varies appreciably among various model complexes and there exists complexes with the same orientation as in the optimised structures [23]. Furthermore, it has been shown that the barrier for the rotation of $O_2$ around the Fe–O bond is low [39]. However, the major difference between the two structures in Figure 2 lies in the orientation of the distal imidazole ring. In the crystal structure, the Fe–O–O–$N^{\varepsilon 2}$ dihedral angle is 138°, whereas in the optimised structure, it is 175°. This means that $N^{\varepsilon 2}$ has moved 236 pm between the two structures, and in the crystal structure, it is fairly close to both oxygen atoms (277 and 295 pm). Moreover, the N–H–O angle is almost straight (178°) in the optimised structure, whereas it is bent in the crystal structure (164°). Yet, as we will see below, these structural differences have rather modest influence on the hydrogen-bond energy.

Interestingly, $O_2\cdots Im$ optimised in vacuum has a straight O–O–H angle, but when oxygen is bound to iron, the angle becomes much smaller. In fact, the angle in the optimised model is 106°, i.e. even less than what could be expected for a superoxide ion, 120°. In the crystal structure, the O–O–H angle is even lower, 91°. It is also notable that the distal imidazole ring in the optimised structure is not in the Fe–O–O plane, but rather perpendicular to it. Thus, the former structure is not optimal, as has been suggested before [23].

Figure 3 shows the structure of ImPorFeCO. As expected, the iron ion is octahedrally coordinated and the Fe–C–O bond is straight. The Fe–$N_{Por}$ and Fe–$N_{His}$ bond lengths are within the range observed experimentally, as can be seen in Table 6. The Fe–C bond (174 pm) is close to what is found in model complexes, but it is significantly shorter than what is observed in protein crystal

structures (177–240 pm [9,15,41-44]). Yet, Parrinello et al. obtain almost the same bond length in their optimisations (172 pm) [39], and Kushkuley and Stavrov have shown with calculations of the vibrational spectra of CO that it is most unlikely that the Fe–C bond is as long as the crystal structure suggests [46]. On the other hand, the C–O bond length (117 pm) is similar to the one found in crystal structures (117-120 pm), but slightly longer than in model complexes (around 112 pm) [9,15,23,41-44,45].

Interestingly, the ImPorFeCO complex also forms a fairly strong hydrogen bond with imidazole. The N–H distance is 210 pm and the bond is almost straight with C–O–H and O–H–N angles of 171° and 179°, respectively. The N–O distance is 312 pm which is similar to the one observed in the myoglobin His64Gln mutants (305-323 pm; in these proteins there is undoubtedly a hydrogen bond between Gln-64 and CO) [11,47]. As for the corresponding oxygen complex, very small changes in the geometry are observed when the hydrogen bond forms (less than 2 pm; the root-mean-squared difference is only 1.3 pm).

In Figure 4, the hydrogen-bonded structure is compared to the neutron structure of CO myoglobin (c.f. also Table 6). Although the general structure is similar, several differences can be seen. First, the neutron structure is disordered, so there are two alternative conformations of CO (differing only in the position of the oxygen atom). In both conformations, the Fe–C–O angle is bent (130-153°) and CO is tilted off the haem normal. Second, as for the oxygen complex, the relative orientation of the imidazole and haem rings differs. Third and most important, the interaction between CO and the distal histidine residue is different. In the neutron structure there is no hydrogen bond between His-64 and the CO molecule. His-64 is protonated on the $N^{\delta 1}$ atom, which is exposed to the solvent, whereas the $N^{\varepsilon 2}$ atom has a lone pair directed towards CO. The N–O and N–C distances are 260 and 311 pm and the N–O–C angle is 104° to the closest conformation of CO. The other conformation of CO is bent away from His-64, with a N–O distance of 348 pm. In the neutron structure, His-64 is much closer to the porphyrin ring than in the optimised structure; as for corresponding $O_2$ complexes, $N^{\varepsilon 2}$ has moved about 260 pm between the two structures. Possible explanations to these differences will be discussed below.

*Hydrogen bond energies*

Tables 5 and 6 show the geometry and hydrogen-bond energy of a number of complexes of imidazole and CO or $O_2$. A free $O_2$ molecule forms very weak hydrogen-bonded complexes, with a binding energy of less about 1 kJ/mole. For CO and imidazole, three types of stable interactions have been found, but all have a hydrogen-bond energy of less than 5 kJ/mole. However, when $O_2$ or CO binds to the iron porphyrin, the hydrogen bonds become much stronger. The energy of the hydrogen bond in the optimised ImPorFe$O_2$···Im structure is 35 kJ/mole. This large increase in the interaction energy is caused by the polarisation of the iron–oxygen bond. In fact, the haem–oxygen complex can best be described as Fe(III)–$O_2^-$ [23,39]. Similarly, the energy of the ImPorFeCO···Im hydrogen bond is 14 kJ/mole. This is a fairly strong energy (of the same size as for the ammonia dimer), which shows that a haem-bound CO is certainly not a bad hydrogen-bond acceptor as frequently has been assumed [1,15].

As was discussed above, the geometry of the distal imidazole group in our optimised complexes is quite different from those observed in the crystal structures. It is therefore conceivable that the hydrogen-bond energies in myoglobin are appreciably lower than in our optimised models. We have tried to estimate how much the hydrogen-bond energy may decrease by the restraints caused by the folding of the protein. First, we calculated the hydrogen-bond energy directly in the crystal structure of oxy-myoglobin. This gave only 10 kJ/mole, but this estimate is strongly affected by uncertainties in the crystal structure. Therefore, we reoptimised the structure (using the $NH_3$DfaFe$O_2$···Im model) keeping the Fe–O–O–N and O–O–N–$C^{\varepsilon 2}$ dihedral angles fixed at the crystal values. Interestingly, the resulting hydrogen-bond energy was only 3 kJ/mole lower than for the fully optimised structure and the geometry is virtually identical to the one found in crystals. Apparently, the hydrogen-bond energy is fairly insensitive to the geometry, and the geometric restraints caused by the crystal probably decrease the hydrogen-bond energy by about 3 kJ/mole.

Similarly, we calculated the interaction energy between the distal imidazole group for the neutron structure of CO–myoglobin. For both conformations of CO, we obtained an *unfavourable* interaction, by –32 kJ/mole for the conformation of CO that is closest to His-64, and by –9 kJ/mole for the other conformation. If His-64 was protonated on the nitrogen atom directed towards CO, the interaction becomes less unfavourable, but only by 6–8 kJ/mole (interaction energies –26 and –1

kJ/mole). This indicates that there may be some problem with the neutron structure or that there are other interactions that stabilise this unfavourable interaction between His-64 and CO. Next, we optimised $NH_3$DfaFeCO···Im with the Fe–C–N, C–N–$C^{\varepsilon 1}$–$N^{\delta 1}$, and Fe–C–N–$C^{\varepsilon 1}$ angles constrained to the value encountered in the neutron structure. This lowered the hydrogen-bond energy slightly more than in the $O_2$ complex, by 6 kJ/mole. Thus, the restraint caused by the protein structure does not seem to decrease the hydrogen-bond energy very much.

*Comparison with experiments*

In conclusion, we estimate the energy of the His-64···$O_2$ hydrogen in myoglobin bond to about 32 kJ/mole and the energy of a putative hydrogen bond between His-64 and CO to about 8 kJ/mole. Thus, the hydrogen bond to the distal histidine residue preferentially stabilises $O_2$ by about 24 kJ/mole. This is reasonably close to the experimental estimate of how much myoglobin favours $O_2$ before CO: 800 times or 17 kJ/mole, and it indicates that the hydrogen bond accounts for the major part of the discrimination between CO and $O_2$. This agrees with estimates based on site-directed mutagenesis, which indicates that the preferential electrostatic stabilisation of $O_2$ accounts for 12–17 kJ/mole of the discrimination between $O_2$ and CO [1,21,48]. It also indicates that steric hindrance plays only a minor role in the discrimination, probably in the lower part of the range suggested by recent experiments, 0.5–7 kJ/mole [1,21,48].

However, our *absolute* hydrogen-bond energies are higher than most experimental estimates. For example, the stabilising effects of the hydrogen bond between His-64 and $O_2$ in myoglobin has been estimated by mutation studies to 8–18 kJ/mole [1,21]. Moreover, if ether linkages in porphyrin models are replaced by amide groups, which may form hydrogen bonds to an oxygen ligand, the oxygen affinity increases only by a factor of about ten, corresponding to an energy of about 6 kJ/mole (yet, a combined quantum chemical and molecular mechanical study of this system indicates that the energy of such an interaction is about 21 kJ/mole [49]). Similarly, our hydrogen-bond energy of the CO complex (8 kJ/mole) is appreciably larger than the one estimated from experiments. For example, Olsen and Phillips estimate this energy to about 1 kJ/mole [1], whereas Ray et al. give a slightly larger estimate: 2.5-2.8 kJ/mole [15]. These differences may be an effect of competing hydrogen-bond interactions with solvent, as has been discussed for myoglobin [1].

However, it is more probable that the major part of the difference is due to entropic effects; the experimental estimates are free energies, whereas our calculated values are enthalpies. Thus it seems that entropic effects partly counteract the hydrogen bonding and that the effect is larger for $O_2$ (about 15 kJ/mole) than for CO (5 kJ/mole). This is in accord with the intuitive interpretation that a hydrogen bond would decrease the mobility of a bound CO or $O_2$ molecule, thereby decreasing the entropy, and that this decrease would be larger the stronger the hydrogen bond is. It

also explains why our estimate of the difference in hydrogen bond energy between CO and $O_2$ is larger than the experimentally measured discrimination.

Yet, another explanation has also to be considered for CO, namely that there may be no hydrogen bond between His-64 and CO in myoglobin under physiological conditions. This is supported by the neutron structure of the protein, showing that His-64 has no proton on $N^{\varepsilon 2}$, the atom closest to CO. Instead His-64 is protonated on $N^{\delta 1}$, and this proton is hydrogen bonded to a crystal water molecule. This indicates that such hydrogen-bond interactions are more favourable than the hydrogen bond to CO. In order to test this possibility, we studied the interaction between imidazole and water with the same quantum chemical method as for the haem complexes. The results in Table 7 show that two conformations are possible. The strongest hydrogen bond (29 kJ/mole) is obtained when water is the hydrogen donor. This conformation is 9 kJ/mole more stable than the interaction observed in the neutron structure, with imidazole as the hydrogen donor. This is quite confusing; according to these results it would be much more favourable if His-64 accepted a hydrogen bond from water and at the same time provided a hydrogen bond to the CO–haem complex (37 kJ/mole), than the structure observed in the neutron structure (less than 20 kJ/mole).

A possible solution to this problem could be that the unprotonated nitrogen atom on His-64 may form a favourable interaction with CO, as for the isolated imidazole–CO complex (see Table 6). However, we have performed extensive geometry optimisations in order to obtain such a structure, using both the ImPorFeCO···Im and $NH_3$DfaFeCO···Im models, and using constrained as well as free optimisations. Yet, all such attempts have been fruitless. No stable interaction between an unprotonated nitrogen atom and the iron-bound CO has been found. Instead the imidazole group either dissociates or it rotates until the protonated nitrogen atom points to the CO molecule. This is also in accord with our observation that the interaction between His-64 and CO in the neutron structure is repulsive. Thus, we can only conclude that our theoretical results are inconsistent with the neutron structure of CO–myoglobin. The issue will be discussed more below in the section about vibrational frequencies.

Still, it should be noted that our estimates of the *relative* strength of the hydrogen bonds are consistent with experimental results. In the deoxy form of myoglobin, a water molecule interacts with $N^{\varepsilon 2}$ of His-64 (but not with the iron ion) and this water molecule has to be displaced when CO

or $O_2$ binds [1]. This hydrogen bond to water has been estimated to be about 4 times (3.5 kJ/mole) stronger than the one to CO and 100 times (11.5 kJ/mole) weaker than the one to $O_2$ [1]. These estimates are close to the difference in hydrogen-bond energy between imidazole and water, CO, or $O_2$, –6 and +15 kJ/mole, respectively. However, the experimental energy of the hydrogen bond between His-64 and a water molecule is again much lower than our estimate, 6 compared to 21 kJ/mole [1], probably due to the entropic effects.

Our calculations allow us also to calculate the binding energies of CO and $O_2$ to the porphyrin model (ImPorFe). The result is 73 kJ/mole for CO and 50 kJ/mole for $O_2$. Thus, the binding energies nicely reflect the stronger affinity of CO to a free porphyrin, and the difference in binding energy, 23 kJ/mole is very close to the observed difference, 25 kJ/mole [1,39]. It should be noticed, however, that our estimate is quite uncertain; calculations with the BP functional gives much larger binding energies and a much larger difference. Moreover, the relaxation effects are large, 24–43 kJ/mole.

*Vibrational frequencies and the CO vibrational substates in myoglobin*

Vibrational spectroscopy of CO–myoglobin has proved to be a powerful technique to study the conformation of CO in the protein and its various mutants [1,12,15,48,50]. In accord with many crystal structures of myoglobin, the vibrational spectra of CO-myoglobin indicate that the CO molecule binds in several different conformations. In mammalian myoglobins, three or four conformers of CO are observed, denoted $A_0$ (1965 $cm^{-1}$), $A_{1,2}$ (1945 $cm^{-1}$; or $A_1$: 1949 $cm^{-1}$, $A_2$: 1942 $cm^{-1}$), and $A_3$ (1932 $cm^{-1}$) [12]. Under physiological conditions, $A_{1,2}$ and $A_3$ dominate, whereas $A_0$ is the dominant state at low pH. However, the correlation between the vibrational states and different coordination geometries of the CO molecule has been elusive. It is widely accepted that the $A_0$ state corresponds to a conformation where His-64 is protonated and removed from the vicinity of the CO molecule [13,15,48]. For the other states, several suggestions have been made, involving tautomers and ring-flip isomers of His-64 (which of the imidazole nitrogen atom is protonated and which atom points towards CO) and solvent water molecule at varying positions [12,15,51].

In order to test the various structural suggestions and to get some clues about the hydrogen-bond structure between His-64 and CO, we have calculated vibrational frequencies for a number of different CO-myoglobin models. The calculated frequencies (Fe–C/O and C/O–O) are shown in Table 8. For the full porphine models, the C–O frequency is found at 1952–1982 $cm^{-1}$, close to the range measured for the various A states of CO myoglobin, 1930-1966 $cm^{-1}$ [1,12,50] and between the frequencies observed for free CO and simple carbonyls (2170 and 1900-2000 $cm^{-1}$, respectively [12]). However, the Fe–C frequency is calculated at a too high value, 570-581 $cm^{-1}$ compared to the experimental values around 510 $cm^{-1}$ [1]. Considering the approximations involved and the strong sensitivity of vibrational frequencies to the geometry and the level of theory, these results are satisfactorily, especially as the *difference* between the frequencies calculated for similar systems can be expected to be appreciably more accurate.

The same seems to be true for our small models. The $NH_3$DfaFeCO···Im models give slightly larger errors compared to experiments, a 33 $cm^{-1}$ too low C–O frequency and a 80 $cm^{-1}$ too high Fe–C frequency. Yet, the *trends* follow those of the full porphyrin model perfectly; a 29 $cm^{-1}$ difference for $\nu_{CO}$ and 13 $cm^{-1}$ for $\nu_{FeC}$ (30 and 11 $cm^{-1}$ for the large model) when a model of His-64 is added. Therefore we used the $NH_3$DfaFeCO···Im model for most of the investigative calculations.

Our results confirm the suggestion that the $A_0$ state involves a conformation in which there is no hydrogen bond between CO and His-64. For all systems, CO without any further interaction exhibits a 26–33 $cm^{-1}$ higher $\nu_{CO}$ frequency than the hydrogen bonded systems, in conformity with the 16–36 $cm^{-1}$ higher $\nu_{CO}$ frequency of the $A_0$ state compared to the other states [15]. This also agrees with the fact that hydrophobic mutants of His-64 have $\nu_{CO}$ frequencies close to 1965 $cm^{-1}$ [12].

To explain the $A_{1,2}$ and $A_3$ states, we have calculated the frequencies of $NH_3$DfaFeCO with imidazole, 2-methyl imidazole, 3-methyl imidazole, water, and with imidazole hydrogen bonded to a water molecule (with the other nitrogen atom). All frequencies are shown in Table 8 and Figure 5. The results show several interesting aspects. First, the points with the same models of the porphyrin and the proximal histidine residue show an approximate linear relation between the $\nu_{CO}$ and $\nu_{FeC}$ frequencies. This correlation has frequently been observed experimentally and it is attributed to a

Fe $d_\pi \varnothing$ CO $\pi^*$ back-donation; as it increases, the Fe-C bond order goes up and the CO bond order goes down [15,52]. Similarly, there is an approximate correlation between the hydrogen-bond strength and the frequencies ($\rho = 0.986$ to $\nu_{CO}$). Strong-hydrogen bonds correlate with a high $\nu_{FeC}$ frequency and a low $\nu_{CO}$ frequency. This shows that distal hydrogen-bond interactions are of crucial importance for the various CO frequencies.

Second, the frequencies of 2-methyl and 3-methyl imidazole (and also unsubstituted imidazole) are almost identical (within 2 $cm^{-1}$). Thus, the two tautomers of histidine (with $H^{\varepsilon 2}$ or $H^{\delta 1}$) inherently give rise to similar interactions. However, since the two nitrogen atoms have different distances to the backbone, steric limitations may change their interactions with CO (and therefore the corresponding frequencies) as has been shown by molecular dynamics simulations [53].

Third, imidazole gives a slightly stronger hydrogen bond than water (by 3 kJ/mole). This is also reflected in the frequencies (a lower $\nu_{CO}$ and a higher $\nu_{FeC}$ for the imidazole complex). If another water molecule forms a hydrogen bond to the imidazole molecule, imidazole is polarised and the hydrogen bond to CO becomes 4 kJ/mole stronger. This is the point that comes highest to the left in the $\nu_{CO}$–$\nu_{FeC}$ diagram.

In Figure 5, the experimental frequencies of the three vibrational states of myoglobin are also included (the figure shows differences in the frequencies relative to the $A_0$ state or the state without any distal interaction, $NH_3$DfaFeCO). If the $A_0$ frequency is interpreted as the complex without any distal interactions, it can be seen that all the calculated frequencies for unconstrained complexes (except $NH_3$DfaFeCO···$NH_3$) fall between the $A_{1,2}$ and $A_3$ states. This is most likely due to that our calculations do not account for the steric restrictions of the active site in myoglobin, nor for the dynamics of the hydrogen bonding. Both these effects would weaken the hydrogen bond and therefore move the points down the correlation line towards the $A_{1,2}$ state. Due to these limitations, it is very hard to make any definitive assignments of the experimentally observed vibrational states (especially as the interaction with water probably is less restricted by the protein folding that the one with imidazole). Therefore, we confine ourselves to note that hydrogen-bond interactions are crucial to explain the vibrational frequencies, that various hydrogen-bond interactions with the

distal histidine residue and water molecules may well explain the $A_{1,2}$ and $A_3$ states observed in myoglobin, and that the $A_{1,2}$ state most likely involves a hydrogen bond between His-64 and CO.

It has been suggested that the $A_3$ state corresponds to the interaction observed in the neutron structure, i.e. a structure where His-64 is protonated on the $N^{\delta 1}$ atom that is exposed to the solvent, whereas the lone-pairs of the $N^{\varepsilon 2}$ atom are directed towards the CO molecule [15]. Such a donor-acceptor interaction has been proposed to induce some $sp^2$ character onto the CO carbon atom, which also could explain the bent conformation of CO observed in the crystal structures [15]. This is quite reasonable, considering that we could optimise a complex of CO and imidazole where the unprotonated nitrogen atom interacted with both atoms in CO (Table 6). However, this complex had a decreased C–O bond length (compared to free CO) and therefore a higher $\nu_{CO}$ frequency, indicating that such a complex would have a higher $\nu_{CO}$ frequency than the $A_0$ state, rather than the lower frequency observed for the $A_3$ state.

As was discussed above, we have made several attempts to model such an interaction between CO and unprotonated nitrogen atom with our myoglobin models, but with no success. First, we calculated the frequencies for $NH_3$DfaFeCO···Im where the six internal coordinates between CO and imidazole were kept fixed to the values encountered in the neutron structure of CO-myoglobin. As can be seen in Figure 5 (Im Lp), this gave rise to a point rather close to $A_0$ but below the $\nu_{CO}$–$\nu_{FeC}$ correlation line. If the O–H distance was allowed to change, imidazole dissociated.

Second, we placed a positive point charge (+0.1–0.3 *e*) 100 pm from the carbon atom in CO (no distal imidazole, Cq0.1 and Cq0.2 in Figure 5). This led to a slightly increased $\nu_{CO}$ frequency and a strongly decreased $\nu_{FeC}$ frequency (i.e. again a point below the correlation line). Thus, the effect of such charges is opposite to the difference between the $A_0$ and $A_3$ frequencies. Similarly, if a point charge (+0.1–0.6 *e*) was placed 100 pm from the oxygen atom in CO at a q–O–C angle of 120°, the $\nu_{CO}$ frequency increased and $\nu_{FeC}$ decreased strongly, but this time the points lie close to the correlation line (Oq0.1; the points for q>0.1 *e* are outside the range shown in Figure 5). Again, the change is opposite to the one observed for $A_3$. However, it is noteworthy that such charges cause the Fe–C–O angle to bend slightly (176° for q=+0.6).

Next, we performed a series of calculations where the distance between the oxygen atom in CO and the unprotonated N atom of imidazole was kept fixed (B250, B300, and B350 in Figure 5).

All other geometric parameters were optimised. This led to structures where imidazole interacts through the $H^{\varepsilon 1}$ atom (i.e. a non-polar hydrogen atom) with CO, forming a weak hydrogen bond (O–H distance around 230 pm). For shorter O–N distances (250–280 pm; it is 260-348 pm in the neutron structure or myoglobin), this led to a higher $\nu_{CO}$ frequency and a lower $\nu_{FeC}$ than for the $A_0$ model (again in the wrong direction). This is most likely an effect of the repulsion between the lone-pairs on nitrogen and oxygen, in accord with the high $\nu_{CO}$ frequency (1984 $cm^{-1}$) of the His64Val+Val68Thr myoglobin double mutant, which has an oxygen lone pair directed towards CO. However, when the O–N distance is larger than 350 pm for the large model (ImPorFeCO···Im) and larger than 280 pm for the small model ($NH_3$DfaFeCO···Im), frequencies in the right direction were obtained. In fact, for the small model, frequencies close to those of $NH_3$DfaFeCO···Im with the protonated nitrogen directed towards CO were obtained when O–N was around 400 pm. It should be noted that in all these complexes, the interaction between imidazole and CO is unfavourable; the energy of the complex decreased as the O–N bond was elongated and if the constraint was released the imidazole group dissociated or rotated.

However, it is unlikely that this interaction may give rise to the $A_3$ state of myoglobin. First, if also the Fe–C–N and Fe–C–N–$C^{\varepsilon 1}$ angles are fixed to the values encountered in the crystal structure, the frequencies are shifted down the correlation line, even past the $A_0$ state (D300 and D350 in Figure 5). Second, the effect in the more realistic ImPorFeCO···Im system is much smaller and in the wrong direction for realistic O–N distances (250-350 pm). Interestingly, there is a hydrogen-bond interaction between a polar hydrogen on the Dfa group and the unprotonated nitrogen atom of imidazole. At large N–O distances, this interaction drags away the imidazole group from CO, but the oxygen atoms partly follow the movement by bending the Fe–C–O moiety. Thus, the $A_3$-like frequencies obtained with these complexes might arise from the bent Fe–C–O angle (172-174°).

In order to test this possibility, we performed some calculations with $NH_3$DfaFeCO forcing the Fe–C–O angle to be bent at different angles. However, as can be seen from Figure 5 (A170 and A165), this gave almost no change in the $\nu_{CO}$ frequency, whereas the $\nu_{FeC}$ frequency *decreased* strongly, leading to points below the correlation line.

In conclusion, our results confirm that the $A_0$ state represents a geometry without any hydrogen bond between CO and His-64, and they indicate that the $A_{1,2}$ and $A_3$ states arise from hydrogen-bond interactions between the protein and CO. However, our results are incompatible with the neutron structure of myoglobin. We predict that the interactions observed in the neutron structure would give rise to frequencies close to the $A_0$ state and probably below the correlation line, and they rule out the possibility that either of the other two states can arise from the interactions observed in the neutron structure. Furthermore, our results indicate that there is no advantage of having the hydrogen atom on the $N^{\delta 1}$ atom rather than on the $N^{\varepsilon 2}$ atom. On the contrary, the $N^{\delta 1}$ atom is exposed to the solvent and a hydrogen bond between water and imidazole is more favourable if water is the donor and imidazole is the acceptor; then the $H^{\varepsilon 2}$ atom would be free to form a hydrogen bond with CO, an interaction of sizeable strength, about 8 kJ/mole. The only possible explanation we can see to the interactions in the neutron structure is that His-64 interacts with a sulphate ion rather than a water molecule (the crystals were grown in saturated ammonium sulphate [9]). If so, such an interaction would be much stronger than the one with water and it would require a protonated $N^{\delta 1}$.

In fact, there are much experimental and theoretical evidence indicating that the protonation status of His-64 observed in the neutron structure is not typical for myoglobin in solution. For example, in the His64Gln myoglobin mutant, the $N^{\varepsilon}$ atom of Gln is positioned at almost exactly the same place as $N^{\varepsilon 2}$ in His-64, although the former atom is undoubtedly protonated, whereas the latter according to the neutron structure is not [9,11,47]. The two proteins also have similar vibrational spectra. Likewise, investigation of the $\nu_{CO}$ frequencies and CO affinity for myoglobin mutants show clear correlation between the frequencies and the presence of distal hydrogen-bond donors [12]. Virtually all theoretical investigations of the CO frequencies in myoglobin have indicated that the $A_{1,2}$ and $A_3$ states arise from polar interactions with CO and that lone-pair interactions would increase $\nu_{CO}$ [46,50,54]. Moreover, molecular dynamics simulations with $N^{\delta 1}$ of His-64 protonated (as in the neutron structure) indicate that the imidazole ring rapidly rotates and exposes the polar hydrogen towards CO [53].

The important conclusion from this discussion is that there are ample evidence indicating that there is a hydrogen bond between His-64 and CO in CO-myoglobin. Therefore, is it reasonable to

compare the strength of the hydrogen bond between His-64 and CO or $O_2$ bound to myoglobin as we have done above, showing that these results can be expected to have relevance to the discrimination between CO and $O_2$ by this protein. Consequently, we can conclude that our results show quantitatively that myoglobin preferentially stabilises bound $O_2$ by about 24 kJ/mole by hydrogen bonding to His-64.

*Acknowledgements*

This investigation has been supported by grants from the Swedish Natural Science Research Council (NFR). It has also been supported by computer resources of the Swedish Council for Planning and Coordination of Research (FRN), Parallelldatorcentrum (PDC) at the Royal Institute of Technology, Stockholm, the National Supercomputer Centre (NSC) at the University of Linköping, and the High Performance Computing Center North (HPC2N) at the University of Umeå.

**References**


1. Olson JS, Phillips Jr GN (1997) J Biol Inorg Chem 2:544-552
2. Coburn RF (1970) Ann NY Acad Sci, 174:11-22
3. Ghosh A (1997) J Biol Inorg Chem 2:515
4. Collman JP, Brauman JI, Halbert TR, Suslick KS (1976) Proc Natl Acad Sci USA, 73:3333-3337
5. Norvell JC, Nunes AC, Schoenborn BP (1975) Science 190:568-570
6. Hanson JC, Schoenborn BP (1981) J Mol Biol 153:117-146
7. Kuriyan J, Wilz s, Karplus M, Petsko GA (1986) J Mol Biol 192:133-154
8. Cheng X, Schoenborn B (1990) Acta Crystallogr B46:195-208
9. Cheng X, Schoenborn B (1991) J Mol Biol 220:381-399
10. Stryer L, In: Biochemistry (1988) 3rd edn. Freeman, NY, p 149
11. Quillin ML, Arduini RM, Olson JS, Phillips GN Jr (1993) J Mol Biol 234:140-155
12. Li T, Quilin ML, Phillips GN, Olson JS (1994) Biochemistry 33:1433-1446.
13. Yang F, Phillips GN Jr (1996) J Mol Biol, 256:762-774
14. Hu S, Vogel KM, Spiro TG (1994) J Am Chem Soc 116:11187-11188
15. Ray GB, Li X-Y, Ibers JA, Sessler JL, Spiro TG (1994) J Am Chem Soc 116:162-176
16. Springer B.A., Sligar S.G., Olson J.S., Phillips G.N Jr (1994) Chem Rev 94:699-714
17. Lim M, Jackson TA, Anfinrud PA (1995) Science 269:962-966
18. Lim M, Jackson TA, Anfinrud PA (1997) J Biol Inorg Chem 2:531-536
19. Ghosh A, Bocian DF (1996) J Phys Chem 100:6363-6367
20. Spiro TG, Kozlowski PM (1997) J Biol Inorg Chem 2:516-520
21. Springer BA, Egeberg KD, Sligar SG, Rohlfs RJ, Mathews AJ, Olson JS (1989) J Biol Chem 264:3057-3060
22. Phillips SEV, Schoenborn BP (1981) Nature 292:81-82
23. Momenteau M, Reed C A (1994) Chem Rev 94:659-698
24. Olson JS, Phillips GN (1996) J Biol Chem 271:17593-17596
25. JE Newton, M.B Hall (1984) Inorg Chem 23:4627-4632
26. Becke AD (1988) Phys Rev A 38:3098-3102

27. Perdew JP (1986) Phys Rev B 33:8822-8824
28. Vosko SH, Wilk L, Nusair M (1980) Canadian J Phys 58:1200-1211
29. Hay PJ., Wadt WR (1985) J Chem Phys 82:270-283
30. Stevens WJ, Krauss M, Basch H, Jasien PG (1992) Can J Chem 70:612-630
31. Here WJ, Radom L, Scheyer PvR, JA Pople (1986) Ab Initio Molecular orbital theory, Wiley-Interscience, New York.
32. Eichkorn K,Treutler O,Öhm H, Häser M, Ahlrichs R (1995) Chem Phys Lett 240:283-290
33. Eichkorn K,Weigend F, Ahlrichs R (1997) Theor Chem Acc 97:119-124
34. Bauschlicher CW (1995) Chem Phys Lett 246:40-44
35. Alrichs R, Bär M, Häser M, Horn H, Kölmel C (1989) Chem Phys Lett 162:165-169
36. Hertwig RH, Koch W (1997) Chem Phys Lett 268:345-351
37. A Schäfer, H Horn and R Ahlrichs (1992) J Chem Phys 97:2571-2577
38. Bytheway I, Hall MB (1994) Chem Rev 94:639-658.
39. Rovira C, Kunc K, Hutter J, Ballone P, Parrinello M (1997) J Phys Chem A 101:8914-8925.
40. van Duijneveldt FB, van Duijneveldt-van de Rijdt JGCM, van Lenthe JH (1994) Chem Rev 94:1873-1885.
41. Phillips, SEV (1980) J Mol Biol 142:531-554
42. Jameson GB, Ibers JA (1994) in Bioinorganic Chemistry (Bertini I, Gray HB, Lippard SJ, Valentine JS, eds.), University Science, pp 167-252
43. Phillips SEV , Schoenborn BP (1981) Nature 292:81-82
44. Kuriyan J, Wilz S, Karplus M, Petsko GA (1986) J Mol Biol 192:133-154
45. Peng S-M, Ibers JA (1976) J Am Chem Soc 98:8032-8036
46. Kushkuley B, Stavrov SS (1997) Biophys J 72:899-912
47. Zhao X, Vyas K, Nguyen BD, Rajarathnam K, Lamar GN, Li T, Phillips GN, Eich R, Olson JS, Ling J, Bochian DF (1995) J Biol Chem 270:20763-20771
48. Sage JT (1997) J Biol Inorg Chem 2:537-543
49. Maseras F (1998) New J Chem 327-332
50. Straub JE, Karplus M (1991) Chem Physics 158:221-248
51. Oldfield E, Guo K, Augspurger JD, Dykstra CE (1991) J Am Chem Soc 113:7537-7541

52. Li X-Y, Spiro TG, J Am Chem Soc., (1988) 110:6024-6033

53. Jewsbury P, Kitagawa T (1994) Biophys J 67:2236-2250

54. Kushkuley B, Stavrov SS (1996) Biophys J 70:1214-1229

55. Frisch MJ, Trucks GW, Schlegel HB, Gill PMW, Johnson BG, Robb MA, Cheeseman JR, Keith T, Petersson GA, Montgomery JA, Raghavachari K, Al-Laham MA, Zakrzewski VG, Ortiz JV, Foresman JB, Cioslowski J, Stefanov BB, Nanayakkara A, Challacombe M, Peng CY, Ayala PY, Chen W, Wong MW, Andres JL, Replogle ES, Gomperts R, Martin RL, Fox DJ, Binkley JS, Defrees DJ, Baker J, Stewart JP, Head-Gordon M, Gonzalez C, and Pople JA, Gaussian 94, Revision D.1, Gaussian, Inc., Pittsburgh PA, 1995

56. Huber KP, Herzberg G (1979) Molecular spectra and Molecular Structure, Vol 4 Constants of diatomic molecules, Van Nostrand Reinhold Company, London

**Table 1.** Comparison of the frequencies of $NH_3DfaFeCO$ and $NH_3DfaFeCO\cdots NH_3$, calculated analytically and with our approximate method (Eqns. 2–4). The frequencies were calculated with the BP method and the DZpdf/6-31G* basis set using the Gaussian-94 software [55].

| | $\nu_{CO}$ | | $\nu_{FeC}$ | |
|---|---|---|---|---|
| | Numerical[a] | Analytical | Numerical[a] | Analytical |
| $NH_3DfaFeCO$ | 1898.3–1903.7 | 1926.7 | 577.0–577.1 | 572.5 |
| $NH_3DfaFeCO\cdots NH_3$ | 1879.0–1882.4 | 1906.8 | 588.2–589.0 | 585.8 |

[a] Estimated using Eqns. 2–4. The two values given are those obtained with positive and negative displacements, respectively.

**Table 2.** The performance of different theoretical methods on the CO···Im complex using the 6-31+G** basis set.

| Method | C–O (pm) | N–H (pm) | H–C (pm) | ΔE (kJ/mole) |
|---|---|---|---|---|
| HF | 111.1 | 99.4 | 269.3 | 3.8 |
| S–VWN | 113.8 | 102.7 | 213.5 | 16.6 |
| BP | 114.6 | 102.1 | 241.7 | 5.2 |
| BP–RI | 114.6 | 102.1 | 241.5 | 5.3 (5.8[a]) |
| B3LYP | 113.5 | 101.2 | 246.2 | 6.0 |
| MP2 | 115.0 | 101.0 | 241.9 | 8.2 |

[a] Energy calculated with the B3LYP method on the BP–RI geometry.

**Table 3.** The performance of effective core potentials (ECPs) on the geometry and hydrogen-bond energies of $NH_3DfaFeCO/O_2\cdots Im$.

| Molecule | Basis set | Fe–C/O (pm) | C/O–O (pm) | N–H (pm) | H–O (pm) | ΔE (kJ/mole) |
|---|---|---|---|---|---|---|
| $NH_3DfaFeCO\cdots Im$ | ECP | 171.2 | 117.9 | 102.5 | 203.4 | –16.7 |
| | 6-31G* | 171.3 | 118.5 | 102.2 | 202.3 | –17.5 |
| $NH_3DfaFeO_2\cdots Im$ | ECP | 184.6 | 131.5 | 103.8 | 184.9 | –38.9 |
| | 6-31G* | 182.0 | 132.4 | 104.0 | 183.9 | –38.2 |

**Table 4.** The geometries and relative energies of the five-coordinate myoglobin models.

| Molecule | $Fe–N_{His}$ (pm) | $Fe–N_{Por}$ (pm) | Fe–Por[a] (pm) | Relative Energy (kJ/mole) |
|---|---|---|---|---|
| $^1NH_3DfaFe$ | 192 | 201–202 | 15 | 243.0 |
| $^3NH_3DfaFe$ | 221 | 199–200 | 21 | 113.1 |
| $^5NH_3DfaFe$ | 217 | 212–215 | 32 | 0.0 |
| $^1ImPorFe$ | 189 | 199 | 15 | 33.6 |
| $^3ImPorFe$ | 222 | 200–201 | 11 | 0.0 |
| $^5ImPorFe$ | 215 | 208–210 | 31 | 0.5 |
| Porphyrin models[b] | 209–216 | 207–209 | 30–43 | |
| Protein structures[c] | 200–222 | 203–210 | 42–63 | |

[a] The distance of the iron ion out of the average porphyrin plane, defined by the four N atoms.

[b] Porphyrin models [15,23,42,45]

[c] Myoglobin crystal structures [9,15,41,42,44]

**Table 5.** The geometry and hydrogen-bond energies of the various $O_2$ complexes.

| Molecule | $Fe–N_{His}$ (pm) | $Fe–N_{Por}$ (pm) | Fe–Por[a] (pm) | Fe–O (pm) | O–O (pm) | N–O (pm) | H–O (pm) | Fe–O–O | $\Delta E$[b] (kJ/mole) |
|---|---|---|---|---|---|---|---|---|---|
| $O_2$ | | | | | 123 | | | | |
| $O_2\cdots$Im | | | | | 123 | 418 | 316 | | –0.3 |
| $NH_3DfaFeO_2$ | 205 | 197–202 | 6 | 183 | 131 | | | 121 | |
| $NH_3DfaFeO_2\cdots NH_3$ | 204 | 196–203 | 7 | 182 | 132 | 315 | 216 | 122 | –21.3 |
| $NH_3DfaFeO_2\cdots$Im | 204 | 196–202 | 6 | 182 | 132 | 287 | 183 | 121 | –38.2 (–36.0) |
| $NH_3DfaFeO_2\cdots$Im[c] | 204 | 196–202 | 6 | 182 | 132 | 286 | 183 | 121 | –35.5 |
| $ImDfaFeO_2\cdots$Im | 203 | 196–201 | 0 | 184 | 132 | 286 | 182 | 121 | –40.4 |
| $ImPorFeO_2$ | 209 | 200–203 | 3 | 181 | 129 | | | 121 | |
| $ImPorFeO_2\cdots$Im | 207 | 200–203 | 3 | 180 | 131 | 293 | 190 | 120 | –36.1 (–34.7) |
| Porphyrin models[d] | 207–211 | 198–200 | –3–11 | 175–190 | 122–124 | | | 129–131 | |
| Protein structures[e] | 205–207 | 195–202 | 12–18 | 180–183 | 121,122[f] | 270–340 | 198 | 115–123 | |

[a] The distance of the iron ion out of the average porphyrin plane, defined by the four N atoms.

[b] $\Delta E$ values in brackets are calculated with Eqn. 1, i.e. they are corrected for both basis superposition error and geometry relaxation effects.

[c] The Fe–O–O–N and O–O–N–$C^{\varepsilon 1}$ dihedral angles were kept fixed at the crystal values 138.3° and –22.6°.

[d] Porphyrin models [15,23,42,45]

[e] Protein crystal and neutron structures [15,41-44]

[f] Free $O_2$ [56]

**Table 6.** The geometry and hydrogen-bond energies of the various CO complexes.

| Molecule | $Fe–N_{His}$ (pm) | $Fe–N_{Por}$ (pm) | Fe–Por[a] (pm) | Fe–C (pm) | C–O (pm) | N–O (pm) | H–O (pm) | Fe–C–O | $\Delta E$[b] (kJ/mole) |
|---|---|---|---|---|---|---|---|---|---|
| CO | | | | | 115 | | | | |
| CO···Im | | | | | 115 | 351 | 249 | | –3.1 |
| OC···Im | | | | | 115 | 345 | 244 | | –5.3 |
| OC···Im[c] | | | | | 115 | 312 | | | –3.3 |
| $NH_3$DfaFeCO | 203 | 203 | 16 | 173 | 118 | | | 180 | |
| $NH_3$DfaFeCO···Im | 203 | 203 | 16 | 172 | 119 | 305 | 202 | 179 | –17.5 (–16.9) |
| $NH_3$DfaFeCO···Im[d] | 203 | 204-206 | 13 | 172 | 118 | 326 | 244 | 177 | –11.5 |
| $NH_3$DfaFeCO···2MeIm | 203 | 204 | 15 | 172 | 118 | 307 | 205 | 180 | –16.5 |
| $NH_3$DfaFeCO···3MeIm | 203 | 204 | 15 | 172 | 118 | 307 | 205 | 179 | –16.6 |
| $NH_3$DfaFeCO···Im···$H_2O$ | 203 | 204 | 15 | 171 | 119 | 301 | 199 | 179 | –21.4 |
| $NH_3$DfaFeCO···Im[e] | 203 | 205 | 15 | 174 | 118 | 348 | | 171 | +12.9 |
| $NH_3$DfaFeCO···$H_2O$ | 203 | 204 | 14 | 172 | 118 | | 205 | 180 | –14.5 |
| $NH_3$DfaFeCO···$NH_3$ | 203 | 203 | 16 | 172 | 119 | 341 | 239 | 178 | –8.2 |
| ImDfaFeCO···Im | 205 | 203 | 9 | 171 | 119 | 305 | 203 | 179 | –18.6 |
| ImPorFeCO | 206 | 204 | 4 | 174 | 117 | | | 180 | |
| ImPorFeCO···Im | 206 | 204 | 4 | 172 | 118 | 312 | 210 | 180 | –16.3 (–14.0) |
| Porphyrin models[f] | 204-210 | 197-205 | –2-2 | 171-181 | 112,113[h] | | | 173-180 | |
| Protein structures[g] | 210-231 | 183-208 | 0-7 | 177-240 | 111-121 | 260-394 | | 120-179 | |

[a] The distance of the iron ion out of the average porphyrin plane, defined by the four N atoms.

[b] $\Delta E$ values in brackets are calculated with Eqn. 1, i.e. they are corrected for both basis superposition error and geometry relaxation effects.

[c] CO interacts with the unprotonated N atom of imidazole as well as the H atom of an adjacent carbon. The H–C, N–C, and N–O distances are 304, 312, and 365 pm, respectively.

[d] The Fe–C–N, C–N–$C^{\varepsilon 1}$–$N^{\delta 1}$, and Fe–C–N–$C^{\varepsilon 1}$ angles were kept fixed at the crystal values 116.6, 150.9, and 36.1°.

[e] The structure was optimised with six constrained internal coordinates between CO and imidazole (from the neutron structure; the deprotonated N atom of imidazole was directed towards CO).

[f] Porphyrin models [15,23,42,45].

[g] Protein crystal and neutron structures [9,11,15,42,44,47].

[h] Free CO [56].

**Table 7.** The geometry and energy of hydrogen bonds between imidazole and water. ΔE values in brackets are calculated with Eqn. 1, i.e. they are corrected for both basis superposition error and geometry relaxation effects.

| | hydrogen bond (pm) | ΔE (kJ/mole) |
|---|---|---|
| imidazole···$OH_2$ | 198 | 20.8 (20.2) |
| imidazole···HOH | 175 | 30.3 (28.7) |

**Table 8.** C–O and Fe–C bond lengths and $\nu_{CO}$ and $\nu_{FeC}$ vibrational frequencies for a number of complexes involving CO. The frequencies were calculated with our approximate quantum chemical method (Eqns. 2–4) and were then used to construct the regression lines in Eqns. 5 and 6.

| | $\nu_{CO}$ ($cm^{-1}$) | $\nu_{FeC}$ ($cm^{-1}$) | C–O (pm) | Fe–C (pm) |
|---|---|---|---|---|
| CO | 2116 | | 113.73 | |
| ImPorFeCO | 1982 | 570 | 116.99 | 173.63 |
| ImPorFeCO···Im | 1952 | 581 | 117.46 | 172.16 |
| ImDfaFeCO···Im | 1909 | 595 | 118.29 | 171.13 |
| $NH_3$DfaFeCO | 1933 | 579 | 117.85 | 172.69 |
| $NH_3$DfaFeCO···$HNH_2$ | 1919 | 584 | 118.11 | 172.19 |
| $NH_3$DfaFeCO···Im | 1905 | 595 | 118.39 | 171.46 |
| $NH_3$DfaFeCO···Im[a] | 1932 | 570 | 117.91 | 173.54 |
| $NH_3$DfaFeCO···Im···$H_2O$ | 1899 | 599 | 118.48 | 171.26 |
| $NH_3$DfaFeCO···2-MeIm | 1903 | 594 | 118.37 | 171.49 |
| $NH_3$DfaFeCO···3-MeIm | 1905 | 595 | 118.35 | 171.52 |
| $NH_3$DfaFeCO···HOH | 1907 | 590 | 118.35 | 171.63 |
| Experiment | 1932–1965[b] 2143[c] | 490–520[b] | | |

[a] The structure was optimised with six constrained internal coordinates between CO and imidazole (from the neutron structure; the deprotonated N atom of imidazole was directed towards CO).

[b] Myoglobin [1,12,15,23,48,50].

[c] Free CO [50]

**Figure Legends**

Figure 1. A comparison of the optimised structures of $ImPorFeO_2$ and $ImPorFeO_2\cdots Im$. The two structures are almost identical with a root-mean-squared deviation of only 2.1 pm.

Figure 2. A comparison of the optimised structures $ImPorFeO_2\cdots Im$ and the crystal structure of oxygenated myoglobin (shaded; taken from the PDB file 1mbo) [41]. Only atoms in the haem group and the proximal and distal histidine residues are shown.

Figure 3. A comparison of the optimised structures of ImPorFeCO and ImPorFeCO···Im. The two structures are almost identical with a root-mean-squared deviation of only 1.3 pm.

Figure 4. A comparison of the optimised structures ImPorFeCO···Im and the neutron structure of CO-myoglobin (shaded; taken from the PDB file 2mb5) [9]. Only atoms in the haem group and the proximal and distal histidine residues are shown.

Figure 5. Calculated $\nu_{CO}$ and $\nu_{FeC}$ frequencies for the $NH_3DfaFeCO\cdots X$ complexes, where *X* represents various models of His-64 as indicated in the figure. The frequencies are given as the difference from those obtained without any *X* molecule (which corresponds to the $A_0$ state). Im Lp denotes an imidazole group with the nitrogen lone-pair directed toward CO. Cq and Oq are point charges (+0.1 or +0.2 *e*) placed 100 pm from the C or O atom of CO. B denotes $NH_3DfaFeCO\cdots Im$, where the O–N distance (to the deprotonated N atom of imidazole) has been fixed to 250, 300, or 350 pm. In the corresponding D-states, the Fe–C–N and Fe–C–N–$C^{\varepsilon 1}$ angles have also been constrained to the values found in crystal structure, 143.2 and 17.5°. A denotes $NH_3DfaFeCO$ where the Fe–C–O angle has been fixed at 170° or 165°. The frequencies were obtained from the corresponding bond lengths using Eqns. 5 and 6. In addition, the three experimentally observed vibrational A states are included [15], as differences from the $A_0$ state (the $A_0$ state is therefore found in the origin together with the $NH_3DfaFeCO$ model without any *X* molecule).