Theoretical study of the discrimination between O(2) and CO by myoglobin.

Sigfridsson, Emma; Ryde, Ulf

Published in:
Journal of Inorganic Biochemistry

DOI:
10.1016/S0162-0134(02)00426-9

2002

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

Theoretical study of the discrimination between O_2 and CO by myoglobin

Emma Sigfridsson & Ulf Ryde

Department of Theoretical Chemistry

Lund University

Chemical Centre, P. O. Box 124

S-221 00 Lund, Sweden

Correspondence to U. Ryde

Tel: +46-46 222 45 02

Fax: +46-46 222 45 43

E-mail: Ulf.Ryde@teokem.lu.se

2017-04-01

Abstract

Combined quantum chemical and molecular mechanics geometry optimisations have been performed on myoglobin without or with O_2 or CO bound to the haem group. The results show that the distal histidine residue is protonated on the N atom and forms a hydrogen bond to the haem ligand both in the O_2 and the CO complexes. We have also re-refined the crystal structure of CO-myoglobin by a combined quantum chemical and crystallographic refinement. Thereby, we obtain the probably most accurate available structure of the active site of this complex, showing a Fe-C-O angle of 174° and a Fe-C bond length of 171 pm. The resulting structures have been used to calculate the strength of the hydrogen bond between the distal histidine residue and O_2 or CO in the protein. It amounts to 31-33 kJ/mole **(xxx update)** for O_2 and 2-3 kJ/mole for CO. The difference in hydrogen-bond strength is 21-22 kJ/mole when corrected for entropy effects. This is slightly larger than the observed discrimination between O_2 or CO by myoglobin, 17 kJ/mole. We have also estimated the strain energy of the active site in the protein. It is 2-4 kJ/mole *larger* for the O_2 complex than for the CO complex, independent of which crystal structure the calculations are based on. Together, these results clearly show that myoglobin discriminates between O_2 and CO mainly by electrostatic interactions, rather than by steric strain

Key words: Myoglobin, QC/MM calculations, Crystallographic refinement, CO/O_2 discrimination, Protein strain.

Introduction

Myoglobin is a small globular protein that reversibly binds O_2. The protein is found in muscle cells of all vertebrates and some invertebrates where it stores oxygen, providing it to the working muscles. Myoglobin is closely related to haemoglobin, the oxygen carrier in red blood cells; haemoglobin is essentially a tetramer of four myoglobin molecules. Both proteins bind O_2 to the ferrous iron ion of a haem group. An imidazole side chain of a histidine residue (the proximal histidine) from the protein completes the octahedral coordination sphere around the iron ion [1].

Myoglobin binds several small gaseous molecules besides O_2, e.g. CO and NO. It is well known that the protein affects the ligand-binding affinity of the haem group; in solution, CO binds to free haem 20 000 times stronger than O_2, but in myoglobin this factor is reduced to 25 [2]. Thus, myoglobin seems to favour the binding of O_2 before CO by ~17 kJ/mole. This discrimination is of vital importance; without it, we would suffocate from CO that is produced during the normal degradation of haem in our bodies. The mechanism of the discrimination is a central question in biochemistry and it has been vividly discussed for a long time [2-8].

Collman et al. [3,8] noted that O_2 and CO have different geometric preferences when bound to the haem iron: The Fe-CO bond is linear with a Fe-C-O angle close to 180 , whereas FeO_2 is bent, with an Fe-O-O angle around 120 . Therefore, they suggested that the discrimination is achieved by steric repulsion, i.e. that myoglobin forces also CO to bind in a bent fashion. Virtually all myoglobins and haemoglobins have a second, so called distal, histidine ligand positioned above the iron ion, but too far to coordinate directly to iron. It is in the right position to affect the geometry of the FeCO complex. This idea was supported by early crystal structures, showing Fe-C-O angles of 120-140° [9-13], and it has made its way into the textbooks [14]. However, the hypothesis has lately been questioned.

Several newer crystal structures show larger Fe-C-O angles (160-175°) [7,15-18] and experimental evidence from vibrational spectroscopy indicate that the Fe-C-O angle is nearly linear [19-26]. Moreover, recent theoretical calculations have indicated that the energy of bending and tilting the haem-CO complex is small for moderate deviations from linearity [27-31]. Therefore, the idea of a strongly bent FeCO unit as the reason for the discrimination is nowadays generally agreed to be incorrect [5].

Another difference between O_2 and CO complexes with haem is that the FeO_2 unit is much more polar than the FeCO unit. In fact, the electronic structure of FeO_2 is close to that of a superoxide anion (O_2^-) bound to a ferric ion [32-33]. Therefore, the protein might discriminate between O_2 and CO by electrostatic interactions [34]. Neutron scattering structures show that there is a hydrogen bond between O_2 and the distal histidine residue [35], whereas no such bond can be seen in the CO complex [13]. Moreover, site-directed mutagenesis and studies of synthetic haem models have shown that electrostatic interactions are of major importance for the relative affinity of O_2 and CO to haem [5,21,27,32,36].

In an earlier paper, we have estimated the energy of hydrogen bonds between the distal histidine residue and O_2 or CO bound to a haem group by quantum chemical methods [6]. These calculations showed that a hydrogen bond to O_2 is 24 kJ/mole stronger than one to CO. Thus, the energy difference is of the same magnitude as myoglobin's discrimination between O_2 and CO, giving strong support to the hydrogen-bond hypothesis. However, these calculations were performed in vacuum, modelling restrictions in the geometry caused by the protein by constraints in two or three dihedral angles.

In this paper, we improve these calculations using better methods. In particular, we optimise the geometry of the haem group with its ligands and a model of the distal histidine residue inside the protein, using combined quantum chemical and molecular mechanics methods. These structures are then used to calculate more realistic hydrogen-bond energies and to estimate strain energies inside the protein. A strong motivation for these studies is a recent crystallographic investigation, which argues strongly for steric strain as the main factor for the myoglobin's discrimination between O_2 and CO [7]. In this paper, we test some of their suggestions and use their raw data to obtain the probably most accurate structure of the haem:CO complex in myoglobin published to date.

Methods

Quantum chemical calculations

All quantum chemical calculations were performed with the density functional B3LYP method. Hybrid density functional methods have been shown to give as good or better geometries as correlated ab initio methods for first-row transition metal complexes [37-39],

and the B3LYP method in particular seems to give the most reliable results among the widely available density functional methods [40]. In the geometry optimisations and frequency calculations we used for iron the double- basis set of Schäfer et al. (62111111/33111/311) [41], enhanced with one *d*, one *f*, and two *p* functions with exponents 0.1244, 1.339, 0.134915, and 0.041843, respectively. For the other atoms, we have employed the 6-31G* basis set, except for the four atoms involved in the hydrogen bond (N_{donor}, H, and O_2 or CO), for which we used the 6-31+G** basis, with an additional set of diffuse function [42]. Earlier calculations have shown that this is necessary to get accurate geometries and energies [6,43]. Only the pure 5 *d* and 7 *f*-type functions were used.

For accurate energy calculations, we used instead the 6-311+G(2d,2p) basis set and enhanced the basis set for iron with one *s* and one *p* function, with exponents 0.01377232 and 0.041843, and replaced the *f* function, with two new *f* functions with exponents 2.5 and 0.8. Geometry optimisations and energy calculations were performed with the Turbomole software [44,45]. Frequencies were calculated with Gaussian-98 [46] and they were scaled by a factor of 0.963 [47]. These were used to calculate entropy and thermodynamic corrections to Gibbs free energy at 300 K and 101.3 kPa pressure using standard methods of statistical mechanics [48]. All calculations were run on SGI Origin 2000 or Octane workstations.

In the quantum chemical calculations, the haem group was modelled by a porphine ring (Por; a haem group without side chains) with a central Fe(II) ion, the proximal and distal histidine residues were modelled by imidazole (Im), and the distal iron ligand was included explicitly (O_2 or CO). Experiments have shown that the Fe(II) ion in myoglobin without any distal ligand is in the high-spin state, whereas the complexes with O_2 or CO are diamagnetic [49]. Therefore, 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 O_2 bound to haem were treated within the unrestricted-spin formalism, because the lowest electronic state is an open-shell singlet formed from the antiferromagnetic coupling of low-spin Fe(III) and a superoxide radical anion [33]. No symmetry restraints were imposed in the calculations.

QC/MM calculations

Combined quantum chemical and molecular mechanics (QC/MM) calculations were used to study the hydrogen-bond strengths in myoglobin. They were performed with the COMQUM-00 software [50-51], which is a combination of Turbomole [44-45] for the quantum chemical calculations and Amber [52] for the molecular mechanics calculations. The haem group without the side chains, the imidazole ring of the distal and proximal histidine, and the oxygen or carbon monoxide molecule were treated by quantum chemistry, using the B3LYP method, as described above.

In the quantum chemical calculations, the quantum system was represented by a wavefunction, whereas all the other atoms were represented by an array of point charges, one for each atom, taken from the Amber libraries. Thereby, the polarisation of the quantum chemical system by the surroundings is included in a self-consistent manner. In the classical force and energy calculations, all atoms were represented by the Amber force field, but without any electrostatic interactions (which are already treated by quantum mechanics). Special action is taken where there is a bond between the classical and quantum chemical systems (a junction, i.e. for the porphyrin side chains and for the proximal and distal histidine residues). The quantum chemical system was truncated by hydrogen atoms, the positions of which are linearly related to the corresponding carbon atoms in the full system [50].

The total energy is calculated as

$$E_{tot} = E_{QC} + E_{MM123} - E_{MM1} \quad (1)$$

Here, E_{QC} is the quantum chemical energy of the quantum system with hydrogen junction atoms, including all the electrostatic interactions. Similarly, E_{MM1} is the classical energy of the quantum system, still with hydrogen junction atoms, but without any electrostatic interactions. Finally, E_{MM123} is the classical energy of all atoms with carbon junction atoms and no electrostatics. The philosophy behind this energy is that the total energy should involve as much quantum chemistry as possible and that terms from hydrogen junction atoms shall cancel out. This approach is similar to the one used in the Oniom method, but electrostatics are treated in a more accurate way [53]. The calculated forces are the negative gradient of this energy. Owing to the different junction atoms, the gradients have to be corrected using the chain rule.

Calculations were performed both with the protein fixed at the crystal structure and with protein atoms within 15 Å of any atom in the quantum system ring allowed to relax by a molecular mechanics calculation in each step of the ComQum optimisation. In the latter calculations, all atoms are represented by standard Amber parameters. Electrostatic interactions are included, using standard charges for the surroundings, but for the quantum system, we used quantum chemical Mulliken charges fitted to Merz-Kollman electrostatic potential charges [54] as described before [51].

The convergence criteria were the default ones in Turbomole: a change of 2.6 J/mole and 0.053 pm, respectively between two cycles in the geometry optimisation. For the calculations with a free enzyme, looser convergence criteria was used, 260 J/mole and 0.53 pm, and these calculations were concluded by an optimisation using a fixed enzyme and the tighter convergence criteria.

The strain energy induced by the protein onto the quantum system was estimated as the difference in energy of the isolated quantum system optimised in the protein and in vacuum (E_1). It should be noted that this energy contains terms that are normally not considered as strain [51,55]. Hydrogen and ligand bond energies were calculated quantum chemically both in vacuum and in the protein (i.e. without and with the point charges) and were corrected for basis set superposition errors using the standard counter-poise method [56].

The ComQum calculations were based on three crystal structures of sperm whale myoglobin, differing in the distal ligand of the haem group: CO, O_2 or no ligand [16]. The coordinates were taken from the Brookhaven protein data bank files 1a6g, 1a6m, and 1a6n (resolution 1.15, 1.00, and 1.15 Å, respectively). Minor alternative conformations were removed from the structures. A few crystal water molecules with partial occupancy, which form close contacts with other water molecules or protein atoms, were also removed.

Hydrogen atoms were added to the protein assuming the normal protonation status at pH 7.0 for the Asp, Glu, Lys, and Arg residues. After a detailed study of the surroundings and possible hydrogen bond networks around the His residues, it was decided that His-24, 48, 81, 93, 113, and 116 are protonated on the N atom, whereas His-36, 82, 97, and 119 are protonated on the N atom. Both protonation states of the distal histidine (His-64) were studied. The proteins were solvated in a sphere of water molecules with a radius of 30 Å,

centred on the iron ion (~2660 water molecules). The hydrogen atoms and the solvation water molecules were optimised by a simulated annealing calculation by molecular dynamics, using the Amber software [52].

The molecular mechanics force field for the haem molecule was that included in the Amber distribution [52]. Since the whole haem ring, except the side chains are included in the quantum system, these parameters mostly cancel out and are therefore of minor importance. Charges for the haem group and its ligands in the initial equilibration were obtained from a standard two-stage RESP fit [57] to quantum chemically calculated electrostatic potentials using the Amber software [52]. The potential points were sampled with the Merz-Kollman method [54] at a high density (20 000 points), and they were calculated with the Gaussian software [46] on the ImFePor(O_2/CO) models.

QC/MM calculations with crystallographic raw data

A set of calculations were also performed with the ComQum-X program, which is a very recent combination of crystallographic refinement and quantum chemical calculations [58]. It can be seen as a standard crystallographic refinement, where the molecular mechanics force field, used to supplement the crystallographic data, especially for bond lengths and angles, has been replaced by a quantum chemical potential for a small part of the structure (thereby improving the accuracy). Alternatively and equivalently, it can be seen as a QC/MM program where the energy term E_{MM123} in Eqn. (1) has been replaced by a standard crystallographic refinement energy function, including terms from the diffraction data. Thus, this approach yields a structure that automatically is consistent with the crystallographic raw data (the structure factors) and not only with the protein coordinates at the beginning of the geometry optimisation. Therefore, the resulting structure is an optimum compromise between crystallography and quantum chemistry, which allows us to interpret and sometimes even improve the crystal structure locally. The results can directly be evaluated in standard crystallographic terms, such as electron density maps and crystallographic *R* factors.

ComQum-X is a combination of Turbomole [44-45] and the freely available crystallographic refinement program CNS (Crystallography and NMR system) [59]. It differs from ComQum-00 only in a few details, besides the change in the molecular mechanics

program from Amber to CNS and the inclusion of a crystallographic energy function (we used the default maximum likelihood amplitude refinement target [60-61]). Following crystallographic custom, hydrogen atoms were not included in the CNS calculations and therefore also no electrostatic interactions (of course, hydrogen atoms are still included in the quantum chemical calculations, but there was no point-charge model of the surroundings). Thereby, we avoid the problem of determining the protonation status of the histidine residues and settling the positions of polar hydrogen atoms, e.g. for water and hydroxyl groups.

The crystallographic penalty function is in arbitrary units and is therefore not directly comparable to the quantum chemical and molecular mechanics energy functions. Therefore, the crystallographic energy function is weighted by a factor, determined to give the same size of the crystallographic and energy-based forces [62-64]. This factor turned out to be 0.010 for the quantum system in these calculations. For the surrounding protein, it was calculated automatically by CNS. Moreover, the CNS force field is obtained from a statistical analysis of crystal structures and its energy terms are typically a factor of 3 larger than terms obtained from quantum chemical calculations [65]. Therefore, the quantum chemical forces and energies were scaled up by a factor of 3 before they were added to the CNS energies and forces.

The ComQum-X calculations were based on the crystal structure of myoglobin in complex with CO determined at 1.15 Å resolution by Bartunik and coworkers (Brookhaven data bank file 1bzr) [7]. We used this structure, instead of that used in the ComQum-00 calculations because the haem:CO unit is not disordered in this structure (the current version of ComQum-X cannot treat alternative configurations in the quantum system) and also because we wanted to study steric distortions suggested to be present in the structure. The structure factors were downloaded from the Brookhaven protein data bank, file r1bzrsf. They were converted to CNS format and a test set consisting of 5 % of the reflections was selected at random (so that a free *R* factor could be calculated).

Alternative configurations of 18 residues were removed (the configuration with the largest occupancy was kept). These residues are on the surface of the protein, quite far from the CO molecule of interest. The quantum system was the same as in the QC/MM calculations: the haem group without side chains, the imidazole rings of the proximal and distal histidine

residues, and the CO molecule. It was constructed by adding hydrogen atoms to the imidazole rings and the meso carbon atoms of the porphyrin ring. The molecular mechanics parameters for the porphyrin ring were the same as in the ComQum-00 calculations, but the force constants were scaled up by a factor of 3.

In the ComQum-X calculations, the protein (all atoms, including crystal water molecules) is always free to relax by one step of crystallographic minimisation refinement and one step of individual *B* factor refinement in each cycle of the geometry optimisation. However, the new coordinates or *B* factors are only accepted if the free *R* factor decreases after the change. The convergence criteria were the default ones in Turbomole (see above), but with the combined energy function.

The B3LYP method is known to overestimate Fe-ligand distances [66-67]. We tried to correct for these systematic errors using the method of offset forces [68]. First, we optimised the geometry of a small model complexes, which contain the relevant ligands and for which accurate crystal data exist: Fe^{II}(octaethylporphyrin)(1-MeIm)CO [69]. It was modelled by ImPorFe^{II}CO. After the optimisation, the Fe-N_{Im}, Fe-N_{Por}, Fe-C, and C-O bond lengths were changed to the experimental ones and the resulting forces were calculated. These forces (-0.0044, 0.0037, 0.112, and 0.0608 a.u., respectively) were then added to all subsequent calculations with ComQum-X. For the model complexes, this approach gave distances within 0.2 pm of the experimental ones (the distances changed by 1-3 pm).

Results and Discussion

QC/MM structures

Three types of complexes were optimised with ComQum-00, differing in the distal ligand: O_2, CO, or none. For the two latter complexes, we tested both possible protonation states of the distal histidine, HIE (where N is protonated and the H atom forms a hydrogen bond with CO or points towards the iron ion) or HID (where N is protonated and H is directed into the solution, whereas the lone-pair orbital of N is directed towards CO or iron). The latter choice is what is observed in the neutron diffraction structure of CO-myoglobin (although this structure has been questioned [5]), whereas the HIE conformer is observed in the diffraction structure of O_2-myoglobin and was therefore the only conformer studied for

that complex. For each model, we performed one set of calculations based on the native crystal structure (i.e. the one with the same distal ligand, CO, O_2, or none) and one based on the crystal structure of the CO complex. Finally, for all systems, we performed one calculation with a fixed protein and one where the protein was allowed to relax. Thus, there are in total fourteen different structures. In addition, we also optimised two structures of the CO and O_2 complexes without the distal histidine in the quantum system (one where the protein was fixed at the crystal structure and one where it was allowed to relax). These structures were based on the native crystal structure and with the distal histidine residue protonated on the N atom.

The optimum geometry of the five-coordinate haem group with the distal histidine protonated on the N atom, based on the crystal structure of the same complex, is shown in Figure 1. It can be seen that the haem ring is distinctly non-planar (domed). Moreover, there are clear steric interactions between the H atom of the distal histidine residue and haem, leading to further distortions of the haem ring. The geometry around the iron ion is what could be expected for a five-coordinate high-spin haem complex, with long Fe-N_{His} and Fe-N_{Por} distances (210-212 pm; cf. Table 1). Moreover, the iron ion is displaced out of the haem plane towards the proximal histidine ligand, but the displacement, 64 pm, is appreciably lager than in vacuum, 31 pm. The reason for this is probably unfavourable steric interactions between the H atom and the iron ion (346 pm distance).

The same complex optimised in the crystal structure of CO-myoglobin is quite similar (Table 1). The haem ring is slightly less distorted and the iron ion is closer to the haem plane, 25 pm, indicating less severe steric interactions between the distal histidine and the haem group.

If the protein is allowed to relax, the structure changes somewhat. As can be seen in Figure 2, the haem group becomes more distorted. Moreover, the Fe-N distances become 2-4 pm longer and the out-of-plane distance of the iron ion increases slightly. These changes probably reflect that B3LYP overestimates Fe-N distances by 2-6 pm [66-67]. Therefore, we expect that the structure with the enzyme fixed at the crystal position gives the most accurate results, provided that the native structure has been used.

Interestingly, if the distal histidine is instead protonated on the N atom, the haem group becomes more planar, the iron ion moves closer to the haem plane (27 pm below the plane) and the Fe-N_{His} bond becomes longer, as can be seen in Figure 3. Again this probably reflects decreased steric interactions between the distal histidine and the haem group.

When CO is bound to myoglobin with the distal histidine residue protonated on the N atom, the haem group becomes more planar, although it still shows a wavelike distortion (Figure 4), probably owing to the interaction between the distal histidine residue and the ring. CO binds to Fe at a distance of 177 pm and the iron ion moves into the haem plane. The Fe-N_{Por} distances are shortened by almost 10 pm, reflecting the transition of the iron ion from high to low spin. However, the Fe-N_{His} distance is not significantly altered (209-210 pm). The Fe-C-O angle is 173°. There is a clear, but weak, hydrogen bond between CO and the distal histidine group; the H -O distance is 252 pm. However, the histidine residue is not positioned for ideal binding. Instead, it is close to the middle of the CO molecule, and the H -C distance is only slightly longer, 264 pm.

This COMQUM structure is very similar to the vacuum structure, obtained with the same quantum chemical method: All Fe-ligand distances are within 1 pm of the distances found in vacuum, but the distance of the iron ion from the porphyrin plane is somewhat shorter (4 pm). All geometric parameters are also within the range found in small inorganic models and in protein structures [6-7,16,32,70-72]. The C-O bond length (116 pm) is somewhat longer than for free CO (114 pm with the same quantum chemical method and 113 pm in experimentals [73]). The N-O hydrogen-bond interaction (312 pm) is close to what was obtained in vacuum (317 pm), but slightly shorter than in crystal structures (316-394 pm). However, the H-O distance (252 pm) is appreciably longer than in vacuum (216 pm), owing to the far-from-ideal hydrogen-bond geometry.

Interestingly, the structure does not change very much if the distal histidine residue is protonated on the N atom instead **(zzz this structure is not converged yet)**. Figure 5 shows that the geometry of the haem group is virtually unchanged. However, the distal histidine rotates slightly, apparently to elongate the interaction between this group and CO; the N -O and N -C distances increase by ~15 pm to 328 and 358 pm. Similarly, the Fe-N_{His}, Fe-N_{Por},

and Fe-C bond lengths increase by 9, 2, and 5 pm, whereas the iron ion remains close to the porphyrin plane. The Fe-C-O angle is similar to that in the other complex, 173°.

The structure of the O_2 complex, Figure 6, is quite similar to that of the CO complex, with the exception of the Fe-O-O angle, which is 120°. The Fe-O distance (187 pm) is 10 pm longer than the Fe-C distance, and therefore, the other Fe-N distances are ~1 pm shorter in the O_2 complex. In the latter complex, the iron ion is also slightly displaced (3 pm) towards the histidine ligand. The H -O hydrogen bond is appreciably stronger in the O_2 complex; it is 185-195 pm long. However, even in this complex it is far from ideal (cf. Figure 6); the distal histidine is too close to the oxygen molecule and therefore it has to interact with the terminal oxygen not from the front as in the ideal vacuum structure but from behind. In fact, the H atom is also close to the oxygen atom bound to iron, 234 pm. Yet, the H-O distance is not very different from what is obtained in vacuum (196 pm).

The COMQUM structure of the O_2 complex is quite similar to the vacuum structure, calculated with the same method. The most significant difference, besides the different hydrogen-bond geometry, is the longer Fe-N_{His} bond (210 pm compared to 205 pm) in the protein. The COMQUM structure is also close to available crystal structures of oxygenated myoglobin. The calculated Fe-O distance is slightly longer than in the crystals (187 pm compared to 180-183 pm), but it is 175-190 pm in small model compounds [32,70-72], and the O-O bond length is too long (130 pm compared to 121-124 pm in protein crystals and model compounds [6,16]). This lengthening in the theoretical structures reflects that the oxygen complex is mainly a superoxide bound to Fe(III); the ideal bond length of O_2 optimised with the same method is 122 pm, i.e. the same as the experimental bond length of free O_2 [73]. The optimum bond length of O_2^- calculated with the same method is 135 pm, whereas the experimental one is 133-134 pm [32,74-75]. Perhaps the density functional calculations overestimate this internal electron transfer somewhat.

If the O_2 complex is instead optimised in the structure of CO-myoglobin, the resulting geometry becomes quite different (Figure 7). **(zzz This structure is not finished yet)** In particular, it can be seen that if the protein is fixed at the crystal position, the distal histidine group comes so far from the O_2 molecule that it is more favourable for it to form a hydrogen bond to the oxygen atom bound to the iron ion than to the distal oxygen atom (although the

latter atom has a more negative charge). In fact, the H -O hydrogen bond distance, 236 pm, is almost the same as the H -$O_{proximal}$ distance in the complex optimised in the native structure (234 pm). It is only the distance to the distal oxygen atom that has changed. This leads to an elongation of all the iron-ligand distances and also of the hydrogen-bond length (cf. Table 1). Thus, there is a significant movement of the distal histidine residue between the O_2 and CO complexes. The hydrogen-bond of this structure is of a similar strength as that optimised in the native crystal structure (c.f. Table 2), indicating that these interactions are almost equivalent. However, the strain energy of this complex is 50 kJ/mole larger than for the other complexes, indicating that this structure is unfavourable for the active site as a whole.

Energies

We have estimated strain energies (E_1) in all the optimised complexes by calculating the difference in energy of the isolated quantum system (i.e. without point charges) at the optimum vacuum geometry and at the geometry optimised by ComQum. For the O_2 and CO complexes, the same complex in vacuum was used as the reference structure. However, for the five-coordinate complex, we instead used the five-coordinate complex without the distal histidine model in vacuum plus the energy of an isolated and optimised imidazole molecule. In vacuum, the most favourable structure of this complex would have both imidazole groups coordinating to the iron ion, leading to a low-spin state (such a complex has a 46 kJ/mole lower energy than the reference we used). However, the change in spin state may significantly affect the accuracy of such a comparison (reduce the cancellation of errors).

With these reference states, the strain energy for all the studied complexes was 37-67 kJ/mole (cf. E_1 in Table 2; two complexes had larger strain energies, but they are discussed separately). It was of a similar magnitude for all three types of complexes, but it was slightly (2-9 kJ/mole) lower when there was no model of the distal histidine in the quantum system. This reduction is a natural consequence of the fact that the quantum system is smaller. Interestingly, the strain energy was often larger for the calculations with free protein than for those with a fixed protein. The reverse is normally observed.

The strain energies are similar to what is found for other proteins studied with the same methods (15-110 kJ/mole for alcohol dehydrogenase, blue copper proteins, iron-sulphur clusters, and ferrochelatase [50-51,76-80]). The calculated strain energies are much larger than strain energies estimated for the binding of CO to myoglobin (less than 10 kJ/mole) [5,31]. The reason for this is that we do not calculate strain energies only for CO molecule, but the whole haem group and the proximal and distal histidine residues also. Moreover, E_1 is strictly speaking not a true strain energy, because it contains several terms that are normally not considered as strain, e.g. electrostatic effects [50,55].

Second, we have calculated hydrogen-bond energies from the COMQUM structure of the quantum system by a standard counter-poise calculation, either in vacuum or by modelling the protein as an array of point charges. The energy of the hydrogen bond to CO was 2-3 kJ/mole. This is smaller than the hydrogen bond energy obtained in vacuum (14 kJ/mole for the optimum geometry and 8 kJ/mole when the energy was corrected for the geometrical constraints in the protein) [6]. The COMQUM values are in good agreement with experiments; for example, Olsen and Phillips have estimated this energy to about 1 kJ/mole [27], whereas Ray et al. gave a slightly larger estimate: 2.5-2.8 kJ/mole [70]. If the electric field from the protein is included in the calculation, the hydrogen bond becomes ~3 kJ/mole stronger (5-7 kJ/mole), but the hydrogen-bond energy is reduced by the same amount when it is calculated with a larger basis set.

The hydrogen bond to O_2 is considerably stronger, 28-30 kJ/mole. When the protein is included in the calculations by point charges, the hydrogen-bond energy again increase by 3 kJ/mole. This shows that electrostatic interactions in the protein matrix polarises and stabilises the hydrogen bonds, but only by 3 kJ/mole and with the same amount for both ligands. The vacuum binding energy was earlier calculated to 35 kJ/mole, which decreased to 32 kJ/mole when corrected for the constraints in the protein [6]. All these hydrogen-bond energies are appreciably larger than experimental estimates. For example, the stabilising effects of the hydrogen bond between the distal histidine residue and O_2 in myoglobin has been estimated by mutation studies to 8-18 kJ/mole [27,34].

This discrepancy is probably caused by entropy effects (we calculate enthalpies, whereas free energies are measured). It is very hard to estimate these, because we then have to

consider interactions between the distal histidine residue and the rest of the protein both before and after binding of the haem ligand. A direct calculation of the thermodynamic corrections from a vibrational analysis of the isolated complexes gives a high disfavouring energy (45 kJ/mole), because such a calculation assumes that the imidazole group is free to move before the hydrogen bond is formed, which is not the case in the protein. If only the vibrational contributions are considered, the energy becomes slightly favourable, ~10 kJ/mole, but this overestimates the binding, because the product complex contains six vibrations more than the reactants. Thus, the absolute contribution of the thermodynamic corrections is quite uncertain. However, the difference between the O_2 and CO is reasonably constant and tend to favour the CO complex by 5 kJ/mole. **Using a larger basis set, the energies change by xxx. Density functional calculations are normally considered to be converged with a basis set of triple- with double polarisation functions [48].**

With this correction, the difference in the hydrogen bond strength to O_2 and CO is 21-22 kJ/mole, both with and without the point-charge model of the protein. This difference is slightly larger than the experimentally determined discrimination of myoglobin between O_2 and CO, 17 kJ/mole. Thus, these calculation confirm our earlier results [6] that this difference in hydrogen-bond strength probably accounts for the majority of myoglobin's discrimination between O_2 and CO.

Very recently, Rovira and coworkers have calculated the hydrogen-bond strength between CO/O_2 and the distal histidine with similar methods [81]. They optimise the complexes with QC/MM methods, but with the distal histidine in the molecular mechanics system and keeping the protein fixed at the neutron diffraction structure of CO-myoglobin [13] or at snapshots taken from molecular dynamics simulations of the same complex [82]. They obtain hydrogen-bond energies of 14 kJ/mole for CO and 21 kJ/mole for O_2. These results are of the same magnitude as ours, but significantly different, probably because of the differences in the optimisation scheme (our approach should be more accurate), and that they use a less accurate density functional [40] and smaller basis sets.

Another way to compare the binding of O_2 or CO to myoglobin, is to calculate directly the binding energy for the ligand. This was done the same way as for the hydrogen-bond energies, i.e. by a counter-poise calculation of the quantum system, either in with or without

the point charges. The resulting binding energies are also shown in Table 2. If the distal histidine is not included in the quantum system, the binding energy is 32-43 kJ/mole lower for O_2 than for CO without a model of the protein. This is similar to the result obtained on the optimum vacuum structures, 43 kJ/mole. However, if the point-charge model of the protein is included, the difference is much smaller. This clearly shows that the protein differentially stabilises the O_2 binding by electrostatic interactions.

The effect of the distal histidine residue is best seen in the calculations with it in the quantum system. For O_2, inclusion of it leads to a 23-kJ/mole increase in the binding energy without the point charges, but hardly any change with them. This indicates that the distal histidine residue stands for almost all the electrostatic stabilisation. For CO, inclusion of the distal histidine in the quantum system leads to a slight (3-9 kJ/mole) decrease in the binding affinity, both with and without the point-charge model. This indicates that the interaction between CO and the histidine residue is poorly described by the molecular mechanics force field (i.e. the binding energies with the distal histidine in the quantum system are more reliable). In the latter calculations, CO binds slightly stronger than O_2 in all except one case. The difference is -3 to 11 kJ/mole, which is close to the experimental difference in binding affinity, 25 times or 8 kJ/mole.

These estimates are sensitive to the theoretical treatment. First, if the energies are calculated with a larger basis set, the binding energies are reduced by 12 kJ/mole. Second, relaxation effects are important, because the change in spin states leads to large changes in the geometry of both the five-coordinate complex and of O_2. These effects are large (24-43 kJ/mole) and decrease the binding energies [6]. Finally, thermodynamic corrections are significant for these binding energies and also tend to decrease them (because a free O_2/CO molecule is bound during the reaction). These corrections (including zero-point energies, the effect of which is small), estimated from from the vibrational frequencies of the isolated complexes, amounts to a decrease of the binding affinity by 50 kJ/mole. In conclusion, the absolute values of these binding energies are not very reliable (compared to experimental values), but the differences can probably be trusted.

Protonation of the distal histidine in CO-myoglobin

As mentioned above, a neutron diffraction structure of CO-myoglobin [13] showed that the distal histidine residue is protonated on the N atom, i.e. the atom directed towards the surrounding solvent, whereas the N atom interacts with the CO molecule with its lone-pair orbital. However, other results have strongly indicated that that the opposite protonation (i.e. with a hydrogen on N , forming a hydrogen bond to CO) is the dominant form for CO-myoglobin in solution. For example, resonance Raman experiments have indicated that there is a hydrogen bond to CO in its complex with myoglobin [83]. Moreover, virtually all theoretical investigations of the vibrational spectrum of CO bound to myoglobin have concluded that the vibrational substates can only be explained by hydrogen-bond interactions between the protein and CO; an interaction between the lone-pair orbital of N and CO would give rise to shifts in the vibrational spectrum opposite to what is observed experimentally [6,5,81,84-88]. Likewise, molecular dynamics simulations with N of His-64 protonated (as in the neutron structure) indicate that the imidazole ring rapidly rotates and exposes the polar hydrogen towards CO [82,89].

There are also ample experimental evidence indicating that the protonation status of the distal histidine residue observed in the neutron structure is not typical for myoglobin in solution. For example, in the His64Gln myoglobin mutant, the N atom of Gln is positioned at almost exactly the same place as N in His-64, although the former atom is undoubtedly protonated, whereas the latter according to the neutron structure is not [13,90-91]. The two proteins have also similar vibrational spectra. Likewise, investigation of the CO frequencies and CO affinity for myoglobin mutants show a clear correlation with the presence of distal hydrogen-bond donors [15].

Our structures support this suggestion: If we compare the crystal structure of CO-myoglobin (without any hydrogen atoms) with the two ComQum-00 calculations with different protonation status of the distal histidine residue, Figure 8, **(zzz update)** it is obvious that the structure with the N atom protonated is less similar to the crystal structure than that with the N atom protonated. In particular, in the former structure, the imidazole ring has rotated about 30° compared to the crystal structure, whereas it remains close to the crystal location in the latter structure. Moreover, the Fe-N_{His} bond is elongated in the HID structure,

which leads to a significant movement of the proximal histidine ligand compared to the crystal structure. The geometry of the haem group and the CO ligand is similar in all three structures. Altogether, this shows that a HIE protonation is consistent with the crystal structure, but not a HID protonation. In other words, it is most likely that the N atom is protonated in the crystal structure.

Moreover, the strain energy (E_1) of the CO complex where the N atom is protonated and the lone-pair orbitals of the N atom is directed towards CO is appreciably higher (94-96 kJ/mole) than that obtained with the opposite protonation (40-45 kJ/mole, cf. Table 2), indicating that this structure is unnatural. Likewise, the interaction between CO and the distal histidine residue in the former structure is repulsive (by 14 kJ/mole). This further strengthens our suggestion that this is not the correct protonation status of the distal histidine residue.

This conclusion is in variance with the neutron structure of CO-myoglobin [13]. However, there are at least two possible reasons for this discrepancy. First, it is conceivable that the distal histidine residue interacted with a sulphate molecule in the neutron structure [85-86]. Such a strong electrostatic interaction could stabilise an otherwise unfavourable protonation of the N atom, especially as the hydrogen bond to CO is rather week [6]. Alternatively, it has been suggested that the neutron structure has been oxidised to met-myoglobin (i.e. Fe^{III} with a coordinated water molecule) during the course of data collection [5,87]. In such a structure, the HID form could be energetically favourable.

The ComQum-X structure of CO myoglobin

The ComQum-00 structures we have studied up to now were obtained by starting at the crystal structure and reoptimising the quantum system (i.e. the haem group, the distal ligand, and the imidazole rings of the proximal and distal histidine ligands) and optionally also the surrounding protein. In such a procedure, there is no guarantee that the final structure remains close to the crystal structure also after the geometry optimisation. In fact, we saw in Figure 8 that this is not always the case. Moreover, it is likely that systematic errors in the quantum chemical calculations may slightly distort the final structure, as our calculations with a relaxed protein indicate (cf. Figure 2).

Since we in many of the previous calculations only wanted to reproduce the crystal structure (to be able to calculate energies), it would be more attractive to directly use the crystal data. However, small errors in the coordinates, primarily in the bond lengths (combined with the systematic error in the calculations), lead to gigantic errors in energies based directly on the crystal coordinates - errors which would totally conceal the interesting trends.

Moreover, the crystal coordinates are not the experimental raw data; they are a product of an involved series of model building, crystallographic refinement, manual examination, and rebuilding [92], possibly involving mistakes in the interpretation. During the refinement, the position of the atoms (and other crystallographic parameters) are optimised with respect to an energy function consisting of an estimate how well the parameters reproduce the crystallographic reflections, supplemented by a molecular mechanics force field. Therefore, the coordinates may be biased by the force field used in the refinement. A better approach would be to use the crystallographic raw data, i.e. the structure factors, instead of the coordinates (these are normally also available from the Brookhaven protein data bank).

We have very recently modified ComQum-00 to use a crystallographic refinement program instead of the Amber for the molecular mechanics calculations [58]. Thereby, we can replace the molecular mechanics force field used in the crystallographic refinement with quantum chemical calculations for the active site, using exactly the method we want to use in the energy calculations. Thereby, we obtain an optimum compromise between quantum chemistry and crystallography and guarantee that the final structure will be fully compatible with the crystallographic reflections. In fact, we can even locally improve the crystal structure beyond what the raw data allows, because the quantum chemical calculations typically have a smaller error in the geometry around a metal site (2-6 pm [66-67]) than a protein crystal structure at low or medium resolution (~10 pm [93]) and that the quantum chemical error is systematic and therefore can be compensated for.

We have studied the structure of CO-myoglobin with this ComQum-X method. The original calculations showed a clear tendency to overestimate the C-O and the iron-ligand distances by a few pm (the first line, "uncorrected" in Table 3). This is a well-known deficiency of the B3LYP method [66-67] and for a crystal structure of such a high resolution,

these systematic errors become significant. Therefore, we also run a set of calculations where we corrected for these errors using the method of offset forces [68].

The resulting structure of CO-myoglobin is compared to the original structure and the electron density in Figure 9. It can be seen that the structure is very similar to the crystal structure and also to the corresponding ComQum structures (Figure 4), even if they are based on different crystal structure. It should also be noted that in the original crystal structure, it is suggested that the distal histidine residue i protonated on the N ^{1} atom [7], whereas the other protonation state has been used in the calculations (based on the results in the previous section). The similarity between the ComQum-X and crystal structures, shows that this was a correct choice and further strengthens the argument in the previous section.

In Table 3, the geometry around the iron ion and the CO ligand is detailed and compared with the two most accurate available crystal structures for the same complex (both at 1.15 Å resolution) [7,16]. It can be seen that the ComQum-X result is close to those of the two crystal structure but closest to the structure, the raw data of which was used in the calculations. The ComQum-X structure, is probably the most accurate estimate of the geometry of the CO-haem complex in myoglobin available. In particular, we predict that the Fe-C distance is around 171 pm, i.e. close to the value in the original crystal structure (173 pm), and appreciably shorter than in the other crystal (182 pm). The improvement of the crystal structure by the quantum chemical calculations is flagged by a decrease in the crystallographic free *R* factor of 0.0003, which is a typical decrease for this global quality factor [58], showing that the structure is locally slightly improved.

As discussed in the introduction, the actual structure of CO-haem moiety has been much discussed. Early crystal structures indicated that the Fe-C-O angle was 120-140° [9-13]. However, later crystal structures show larger Fe-C-O angles (160-175°) [7,15-18] and vibrational spectroscopy investigations also indicate that the Fe-C-O angle is nearly linear [19-25]. The reason for this discrepancy in the crystal structure is probably that the earlier structures have been oxidised to met-myoglobin during the data collection [5,87]. ComQum-X predicts a Fe-C-O angle of 174°, slightly larger than both in the two crystal structures (171°), and in the ComQum-00 calculations (173°), but smaller than in vacuum calculations (179-180°).

The hydrogen-bond geometry, as described by the N -O and N -C distances is very similar in the original crystal structure and in the ComQum-X structures (within 1 pm). In the other crystal structure in Table 3, the distal histidine residue exhibits two alternative configurations, one with a short N-O interaction, suggested to represent the HID form of the histidine residue, and one with a longer N-O distance, suggested to describe the HIE form. Quite satisfactorily, the ComQum-X structure is close to the latter form, even if the discrepancy is slightly larger than for the other crystal structure (up to 11 pm).

The strain energy (E_1) of the ComQum-X structures is 39 kJ/mole. This is similar (1 kJ/mole lower) to what was found for the ComQum-00 structure. Likewise, the hydrogen-bond energy (2.8 kJ/mole) is not different from that obtained with ComQum-00. Thus, there is no significant difference between the ComQum-00 and ComQum-X structures, although they are based on different methods and crystal structures.

Concluding remarks

In this article we have studied the structure of deoxy, O_2, and CO-myoglobin. Starting from the crystal structures, we have used combined quantum chemical and molecular mechanics or crystallographic refinement calculations to determine the optimum structure of the haem group and its ligands in the protein. The optimised structures stay close to the crystal structures, and the ComQum-X structure provides probably the most accurate picture of the CO complex available.

We have investigated different protonation states of the distal histidine residue and the results clearly show that in the crystal structure, this residue is protonated on the N atom, which forms a weak hydrogen bond to CO. This is in accordance with ample experimental and theoretical evidence [6,5,15,83-87,89-91], but not with the neutron structure of CO-myoglobin [13].

The structures have been used to calculate hydrogen-bond energies between the distal histidine residue and O_2 and CO. These calculations indicate that a hydrogen bond to O_2 is 21-22 kJ/mole stronger than one to CO. These results are in good agreement with our previous vacuum calculations, 21 kJ/mole for the pure vacuum structures, and 24 kJ/mole if the geometries were forced to be similar to the protein structure by constraining two or three

dihedral angles [6]. Thus, reasonable energies can be obtained already in vacuum with appropriate constraints.

The calculated difference in the hydrogen-bond strength is slightly larger than myoglobin's discrimination between O_2 and CO, 17 kJ/mole. This strongly support the suggestion that electrostatic interactions is the main source to this discrimination, rather than steric interactions. This conclusion is in accordance with several experimental and theoretical investigations [2,6,21,27-32,34,36].

However, in a recent crystallographic investigation, Bartunik and coworkers suggest that steric strain is the main source for myoglobin's discrimination between O_2 and CO [7]. More precisely, they observe a concerted movement of the E and F helices of myoglobin when comparing the structures of deoxy and CO-myoglobin. This motion leads to altered positions of the haem group and the iron ion and gives more room for the CO ligand, which nevertheless makes repulsive interactions with the distal histidine residue and the nearby Val-68 (E11) residue. It must be noted, however, that these geometric observations give no information about the corresponding energies and therefore provide very week support for a steric mechanism (it may well be that the two configurations of the protein have almost the same energy). Moreover, the observation must be complemented by a study of the O_2 structure, before anything can be said about the discrimination between O_2 and CO.

Our calculations give no support to the steric mechanism. We observe a significant difference in the position of the distal histidine residue between the O_2 and CO structures: If O_2 is optimised in the CO structure, the distal histidine prefers to form a hydrogen bond to the proximal, rather than the distal, oxygen atom. This gives rise to a 52-kJ/mole increase in the strain energy, but the hydrogen-bond energy is of the same size as in the native structure, and if the protein is allowed to relax, the strain energy is reduced by 15 kJ/mole. The structure does not converge back to the native structure, however, which shows that conformation of the protein observed in CO-myoglobin is a stable local minimum also in O_2-myoglobin and that the difference between the two structures is rather large.

We have no reliable estimate of the difference in strain energy *in the protein* between the two structures. However, for the active site (the haem group, the haem ligand, and the two histidine residues) we have estimated the strain energies in the two complexes. Interestingly,

the strain energy (E_1 in Table 3) is 2-4 kJ/mole *lower* for the CO complex than for the O_2 complex. This convincingly shows that the haem-CO complex is not more strained than the haem-O_2 complex in the protein, contrary to Bartunik's suggestion. This conclusion remains true even in the COMQUM-X calculations, which are based on Bartunik's crystal structure [7], showing that there is no significant difference between the two crystal structures.

It might be argued that there could be an even larger strain energy in the surrounding protein, which is greater for CO than for O_2, but there are at least three lines of evidence against such a suggestion. First, mutations of the Val-68 residue to a smaller alanine residue selectively increases the CO binding constant, but only by a factor of 2 [2]. This shows that Val-68 contributes to the discrimination between O_2 and CO by less than 2 kJ/mole [5]. Second, free energy perturbation calculations have shown that Van der Waals interactions contribute by only ~2 kJ/mole to the discrimination between O_2 and CO [95]. Third, a larger strain energy in the protein would give rise to larger strain energies in the active site of CO-myoglobin than in O_2-myoglobin, contrary to our estimates, because strain arises as an effect of opposing forces with equal magnitude (and there is no reason why one of the systems should be significantly more stiffer than the other) as we have discussed before [55].

Therefore, it seems safe to conclude that electrostatics is the major source of myoglobin's discrimination between O_2 and CO. Apparently, it is easier for a protein to exploit the polarity difference between FeO_2 and FeCO than their differing geometries, probably because a protein is not rigid enough to induce small geometric changes onto a bound molecule (the protein, rather than the molecule, will distort if their geometric preferences differ). Such a conclusion is in accordance with our and others previous findings on various proteins, such as blue copper proteins, vitamin B_{12}-dependent enzymes, and lysozyme [55,96-99].

Acknowledgements

This investigation has 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 High Performance Computing Center North (HPC2N) at the University of Umeå, and Lunarc at Lund university.

References

1. S.J. Lippard & J.M. Berg, Principles of bioinorganic chemistry, (1994) University Science Books, Mill Valley.
2. Springer B.A., Sligar S.G., Olson J.S., Phillips G.N Jr (1994) Chem Rev 94:699-714
3. Collman JP, Brauman JI, Halbert TR, Suslick KS (1976) Proc Natl Acad Sci USA, 73:3333-3337
4. Ghosh A (1997) J Biol Inorg Chem 2:515
5. Spiro TG and Kozlowski PM, Acc. Chem. Res. 34 (2001)137-144
6. Sigfridsson E. and Ryde U., J. Biol. Inorg. Chem., 4 (1999) 1, 99-110
7. G.S. Kachalova, A.N. Popov, H.D. Bartunik, Science 284 (1999) 473-476.
8. Collman, J. P. (1997) Inorg. Chem, 36, 5145-5155
9. Norvell JC, Nunes AC, Schoenborn BP (1975) Science 190:568-570
10. Hanson JC, Schoenborn BP (1981) J Mol Biol 153:117-146
11. Kuriyan J, Wilz s, Karplus M, Petsko GA (1986) J Mol Biol 192:133-154
12. Cheng X, Schoenborn B (1990) Acta Crystallogr B46:195-208
13. Cheng X, Schoenborn B (1991) J Mol Biol 220:381-399
14. Stryer L, In: Biochemistry (1988) 3rd edn. Freeman, NY, p 149
15. Li T, Quilin ML, Phillips GN, Olson JS (1994) Biochemistry 33:1433-1446.
16. J. Vojtechovský, K. Chu, J. Berendzen, R. M. Sweet and I Schlichting, Biopys. J. 1999, 77, 2153-2174
17. Quillin ML, Arduini RM, Olson JS, Phillips GN Jr (1993) J Mol Biol 234:140-155
18. Yang F, Phillips GN Jr (1996) J Mol Biol, 256:762-774.
19. Hu S, Vogel KM, Spiro TG (1994) J Am Chem Soc 116:11187-11188
20. Ray GB, Li X-Y, Ibers JA, Sessler JL, Spiro TG (1994) J Am Chem Soc 116:162-176
21. Springer B.A., Sligar S.G., Olson J.S., Phillips G.N Jr (1994) Chem Rev 94:699-714
22. Lim M, Jackson TA, Anfinrud PA (1995) Science 269:962-966
23. Lim M, Jackson TA, Anfinrud PA (1997) J Biol Inorg Chem 2:531-536
24. D. Ivanov, J.T. Sage, M. Keim, J.R. Powell, S.A. Asher, P.M. Champion, J. Am. Chem. Soc. 116 (1994) 4139.

25. M.T. McMahon, A.C. deDios, N. Godbout, R. Salzmann, D.D. Laws, H. Le, R.H. Havlin, and E. Oldfield, J. Am. Chem. Soc. 120 (1998) 4784.
26. Sanders, L. K., Arnold, W. D. & Oldfiedl, E. (2001) J. Porphyrins. Ptalalocyanins (2001) 5, 323-333.
27. Olson J.S., Phillips G.N.Jr., J Biol Inorg Chem 2 (1997) 544
28. Ghosh A, Bocian DF (1996) J Phys Chem 100:6363-6367
29. Spiro TG, Kozlowski PM (1997) J Biol Inorg Chem 2:516-520
30. T. Vangberg, D.F. Bocian, A. Ghosh, J. Biol. Inorg. Chem. 2 (1997) 526-530.
31. P.M. Kozlowski, K.M. Vogel, M.Z. Zgiersky, T.G. Spiro, J. Porph. Phtal. 5 (2001) 312-322.
32. Momenteau M, Reed C A (1994) Chem Rev 94:659-698.
33. Rovira C, Kunc K, Hutter J, Ballone P, Parrinello M (1997) J Phys Chem A 101:8914-8925.
34. Springer BA, Egeberg KD, Sligar SG, Rohlfs RJ, Mathews AJ, Olson JS (1989) J Biol Chem 264:3057-3060
35. Phillips SEV, Schoenborn BP (1981) Nature 292:81-82
36. Olson JS, Phillips GN (1996) J Biol Chem 271:17593-17596
37. Holthausen, M. C. , Mohr, M. & Koch, W. Chem. Phys. Lett. 1995, 240, 245-252.
38. Ricca, A. & Bauschlicher, C. W. J. Phys. Chem. 1994, 98, 12899-12903.
39. Ricca, A. & Bauschlicher, C. W. Theor. Chim. Acta 1995, 92, 123-131.
40. Bauschlicher C. W. Chem. Phys. Lett. 1995, 246, 40.
41. Schäfer, A., C. Huber & R. Ahlrichs, J. Chem. Phys, 100 (1994) 5829.
42. Hehre, W. J., Radom, L., Schleyer, P. v. R. & Pople, J. A. Ab initio molecular orbital theory; Wiley-Interscience; New York; 1986.
43. G. Barea, F. Maseras & A. Lledós (2001) Internat. J. Quantum Chem. 85: 100-108.
44. Hertwig R. H.; Koch W. Chem. Phys. Lett. 1997, 268, 345-351.
45. Treutler D, Ahlrichs R J. Chem. Phys. 1995, 102, 346
46. Frisch M. J., Trucks G. W., Schlegel H. B., Scuseria, G. E., Robb M. A., Cheeseman J. R., Zakrzewski V. G., Montgomery J. A., Stratmann, R. E., Burant, J. C., Dapprich, S., Millam, J. M., Daniels, A. D., Knudin, K. N., Strain, M. C, Farkas, O.,

Tomasi, J. Barone, V., Cossi, M., Cammi, R., Mennucci, B., Pomelli, C., Adamo, C., Clifford, S., Ochterski, J., Petersson G. A., Ayala P. Y., Cui, Q., Morokuma, K., Malick, D. K., Rabuck, A. D., Raghavachari K., Foresman J. B., Cioslowski J., Ortiz J. V., Stefanov, B. B., Liu, G., Liashenko, Al, Piskorz, P, Komaromi, I. Gomperts R., Martin R. L., Fox D. J., Keith T., Al-Laham M. A., Peng C. Y., Nanayakkara A., Gonzalez C., Challacombe M., Gill P. M. W., Johnson B. G., Chen W., Wong M. W., Andres J. L., Head-Gordon M., Replogle E. S., Pople J. A. 1998. Gaussian 98, Revision A.5, Gaussian, Inc, Pittsburgh PA.

47. Rahut, G.; Pulay, P. (1995) J. Phys. Chem. 99, 3093-3100.
48. Jensen, F. , Introduction to computational chemistry, Wiley, Chichester, 1999.
49. Bytheway I, Hall MB (1994) Chem Rev 94:639-658.
50. Ryde, U. (1996) J. Comp.-Aided Mol. Design. 10, 153–164.
51. Ryde U, and Olsson MHM (2001), Intern. J. Quant. Chem., 81, 335-347
52. D.A. Case, D.A. Pearlman, J.W. Caldwell, T.E. Cheatham III, W.S. Ross, C.L. Simmerling, T.A. Darden, K.M. Merz, R.V. Stanton. A.L. Cheng, J.J. Vincent, M. Crowley, D.M. Ferguson, R.J. Radmer, G.L. Seibel, U.C. Singh, P.K. Weiner & P.A. Kollman (1997), Amber 5, University of California, San Francisco.
53. M. Svensson, S. Humbel, R.D.J. Froese, T. Matsubara, S. Sieber, K. Morokuma, J. Phys. Chem. 100 (1996) 19357.
54. B. H. Besler, K. M. Merz and P. A. Kollman, J. Comput. Chem. 11, 431 (1990
55. U. Ryde (2001) "On the role of covalent strain in protein function", in Recent Research Developments in Protein Engineering, Research Signpost, Trivandrum, in press. http://signe.teokem.lu.se/~ulf/strain.pdf
56. van Duijneveldt FB, van Duijneveldt-van de Rijdt JGCM, van Lenthe JH (1994) Chem Rev 94:1873-1885.
57. C. I. Bayly, P. Cieplak, W.D. Cornell, and P.A. Kollman, J. Phys. Chem. 97, 10269 (1993).
58. U. Ryde, L. Olsen, K. Nilsson (2001) "Quantum chemical geometry optimisations in proteins using crystallographic raw data", J. Comput. Chem., submitted.

59. A.T. Brunger, P.D. Adams, G.M. Clore, W.L. Delano, P. Gros, R.W. Grosse-Kunstleve, J.-S. Jiang, J. Kuszewski, M. Nilges, N.S. Pannu, R.J. Read, L.M. Rice, T. Simonson, G.L. Warren, (2000) Crystallography & NMR System (CNS), Version 1.0, Yale University.
60. Pannu, N. S.; Read, R. J. Acta Crystallogr. 1996, A52, 659-668.
61. Adams, P. D.; Pannu, N. S.; Read R. J.; Brünger, A. T. Proc. Natl. Acad. Sci. USA 1997, 94, 5018-5023.
62. Jack, A; Levitt, M. Acta Crystallogr. 1978, A34, 931.
63. Brünger, A. T.; Karplus, M.; Petsko, G. A. Acta Crystallogr. 1989, A45, 50.
64. Brünger, A. T. Methods Enzymol., 1997, 277, 243-269.
65. Engh, R. A.; Huber, R. Acta Cryst. 1991, A47, 392-400.
66. E. Sigfridsson, M.H.M. Olsson & U. Ryde, J. Phys. Chem. B, 105 (2001) 5546-5552.
67. M. H. M. Olsson & U. Ryde (2001), J. Am. Chem. Soc., 123, 7866-7876.
68. Fogarasi, G, Zhou, X, Taylor, P W & Pulay, P (1992) J Am Chem Soc 114:8191-8201
69. Salzmann, R., McMahon, M. T., Godbout, N., Sanders, L. K., Wojdelski, M. & Oldfield, E. (1999), J. Am. Chem. Soc. 121, 3818-3828.
70. Ray GB, Li X-Y, Ibers JA, Sessler JL, Spiro TG (1994) J Am Chem Soc 116:162-176
71. Jameson GB, Ibers JA (1994) in Bioinorganic Chemistry (Bertini I, Gray HB, Lippard SJ, Valentine JS, eds.), University Science, pp 167-252.
72. Peng S-M, Ibers JA (1976) J Am Chem Soc 98:8032-8036
73. Huber KP, Herzberg G (1979) Molecular spectra and Molecular Structure, Vol 4 Constants of diatomic molecules, Van Nostrand Reinhold Company, London.
74. Lippard, Principles of bioinorganic chemistry, p. 287.
75. Collman, Inorg. Chem. 36(97)5145.
76. E. Sigfridsson, M. H. M. Olsson & U. Ryde, Inorg. Chem., 40 (2001) 2509-2519.
77. U. Ryde (1995), Protein Science 4, 1124-1132.
78. U. Ryde (1996), Eur. J. Biophys. 24, 213-221.
79. U. Ryde & L. Hemmingsen (1997) J. Biol. Inorg. Chem. 2, 567-579.

80. E. Sigfridsson, U. Ryde, "The importance of porphyrin distortions for the reaction of ferrochelatase", to be submitted to ChemBioChem.

81. Rovira, C., Schulze, B., Eichinger, M., Evanseck, J. D & Parrinello, M. (2001) Biophys. J. 81, 435-445.

82. Schulze, B. & Evanseck, J. D (1999) J. Am. Chem. Soc. 121, 6444-6454.

83. Unno, M., Christina, J.F., Olson, J.S., Sage, J.T., Champion, P.M., J. Am. Chem. Soc. 120 (1998) 2670-2671.

84. Straub JE, Karplus M (1991) Chem Physics 158:221-248.

85. Kushkuley B, Stavrov SS (1997) Biophys J 72:899-912.

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

87. Philips, G.N., Teodoro, M.L., Li, T., Smith, B., Olson, J.S, J. Phys. Chem. B 103 (1999) 8817-8829.

88. Cui, Q & Karplus, M. (2000), J. Chem.Phys. 112, 1133-1149.

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

90. Quillin ML, Arduini RM, Olson JS, Phillips GN Jr (1993) J Mol Biol 234:140-155.

91. 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.

92. G. J. Kleywegt, T. A. Jones", Structure, 3 (1995) 535.

93. Cruickshank, D.W. Acta Crystallogr. D55 (1999) 583.

95. Lopez, M.A., Kollman, P.A. Prot. Sci. 2 (1993) 1975.

96. Warshel A (1991) In Computer modelling of chemical reactions in enzymes and solutions, pp 155-158, 209-211, J Wiley, Sons, New York

97. Levitt M (1974) In Peptides, polypeptides and proteins, (Blout, E R, Bovey, F A, Goodman, M, Lotan, N eds), p 99-102, Wiley, New York

98. Fersht A (1985) In Enzyme Structure and Mechanisms, pp 341-342, W H Freeman, Co, New York

99. U. Ryde, M. H. M. Olsson, B. O. Roos, K. Pierloot & J. O. A. De Kerpel (2000) J. Biol. Inorg. Chem., 5, 565-574.

100. K. P. Jensen, U. Ryde, J. Biol. Inorg. Chem. (2001), submitted.

Table 1. Important bond lengths (pm) and angles (degreees) in the COMQUM-00 calculations.

Ligand	Crystal structure	His64	Protein relaxed	Fe-N_{His}	Fe-N_{Por}	Fe-Por	Fe-O/C	O/C-O	N -O[a]	H -O	Fe-C/O-O
-	-	HIE	No	211.6	210.5	-63.8	-	-	-	-	-
			Yes	210.9	213.4	-50.4	-	-	-	-	-
		HID	No	217.7	209.3	-26.7	-	-	-	-	-
			Yes	220.9	209.9	-33.6	-	-	-	-	-
	CO	HIE	No	211.6	208.0	-24.8	-	-	-	-	-
			Yes	216.0	210.2	-32.1	-	-	-	-	-
		vacuum		220.5	209.4	-30.5	-	-	-	-	-
CO	CO	HIE	No	209.3	203.2	1.1	177.1	115.6	311.7 344.6	251.6 263.9	172.7
			Yes	210.1	203.3	1.3	177.2	115.8	314.4 348.5	254.6 268.4	172.6
		HID	No	218.6	205.4	0.6	182.3	113.3	328.0 357.5	-	173.1
		vacuum		208.5	202.7	4.2	177.7	115.7	317.1	216.0	179.4
	porphyrin models [6,69]			204-210	197-205	-2 to 2	171-181	112-116			173-180
	protein crystals [6,7,16]			206-231	183-208	0-7	173-240	109-121	316-394	-	120-179
O_2	O_2	HIE	No	204.7	202.1	-3.4	186.8	130.1	283.5	185.4	120.2
			Yes	208.9	202.6	-3.3	187.0	129.9	292.9	195.1	122.0
	CO	HIE	No	217.2	205.9	-2.1	189.9	129.7	328.7[b]	235.5[b]	120.8
			Yes	222.1	207.1	-3.1	189.9	129.2	339.6[b]	245.2[b]	120.9
		vacuum		210.3	201.7	-1.2	186.8	130.0	295.0	196.0	118.7
	porphyrin models [6]			207-211	198-200	-3 to 11	175-190	122-124			129-131
	protein crystals [6,16]			205-207	195-202	-2 to 18	180-183	121-124	270-340	198	115-123

[a] For complexes with CO, two figures are given. The first is the N -O distance, the second is the N -C distance.

[b] The hydrogen bond is in this structure to the proximal oxygen atom.

Table 2. Energetic results (kJ/mole) of the COMQUM-00 calculations.

Ligand	Crystal structure	His-64	Protein relaxed	E_1	E_{Hbond}		E_{bind}	
					vacuum[a]	protein[b]	vacuum[a]	protein[b]
-	-	HIE	No	37.2	-	-	-	-
			Yes	62.6	-	-	-	-
		HID	No	67.2	-	-	-	-
			Yes	59.6	-	-	-	-
	CO	HIE	No	37.7	-	-	-	-
			Yes	49.0	-	-	-	-
CO	CO	HIE	No	40.0	-2.2	-5.3	-99.1	-104.4
			Yes	44.6	-3.3	-7.5	-97.5	-103.8
		HID	No	95.9	14.1			
			Yes	93.5	13.8			
		- [c]	No	31.4	-	-	-102.8	-107.3
		- [c]	Yes	42.2	-	-	-105.4	-113.3
O_2	O_2	HIE	No	42.4	-27.9	-31.4	-94.3	-107.4
			Yes	48.8	-30.1	-33.4	-86.3	-96.8
	CO	HIE	No	94.8	-30.6			
			Yes	80.2	-20.7			
		- [c]	No	37.9	-	-	-71.9	-111.8
		- [c]	Yes	39.8	-	-	-62.2	-96.6

[a] This energy was calculated for the quantum system in vacuum.

[b] This energy was calculated for quantum system including the field of the point charges of the protein.

[c] In these calculations, His-64 is in the HIE form, but it is not included in the quantum system.

Table 3. Geometric parameters (distances in pm, angles in degrees) of the ComQum-X structure of CO-myoglobin, without or with a correction of the systematically too long Fe-ligand distances with offset forces, compared with the two most accurate structures of this protein complex (both at 1.15 Å resolution) [7,16].

Structure	Fe-N_{His}	Fe-N_{Por}	Fe-$N_{Por,av}$	Fe-Por	Fe-C	C-O	N - O	N - C	Fe-C-O
Uncorrected	211.0	200-205	202.5	3.0	176.5	115.5	312.3	352.2	173.1
Corrected	212.1	199-203	200.5	3.8	171.0	114.8	310.6	352.6	174.1
1bzr [7]	211.2	199-203	200.5	1.6	173.1	112.6	311.2	353.2	171.3
1a6g [16]	206.2	194-202	198.4	1.6	182.2	109.2	315.6[a]	341.7[a]	171.2

[a] The distal histidine residue is disordered in this structure.The table describes the conformation which is suggested to have the N atom protonated (20% occupancy). The other conformation is suggested to represent a structure where the N atom is protonated (60% occupancy). It has N -O and N -C distances of 274 and 318 pm [16].

Figure 1. The ComQum-00 structure of the active site of myoglobin without any distal ligand and with the distal histidine protonated on the N atom. The complex was optimised in the native protein, which was fixed at the crystal structure.

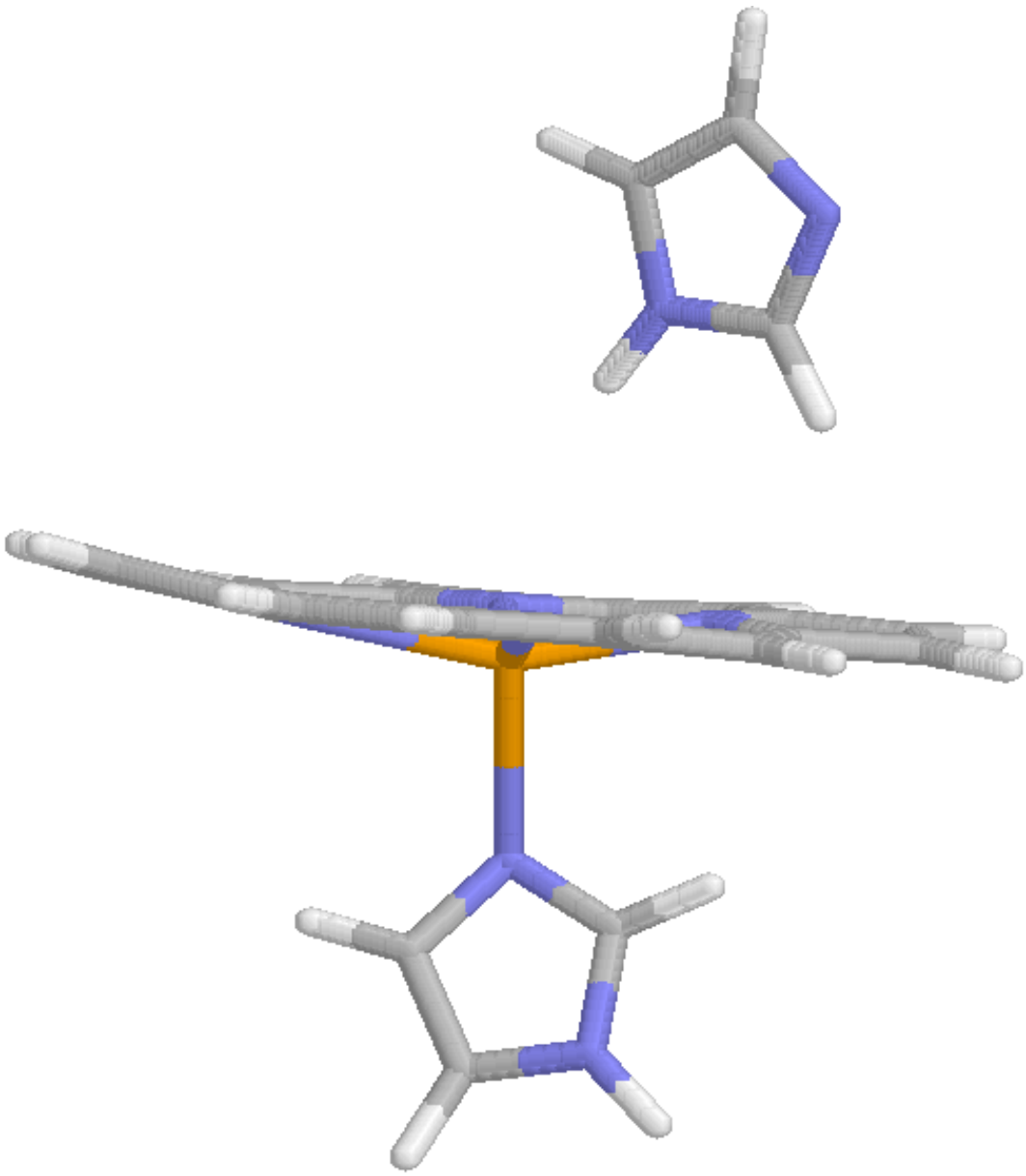


Figure 2. The ComQum-00 structure of the active site of myoglobin without any distal ligand and with the distal histidine protonated on the N atom. The complex was optimised in the native protein, which was allowed to relax during the optimisation.

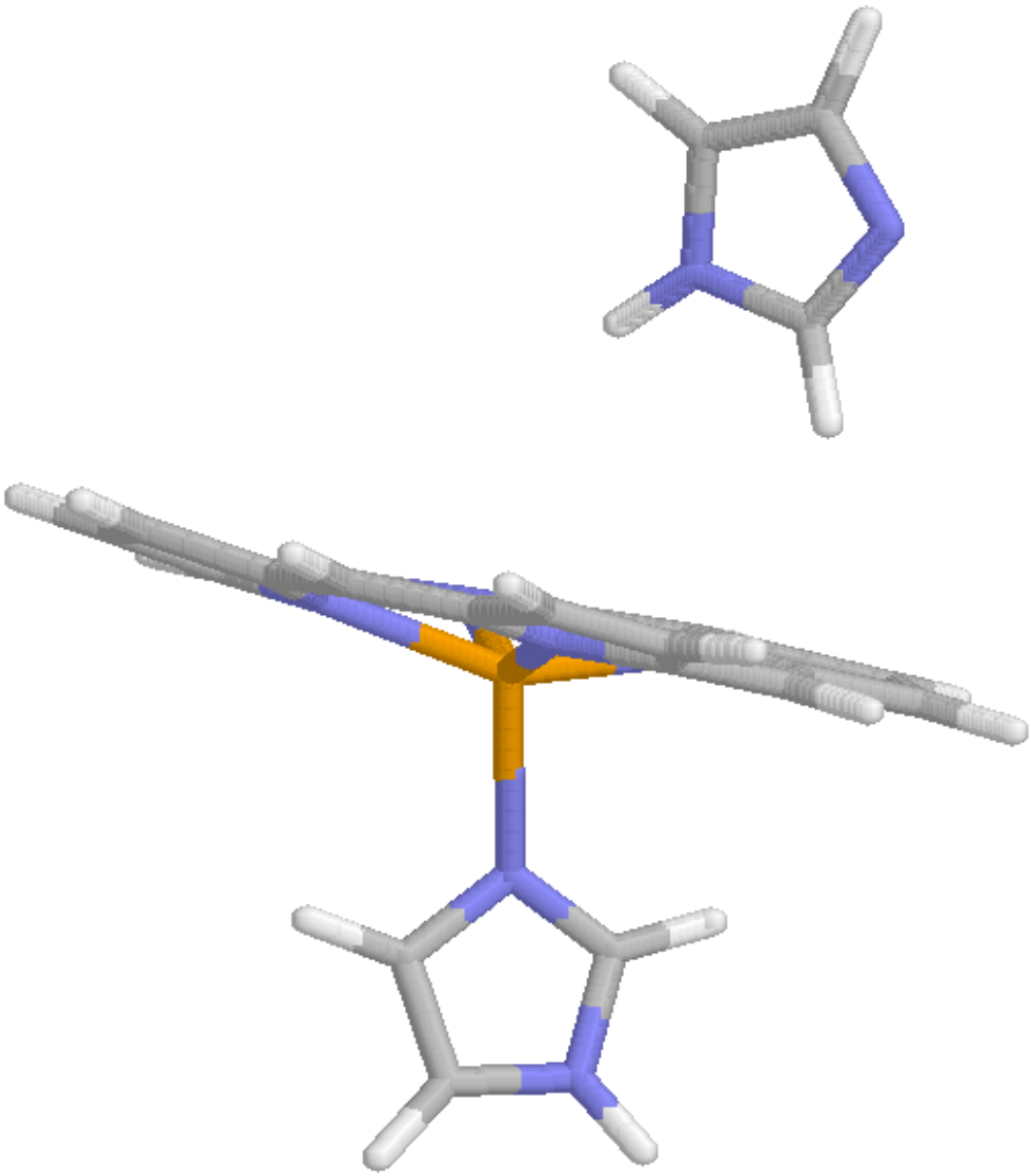


Figure 3. The ComQum-00 structure of the active site of myoglobin without any distal ligand and with the distal histidine protonated on the N atom. The complex was optimised in the native protein, which was fixed at the crystal structure.

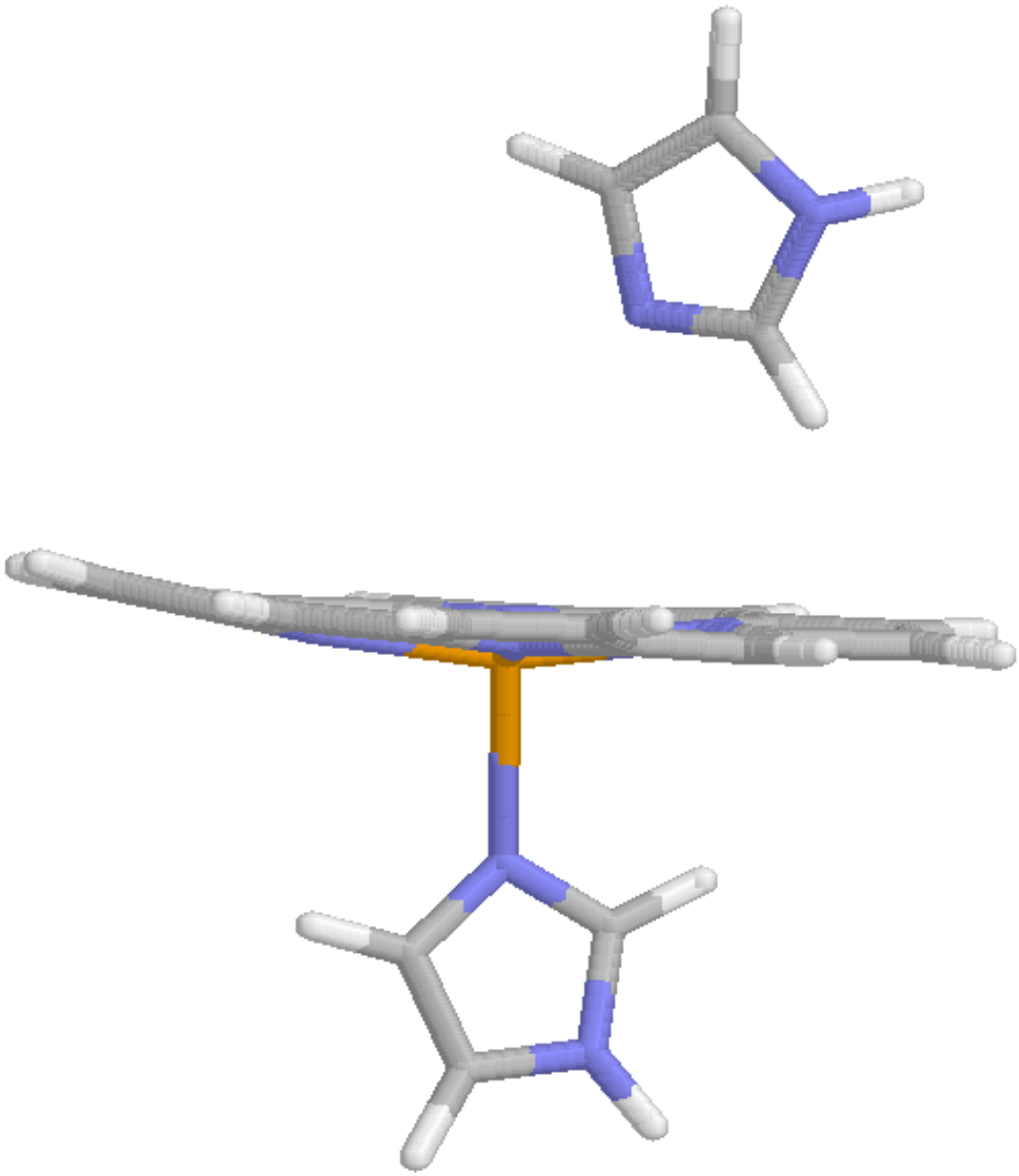


Figure 4. The ComQum-00 structure of the active site of myoglobin binding a CO ligand and with the distal histidine protonated on the N atom. The complex was optimised in the native protein, which was fixed at the crystal structure.

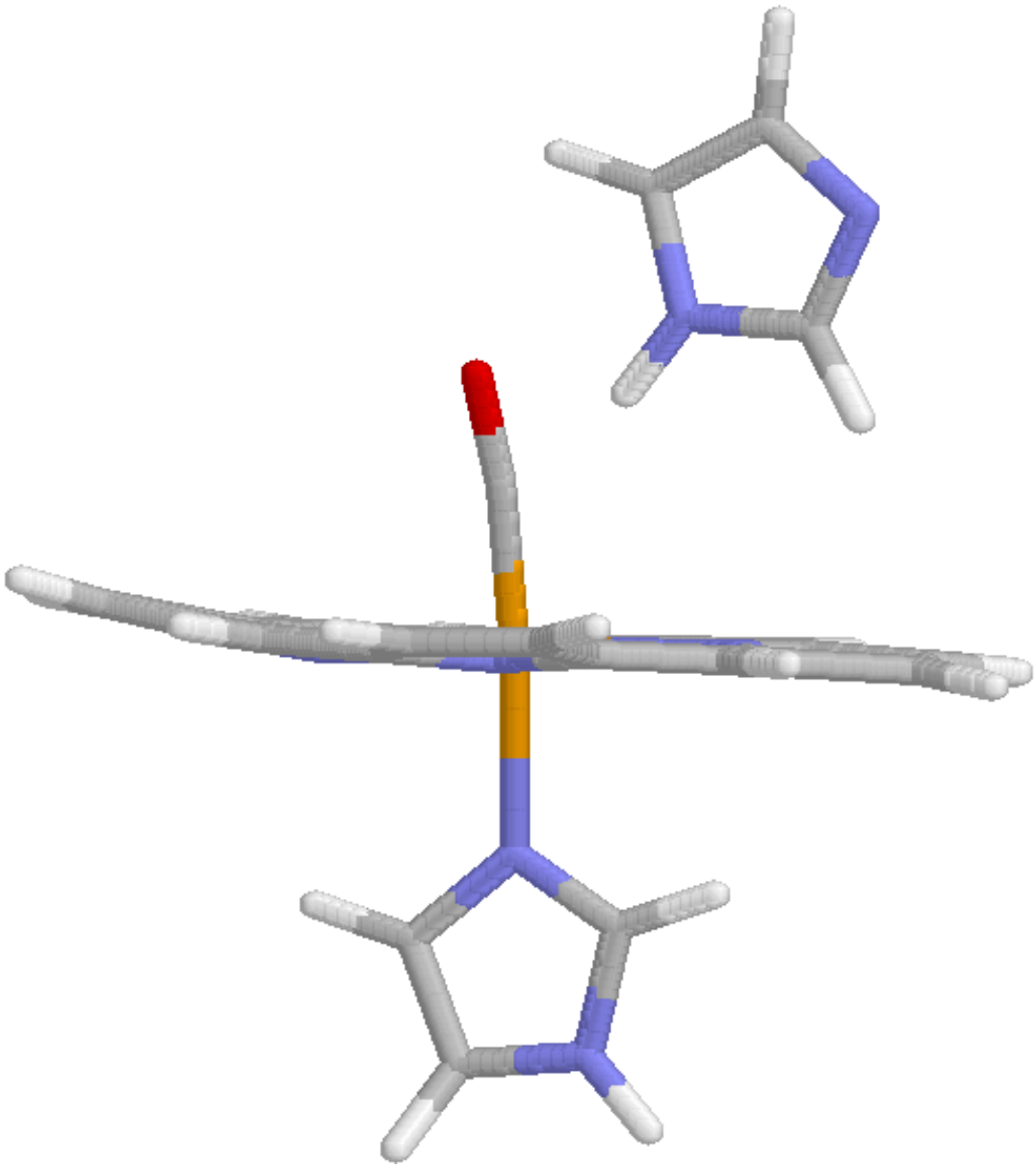


Figure 5. The ComQum-00 structure of the active site of myoglobin binding a CO ligand and with the distal histidine protonated on the N atom. The complex was optimised in the native protein, which was fixed at the crystal structure.

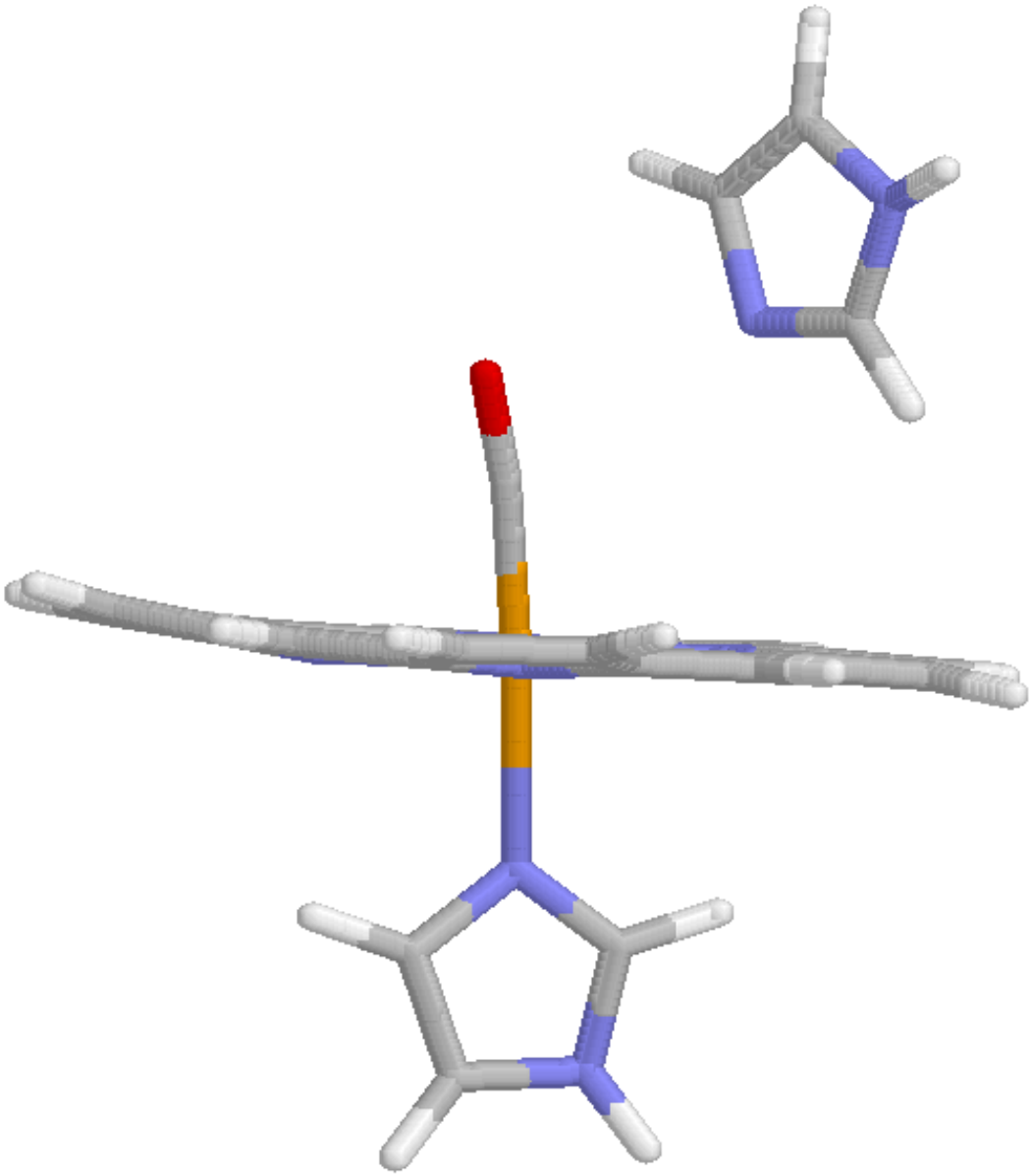


Figure 6. The ComQum-00 structure of the active site of myoglobin binding a O_2 ligand and with the distal histidine protonated on the N atom. The complex was optimised in the native protein, which was fixed at the crystal structure.

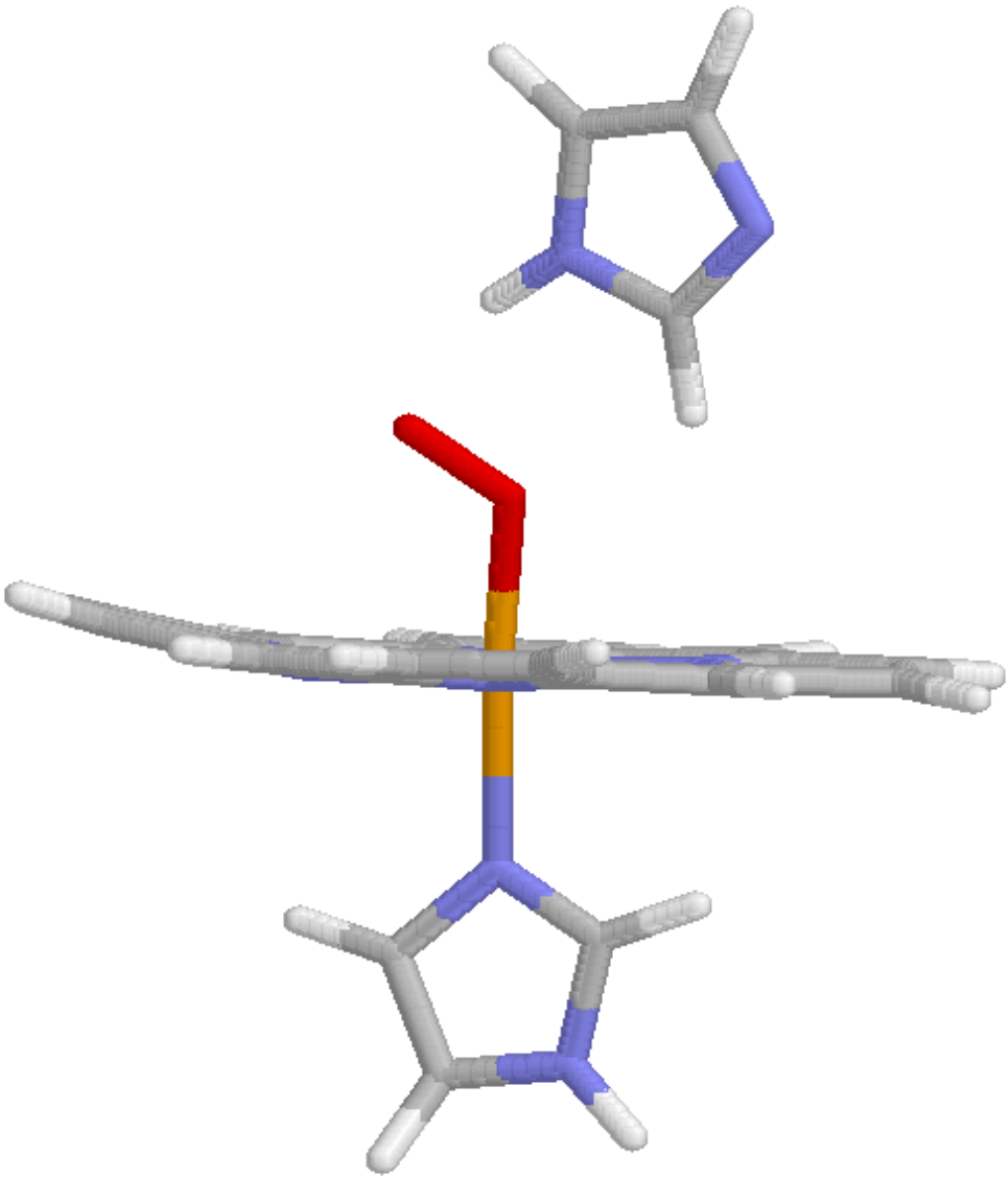


Figure 7. The ComQum-00 structure of the active site of myoglobin binding a O_2 ligand and with the distal histidine protonated on the N atom. The complex was optimised in CO-myoglobin, which was fixed at the crystal structure.

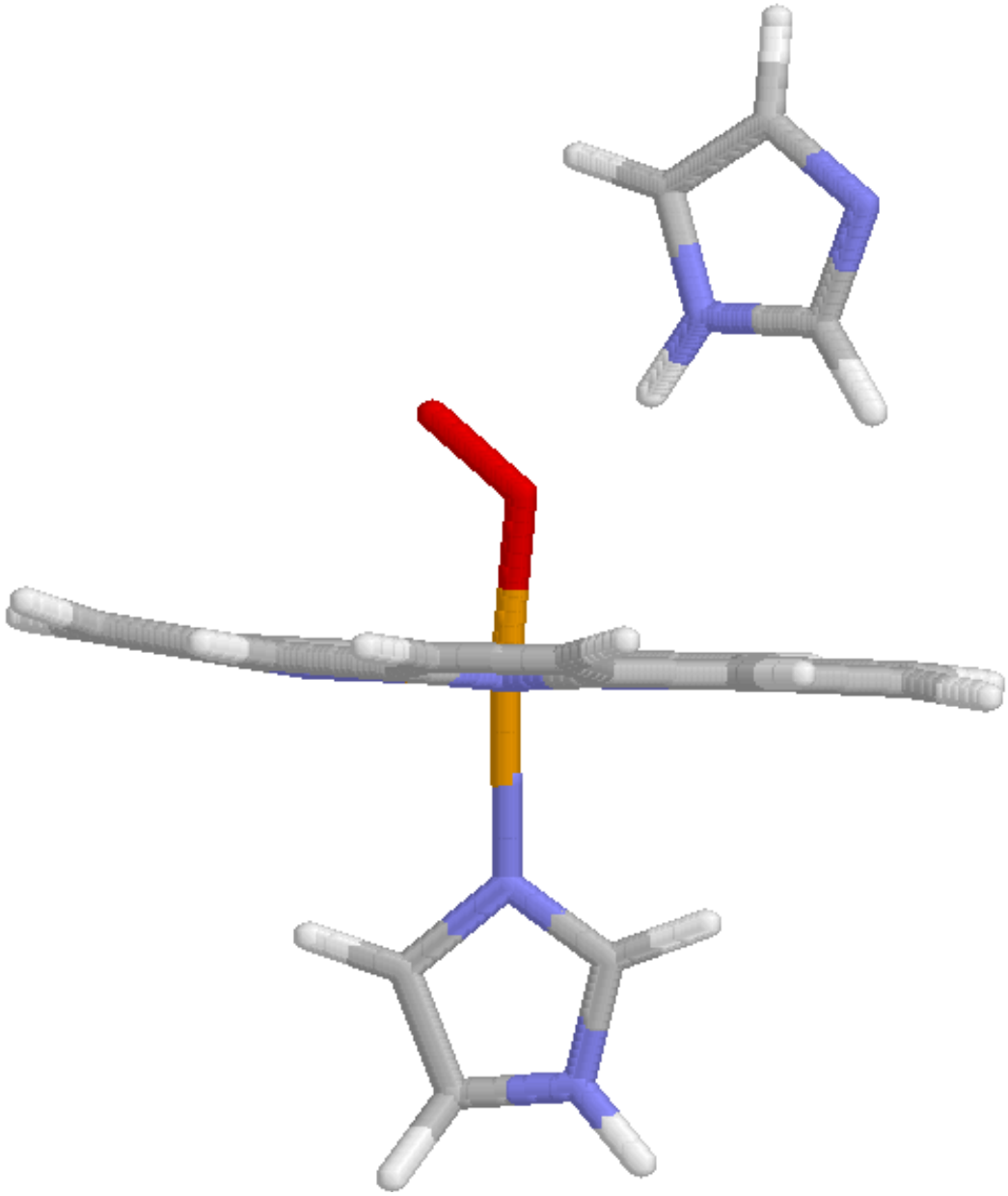


Figure 8. A comparison of the crystal structure of CO-myoglobin (thick lines and no hydrogen atoms) [16] with the ComQum-00 structures of the same complex with either the distal histidine protonated on the N atom (thin lines) or on the N atom (intermediate lines). The calculations were performed with the protein fixed at the crystal structure.

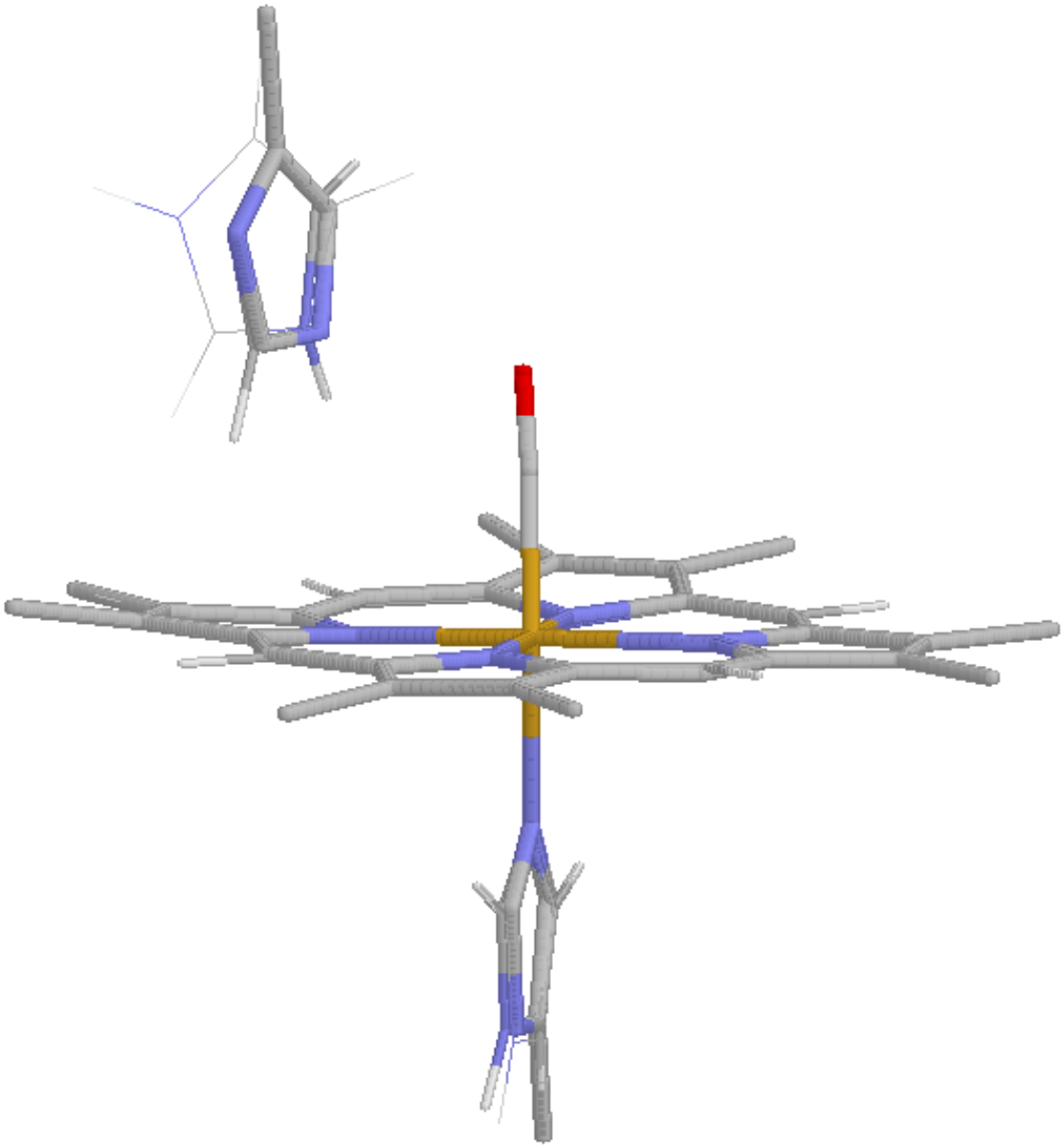
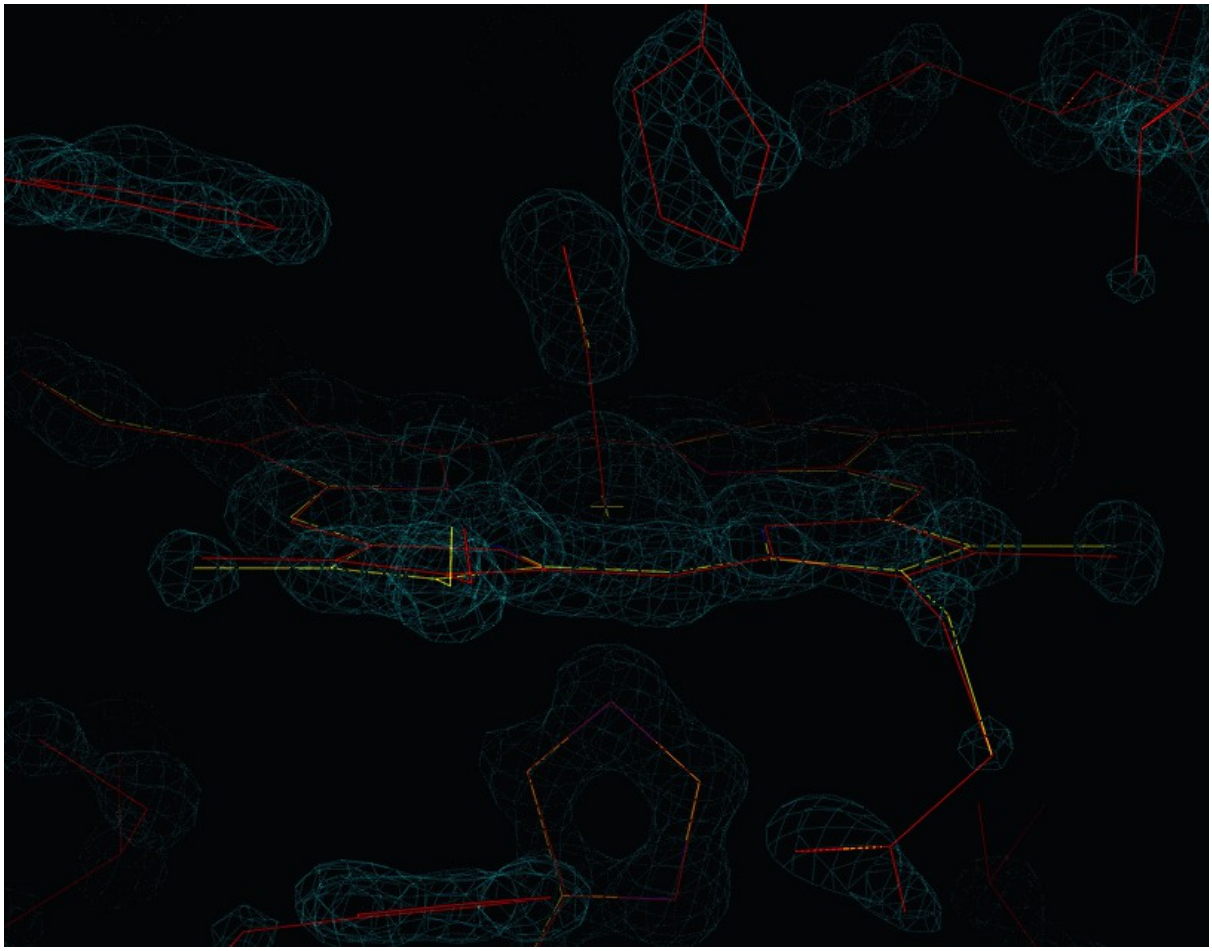


Figure 9. The ComQum-X structure of the active site of myoglobin binding a CO ligand (red) compared to the original crystal structure (yellow) [7] and the electron density (a $3f_o$-$2f_c$ map at the 2.5 level).



Not completed
CO.HID.fix and free
O2 in CO.fix and free
ComQum-X Offset f
PorO2 and PorO2Im: closed-shell singlets
PorO2Im, vacuum, new

Energies with large basis sets
PorO2Im

Frequencies
Start correct on husmodern (talk to Daniel first)

Plan:
1. Energies with large basis sets
   PorO2Im
2. Submit

Remaining jobs:
1. Update abstract with latest ComQum-X results
2. Update the CO-HID structure (T1, T2, F5, F8, p. 14, 20)
3. Update the O2-CO structure (T1, T2, F7, p. 15)