Accurate metal-site structures in proteins obtained by combining experimental data and quantum chemistry

Ryde, Ulf

Published in:
Dalton Transactions

DOI:
10.1039/b614448a

2007

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

Accurate metal-site structures in proteins obtained by combining experimental data and quantum chemistry

Ulf Ryde *

Department of Theoretical Chemistry, Lund University, Chemical Centre, P. O. Box 124,

SE-221 00 Lund, Sweden

Correspondence to Ulf Ryde, E-mail: Ulf.Ryde@teokem.lu.se,

Tel: +46 – 46 2224502, Fax: +46 – 46 2224543

2017-04-02

Abstract
The use of molecular mechanics calculations to supplement experimental data in standard X-ray crystallography and NMR refinements is discussed and it is shown that structures can be locally improved by the use of quantum chemical calculations. Such calculations can also be used to interpret the structures, e.g. to decide the protonation state of metal-bound ligands. They have shown that metal sites in crystal structures are frequently photoreduced or disordered, which makes the interpretation of the structures hard. Similar methods can be used for EXAFS refinements to obtain a full atomic structure, rather than a set of metal–ligand distances.

Keywords: crystallographic refinement, NMR structure refinement, EXAFS structures, density functional theory, molecular mechanics methods, QM/MM methods, quantum refinement, metal sites, metalloproteins.

Biography

Ulf Ryde was born in 1963 and obtained a Ph.D. in biochemistry (enzyme kinetics and control theory) at Lund University, Sweden, 1991, under the supervision of Prof. Gösta Pettersson. As a postdoctoral fellow, he moved to the department of Theoretical Chemistry at the same university, where he was appointed as a docent 1996 and professor 2004. He studies the relation between the structure and function of metalloproteins with theoretical methods and their combinations with experimental data. He has studied many different proteins, e.g. alcohol dehydrogenase, blue copper proteins, hydrogenase, superoxide dismutases, as well as haem and other tetra-pyrrole proteins.

Introduction

Protein structures are invaluable for biochemical studies and for the understanding of how proteins work. The most common methods to obtain protein structures are X-ray crystallography and nuclear magnetic resonance (NMR) refinement. For metalloproteins, a third method to obtain local information about the metal site is extended X-ray absorption fine structure (EXAFS). On the other hand, theoretical methods, in the form of quantum mechanical or molecular mechanics calculations have become strong alternative and complement to experimental methods to obtain insight in biochemical processes. In particular, they are powerful methods to obtain and compare reaction and activation energies for suggested enzyme mechanisms and they can also provide atomic details about the protein dynamics.[1]

In this paper, I will discuss these three experimental methods to obtain structures and their relation to theoretical methods. In particular, I will describe how they can be supplemented and improved by using quantum chemical calculations. I will start with a short introduction of the available theoretical methods for structure determination and then discuss the three experimental methods and their possible combination with quantum chemical methods, including a number of typical applications.

Theoretical Methods

A method is called theoretical if it obtains its results from calculations rather than from experiments. In fact, the laws of quantum mechanics (QM), in particular the Schrödinger equation, allow us to calculate any measurable property of a defined set of atoms.[2] The only problem is that the equations are so complicated that they cannot be exactly solved for any system, except the very simplest ones, such as the hydrogen atom. Instead, approximations have to be made and the equations are solved numerically by computers. In general, the computational load rises rapidly with the number of atoms and electrons considered and therefore, only rather small systems can be studied accurately. However, there are methods that allow larger systems to be studied by the use of statistical mechanics methods and continuum theory.[3] As an effect of these problems, a great number of theoretical methods exists. However, in the following, I will restrict the discussion to the two types of methods that are most appropriate for metalloproteins, viz. density functional theory and molecular mechanics.

The prime result of theoretical calculations is the energy of the studied system. By calculating the first derivative of the energy with respect to the coordinates, the forces are obtained, and these can be used to obtain the optimised geometry or to follow the time-evolution of the system, molecular dynamics (MD). Many other properties of the system can also be obtained (electric properties, vibrational and electronic spectra, NMR and EPR parameters, etc.), provided that somebody has written the corresponding computer code.

DFT methods

Density functional theory (DFT) provides a powerful alternative to the direct solution of the Schrödinger equation.[3,4] The advantage of the DFT methods is that they consider only the electron density (a function of the three spatial coordinates), whereas the Schrödinger equation involves the coordinates of all electrons in the system. Unfortunately, the exact form of the DFT equations is not known; therefore, a plethora of DFT methods is available. Although physicists have used DFT methods for over 40 years, it was not until modern gradient-corrected functionals became available in the late 1980ies that good results were also obtained for chemical systems. During the last ten years, DFT methods have replaced nearly all other QM methods for biochemical and bioinorganic applications, because they are cheap and accurate.[4] They can be applied to systems with over 100 atoms, and the first applications

to whole proteins have started to appear.[5]

Although it has recently become more and more obvious that different DFT functionals may give widely different relative energies for transition-metal complexes,[4,6] geometries are much more stable[4,7] In fact, state-of-the-art DFT methods provide bond lengths for molecules of first- and second-row atoms to within 0.02 Å.[8] Metal–ligand bonds are harder to calculate, but our experience show that they can be calculated to within 0.07 Å, compared to atomic-resolution crystal structures of small inorganic complexes and proteins,[4,9,10,11,12,13,14] as is shown in Table 1, with the exception for very weak interactions, e.g. the axial Cu–S_{Met} bond in the blue copper proteins.[15] In fact, this is appreciably better than what is obtained in medium-resolution protein crystal structures, for which the *average* error in the bond lengths is ~0.10 Å, and errors of over 0.3 Å are frequently observed (as we will see below).[16,17] Of course, this is the basis for the use of DFT methods to supplement experimental data. Likewise, it is also the basis of the use of vacuum QM geometries as the reference also for structures in proteins – we will see that the geometry of metal sites in proteins is very close to that obtained in vacuum (again with the exception of very weak interactions). The is in accordance with the observation that a protein cannot significantly distort the structure of a bound ligand and that mechanical strain typically has little influence on biochemical reactions.[18]

MM methods

As mentioned above, the first DFT calculations of a full protein have recently started to appear. However, it is still not possible to study large proteins and the surrounding solvent in a systematic manner. Therefore, more approximate methods have been developed, which completely ignore the electrons. Instead, a molecule is viewed as a collection of balls, connected with springs. Thus, the interactions within and between different molecules are described by an empirical potential (a force field), instead of solving the Schrödinger equation.[3] This is the molecular mechanics (MM) or force-field based methods. Of course, the efficiency and accuracy of the methods depend on the form and accuracy of the empirical potential.

Many such potentials have been developed, but the most common functional form for protein simulations is the following:[3]

$$E_{MM} = \sum_{\text{bonds}} C_i (b_i - b_{i0})^2 + \sum_{\text{angles}} D_i (\alpha_i - \alpha_{i0})^2 + \sum_{\text{dihedrals}} \sum_{j=1} E_{ij} (1 + \cos(j\phi_i + \delta_j)) + \sum_{\text{atom pairs}} \frac{q_i q_j}{4\pi\epsilon_0 r_{ij}} + \frac{A_{ij}}{r_{ij}^{12}} + \frac{B_{ij}}{r_{ij}^{6}} \quad (1)$$

In this equation, the energy of a bond depend harmonically on the actual bond length (b_i; b_{i0} is the corresponding equilibrium bond length and C_i is the corresponding force constant; i.e. the first term of a Taylor expansion of the energy). The same applies to the angles (with the actual and equilibrium angles $_i$ and $_{i0}$, and the force constant D_i), whereas the dihedral angles (the torsions) are assumed to be described by a cosine function with a periodicity (j) of 1, 2, or 3 (or perhaps even 4 or 6; i.e. the first terms in a Fourier series). E_{ij} is the corresponding force constant and $_{ij}$ is a phase factor, which determines the position of the minimum. Non-bonded interactions are described by Coulomb's law between partial charges on each atom (q_i) and a Lennard–Jones potential between each pair of atoms that are more than two bonds apart, using the constants A_{ij} and B_{ij}. Consequently, a large number of empirical parameters (C_i, b_{i0}, D_i, $_{i0}$, E_{ij}, $_{ij}$, q_i, A_{ij}, and B_{ij} for all atoms or pairs of atoms) need to be determined, even if this number is reduced by dividing the atoms into types that share the same parameters,[3] and this constitutes the force field.

The accuracy of MM calculations strongly depend on the force field and it should be noted that Eqn. 1 is a rather crude approximation to the true solution to the Schrödinger equation, especially for electrostatic interactions, for which dipoles and higher moment, as

well as electronic polarisation are completely ignored. In general, a good force field of the type in Eqn. 1 gives good bond lengths and angles of standard amino-acid residues and nucleic acids (typically as good or better than what is obtained with DFT methods).[19] However, for more unusual molecules, e.g. substrates, inhibitors, and metal sites, parameters are typically missing or quite inaccurate. Moreover, some systems, e.g. transition metals and transition states, are hard to study with MM methods.[3]

QM/MM methods

We have seen that some systems are hard to study by MM methods and that the accuracy of such methods is typically rather low, whereas QM methods are too expensive to use for a full protein. A common solution to this dilemma is to combine the two methods, i.e. to use QM for a small, but interesting part, whereas MM is used for the rest of the protein (and parts of the surrounding solvent), as is shown in Figure 1. This is the combined QM and MM (QM/MM) approach.[20] This can be accomplished by simply adding the energies:

$$E_{QM/MM} = E_{QM1} + E_{MM12} - E_{MM1} \quad (2),$$

where E_{QM1} is the QM energy of the small QM system, whereas E_{MM12} is the MM energy of the full system. Of course, we have to remove the MM energy of the QM system (E_{MM1}) to avoid double counting of interactions within the QM system.[21]

X-ray Crystallography

X-ray crystallography is the prime source of structural information in biochemistry with over 32 000 structures in the Protein Data Bank (August, 2006). However, it is important to remember that the coordinates obtained from the Protein Data Bank are not the experimental raw data. On the contrary, they are the result of an involved cycle of data processing and interpretation, which is illustrated in Figure 2.[22] After purification and crystallisation of the protein of interest, the crystal is inserted into a beam of X rays and diffracted beams are recorded by a detector. The intensities of these reflections, called the *structure factors*, are the actual raw data of the experiment and they are available in the Protein Data Bank for about 2/3 of the structures.

Unfortunately, the structure factors cannot directly be interpreted as a structure, because the phase of the reflections is not known. There are various methods to determine the phase (e.g. using a similar known structure or heavy-atom derivatives), but once it is known, an electron-density map can be constructed. However, as is shown in Figure 3, the electron density still does not unambiguously determine the position of the atoms, except at the highest resolution (<~1 Å).

Now, the crystallographer tries to build an atomic model into the electron density, typically using the primary sequence, if available. Although powerful software are available to help this model building, this is an error prone procedure and therefore, the structure is *refined* by minimising the difference between the experimental structure-factor amplitudes (f_o) and amplitudes calculated from the current atomic model (f_c) for all reflections, measured by the crystallographic R factor:

$$R = \frac{\sum ||f_o| - |f_c||}{\sum |f_o|} \quad (3)$$

In practice, it is more favourable to minimise other related quantities,[23] but the R factor is still used as one of the most important criteria for the quality of the crystal structure. However, it has been found that there is a risk of overfitting the structure if you only consider the R factor. Therefore, a random set of reflections (typically ~5 %) is set aside from the beginning and is not used during the refinement. They are used only in the cross-validation of the structure, by calculating a separate R factor for these reflections, called R_{free}.[24] Ideally, both the R and R_{free}

factors should be as small as possible. Moreover, $R_{free} > R$, but the difference should be as small as possible.

For all structures except those at the highest resolutions, the experimental data are not accurate enough to give the exact position of all atoms in the structure. Therefore, the experimental data are supplemented by a MM energy function, similar to that in Eqn. 1:

$$E_{total} = E_{MM} + w_A E_{Xref} \quad (4),$$

where E_{Xref} is describes how close the structure is to the experimental data (e.g. the R factor).[22] Now, E_{MM} is in energy units, whereas E_{Xref} is in arbitrary units. Therefore, the two terms need to be weighted with the factor w_A, which determines the relative importance of the two terms, i.e. the relative influence of the experimental and theoretical data on the final structure. Ideally, this factor can be determined by varying it and selecting the value that gives the lowest R_{free} factor. However, in practice, w_A is normally determined so that the X-ray and MM forces have equal root-mean-squared magnitudes in a short MD simulation.[25] *Thus, most crystal structures are actually 50 % theoretical.* Moreover, the method used in the refinement is a standard geometry optimisation, using the energy function in Eqn. 4.

The MM restraints are needed to give a chemically reasonable structure. In general, the X-ray data determine the fold and the dihedrals of the protein structure, whereas the MM restraints determine the exact bond lengths and bond angles (i.e. the detailed position of the atoms). This is illustrated in Figure 4, where the same group is refined without or with the MM restraints.[17,26] It can be seen that without the restraints, the structure is not chemically reasonable (owing to a poor electron density for this group). However, the MM restraints make the structure chemically reasonable, by moving the atoms slightly away from their ideal positions in the electron density. Consequently, the MM restraints will always increase the R factor. For the single group in Figure 4, (18 out of 22224 atoms), the R factor increases only from 0.2273 to 0.2274, illustrating that the R factor is a global measure that is insensitive to small details of the structure. However, if all MM restraints in the whole structure are removed, the R factor decreases from 0.227 to 0.218.

The inclusion of MM restraints of course means that the final structure will depend on what MM potential was used, i.e. the quality of the final structure depends on the quality of the MM potential. This is a relatively small problem for standard amino-acid residues and nucleic acids, for which accurate force fields are available[19] and the force fields used are specified in the PDB file. However, for metal sites, substrates, and inhibitors (so-called hetero-compounds), i.e. often the most interesting part of the structure, no standard force field is available and the crystallographer typically has to construct it himself (i.e. a experimental chemist doing quite advanced theoretical chemistry), an error-prone procedure.[27] Even worse, non-standard force fields for the hetero-compounds are rarely specified, so the errors cannot be identified without re-refining the structure. These problems are illustrated in Figure 5, where the adenosyl group in Figure 4 is refined with two different force fields, an unspecified force field in the original investigation,[26] giving a quite strange structure with a severe steric clash, and the Amber force field,[28] giving a much more reasonable structure and a change in a non-bonded C C distance of 1.7 Å![17]

Metal sites provide one of the hardest problem in crystallography, because there are no accurate general force field for metals. It is therefore often said in the refinement procedure that the metal site was treated without any MM restraints. This typically means that the crystallographer has not defined any explicit bonds between the metal and its ligands (which has to be done by hand). However, that does not mean that there are no interactions between the metal and the ligands. On the contrary, there will be a van der Waals interaction between them (i.e. the last two terms in Eqn. 1; the Coulomb term is normally ignored). With standard van der Waals parameters, such an interaction has a minimum around 2.5 Å, so there is actually a quite strong restraint to increase the distances. Therefore, metal–ligand distances are often too long in crystal structures. The only way to have a truly unrestrained metal–ligand interaction is to define bonds between the metal and its ligands and then zeroing the

corresponding force constants (C_i in Eqn. 1).

Quantum refinement

We have seen that standard X-ray crystallography may have problems around hetero-compounds. A solution to these problems, and also a possibility to improve the crystal structure locally, is to replace the relatively inaccurate MM potential in Eqn. 4 with more accurate QM calculations for a small part of the protein, in the same manner as in the QM/MM methods (cf. Figure 1), i.e. to use the energy function:

$$E_{ComQum\text{-}X} = E_{QM1} + E_{MM12} - E_{MM1} + w_A E_{Xref} \quad (5)$$

(this is equivalent to adding a restraint towards the X-ray structure to the QM/MM energy function in Eqn. 2). This energy function has been implemented in the program ComQum-X,[29] as a simple interface between the QM program Turbomole[30] and the crystallographic refinement program CNS[31] (Crystallography and NMR system).

Test calculations on cytochrome c_{553}

The ComQum-X method was first calibrated and tested by the re-refinement of the inhibitor *N*-methylmesoporphyrin in the enzyme ferrochelatase (Figure 3a).[29] However, the most convincing results come from an application on the haem group in cytochrome c_{553}.[11] It is quite hard to show that ComQum-X actually improves a crystal structure, because the MM or QM restraints make the fit to the experimental data worse (in terms of the *R* factor), as we saw above. A way to solve this dilemma is to find a protein that has been studied at both atomic resolution (where geometric restraints have a small influence on the structure) and at a lower resolution (where they are important), but otherwise at as similar conditions as possible.

There is such a pair with an appropriate hetero-compound in the Protein Data Bank: Cytochrome c_{553} from *Bacillus pasteurii* has been solved at 0.97 Å resolution with ab initio phasing and independently by the same group at 1.7 Å resolution in an multiple anomalous dispersion experiment.[32] The crystals were obtained at the same conditions. This protein contains a haem group, where the central Fe^{III} ion binds also to a histidine and a methionine residue from the protein (Figure 3b). The two crystal structures show an appreciable difference in the iron–ligand bond lengths, as can be seen in Table 2.

We have optimised the structure of this haem group with ComQum-X, using the low-resolution data.[32] The results are included in Table 2 and show that ComQum-X brings the structure much closer to the high-resolution structure: The difference in the $Fe–N_{His}$ bond length compared to the high-resolution structure is reduced from 0.32 Å in the low-resolution structure to 0.00 Å in the re-refined structure, that of the $Fe–S_{Met}$ bond is reduced from 0.12 to 0.02 Å, and those of the $Fe–N_{Por}$ bonds are reduced from 0.03–0.09 to 0.01–0.02 Å. This is of course a manifestation of the excellent performance of density functional theory for this metal site; already the vacuum structure gives differences to the high-resolution structure of less than 0.03 Å.

This improvement can also be seen for the *R* factors. Unfortunately, the selection of the test set of the reflections is not available in the databank. Therefore, we can only measure how much the *R* factor is reduced by ComQum-X compared to the low-resolution structure (R_{low} in Table 2). It can be seen that it is improved by 0.0055. We can calculate a similar *R* factor based on the high-resolution reflections. These are also given in Table 2 (R_{high}) and it shows a similar improvement.

The improvement is even clearer when we compare the ComQum-X structure (orange) with the low- (magenta) and high-resolution (green) crystal structures in Figure 3b. It is mainly the iron ion (0.19 Å), the N^{1} atom (0.27 Å), and one of the ethyl side chains (2.10 Å) that move (the average movement of all atoms is 0.12 Å), but the ComQum-X structure is consistently closer to the high- than to the low-resolution structure. The high-resolution

electron-density map also confirms that ComQum-X has improved the low-resolution structure.

Protonation of metal-bound solvent molecules

Thus, we have shown that we can obtain improved structures using ComQum-X. Next, we tested if ComQum-X can also be used to *interpret* crystal structures, i.e. to decide exactly what chemical species are present in the structure. Protons are normally not seen in protein structures.[22] Therefore, it is interesting to see if ComQum-X can be used to decide where they are, e.g. to determine the protonation status of various molecules in the structure. For example, it would be highly interesting to if we could determine the protonation status of metal-bound solvent molecules using ComQum-X, because they are often important for the function of the proteins.

At first, it may seem strange that this cannot be done directly from crystal structures, because a metal–OH^- bond is typically ~0.3 Å shorter than the corresponding bond to water (cf. Table 3). However, metal–ligand bond lengths depend on the other ligands of the metal and they are normally not available to the crystallographer, whereas they can be calculated by density functional theory. Moreover, we have seen that there may very well be errors of this size in the crystal structures.

In order to calibrate the method, we first needed some structures for which the protonation status is known. Alcohol dehydrogenase is such a case, where the pK_a of the solvent molecule bound to the catalytic zinc ion is known from kinetic measurements.[33] We have studied a complex between alcohol dehydrogenase, NAD^+, and trifluoroethanol (at 2.0 Å resolution),[34] in which the alcohol should have a pK_a of ~6.[33] This is well below the pH at which the crystal were grown, 8.4, which means that it should contain a deprotonated alkoxide ion.

We have calculated ComQum-X structures of this complex with both an alkoxide and a protonated alcohol.[35] The results in Table 3 show that the alkoxide structure fits the experiment data better by at least three independent criteria. First, the alkoxide gives the lowest value for the R_{free} factor, calculated with the same value of w_A (3 or ~1.8), as well as with the optimum values of w_A (in terms of the R_{free} factor; 10 for the alcohol and 0.3 for the alkoxide).

Second, the alkoxide gives a lower strain energy (E_{QM1}, i.e. QM energy difference of the QM system optimised in vacuum and in the protein) than the alcohol for all values of w_A. This indicates that the alkoxide fits the electron density better than the alcohol. Third, the Zn–O distance in the alkoxide structure is close to that found in vacuum calculations (1.93 Å) at all values of w_A, whereas for the alcohol, the Zn–O distance (2.01–2.04 Å) is far from the vacuum value (2.29 Å) and actually converges towards the vacuum value of the *alkoxide* complex. This clearly indicates that the Zn–O bond length preferred by the crystal data is closer to that expected for the alkoxide than for the alcohol.

We have also studied another crystal structure of alcohol dehydrogenase,[36] for which the experimental data indicate that the zinc-bound water molecule is protonated. With the same three criteria as above, we showed that the a zinc-bound water molecule fits the crystal data better than a hydroxide ion.[35] Thus, it is clear that ComQum-X can discern the two protonation states of zinc-bound solvent molecules in alcohol dehydrogenase.

Thus, ComQum-X can the determine protonation status of groups in proteins. We have applied this method to the protonation status of *N*-methylmesoporphyrin in ferrochelatase,[37] to the iron-bound water molecule in both iron and manganese superoxide dismutases,[17,37,38,39] to compound II in myoglobin,[40] to copper nitrite reductase from various sources,[41] and to various states of [Ni,Fe] hydrogenase.[14] In general, it seems to be possible to determine the protonation status of metal-bond ligands if the crystal structure represents a well-defined state (see below), whereas it is hard to determine the protonation of organic groups in proteins,

because the difference in geometry of the two protonation states is much smaller than for metal–ligand distances.

Photoreduction

Unfortunately, in many of our applications of ComQum-X, we have run into severe problems. For example, we have studied atomic-resolution structures (0.9 Å) of manganese superoxide dismutase (MnSOD) in the oxidised (Mn^{III}) and reduced (Mn^{II}) states.[42] MnSODs are found in bacteria and in the mitochondria of eukaryotes, where they are believed to protect DNA from endogenous oxidative stress.[43] They catalyse the disproportionation of two molecules of the poisonous superoxide radical to molecular oxygen and hydrogen peroxide. The active site consists of a Mn ion bound to one aspartate (Asp) and three histidine (His) residues (Figure 6). A solvent molecule completes the trigonal bipyramidal structure, in which the solvent molecule and one of the His ligands occupy the axial positions. The solvent molecule can be either water or a hydroxide ion, and it is believed that the water is deprotonated when the enzyme is oxidised.[43,44] This suggestion has gained support from computational studies,[45] but it has not been shown by any direct methods.

Therefore, we re-refined these structures with ComQum-X.[38] For the reduced protein, the results are quite conclusive (Table 4): The structure with a Mn-bound water molecule fits the experimental data best according to all criteria: It gives an almost three times smaller strain energy (E_{QM1}) than the hydroxide structure. Second, all five distances in the water structure are within 0.07 Å of those obtained in vacuum, whereas the OH^- structure gives much larger differences (0.12–0.17 Å). Third, the metal–ligand bond lengths in water structure are very close to those in the original crystal structure (maximum difference 0.02 Å), whereas in the OH^- structure, the Mn–O_{Sol} distance differs by 0.12 Å from the original crystal structure. The same applies to the R and R_{free} factors, and also to the real-space (residue) R factor (i.e. an R factor calculated for only a small part of the structure, e.g. a single residue),[46] which shows a much larger difference between the two structures: The real-space R factor of the solvent molecule calculated from an omit map of the protein without the metal and its ligands is 0.051 for the water structure, but 0.056 for the OH^- structure. Thus, we can safely conclude that the reduced crystal structure contains a metal-bound water molecule.

However, for the oxidised crystal structure, the results are much harder to interpret. For example, the R_{free} factors of both the water and OH^- structures increased compared to the original structure, indicating that the structures are not improved by QM. Likewise, both structures showed quite large deviation in the metal–ligand distances from those obtained in vacuum. This indicates a significant misfit between the crystal structure and the QM calculations.

A possible explanation for the misfit is that the crystal structure has been partly reduced during data collection. A significant amount of the X-ray photons deposit their energy into the crystal lattice, giving rise to secondary electrons that may change the redox-state of metalloproteins.[47] In fact, it has been suggested before that oxidised structures of MnSOD are partly reduced.[45,48] Therefore, we also re-refined the oxidised structure with Mn^{2+} and either water or OH^- in the QM system.

These new structures gave somewhat better results. For example, R_{free} was reduced to 0.2122 in the $Mn^{II}H_2O$ structure, and the difference in the Mn–O_{Sol} distance between crystal and vacuum was reduced to 0.04 Å. This structure also had the lowest deviation form the original crystal structure, but both the maximum deviation (0.06 Å) and the absolute sum of the deviations (0.11 Å) were 2–3 times larger than for the reduced crystal structure. Moreover, the other criteria pointed out different structures as the best ones: The $Mn^{II}OH^-$ structure had the lowest value for the R and real-space R factors, whereas the $Mn^{III}OH^-$ structure had the lowest sum of absolute deviations in the metal–ligand distances. This quite strongly indicates that the crystal structure is a mixture of oxidation and protonation states, as can be expected if

the structure is successively reduced during data collection. Most likely, it is a mixture of the $Mn^{III}OH^-$ and the (dominant) $Mn^{II}H_2O$ structures.

This shows it is hard to interpret with ComQum-X crystal structures that contain a mixture of different oxidation states (as an effect of photoreduction). On the other hand, ComQum-X can detect such reduction, which is highly important for the interpretation of the structure: If there is a mixture of oxidation states in the active site, it means that the crystal structure is unreliable, i.e. that the details of the structure cannot be trusted. This is nicely illustrated by the present structure, which shows a $Mn–O_{Sol}$ distance of 2.12 Å, i.e. in between the expected distances for $Mn^{III}OH^-$ (1.78 Å) and $Mn^{II}H_2O$ (2.23–2.27 Å).

Structural disorder

A related problem is illustrated by our applications on [Ni,Fe] hydrogenase.[14] This enzyme is found in a large number of bacteria and archaea, as well as in some simple eukaryotes, and it catalyses the reaction $H_2 \quad 2\ H^+ + 2\ e^-$.[49] The active site contains a Ni ion bound to four cysteine ligands (cf. Figure 7). Two of these are terminal ligands, whereas the other two form bridges to the Fe ion, which also coordinates one CO and two CN^- ligands. In oxidised structures, there is an additional bridging solvent species, the nature of which is a subject of debate – it may be H_2O, OH^-, O^{2-}, or even a peroxide ion.[50] We therefore started a ComQum-X investigation[14] of the structure with the lowest resolution at that time (1.81 Å).[51] It is possible that some of the Cys ligands are protonated and the oxidation state of Ni may be either +II or +III. Therefore, we tested 47 different structures[14] However, in no case, we obtained any reasonable structure with reduced *R* factors and the maximum change in the metal–ligand distances compared to the corresponding vacuum structures was 0.13 Å, or more, i.e. twice as large as expected. After much effort we realised that the reason for this is that the structure is a mixture of several states. In particular, at least two of the cysteine residues were partly oxidised (to Cys-SO^- groups). As an effect of this oxidation, a neighbouring Glu group moves by over 1 Å (Figure 8).

This shows that ComQum-X is also sensitive to structural disorder in the quantum system. However, once again, ComQum-X can also be used to detect such disorder and if it is present, it shows that the original structure is not reliable. Still, we were able to show that the bridging ligand in [Ni,Fe] hydrogenase is most likely a hydroxide ion, if it is mononuclear and that at least in the studied crystal structure,[51] a structure with OH^- fits the experimental data better than any form of peroxide.[14] Moreover, although the suggested peroxide structures are chemically reasonable, they are over 200 kJ/mole less stable than the corresponding structures with oxidised Cys residues. This makes it questionable if peroxide structures actually are observed for [Ni,Fe] hydrogenases and not complicated mixtures with various oxidised Cys ligands.

High-resolution structures and offset forces

Finally, we have also studied CO–myoglobin with ComQum-X.[52] We compared two crystal structures of this complex. Although both structures were of a high resolution (1.15 Å),[53] they still differed by 0.09 Å in the Fe–C bond length (cf. Table 5). Therefore, we re-refined one of the crystal structures using ComQum-X.[52] At this high resolution it is possible that the systematic overestimation of metal–ligand distances by density functional methods becomes significant.[9] Therefore, we investigated if such systematic errors can be corrected by the method of offset forces.[54] Interestingly, the results in Table 5 show that the ComQum-X results are insensitive to the correction. ComQum-X gives a Fe–C–O angle similar to that in both crystal structures, a $Fe–N_{His}$ distance intermediate between the two structures, $Fe–N_{Por}$ distances closer to the structure used in the refinement, but a Fe–C distance that is shorter (1.70 Å, compared to 1.73 and 1.82 Å) and a C–O distance that is longer (1.16 Å, compared

to 1.09 and 1.13 Å) than those in the two crystal structures. On the other hand, these two distances are closer to what is observed in a small inorganic model of a similar complex (1.74 and 1.16 Å).[55] The improvement of the crystal structure is flagged by a decrease in the R_{free} factor of 0.0013. Thus, appreciable errors (>0.1 Å) in metal–ligand bond lengths can be observed also in high-resolution crystal structures.

NMR refinement

Structure determination using nuclear magnetic resonance (NMR) techniques is the second most important source of protein structures (5600 structures in the Protein Data Bank in August, 2006). In principle, the method is similar to that of X-ray structure determination: The raw data consist of a number of estimated distances between pairs of atoms, dihedral angles, and hydrogen bond lengths.[56] These are converted to a three-dimensional structure by the use of distance geometry methods or restrained molecular dynamics (MD), followed by energy minimisation. In the latter case, which today is most common, the NMR data are supplemented by a MM force field, so that the energy function is of the form:

$$E_{total} = E_{MM} + w_N E_{NMR} \quad (6),$$

which is completely analogous to that used in X-ray crystallography (Eqn. 4), but with a restraint to the NMR data (E_{NMR}) instead. Therefore, it is straightforward to introduce QM data also in the NMR refinement, using the energy function:

$$E_{ComQum\text{-}N} = E_{QM1} + E_{MM12} - E_{MM1} + w_N E_{NMR} \quad (7),$$

especially as the CNS program also performs NMR refinements.[31] This has been implemented in the ComQum-N program.[57]

As with quantum refinements, the aim of ComQum-N was to provide a more detailed and accurate description of parts of the structure that is not well-described by the NMR data, e.g. around metal sites.

Application on EGF34

We have applied the ComQum-N method to the two calcium sites in the epidermal growth factor-like (EGF) domains 3 and 4 in vitamin K dependent protein S (EGF34), which is involved in the blood clotting system. EGF domains often bind Ca^{2+} ions in 6–7 coordinate sites and such domains often come in tandem repeats, in which the calcium affinity is larger than in the isolated domains. For example, in the studied dimer, domain 4 has a 8600 times larger Ca affinity than when it is isolated.[58] Many crystal and NMR structures of Ca-binding EGF domains are known.[59] One of the best is a 1.5 Å crystal structure of the EGF-like domain in human clotting factor IX.[60] It shows two EGF domains, each binding a Ca ion in a pentagonal bipyramidal manner with two carboxylate ligands (one binding bidentately), one amide oxygen, one water molecule and two back-bone carbonyl oxygen atoms (cf. Figure 9).

The largest difference between crystal and NMR structures is that the former gives a single structure, whereas NMR structures normally are presented as an ensemble of structures (from MD simulations with different starting conditions), which all satisfy the NMR restraints approximately equally well. They are typically presented as superimposed structures. In EGF34, the two EGF domains are connected by a flexible linker, so the two domains have to be superimposed in separate figures. This is illustrated in Figure 10, which also shows that the general structure is well determined by the NMR restraints.

However, the NMR restraints contain little information about the Ca sites, which therefore are poorly defined, as can be seen in Table 6 (first row). If external information is available, e.g. the crystal structure[60] in the present case, this information is normally included as separate restraints in the MD simulations (the Ca ligands are restrained to be close to each other). However, even such a procedure often gives quite poor results,[61] as can be seen in Table 6 (O–O and Ca–O restraints) and in Figure 11.

A better description of the Ca sites can be obtained by using a MM potential for the Ca ion and its ligands. This can be automatically obtained from QM calculations on models of the metal sites.[17,57] This gives excellent Ca sites as can be seen in Table 6 (Ca MM). However, because there is little NMR information about these sites, the sites will be entirely determined by the MM potential. A problem with Ca is that it has a flexible coordination: It may be 6–8 coordinate with no clear preference and the bond lengths vary widely.[62] This is very hard to model by standard MM methods, where the number of bonds must be fixed and the equilibrium bond lengths are input parameters.

The is a related complication also for the carboxylate ligands: They may bind to Ca with either one or both of the carboxylate atoms (monodentate and bidentate binding; cf. Figure 9). The two coordination modes have a similar energy and the barrier between them is typically only a few kJ/mole.[63] The best thing that can be done is to test the different binding modes in separate NMR refinements, using different force fields (for mono- or bidentate binding of the two carboxylate ligands) and then to compare the energies and how well the NMR restraints are satisfied. This is illustrated in Table 6, where three different MM force fields for Ca are compared, one with only monodentate ligands, one with the Glu ligand binding bidentately, and the third with the Asp ligand binding bidentately. The results indicate that the latter gives the most favourable energies.

However, a much better strategy (although computationally more demanding) is to treat both the calcium sites with QM methods, using ComQum-N.[57] The results of such refinements are shown in Table 6 and in Figure 12. They show that the general structure has not changed, but the Ca sites are much better defined. In particular we get a more detailed information about the site geometries. For example, out of the 60 ComQum-N structures, 14 are six-coordinate with only monodentate ligands, 5 and 11 are seven-coordinate with either of the two carboxylate groups binding in a bidentate manner, 7 and 12 are six-coordinate with either of the two carboxylate groups binding in a bidentate manner, two are seven-coordinate with both groups binding in a bidentate manner, and 9 sites are five coordinate (the three latter types of structures have one missing ligand). This statistics, in combination with the energy and degree of satisfaction of the NMR restraints, indicate that the calcium sites in EGF34 probably are bidentate a significant amount of the time. Interestingly, the results indicate that it is preferably the Asp ligand that binds bidentately, in contrast to the crystal structure (of a different EFG domain). This shows that the preferential binding of the carboxylate ligands depends on the detailed structure of the metal site.

EXAFS refinement

A third method to obtain structural data for metal sites in proteins is extended X-ray absorption fine structure (EXAFS). The X-ray spectra close to the absorption edge of the core electrons show small oscillations, which are caused by scattering of the photoelectron by the surrounding atoms.[64,65] Therefore, these oscillations contain information about the distance between the metal and the neighbouring atoms, the number of neighbouring atoms at a certain distance, and the atomic number of these atoms.

There are several advantages with EXAFS: It can be used for nearly any element, it is selective (you can study a single metal site in a large protein), and it can be used for non-crystalline samples. In addition, it gives very accurate distance estimates, typically ±0.02 Å,[64] which is ~5 times better than for medium-resolution crystal structures.[16] Naturally, there are also some disadvantages with EXAFS: It cannot distinguish bond lengths that differ by less than ~0.1 Å, the coordination numbers are not determined better than about ±20 %, scattering atoms of similar atomic numbers (e.g., N, O) cannot be distinguished, it gives only an average structure if there are several absorbers in the sample or if it is heterogeneous, and many topological arrangements of the scattering atoms may agree equally well with the EXAFS data.

However, the prime problem with EXAFS is its limited information content, typically restricted to a few atomic distances, which is too little to give a detailed structure (i.e. coordinates of all atoms involved), except for the most simple and symmetric cases. This problem is inherent in the technique, because the EXAFS spectrum typically involves only 20–70 independent degrees of freedom.[64] This is similar to the problem of missing detailed data in X-ray crystallography and NMR structure determination, for which it was solved by inclusion of a MM force field. It may therefore seem somewhat unexpected that this has not been tested for EXAFS. The reason is most likely that there is no general force field for metal sites and that the structural information is so limited in the EXAFS experiments. However, some attempts in this direction have been made. For example, it has been shown that the inclusion of rigid or partly flexible imidazole groups in the refinement of EXAFS spectra improves the fit of histidine ligands in proteins.[66] The reason for the better fit is that multiple-scattering paths (i.e. scattering involving more atoms than the metal and its ligand) provide some angular information about the absorbing site.

The standard EXAFS method is illustrated in Figure 13a. Based on a guess of the structure, a set of scattering amplitudes (f_i) and phase shifts ($_j$) are calculated by a simple QM approach.[67] These are then used in the EXAFS master equation

$$\chi(k)=\sum_j \frac{N_j f_j(k)}{k R_j^2} \mathrm{e}^{-2k^2\sigma_j^2} \sin\left(2kR_j+\delta_j(k)\right) \qquad (8)$$

to obtain the coordination number (N_j) and the distance to each of the scattering atoms (R_j) by a numerical fit to the experimental spectrum. However, the equation includes one additional quantity, the Debye–Waller factors (DWFs, $_i^2$), which describe the structural and thermal disorder of the scattering atoms (cf. the B factors in X-ray crystallography). If the number of scattering atoms are low, they are typically also fitted, but if all neighbouring atoms and paths involving several scattering atoms are considered, they become too many to be fitted, and they have to be calculated by some approach, e.g. from vibrational frequencies obtained by QM calculations.[68,69]

The EXAFS/CC method

We decided to test if it is possible to obtain a full atomic structure by combining EXAFS data with computational chemistry (CC) in the form of a MM or a QM calculation.[69] Thus, EXAFS should give accurate metal–ligand distances, whereas CC should provide the rest of a full atomic structure. In analogy with crystallographic and NMR refinement, we therefore use an energy function of the form

$$E_{EXAFS/CC} = E_{CC} + w_{EXAFS} E_{EXAFS} \qquad (9),$$

where E_{CC} is the standard CC energy function and E_{EXAFS} is a pseudo-energy of the EXAFS fit, representing how well the current model (i.e. the coordinates and DWFs) fit the EXAFS data. There are several widely used quality measures in EXAFS and we have, quite arbitrarily, selected ^{2} as E_{EXAFS}. As before, the weight w_{EXAFS} is necessary because E_{CC} and E_{EXAFS} have different units.

Once the energy function is defined, a standard CC geometry optimization program can be used to obtain the optimum structure. The EXAFS/CC approach is illustrated in Figure 13b. A technical complication is that analytical EXAFS forces are not used in standard EXAFS refinements (in variance to crystallography and NMR refinement). Therefore, the EXAFS forces had to be obtained by numerical differentiation.[69]

Test calculations on model systems

The EXAFS/CC method has been calibrated and tested on five different model complexes for which both accurate EXAFS spectra and crystal structures are available: the highly

symmetric $Ni(NH_2CH_2CH_2NH_2)_3^{2+}$ model ($Ni(en)_3$), a less symmetric $Cu(II)S_2N_2$ complex, Ni(II)tetraphenylporphyrin, and the macrocyclic Cu(biphenyldiiminodithioether) complex in both its Cu(I) and Cu(II) states.[69]

Typical results for the $Ni(en)_3$ complex (Figure 14) are shown in Table 7.[70] The crystal structure of this complex shows six Ni–N distances of 2.12 Å and six Ni–C distances of 2.98 Å. A simple standard EXAFS fit of only the Ni–N distance and the corresponding DWF (as well as the absorption edge energy, E_0) gives slightly longer Ni–N distances of 2.13 Å (Fit 2 in Table 7). Such a fit gives a $^2_{red}$ (a quality factor, presenting how well the structure fits the EXAFS data, but also corrected for the number of fitted variables) of 598. The Ni–N distance returns to 2.12 Å if the Ni–C distance is also fitted (together with the corresponding DWF), but the Ni–C distance is slightly shorter than in the crystal structure, 2.96 Å (Fit 3). At the same time, $^2_{red}$ is reduced to 169, showing that the Ni–C distances strongly influence the spectrum. This is the best result that can be obtained without any structural information.

If we employ the crystal structure and include all paths involving up to four scattering legs and with a half-length of up to 10 Å, fitting the Ni–N and Ni–C distances, as well as three DWFs (one for each of the distances and one for all the multiple-scattering paths), we obtain Ni–N and Ni–C distances of 2.13 and 2.95, and a $^2_{red}$ of 147 (Fit 4). This shows that the multiple-scattering paths also slightly influence the EXAFS spectrum.

Alternatively, and theoretically more satisfactorily, we can employ individual calculated DWFs for all the 52 considered paths (Fit 5). This gives the same distances, but a slightly higher $^2_{red}$ (195), showing that the fit quality is strongly sensitive to the DWFs (but it is also possible that the fit is improved by fitting noise).

A standard DFT calculation also gave good structures, although the bond lengths are 0.03–0.05 Å too long, a typical result for such calculations.[4,9,10,11,12,13,14] However, these small errors have a strong influence on the EXAFS spectrum: $^2_{red}$ becomes as high as 1036 (fitting only E_0). The aim of the EXAFS/CC method is to correct this systematic error and at the same time provide a full structure to the standard EXAFS fit without using any crystal structure.

The QM calculations also gives us the opportunity to estimate how chemically reasonable a structure is by comparing its QM energy with that of a structures optimised with the QM method (E_{QM1}). From Table 7, it can be seen that E_{QM1} is high for both the crystal and EXAFS structures (72 and 79 kJ/mole), because the positions of the hydrogen atoms are poorly determined in the crystal structure. This is typical for experimental structures: If employed directly in QM calculations, they give so high strain energies that they are useless for functional studies (activation and reaction energies are typically of this size). Therefore, experimental structures always need to be relaxed by the QM method before they can be used, with the risk of introducing systematic errors.

This problem is also cured by the EXAFS/CC method. However, we first need to determine a proper value of the w_{EXAFS} weight factor in Eqn. 9. This can be done by varying it over a range of values that give QM and EXAFS forces of a similar size and a proper convergence in the geometry optimisation. We have found[69] that the best results are obtained when the distances are converged to 0.01 Å but the E_{QM1} energy is still low (<10 kJ/mole), because the EXAFS method gives more accurate distances (±0.02 Å[64,65]) than does QM (±0.07 Å[9,10,11,12,13,14]).

The results of the EXAFS/QM method are given in Table 7. If we consider EXAFS forces of only the N atoms (Fit 7), the Ni–N distances become identical to those in the best EXAFS fits, but the Ni–C distance is still slightly too large (2.97 Å). However, the fit is already the best obtained in this investigation ($^2_{red}$ = 118), i.e. appreciably better than in the QM calculations with constraints. The E_{QM1} energy is only 4 kJ/mole. If EXAFS forces are calculated also for the C atoms, the Ni–C distance also reproduce the best EXAFS results and $^2_{red}$ becomes even lower (98), whereas E_{QM1} increases to 7 kJ/mole.

Finally, we have also tested similar calculations with a MM method, using the universal

force field (UFF),[71] which has parameters for all elements. From Table 7, it can be seen that the pure UFF method gives a very poor geometry for the $Ni(en)_3$ complex, with a Ni–N distance of 1.84 Å and $^2_{red}$ of 9600. However, the EXAFS/UFF can fully compensate for this and gives an excellent final structure with $^2_{red}$ = 98, i.e. as good as for EXAFS/QM. On the other hand, E_{QM1} cannot be used to determine if the structure is reasonable (E_{QM1} = 887 kJ/mole), because of the poor UFF structure. Moreover, for some of the other model systems, the UFF structure was so poor that EXAFS/UFF could not obtain any reasonable structure.[69]

On the other hand, the EXAFS/QM method gave excellent results for all the five tested model complexes.[69] Therefore, we can conclude that EXAFS/QM is a promising method for the improvement of EXAFS structures, providing a full atomic structures of the complexes and employing all the available information in the experimental data in an ideal way.

Applications on sitting-atop complexes

A typical application of the EXAFS/QM method is if EXAFS data are available for a complex, but the structure is not known, although there are a number of reasonable alternatives. An example of this is the putative sitting-atop complex of copper porphyrins: The insertion of a metal into a porphyrin molecule in solution is believed to take place by the following steps:[72] deformation of the porphyrin ring, outer-sphere association of the solvated metal ion and the porphyrin, exchange of a solvent molecule with the first porphyrin N atom, chelate-ring closure with the expulsion of more solvent molecules, first deprotonation of a porphyrin N atom, and second deprotonation of the other N atom, which will lead to the formation of the metalloporphyrin. The intermediate formed after the chelate-ring closure is called a *sitting-atop* (SAT) *complex* [14]. Thus, SAT is a complex of a doubly protonated porphyrin ring with a metal ion, where the latter coordinates to two of the porphyrins N atoms. The protons on the other two porphyrin N atoms prohibits the metal ion to go into the centre of the porphyrin plane; instead, it will reside above the ring plane and form bonds to some solvent molecules. This complex has been much debated: SAT complexes of porphyrins with Pt^{2+}, Cu^{2+}, and Rh^{+} have been reported[73] and kinetic evidence indicates that it exists for some other ions.[74] On the other hand, no crystal structure of a SAT complex has been presented.

However recently, SAT complexes of Cu^{2+} with various porphyrins in acetonitrile have been characterised by kinetic measurements, EXAFS and NMR spectroscopy .[75] The data were interpreted as a six-coordinate complex with three kinds of Cu–N interactions, viz. bonds to the porphyrin N atoms (2.05 Å), to the equatorial (1.98 Å), as well as the axial (2.32 Å) acetonitrile N atoms (trans-4 in Figure 15). However, the results has been challenged by other groups, suggesting that other species have been observed, e.g. a porphyrin molecule with four protons, as well as the reactant and product of the reaction (copper–acetonitrile complexes and the copper porphyrin).[76]

We have performed a QM study of SAT complexes of Mg^{2+}, Fe^{2+} and Cu^{2+} with porphine in water and acetonitrile solution, which showed that there are a large number of possible structures for the SAT complex, with 1–5 solvent molecules, one or two metal ion, and cis or trans protonation of the porphyrin ring (some examples are shown in Figure 15).[77,78] A structure with only one bond between the porphyrin and copper and 3–5 acetonitrile molecules (1N-3/4/5 in Figure 15) or a structure with two bonds between the cis-protonated porphyrin and copper and 2–4 acetonitrile molecules (cis-2/3/4 in Figure 15) fitted the experimental EXAFS and NMR data best and were also low in energy, whereas the originally suggested structure[75] was strongly strained and very unlikely. However, we could not exclude that only the reactants ($CuAn_{4/5/6}$ in Figure 15) and products (CuPor or $CuPorAn_2$) were observed.

Therefore, we decided to use the EXAFS raw data to see if we could better discriminate between the various structures.[79] We restricted the investigation to the 15 most probable

structures, shown in Figure 15. First we performed an EXAFS fit directly on the vacuum structure, without fitting anything else than E_0 (we used DWFs calculated from QM vibrational frequencies). This indicated that the cis-2 structure fitted the experimental data best, but also the cis-3, cis-4, trans-3, and 1N-3 structures gave reasonable results (cf. Table 8, first column).

Next we extended the EXAFS fit to also all the Cu–N distances (second column in Table 8). This did not change the results much: The same complexes were still best, with cis-2 on the top. However, if we tried to fit more distances, we soon obtained distances that were chemically unreasonable, indicating a misfit between the EXAFS data and the tested structures.

Therefore, we also tested the EXAFS/QM method, because it allows the structure to reorganise if the data do not fit the structure. These calculations are also shown in Table 8. Some of the structures actually reorganised, but only to other structures in the tested set, indicating that this set was complete. It can be seen that the cis-2 and cis-3 structures gave the best results with $^2_{red} = 1.6–1.7$ and with ΔE_{QM} energies of 1–3 kJ/mole. It can also be seen that there is a quite extensive variation in the optimised distances, which illustrates that there are multiple minima in the EXAFS fits and that the estimated EXAFS accuracy of ±0.02 Å[64] only applies when the correct structure is used for the fit.

It can be seen that the reactants and products ($CuAn_{4/5/6}$ and $CuPorAn_{0/2}$) fitted the EXAFS data quite poorly. Therefore, we also tested a number of mixtures between reactants and products (fitting the composition, as well as E_0 and the Cu–N distances).[79] The best results was obtained for a mixture of 55% $CuAn_4$ and 45% $CuPorAn_2$, which gave a $^2_{red}$ of 2.5, i.e. rather close to the cis-2 structure (the best SAT complex, with $^2_{red} = 2.2$, which was not significantly improved if any reactants were added). Moreover, if also some DWFs were fitted, the two cases gave an identical $^2_{red}$ of 1.6 (Figure 16). Thus, we could still not with certainty conclude whether a SAT complex was observed with Cu^{2+} in acetonitrile or if only reactants and products are observed. However, if there is a SAT complex, it is certainly of the cis-2 or cis-3 type.

Multi-copper oxidases and the EXAFS/QM/MM method

Finally, we have also run an application on a metal site in a protein.[80] Of course, a whole protein is too large to treat by QM methods. Therefore, we had to develop a QM/MM variant of the EXAFS/QM technique, viz. EXAFS/QM/MM with the energy function:

$$E_{EXAFS/QM/MM} = E_{QM} + E_{MM12} - E_{MM1} + w_{EXAFS} E_{EXAFS} \quad (10),$$

in which E_{CC} in Eqn. 9 is replaced by the QM/MM energy function in Eqn. 2.

This method was applied to two intermediates in the reaction cycle of the multi-copper oxidases (MCO). The MCOs are a group of enzymes that couple the four-electron reduction of molecular oxygen to water with four one-electron oxidations of various substrates.[81] Typical examples are laccase, ascorbate oxidase, ceruloplasmin, and CueO.[82,83] The MCOs contain at least four copper ions: One isolated type 1 Cu near the substrate-binding site and a trinuclear cluster, consisting of a type 2 Cu ion (Cu2) and a pair of type 3 (T3) Cu ions in a triangular arrangement. Crystal structures of several MCOs are known.[82,83] They show that Cu2 is bound to two His ligands and a solvent molecule, whereas the two T3 Cu ions are bound to three His ligands each (cf. Figure 17). Depending on the oxidation state of the Cu ions, there may be a bridging ligand between the two T3 Cu ions, a solvent molecule, or perhaps a derivative of the O_2 substrate.

Extensive kinetic and spectroscopic studies have been performed to deduce the mechanism of the MCOs.[84,85] Four distinct states have been identified in the mechanism: the fully oxidized resting state, a fully reduced state, the peroxy intermediate (PI), and the native intermediate (NI). The PI arises after the binding of O_2 to the reduced state, leading to the two-electron reduction of O_2 to the peroxide level. In the NI, O_2 is reduced by two additional

electrons (to the level of two water molecules) and all Cu ions are oxidized.

It is clear that O_2 binds to the trinuclear cluster, but the details of the binding are still unclear. On the basis of the spectroscopic evidence, two possible structures have been suggested for the PI: One with O_2^{2-} bridged to the three coppers in the centre of the cluster (C_3), and one with HO_2^- bridged to two coppers on the side of the cluster (S_{23}) and with an additional OH^- bridging the two T3 coppers.[84] In a recent QM/MM study, we have refined these two structures (Figure 17).[86] Energetic results indicated that the C_3 structure was more stable, but the multiplicity of the electronic ground state was not in accordance with experimental data. Both structures fitted EXAFS data equally well: EXAFS indicates that there is a short Cu–Cu interaction of 3.4 Å in the PI[84] and the best computational structures of the C_3 and S_{23} states had Cu–Cu distances of 3.45–3.57 and 3.42–3.53 Å, respectively.

Therefore, we decided to test also our new EXAFS/QM(/MM) methods. For technical reasons, we have used the EXAFS data from the peroxy adduct (PA), which is the complex of H_2O_2 with the fully oxidized resting state of the protein.[81,87] Spectroscopic and computational studies have shown that the structures of the PI and the PA are very similar, although the T2 Cu is reduced in the former, but oxidized in the latter.[84,87]

To start with, we optimized the two putative structures of PA in vacuum with QM and made a standard EXAFS fit, in which only E_0 was free to vary. The results of these calculations are collected in Table 9 and show that the C_3 model fits the EXAFS data better than the S_{23} model. If three Cu–ligand distances were also fitted (QM+EXAFS fit), the difference between the two structures is reduced (the shortest Cu–Cu distance is 3.41–3.42 Å in both structures), but the C_3 structure is still better. The EXAFS/QM results show that the two structures give a similar Cu–Cu distance (3.42–3.44 Å). They are both strongly improved by the optimization, but the C_3 structure is still best. The E_{QM1} energies are 4 and 11 kJ/mole, respectively.

Next, we made a similar investigation of our QM/MM structures.[86] Interestingly, the QM/MM structures fitted the EXAFS data quite poorly and the C_3 structure gave worse results than the S_{23} structure, although the shortest Cu–Cu distance, 3.52 Å, is closer to 3.4 Å than that in the S_{23} structure, 3.72 Å. This illustrates the problem of comparing only distances in QM calculations with those obtained in EXAFS fits. However, a standard EXAFS fit of three distances (QM/MM+EXAFS fit) made the C_3 structure better than the S_{23} structure again.

Finally, the EXAFS/QM/MM approach (last two rows in Table 9) shows that the C_3 structure is strongly improved, much more than the S_{23} structure ($^2_{red}$ = 38 and 74). The resulting structure has a Cu–Cu distance of 3.41 Å, i.e. close to what was found in all the other structures that included EXAFS data in the fits. However, the other two Cu–Cu distances are quite far from what was obtained in the EXAFS/QM structure. The E_{QM1} energies are 10 and 5 kJ/mole, respectively. This shows that both the EXAFS and QM energies are insensitive to those distances. Thus, we can conclude that the PA most likely has O_2^{2-} in the centre of the cluster and that the EXAFS/QM/MM approach works properly.

Concluding remarks

Nobody believes in theoretical calculations,
except the one who did them.
Everybody believes in experimental results,
except the one who did them.

This is a well-known quotation, sometimes attributed to Albert Einstein. However, we have seen in this review that, at least for structural methods, the distinction between experimental and theoretical methods is less clear today. In fact, nearly all "experimental" crystal and NMR structures are actually to a great part theoretical, and the quality of the structures strongly depends on the calculations. In particular, substrates, inhibitors, and metal

sites may have a lower accuracy than the protein, because there are no accurate force fields for these sites.

We have seen that this problem can be solved by using QM methods for the interesting parts of the protein. Thereby, we can actually improve the crystal structures locally. Moreover, we can interpret the structures, i.e. determine what atoms are actually seen. In particular we can often determine the position of protons and electrons (i.e. the protonation and oxidation states) in the structures. Likewise QM calculations can be used in NMR structure refinement to obtain detailed structures also for parts of the structure for which little experimental information is available. Finally, we have shown that computational chemistry, in the form of either QM or MM calculations, can be used in EXAFS refinement to supplement the experimental data so that a full atomic structure can be obtained, rather than only a set of metal-ligand distances.

Naturally, there are alternatives to the presented approaches. For example, Merz and coworkers have showed that linear-scaling semiempirical methods can be used instead of MM for X-ray structure refinement.[88] Unfortunately, semiempirical methods are quite inaccurate, with much larger systematic errors than DFT methods. Therefore, it is not clear that semiempirical methods actually will improve the structure – on the contrary, for the protein residues, it is likely that the standard MM force field will perform better than semiempirical methods, at least when electrostatic effects are small.

Alternatively, QM calculations can be used to construct a MM force field. This is a quite tedious task, if an accurate force field is wanted.[89] However, for more approximate force fields, automatic methods are available that extract a force field directly from a frequency calculation obtained at any CC level.[17] Such an approach has successfully been used both for X-ray crystallography and NMR refinement.[17,57] It has the advantage of speed, once the force field is constructed, but it is appreciably less accurate than methods involving explicit calculations with QM. Moreover, the parametrisation needs to be redone every time a new system is studied and the frequency calculations have higher demands on computer memory and time than a normal geometry optimisation. In addition, there are several severe limitations of the MM methods (e.g. that the coordination number of an atom cannot change during the simulation), which are not present for QM methods. Therefore, the direct QM methods are normally preferable, except when many calculations on the same system are needed.

Another use of the combinations of experimental and QM methods is for QM investigations of proteins. In fact, the methods were originally developed for this purpose: We realised that all our protein QM projects started with an reoptimisation of a crystal structure (because, the crystal structures cannot be used directly, owing to systematic errors in both the experimental and theoretical methods, as we saw above). We decided to try to design a method that is guaranteed to stay close to the experimental data, and the result was the quantum refinement, which provides an ideal compromise between the experimental and QM data (it is close to the crystallographic raw data, but still compatible with QM calculations).

By these methods, structural data can directly be obtained. However, accurate energies, which in principle are more interesting, are much harder to obtain.[14,40,41] The reason for this is that structures are local properties, which depend only on the nearest neighbours of the site of interest. On the other hand, energies are global, which depend on the position and interaction of all atoms in the system (the energy will change significantly if a single hydrogen bond between two water molecules far from the site of interest forms or breaks). Moreover, long-range electrostatics and solvent effects are hard to treat in an accurate and consistent way. Therefore, specialised methods are needed to obtain accurate (free) energies in proteins, preferably based on some type of conformational sampling.[90,91]

Of course, the presented methods still can be improved. The most important short-coming is the ignorance of electrostatic interactions in ComQum-X. These interactions are also ignored in standard crystallographic refinement, because they normally do not improve the structure,[22] but also because they depend on the positions of the hydrogen atoms, which are

not discerned in the crystal structures.[29] Thus, if we want to consider electrostatic interactions

1 P. E. M. Siegbahn and M. R. A. Blomberg, Annu. Rev. Phys. Chem., 1999, 50, 221–249; P. E. M. Siegbahn and M. R. A. Blomberg, Chem. Rev., 2000, 100, 421–437.
2 A. Szabo and N. S. Ostlund, Modern Quantum Chemistry, Dover publications, Inc, Mineola, 1996.
3 C. J. Cramer, Essentials of Computational Chemistry, J. Wiley & Sons, Ltd., Chichester, 2002.
4 F. Neese, J. Biol. Inorg. Chem. 2006, 11, 702-711.
5 M. Sulpizi, J. Vondele, R. Ayala, M. Sprik, in press.; T. Ishida, D. G. Fedorov, K. Kitaura, J. Phys. Chem. B, 2006, 110, 1457-1463; P. Canfield, M. G. Dahlbom, N. S. Hush, and J. R. Reimers, J. Chem. Phys., 2006, 124, 024301.
6 E. I. Solomon, Szilagy, R. K., DeBeer George, S., Basumallick, L. Chem. Rev. 2004, 104, 419-458; Rode M. F., Werner, H.-J. Theor. Chim. Acc. 2005, 114, 309-317; Cramer, C. J., Wloch, M., Pieuch, P., Puzzarini, C., Gagliardi, L. J. Phys. Chem. A, 2006, 110, 1991-2004; .K. P. Jensen and U. Ryde, J. Phys. Chem. B, 2003, 107, 7539-7545.
7 U. Ryde, Y.-W. Hsiao, L. Rulíšek and E. I. Solomon, J. Am. Chem. Soc., submitted.
8 C. J. Cramer, Essentials of Computational chemistry, J. Wiley & Sons, Chichester, 2002, pp. 265-268.
9 E. Sigfridsson, M. H. M. Olsson and U. Ryde, J. Phys. Chem. B, 2001, 105, 5546-5552.
10 E. Sigfridsson and U. Ryde, J. Inorg. Biochem., 2002, 91, 101-115.
11 U. Ryde and K. Nilsson, J. Am. Chem. Soc., 2003, 125, 14232-14233.
12 M. H. M. Olsson and U. Ryde, J. Am. Chem. Soc., 123 2001, 7866-7876.
13 Y. Shen and U. Ryde, J. Inorg. Biochem., 2004, 98, 878-895.
14 P. Söderhjelm and U. Ryde, J. Mol. Struct. Theochem, 2006, in press.
15 K. Pierloot, J. O. A. De Kerpel, U. Ryde, M. H. M. Olsson and B. O. Roos 1998, J. Am. Chem. Soc., 120, 13156-13166.
16 B. A. Fields, H. H. Bartsch, H. D. Bartunik, F. Cordes, J. M. Guss and H. C. Freeman, Acta Cryst., D50 1994, 709-730; D. W. J. Cruickshank, Acta Cryst. D, 1999, 55, 583-601.
17 K. Nilsson, D. Lecerof, E. Sigfridsson and U. Ryde, Acta Cryst. D, 2003, 59, 274-289.
18 J. Boström, Norrby P-O, Liljefors T, J. Comp.-Aided Mol. Design., 1998, 12, 383-396; A. Warshel, Computer modeling of chemical reactions in enzymes and solution, J. Wiley & Sons, Inc., Chichester, 1991, pp.153-168; U. Ryde, in Recent Research Developments in Protein Engineering, 2, pp. 65-91; Research Signpost, Trivandrum, 2002.
19 R. A. Engh and R. Huber, Acta Cryst. A, 1991, 47, 392-400.
20 A. J. Mulholland, in *Theoretical biochemistry – processes and properties of biological systems*, Theoretical and computational chemistry, Vol. 9. L. A. Eriksson, ed. Amsterdam: Elsevier Science, 2001, 597-653; U. Ryde, 2003, Curr. Opin. Chem. Biol., 7, 136-142.
21 M. Svensson, Humbel, S., Froese, R. D. J., Matsubara, T., Sieber, S., Morokuma, K. *J. Phys. Chem.* **1996**, *100*, 19357-19363; U. Ryde, J. Comput.-Aided Mol Design, 1996, 10, 153-164.
22 G. J. Kleywegt and T. A. Jones, Meth. Enz. 1997, 277, 208-230; A. T. Brünger and L. M. Rice, Meth. Enz. 1997, 277, 243-269; G. Rhodes, Crystallography made crystal celar, Academic Press, London, 2000.
23 N. S. Pannu and R. J. Reed, Acta Cryst. A, 1996, 52, 659-668.
24 A. T. Brünger, Acta. Cryst. D, 1993, 49, 24-36; G. J. Kleywegt and T. A. Jones, Structure, 3,1995, 535-540.
25 A. Jack and M Levitt, Acta Cryst. A, 1978, 34, 931; Brünger, A. T., Karplus, M., Petsko, G. A. Acta Cryst 1989, A45, 50; Brünger, A. T. Methods Enzymol 1997, 277, 243-269.
26 F. Mancia, G. A. Smith and P. R. Evans, Biochemistry, 1999, 38, 7999-8005.
27 G. J. Kleywegt and T. A. Jones, Acta. Cryst. 1998, D54, 1119-1131.
28 W. D. Cornell, P. Cieplak, Ci. I. Bayly, I. R. Gould, K. M. Merz, D. M. Ferguson, D. C. Spellmeyer, T. Fox, J. W. Caldwell and P. A. Kollman, J. Am. Chem. Soc. 1995, 117, 5179-5197.
29 U. Ryde, L. Olsen and K. Nilsson, J. Comp. Chem., 2002, 23, 1058-1070.
30 D. Treutler and R. Ahlrichs, J. Chem. Phys., 1995, 102, 346.
31 A. T. Brunger, P. D. Adams, G. M. Clore, W. L. Delano, P. Gros, R. W. Grosse-Kunstleve, J.-S. Jiang, J. I. Kuszewsk, M. Nilges, N. S. Pannu, R. J. Read, L. M. Rice, T. Simonson and G. L. Warren, Crystallography & NMR System CNS, Version 1.0, Yale University, 2000.
32 S. Benini, A. González, W. R. Rypniewski, K. S. Wilson, J. J. Van Beeumen and S. Ciurli, Biochemistry, 2000, 39, 13115-13126.
33 G. Pettersson, CRC Crit. Rev. Biochem., 21, 1987, 349-389.
34 B. J. Bahnson, T. D. Colby, J. K. Chin, B. M. Goldstein and J. P. Klinman, Proc. Natl. Acad. Sci. USA, 1997, 94, 12797-12802.
35 K. Nilsson and U. Ryde, J. Inorg. Biochem., 2004, 98, 1539-1546.
36 J. K. Rubach, S. Ramaswamy and B. V. Plapp, Biochemistry, 2001, 40, 12686-12694.
37 K. Nilsson, "Quantum chemical interpretation of protein crystal structures", Ph.D. Thesis, Lund University, 2003.
38 L. Rulíšek, U. Ryde, J. Phys. Chem. B, 2006, 110, 11511-11518.
39 U. Ryde and Nilsson, K. *J. Mol. Struct.* **2003**, *632*, 259-275.
40 K. Nilsson, H.-P. Hersleth, T. H. Rod, K. K. Andersson and U. Ryde, Biophys. J., 2004, 87, 3437-3447.

in the refinement, we need to speculate on the position of the hydrogen atoms. Most of these positions are well-determined from the positions of the heavy atoms, but for some residues (e.g. Ser, Thr, Tyr), there are some uncertainty in the positions, for water molecules the first protons can be put anywhere on a sphere around the oxygen atom, and for the His residues, it is not even known how many protons that should be added (one or two) or where to put it (on N^{1} or N^{2}). Even if this problem also arises in any QM(/MM) study of a protein and therefore there are standard solutions of the problem, there is a significant risk that we will introduce errors in the structure by adding all the invisible protons.

On the other hand, it is quite clear that there are problems in the quantum refined structures that are cause by the ignorance of electrostatic interactions, e.g. that the charged propionate side chains of porphyrins tend to curl back towards the porphyrin ring.[29] The reason for this is that electrostatics interactions are explicitly considered in the QM calculations. If they are missing in the MM calculations, this means that peripheral atoms in the QM system will try to interact with other atoms in the QM system, instead of forming their real interactions with the surrounding protein.

We have made some preliminary attempts to include electrostatics in the quantum refinement by adding a point-charge model of the surrounding protein in the QM calculations (as in standard QM/MM methods).[14,41] However, this never led to any improvement in the structures. Instead, the results indicate an exaggerated effect of the surrounding protein (i.e. that the point charges are too large). The reason for this is most likely that the system is not properly solvated (solvation tend to screen the charges). However, it remains to solve this problem in a consistent and automatic way. Similar problems apply to NMR refinements.[57]

Finally, we note that the present methods easily can be combined. For example, it is possible to combined both X-ray crystallographic (or NMR) and EXAFS data with QM/MM methods to obtain accurate metalloprotein structures. Such an approach would be highly interesting for several systems, e.g. for the oxygen-evolving complex in photosystem II, for which only low-resolution crystal structures are available,[92] but also accurate EXAFS data.[93] Thus, we still expect some interesting development of these versatile combinations of experimental and QM methods for obtaining structural information of proteins.

Acknowledgements

This investigation has been supported by grants from the Swedish research council and by the Crafoord Foundation. It has also been supported by computer resources of Lunarc at Lund University.

References

41 N. Källrot, K. Nilsson, T. Rasmussen and U. Ryde,Intern. J. Quant. Chem., 2005, 102, 520-541.

42 B. F. Anderson, Edwards RA, Whittaker MM, Whittaker JW, Baker EN, Jameson GB, to be published, PDB files 1ix9 and 1ixb.

43 M. E. Stroupe, DiDonato M, Tainer JA In Handbook of Metalloproteins, Messerschmidt A, Huber R, Wieghardt K, Poulos T, Eds, John Wiley & Sons, Chichester, 2001, pp 941-951; Chance B, Sies H, Boveris A, Physiol Rev 1979, 59, 527-605.

44 W. C. Stallings, Metzger AL, Pattridge KA, Fee JA, Ludwig ML, Free Rad Res. Commun., 1991, 12-13, 259-268; MM Whittaker, CA Ekberg, RA Edwards, EN Baker, GB Jameson, JW Whittaker, J. Phys. Chem. B, 1998, 102, 4668-4677; Miller A-F, Padmakumar K, Sorkin DL, Karapetian A, Vance CK, J. Inorg. Biochem., 92003, 3, 71-83.

45 C. L. Fisher, Chen J-L, Li J, Bashford D, Noodleman L, J. Phys. Chem., 1996, 100, 13498-13505; Li J, Fisher CL, Konecny R, Bashford D, Noodleman L, Inorg. Chem., 1999, 38, 929-939; Han W-G, Lovell T, Noodleman L, Inorg. Chem., 2002, 41, 205-218.

46 T. A. Jones, Zou, J.-Y., Cowand, W. E., Kjeldgaard, M. *Acta Cryst. A,* **1991**, *47*, 110-119.

47 C. M. Wilmot, Sjögren, T., Carlsson, G. H., Berglund, G. I., Hajdu, J. *Methods Enzymol.* **2002**, *353*, 301-318.

48 J. W. Whittaker, Whittaker, M. M. *J. Am. Chem. Soc.* **2002**, *113*, 5528-5540; Jackson, T. A., Xie, J, Yikilmaz,

E, Miller, A.-F., Brunold, T. C. *J. Am. Chem. Soc.* **2002**, *124*, 10833-10845.
49 P. M. Vignais, B. Billoud and J. Meyer, FEMS Microbiol Rev. 25, 2001, 455-501.
50 A. Volbeda, Martin L, Cavazza C, Matho M, Faber BW, Rosenboom W, Albracht SPJ, Garcin E, Rousset M, Fontecilla-Camps JC, J. Biol. Inorg. Chem., 2005, 10, 239-249; H. Ogata, S. Hirota, A. Nakahara, H. Komori, N. Shibata, T. Kato, K. Kano, Y. Higuchi, Structure, 2005, 13, 1635-1642.
51 A. Volbeda, Montet, Y., Vernède, X., Hatchikian, E. C., Fontecilla-Camps, J. C. *Intern. J. Hydrogen Energy* **2002**, *27*, 1449-1461.
52 E. Sigfridsson and U. Ryde, J. Inorg. Biochem., 2002, 91, 116-124.
53 G. S. Kachalova, A. N. Popov and H. D. Bartunik, Science 1999, 284, 473-476; J. Vojtechovský, K. Chu, J. Berendzen, R. M. Sweet and I. Schlichting, Biopys. J., 1999, 77, 2153-2174.
54 G. Fogarasi, X. Zhou, P. W. Taylor and P. Pulay, J. Am. Chem. Soc.1992, 114, 8191-8201.
55 R. Salzmann, M. T. McMahon, N. Godbout, L. K. Sanders, M. Wojdelski and E. Oldfield, J. Am. Chem. Soc. 1999, 121, 3818-3828.
56 J. Cavanagh, W. J. Fairbrother, A. G. Palmer and N. J, Skelton, Protein NMR spectroscopy. Principles and Practice, Academic Press, London, 1996.
57 Y.-W. Hsiao, T. Drakenberg and U. Ryde, J. Biomol. NMR, 2005, 31, 97-114.
58 Y. Stenberg, S. Linse, T. Drakenberg and J. Stenflo J. Biol. Chem. 1997, 272, 5546-5552; Y. Stenberg, A. Muranyi, C. Steen, E. Thulin, T. Drakenberg and J. Stenflo J. Mol. Biol. 1997, 294, 653-665.
59 J. Stenflo and Y. Stenberg and A. Muranyi, Biochim. Biophys. Acta, 2000, 1477, 51-63.
60 Z. Rao and P. Handford and M. Mayhew and V. Knott and G. G. Brownlee and D. Stuart, Cell, 1995, 82, 131-141.
61 A. K. Downing, V. Knott, J. M. Werner, C. M. Cardy, L. D. Campbell and P. A. Handford Cell, 1996, 85, 597-605; S. Saha, J. Boyd, J. M. Werner, V. Knott, P. A. Handford, L. D. Campbell and A. K Downing Structure, 2001, 9, 451-456; H. Tossavainen, P. Permi, A. Annila, A. Kilpeläinen and T. Drakenberg Eur. J. Biochem. 2003, 270, 2505-2512; X. Wang and M. X. Li and L. Spyracopoulos, N. Beier, M. Chandra, R. J. Solaro and B. D. Syske J. Biol. Chem. 2001, 276, 25456-25466.
62 J. J. R. Frausto da Silva and R. J. P. William, The biological chemistry of the element, Clarondon, Oxford, 1991, pp. 182-184.
63 M. Torrent and D. G. Musaev and K. Morokuma, J. Phys. Chem. B 2001, 105, 322-327; U. Ryde, Biophys. J., 1999, 77, 2777-2787.
64 J. E. Penner-Hahn, Coord. Chem. Rev. 1999, 190-192, 1101-1123.
65 R. A. Scott, Meth. Enzym. 1985, 117, 414-459.
66 M. S. Co, R. A. Scott and K. O. Hodgson, J. Am. Chem. Soc. 1981, 103, 986-988; R. W. Strange, N. J. Blackburn, P. F. Knowles and S. S. Hasnain, J. Am. Chem. Soc. 1987, 109, 7157-7162; S. Wang, M. H. Lee, R. P. Ausinge, P. A. Clark, D. E. Wilcox and R. A. Scott, Inorg. Chem. 1994, 33, 1589-1593.
67 A. Ankudinov, C. Bouldin, J. J. Rehr, J. Sims, and H. Hung, Phys. Rev. B, 2002, 65, 104107.
68 A. V. Poiarkova, X-ray absorption spectroscopy fine structure Debye–Waller factors, Ph.D. Thesis, University of Washington, 1999.
69 Y.-W. Hsiao, Y. Tao, J. E. Shokes, R. A. Scott and U. Ryde, Phys, Rev. B, 2006, in press.
70 J. D. Korp, I. Bernal and R. A. Palmer, Acta Cryst. B, 1980, 36, 560-565.
71 A. K. Rappé, C. J. Casewit, K. S. Colwell, W. A. Goddard and W. M. Skiff, J. Am. Chem. Soc. 1992, 114, 10024-10038.
72 P. Hambright, J. Am. Chem. Soc., 1974, 96, 3123-3131; D.K. Lavallee, Coord. Chem. Rev. 1985, 61, 55-96; S. Funahashi, Inada, Y. Inamo, M. Analyt. Sci. 2001, 17, 917-927.
73 J. P. Marcqet, T. Theophahanides, Can. J. Chem. 1973, 51, 219; K. Letts, R.A. Mackay, Inorg. Chem. 14, 1975, 2993; E.B. Fleischer, F. Dixon, Bioinorg. Chem. 1977, 7, 129.
74 T. P. G. Sutter, P. Hambright, Inorg. Chem. 1992, 31, 5089-5093; Y. Inada, Y. Sugimoto, Y.Nakano, S. Funahashi, Chem. Lett., 1996, 881-882; J. Takeda, T. Ohya, M. Sato, Inorg. Chem. 1992, 31, 2877-2880; S. Funahashi, Y. Yamaguchi, M. Tanka, Bull. Chem. Soc. Jpn, 1984, 57, 204-208.
75 Y. Inada, Y. Sugimoto, Y. Nakano, Y. Itoh, S. Funahashi, Inorg. Chem. 1998, 37, 5519-5526; M. Inamo, N. Kamiya, Y. Inada, M. Nomura, S. Funahashi Inorg. Chem. 2001, 40, 5636-5644; Y. Inada, Y. Nakano, M. Inamo, M. Nomura, S. Funahashi Inorg. Chem. 2001, 39, 4793-4801.
76 C.-H. Tsai, J.-Y. Tung, J.-H. Chen, F.-L. Liao, S.-L.Wang, S.-S.Wang, L.-P. Hwang, C.-B. Chen, Polyhedron, 2000, 19, 633-639.
77 Y. Shen and U. Ryde, J. Inorg. Biochem., 2004, 98, 878-895.
78 Y. Shen and U. Ryde, Chem. Eur. J., 2005, 11, 1549-1564.
79 Y.-W. Hsiao and U. Ryde, Inorg. Chim. Acta, 2005, 359, 1081-1092.
80 U. Ryde, Y.-W. Hisao, L. Rulíšek and E. I. Solomon, J. Am. Chem. Soc., submitted.
81 E. I. Solomon., Sundaram, U. M., Machonkin, T. E. *Chem. Rev.* **1996**, *96*, 2563-2605.
82 A. Messerschmidt in *Handbook of Metalloproteins*; Messerschmidt, A., Huber, R., Wieghardt, K., Poulos, T, Eds.; Wiley, New York, 2001, pp. 1345-1357; G. J, Davies and V. Ducros, In *Handbook of Metalloproteins*; Messerschmidt, A., Huber, R., Wieghardt, K., Poulos, T, Eds.; Wiley, New York, 2001, pp. 1359-1368; P. F.

Lindley, In *Handbook of Metalloproteins*; Messerschmidt, A., Huber, R., Wieghardt, K., Poulos, T, Eds.; Wiley, New York, 2001, pp. 1369-1379.
83 S. A. Roberts, Weichsel, A., Grass, G., Thakali, K, Hazzard, J. T., Tollin, G., Rensing, C., Montfort, W. R. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 2766-2771.
84 W. Shin, Sundaram, U. M., Cole, J. L., Zhang, H. H., Hedman, B, Hodgson, K. O., Solomon, E. I. *J. Am. Chem. Soc.* **1996**, *118*, 3202-3215.
85 A. E. Palmer, Lee, S.-K., Solomon, E. I. *J. Am. Chem. Soc.* **2001**, *123*, 6591-6599.; Palmer, A. E., Quintanar, L., Severance, S., Wang, T.-PI., Kosman, D. J., Solomon, E. I. *Biochemistry* **2002**, *41*, 6438-6448.; Lee, S.-K., George, S. D., Antholine, W. E., Hedman, B, Hodgson, K. O., Solomon, E. I. *J. Am. Chem. Soc.* **2002**, *124*, 6180-6193.; Torres, J., Svistunenko, D., Karlsson, B., Cooper, C. E., Wilson, M. T. *J. Am. Chem. Soc.* **2002**, *124*, 963-967.; Solomon, E. I., Chen, PI., Metz, M., Lee, S.-K., Palmer, A E. *Angew. Chem. Int. Ed.* **2001**, *40*, 4570-4590.
86 L. Rulíšek, Solomon, E. I., Ryde, U. *Inorg. Chem.* **2005**, *44*, 5612-5628.
87 U. M. Sundaram, Zhang, H. H., Hedman, B., Hodgson, K. O., Solomon, E. I. *J. Am. Chem. Soc.* **1997**, *119*, 12525-12540.
88 N. Yu, Yennawar, H. P., Merz, K. M. Acta Crystallogr. D 2005, 61, 322; N. Yu, S. A. Hayik, B. Wang, N. Liao, C. H. Reynolds, K. M. Merz, J. Chem. Theory Comput. 2006, 2, 1057-1069.
89 P.-O. Norrby and T. Liljefors, J. Comput. Chem. 1998, 19, 1146-1166.
90 A. Warshel, Computer modeling of chemical reactions in Enzymes and solutions, J. Wiley & Sons, Inc. Chichester, 1991; D. Riccardi, P. Schaefer, Q. Cui J. Phys. Chem. B 2005, 109, 17715-17733.
91 T. H. Rod and U. Ryde, Phys. Rev. Lett., 2005, 94, 138302; T. H. Rod and U. Ryde, J. Chem. Theory Comput., 2005, 1, 1240-1251.
92 K. N. Ferreira, T. M. Iverson, K. Maghlaoui, J. Barber and S. Iwata Science 2004, 303, 1831-1838; B. Loll, J. Kern, W. Saenger, A. Zoundi and J. Biesiadka Nature 2005, 438, 1040-1044.
93 J. Yano, Y. Pushkar, P. Glatzel, A. Lewis, K. Sauer, J. Messinger, U. Bergman and V. Yachandra J. Am. Chem. Soc. 2006, 127, 14974-14975.

Table 1. Errors in metal–ligand bond lengths of two DFT methods compared to crystal structures of small inorganic complexes or atomic-resolution crystal structures. The methods are either B3LYP or Becke–Perdew86, both with the 6-31G* basis set.

Model	Method	Error in distance (Å)		
Cytochrome models[9,10,11]		$Fe–N_{Por}$	Fe–Ax1	Fe–Ax2
$Fe^{II}(SR_2)_2$	B3LYP	0.02	0.06	0.06
	BP86	0.01	-0.02	-0.02
$Fe^{III}(SR_2)_2$	B3LYP	0.04	0.07	0.05
	BP86	0.03	0.03	0.01
Fe^{II}(imidazole)$_2$	B3LYP	0.02	0.04	0.04
	BP86	0.01	0.02	0.02
Fe^{III}(imidazole)$_2$	B3LYP	0.01–0.02	0.06	0.04
	BP86	0.01–0.02	-0.03	0.00
Fe^{II}(His)(Cys)	B3LYP			0.03
Fe^{II}(His)(CO)	B3LYP	0.01	0.03	0.06
	BP86	-0.01	0.02	0.01
Fe^{III}(His)(Met)	B3LYP	0.01–0.04	0.01	0.10
	BP86	0.01–0.03	0.00	0.00
Cu_A model[12]		Cu–Cu	Cu–S	Cu–N
	B3LYP	0.11–0.14	0.05–0.07	0.08
Cu models[13]		Cu–N	Cu–N(ax)	
Cu(porphine)	BP86	0.00		
Cu^{+}(acetonitrile)$_4$	BP86	-0.02		
Cu^{2+}(acetonitrile)$_6$	BP86	0.01	-0.01 to -0.02	
Hydrogenase model[14]		Ni–Ni	Ni–S	Ni–N
	BP86	0.05	0.03	0.05

Table 2. Fe–ligand distances, strain energies (E_{QM1}, kJ/mole, i.e. the difference in energy of the quantum system optimised in vacuum or with ComQum-X), and *R* factors for the haem group in cytochrome c_{553} calculated with ComQum-X using the 1.7 Å (low-resolution) data.[11] For comparison, the low- and high-resolution crystal structures[32] and the result of a QM vacuum calculation are also included. The quantum systems consisted of Fe^{III}(porphyrin)(imidazole)($S(CH_3)_2$) in the low-spin doublet state, and it was studied with the density functional Becke–Perdew86/6-31G* method.[11]

Structure	Distance to Fe (Å)			E_{QM1}	R_{low}	R_{high}
	N_{His}	S_{Met}	N_{Por}			
Low-resolution	2.31	2.21	2.02–2.08		0.0000	0.0000
High-resolution	1.99	2.33	1.97–2.00			
ComQum-X	1.99	2.31	1.99	51	-0.0055	-0.0057
Vacuum	1.99	2.33	2.00–2.01	0		

Table 3. Zn–ligand distances, strain energies (E_{QM1}, kJ/mole, i.e. the difference in energy of the quantum system optimised in vacuum or with ComQum-X), and the R_{free} factor for the catalytic zinc ion in alcohol dehydrogenase in complex with NAD^+ and trifluoroethanol (the ligands are histidine, two cysteines, and the alcohol) calculated with ComQum-X using different values of the relative weight of the crystallographic and theoretical energy terms (w_A, defined in Eqn. 5).[35] For comparison, the original crystal structure (at 2.0 Å resolution)[34] and the result of QM vacuum optimisations are also included. The QM calculations included Zn(imidazole)$(SCH_3)_2(H_2O/OH)$ and used the Becke–Perdew86/6-31G* method.

Ligand	w_A	Distance to Zn (Å)			E_{QM1}	R_{free}
		N	S	O		
ROH	Vacuum	2.09	2.25–2.29	2.29	0	
RO^-	Vacuum	2.24	2.33	1.93	0	
ROH	10	2.18	2.33–2.37	2.01	119	0.2283
ROH	3	2.15	2.31–2.34	2.04	93	0.2283
ROH	1.76	2.14	2.30–2.32	2.07	87	0.2285
RO^-	3	2.28	2.36	1.90	62	0.2280
RO^-	1.85	2.28	2.34–2.36	1.90	54	0.2280
RO^-	0.3	2.25	2.31–2.33	1.92	54	0.2279
	Crystal	2.13–2.20	2.13–2.29	2.05–2.07		0.2303

Table 4. The results of quantum refinements on manganese superoxide dismutase.[38] E_{QM1} is the energy difference (in kJ/mole) between structures optimised in the protein and in vacuum. *r* is the sum of the unsigned difference in the Mn–ligand distances (in Å) between the quantum refined structure and the optimised vacuum structure or the original crystal structure. Ligand *R* is the real-space *R* factor[46] of the solvent ligand.

Mn state	Ligand	R_{free}	*R*	Ligand *R*	E_{QM1}	*r* vacuum	*r* crystal	Distance (in Å) from Mn to N_{His1}	N_{His2}	N_{His3}	O_{Asp}	O_{Sol}
Reduced structure												
II	H_2O	0.2155	0.2078	0.051	33.7	0.19	0.05	2.18	2.13	2.16	2.04	2.27
II	OH^-	0.2157	0.2079	0.056	93.5	0.62	0.16	2.19	2.14	2.16	2.06	2.15
1ixb[42]	subunit 1	0.2158	0.2077					2.17	2.13	2.14	2.05	2.27
	subunit 2							2.18	2.15	2.13	2.04	2.26
Oxidised structure												
III	H_2O	0.2128	0.2081	0.052	54.4	0.35	0.25	2.12	2.12	2.10	2.03	1.96
III	OH^-	0.2130	0.2082	0.057	43.2	0.18	0.32	2.13	2.13	2.11	1.99	1.87
II	H_2O	0.2122	0.2080	0.041	40.0	0.24	0.11	2.16	2.14	2.12	2.03	2.18
II	OH^-	0.2125	0.2079	0.041	74.3	0.56	0.12	2.17	2.15	2.13	2.05	2.05
1ix9[42]	subunit 1	0.2125	0.2079					2.15	2.14	2.12	2.02	2.12
	subunit 2							2.14	2.14	2.12	2.03	2.16

Table 5. Geometric parameters (distances in Å, angle in degrees) of the ComQum-X structure of CO–myoglobin (B3LYP/6-31G* method), without or with a correction of the Fe–ligand and C–O distances with offset forces.[52] The results are compared with the two most accurate structures of this protein complex (both at 1.15 Å resolution).[53] The quantum system included Fe(porphine)(imidazole)$_2$(CO), where the second imidazole group is a model of the distal His residue.

Method	Fe–N_{His}	Fe–N_{Por}	Fe–C	C–O	Fe–C–O
Uncorrected	2.08	1.99–2.03	1.71	1.16	170.8
Corrected	2.08	1.99–2.03	1.70	1.17	170.9
1bzr[53]	2.11	1.99–2.03	1.73	1.13	171.3
1a6g[53]	2.06	1.94–2.02	1.82	1.09	171.2

Table 6. Summary of the various NMR results for the two Ca sites in EGF34.[57] The total and NMR energy, as well as and maximum Ca–O bond length to the putative Ca ligands in the two sites are listed (as averages of the 20 structures with the lowest energies). For ComQum-N, no total energy is given, because it is not comparable to the other energies. The calculations were performed with the Becke–Perdew86/6-31G* method.

Method	Energy terms (kJ/mole)		Maximum Ca–O distance (Å)	
	Tot	NMR	Site 1	Site 2
Only NMR	654	92	10.0	6.3
O–O restraints	859	134	3.9	3.8
Ca–O restraints	827	131	3.1	3.1
Ca MM, monodentate	946	144	2.6	2.7
Ca MM, Glu bidentate	986	155	2.8	2.9
Ca MM, Asp bidentate	880	121	2.8	2.9
ComQum-N		125	2.6	2.7

Table 7. The results of the EXAFS/CC methods for the $Ni(en)_3$ complex, compared to the crystal structure and standard EXAFS fits (distances in Å).[69] $^2_{red}$ shows how well the structure fits the EXAFS data and E_{QM1} is the energy difference (in kJ/mole) between structures optimised in the protein and in vacuum. The Becke–Perdew86/SVP method was used.

Fit	Method	Ni–N	Ni–C	$^2_{red}$	E_{QM1}
1	Crystal	2.12	2.98	312	72
2	EXAFS 1-shell [a]	2.13		598	
3	EXAFS 2-shell [a]	2.12	2.96	169	
4	EXAFS 3-shell [a]	2.13	2.95	147	79
5	EXAFS 2-shell	2.13	2.95	195	79
6	QM	2.17	3.00	1036	0
7	EXAFS/QM [b]	2.13	2.97	118	4
8	EXAFS/QM [c]	2.13	2.95	98	7
9	UFF	1.84	2.79	9598	0
10	EXAFS/UFF [b]	2.13	2.94	98	887

[a] With fitted, rather than calculated DWFs.
[b] EXAFS forces were only calculated for the N atoms.
[c] EXAFS forces were calculated for both the N and C atoms.

Table 8. Results of the EXAFS/QM fits of our 15 models of the putative SAT complex of Cu^{2+} in acetonitrile.[79] $^2_{red}$ shows how well the structure fits the EXAFS data, E_0 is the absorption edge energy, and E_{QM1} is the energy difference between structures optimised in the protein and in vacuum. The Becke–Perdew86/6-31G* method was used.

Complex	$^2_{red}$	$^2_{red}$	$^2_{red}$	E_0	Cu – N distance (Å)						E_{QM1}
	1[a]	2[b]	3[c]	eV	N1	N2	N3	N4	N5	N6	kJ/mole
cis-2	2.3	2.2	1.7	11.2	1.94	2.01	2.02	2.03			1
cis-3	3.8	2.7	1.6	12.7	1.99	2.00	2.03	2.03	2.34		3
cis-4	3.4	3.5	2.7	13.0	1.99	2.02	1.97	2.05	2.33	2.34	11
trans-2	11.7	8.4	5.3	15.4	2.00	2.00	2.05	2.05			15
trans-3	3.3	3.5	3.0	11.2	1.96	1.97	2.03	2.04	2.31		8
trans-4	4.9	5.2	4.9	13.8	1.97	2.04	1.98	2.05	2.27	2.65	0
1N-3	4.0	4.0	3.7	12.2	2.01	2.01	2.01	2.01			1
1N-4	7.2	5.3	3.0	11.6	2.00	2.00	2.01	2.51	2.68		2
1N-5	4.8	5.0	3.8	13.2	1.96	2.02	2.03	2.03	2.23	2.43	1
CuPor	22.6	23.2	22.0	15.2	2.01	2.01	2.01	2.01			1
$CuPorAn_2$	29.4	23.5	7.7	20.8	2.00	2.06	2.24	2.06	2.85	2.85	20
$CuAn_4$	19.1	19.7	18.2	12.2	1.99	1.99	1.99	1.99			2
$CuAn_5$ tbp	14.4	14.9	8.6	13.2	2.02	2.02	1.97	1.98	2.29		3
$CuAn_5$ sqpy	13.5	13.8	7.8	13.9	2.03	1.95	2.03	2.01	2.30		5
$CuAn_6$	11.1	11.1	6.7	15.4	1.98	2.08	2.02	2.00	2.29	2.39	2

[a] The result of an EXAFS fit of only E_0, based on the QM structures, optimised in vacuum.
[b] The result of an EXAFS fit of E_0, and all Cu–N distances based on the QM structures, optimised in vacuum.
[c] The result of EXAFS/QM calculations.

Table 9. Results of the combined QM and EXAFS calculations[80] for the two peroxide adduct models of the multi-copper oxidases, using six different methods: An EXAFS fit (of only E_0) directly to the QM or QM/MM structure, a standard EXAFS fit of three distances (in addition to E_0), starting from the QM or QM/MM structures, and the combined EXAFS/QM and EXAFS/QM/MM methods. $^2_{red}$ shows how well the structure fits the EXAFS data, E_0 is the absorption edge energy, and E_{QM1} is the energy difference between structures optimised in the protein and in vacuum. All QM calculations were performed with the Becke–Perdew86/6-31G* method.

Calculation	Model	E_0	$^2_{red}$	E_{QM1}	Cu–Cu distances (Å)		
		eV		kJ/mole	2–3	2–3'	3–3'
QM	C_3	5.5	76		3.41	3.80	4.63
	S_{23}	3.4	113		3.34	3.81	3.75
QM+EXAFS fit	C_3	4.2	72		3.41	3.80	4.63
	S_{23}	3.8	77		3.42	3.78	3.72
EXAFS/QM	C_3	4.8	48	3.6	3.44	3.71	4.60
	S_{23}	3.5	50	10.9	3.42	3.85	3.71
QM/MM	C_3	2.7	218		3.52	4.04	4.49
	S_{23}	6.3	134		3.72	4.03	3.72
QM/MM+EXAFS fit	C_3	3.6	83		3.42	4.04	4.49
	S_{23}	3.3	86		3.85	3.95	3.86
EXAFS/QM/MM	C_3	3.5	38	9.9	3.41	4.02	4.44
	S_{23}	4.1	74	5.0	3.70	4.00	3.77

Figure 1. The division of a protein into a QM and an MM system in the QM/MM approach.

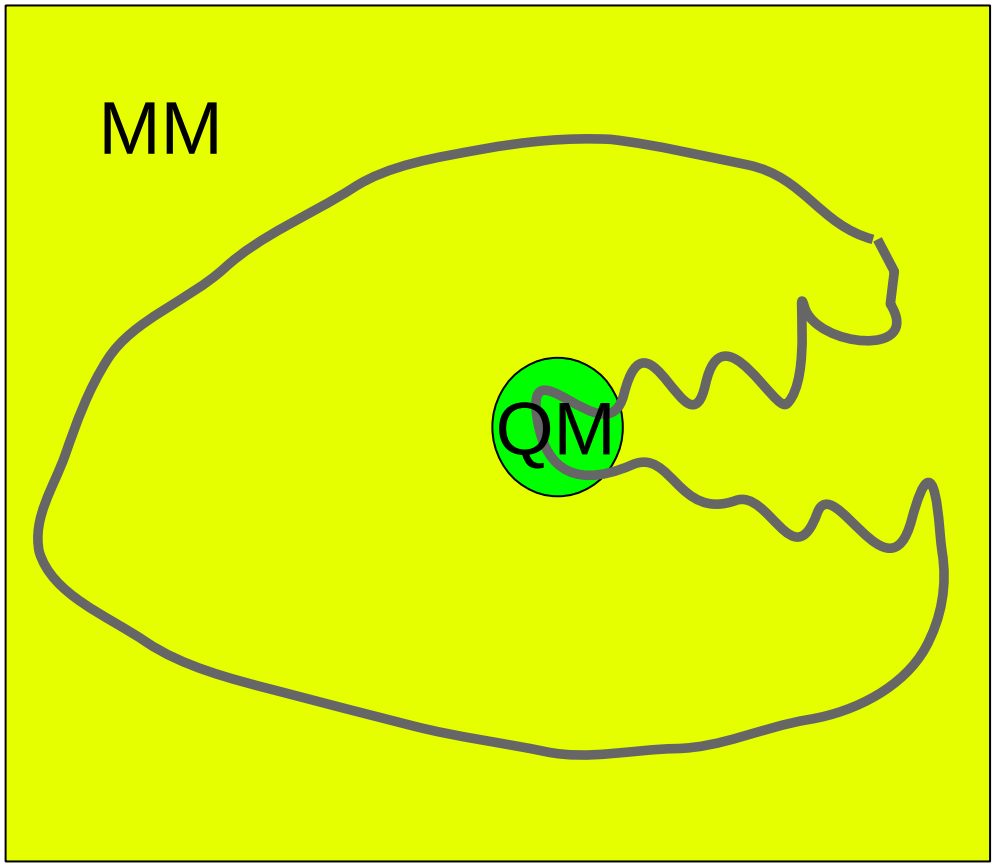


Figure 2. A flow scheme of standard X-ray crystal structure determination.[22]

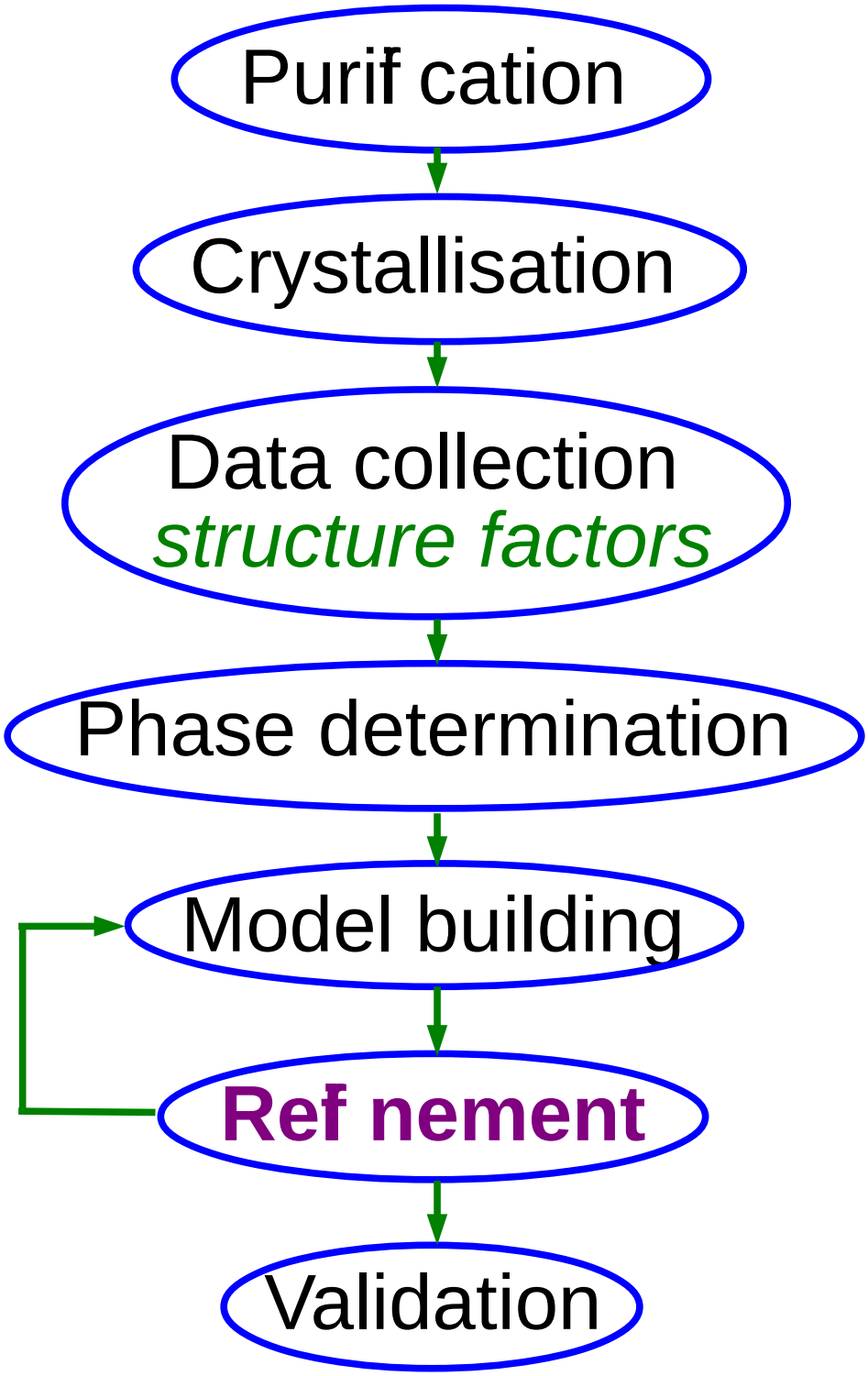


Figure 3. Electron density maps of metal sites at various levels of resolution: *N*-methylmesoporphyrin in ferrochelatase at 1.7 Å resolution[29] (in which the electron density is not continuous and it is not always clear where to put all atoms) and the haem active site in cytochrome c_{553} at 0.97 Å resolution (where all atom positions are well-defined).[11,32] In the latter, the low- (magenta) and high-resolution (green) crystal structures[32] are shown together with the ComQum-X structure (orange).[11]

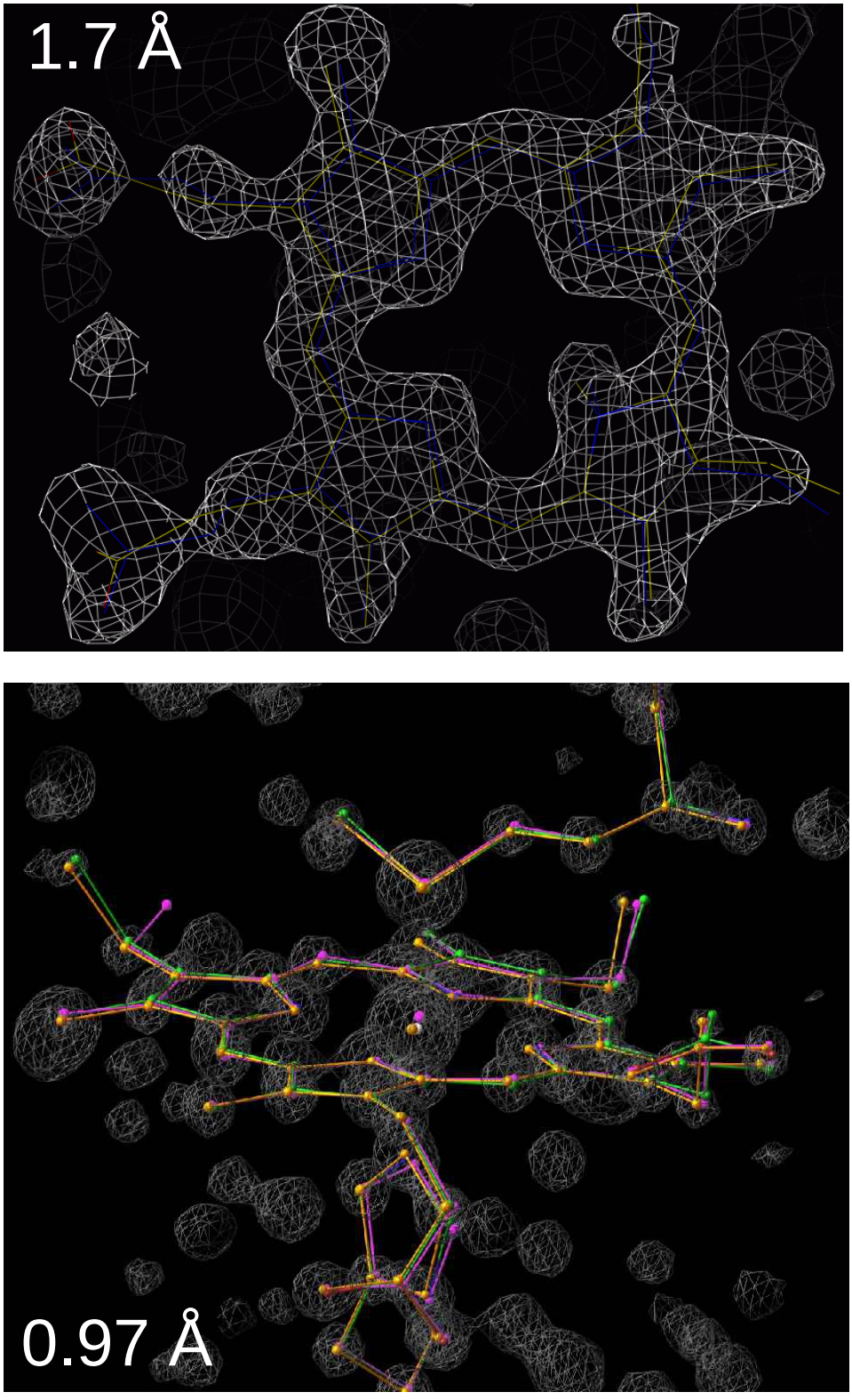


Figure 4. The adenosyl group in methylmalonyl coenzyme A mutase at 2.2 Å resolution, refined without and with MM restraints.[17,26]

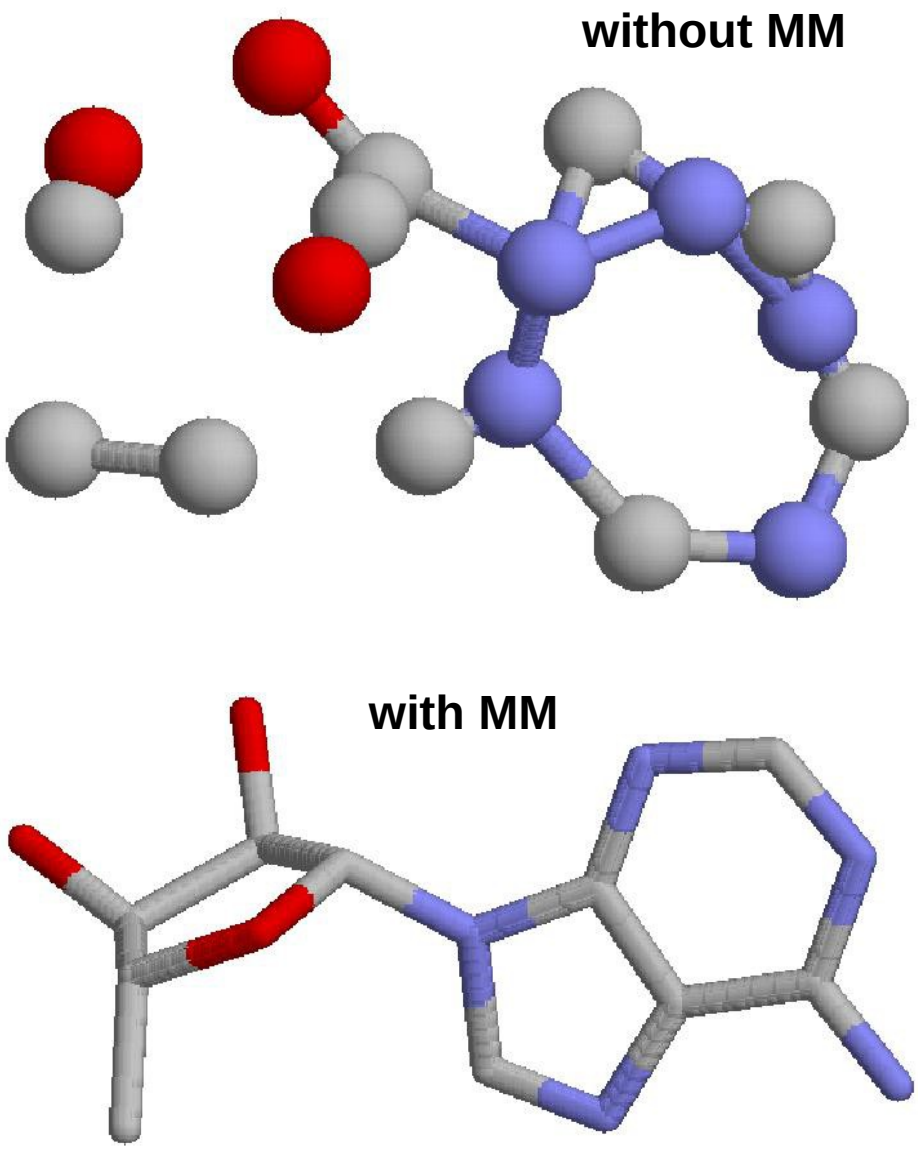


Figure 5. The adenosyl group in methylmalonyl coenzyme A mutase, refined with an unspecified force field[26] and with the Amber force field.[17]

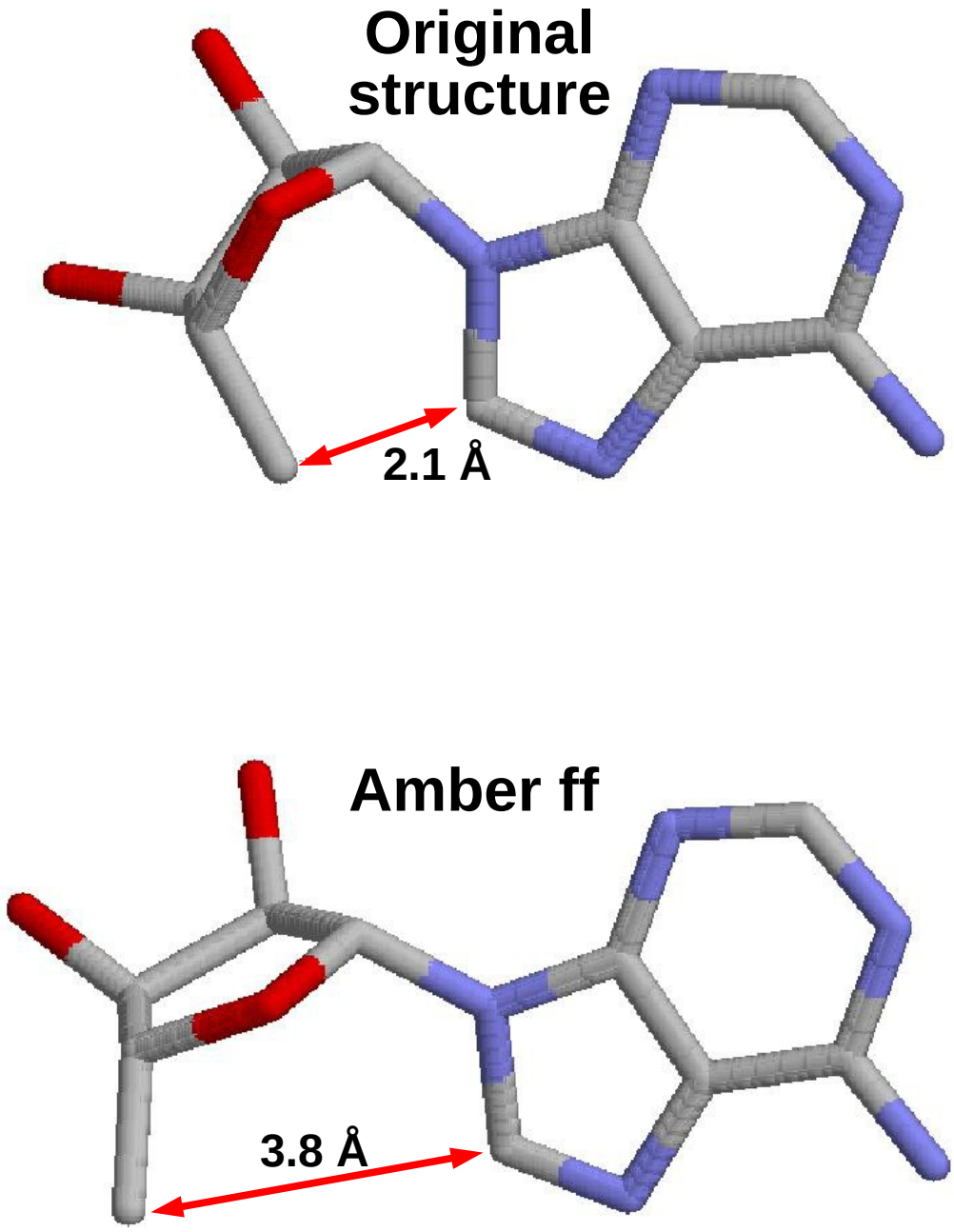


Figure 6. The active site of Mn superoxide dismutase,[42] illustrated by the QM system used in our calculations.[38]

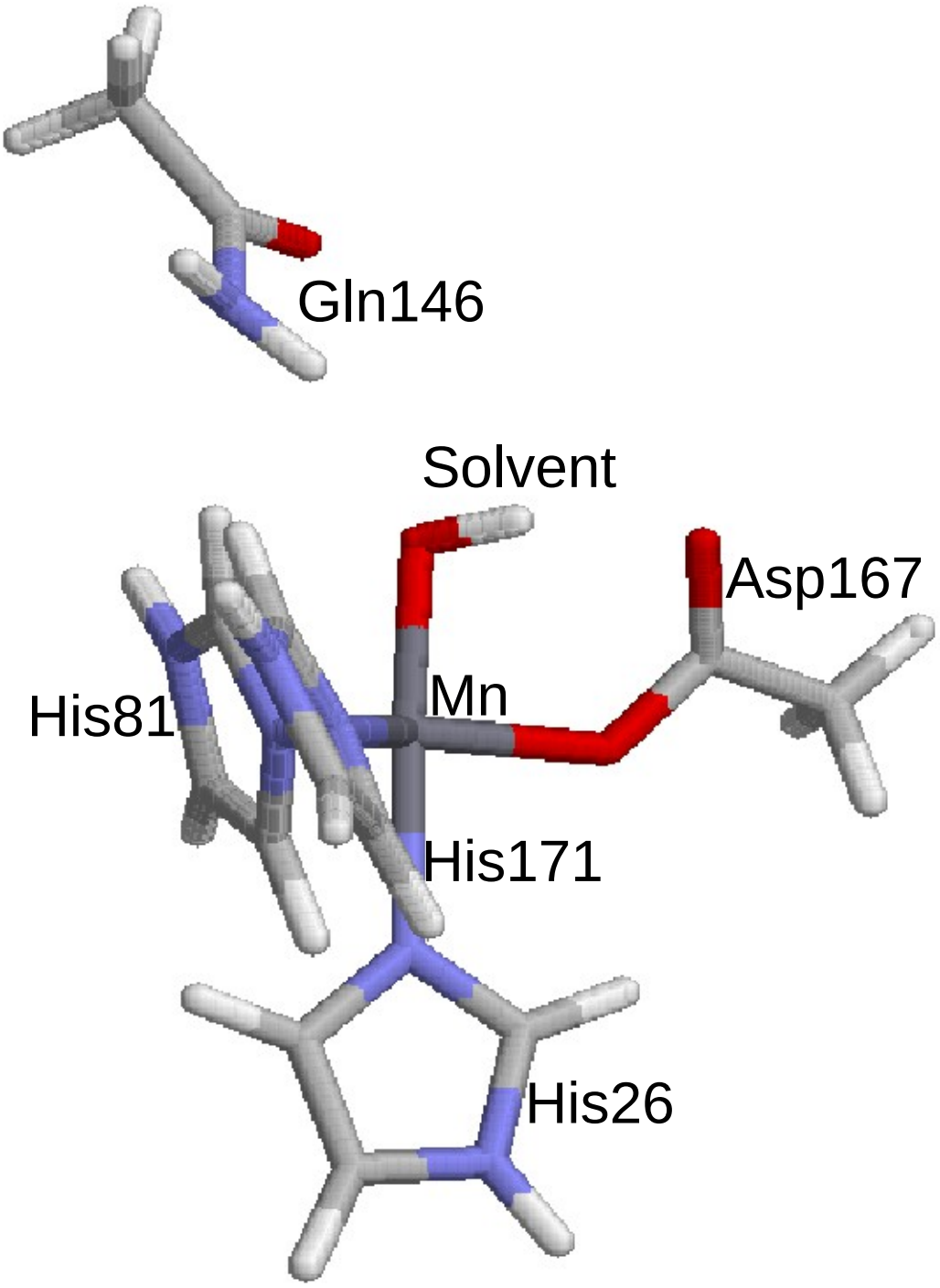


Figure 7. The active site of [Ni,Fe] hydrogenase,[51] illustrated by the QM system used in most of our quantum-refinement calculations.[14]

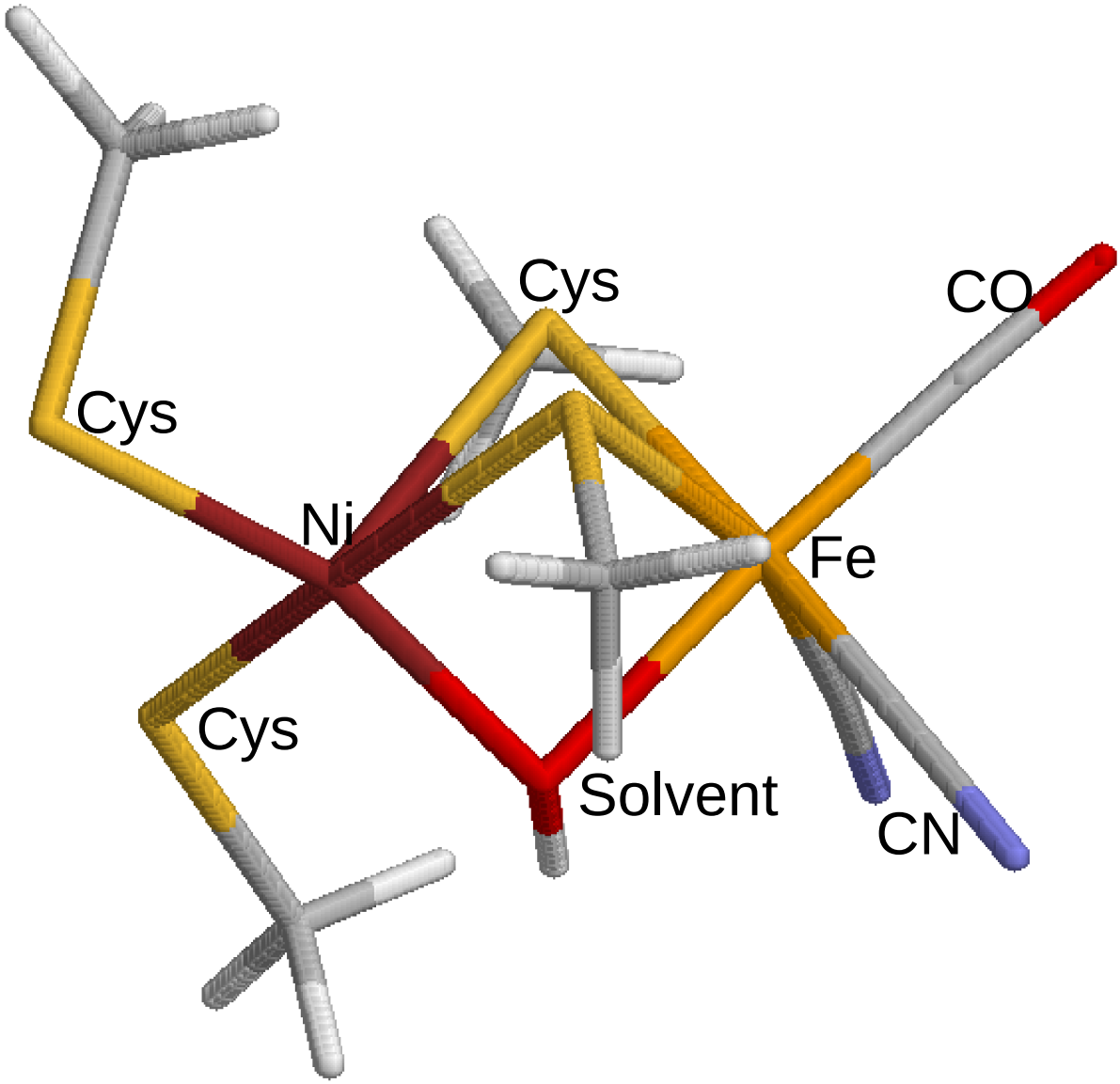


Figure 8. A comparison of the crystal structure of [Ni,Fe] hydrogenase (green) with a quantum-refined structure with Cys-543 oxidised. It can be seen that the oxidised structure explains the alternative position of the S^{γ} atom of Cys-543 ($S_{alt)}$ and it also shows that Glu-25 follows this movement, which explains features in the electron-density map.[14] The calculations were performed with the DFT Becke–Perdew-1986 method and the 6-31G* basis set.

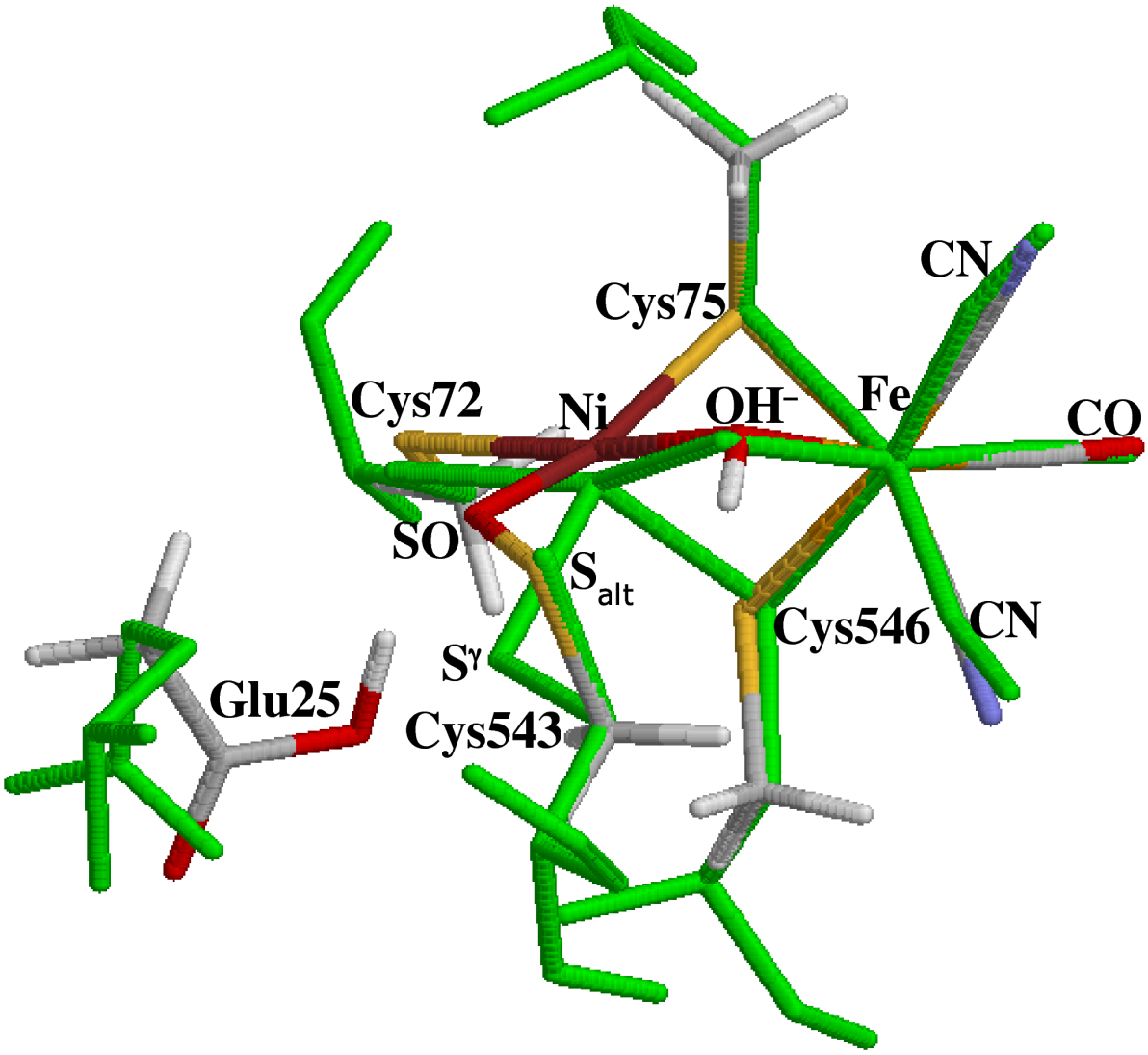


Figure 9. A ComQum-N structure of the Ca site in EFG34, illustrating the QM system used and that the carboxylate ligands can be either bidentately (backside) or monodentately (upper left).

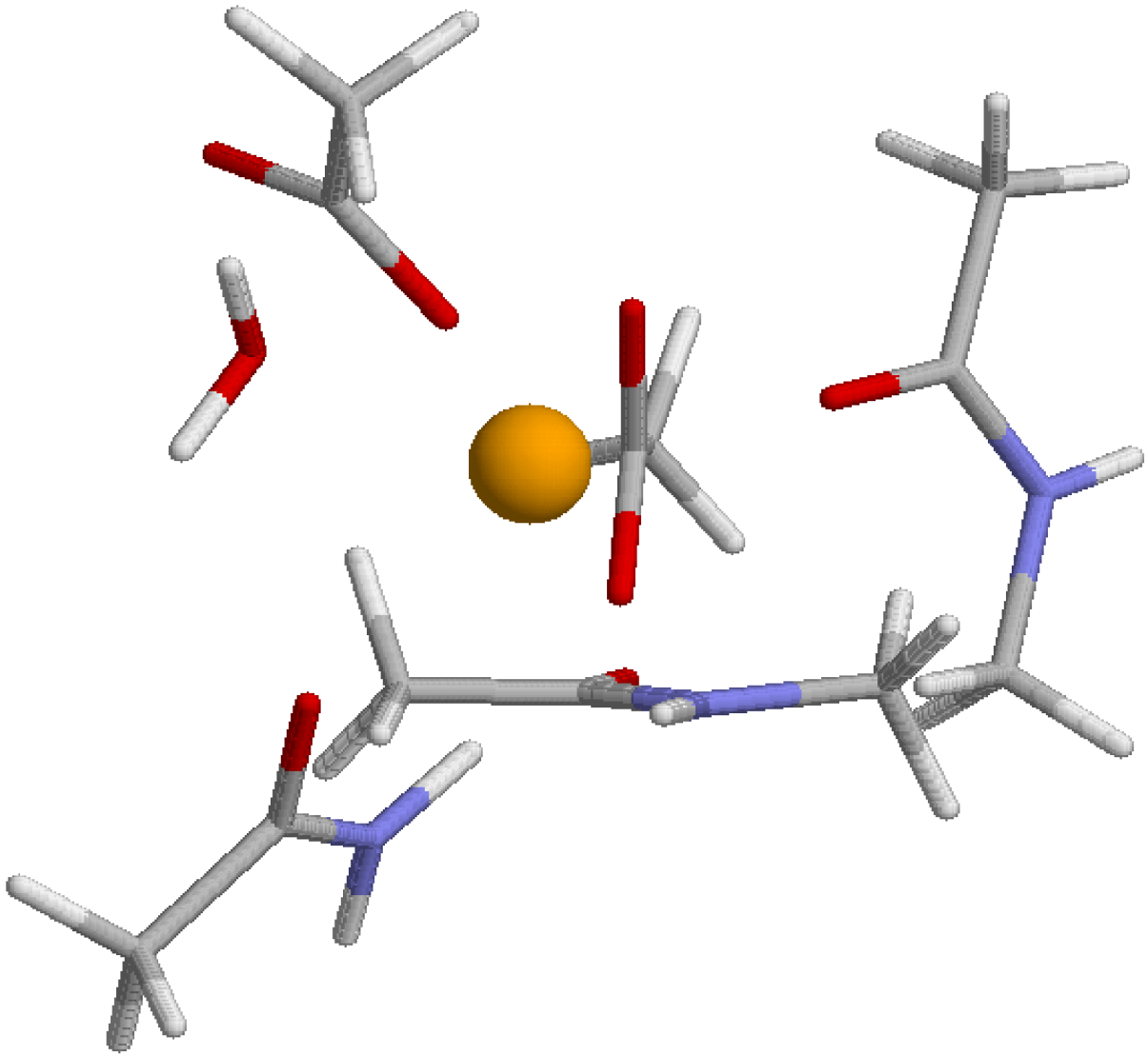


Figure 10. NMR structure of EFG34 without any information about the Ca sites. The 20 structures with the lowest energy are shown. The left (right) image was obtained by superimposing domain 3 (4; in red). The other domain is shown in blue. The two domains are connected by a flexible hinge. The Ca ligands are shown in green.

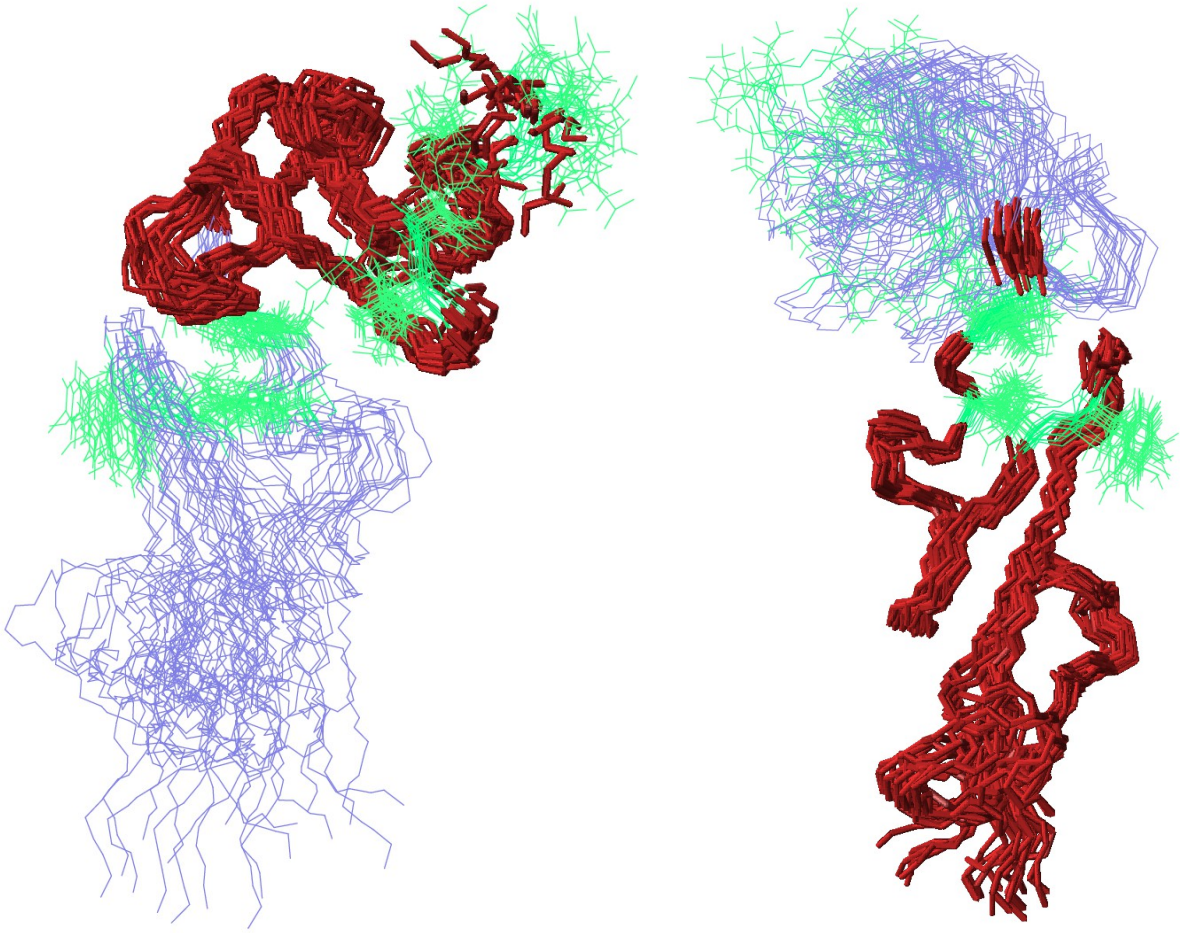


Figure 11. One of the Ca sites in EFG34 obtained with a NMR refinement using O–O restraints for the Ca ligands (i.e. a set of ten O–O restraints between the five putative protein Ca^{2+} ligands, defined with a flat-bottomed harmonic potential, between 3.0 and 5.1 Å, i.e. in a similar way as the normal NMR restraints). It can be seen that this gives O–O bond lengths around the upper limit of the flat-bottomed potential and that some of the ligands are not directed towards a putative Ca^{2+} in the centre.

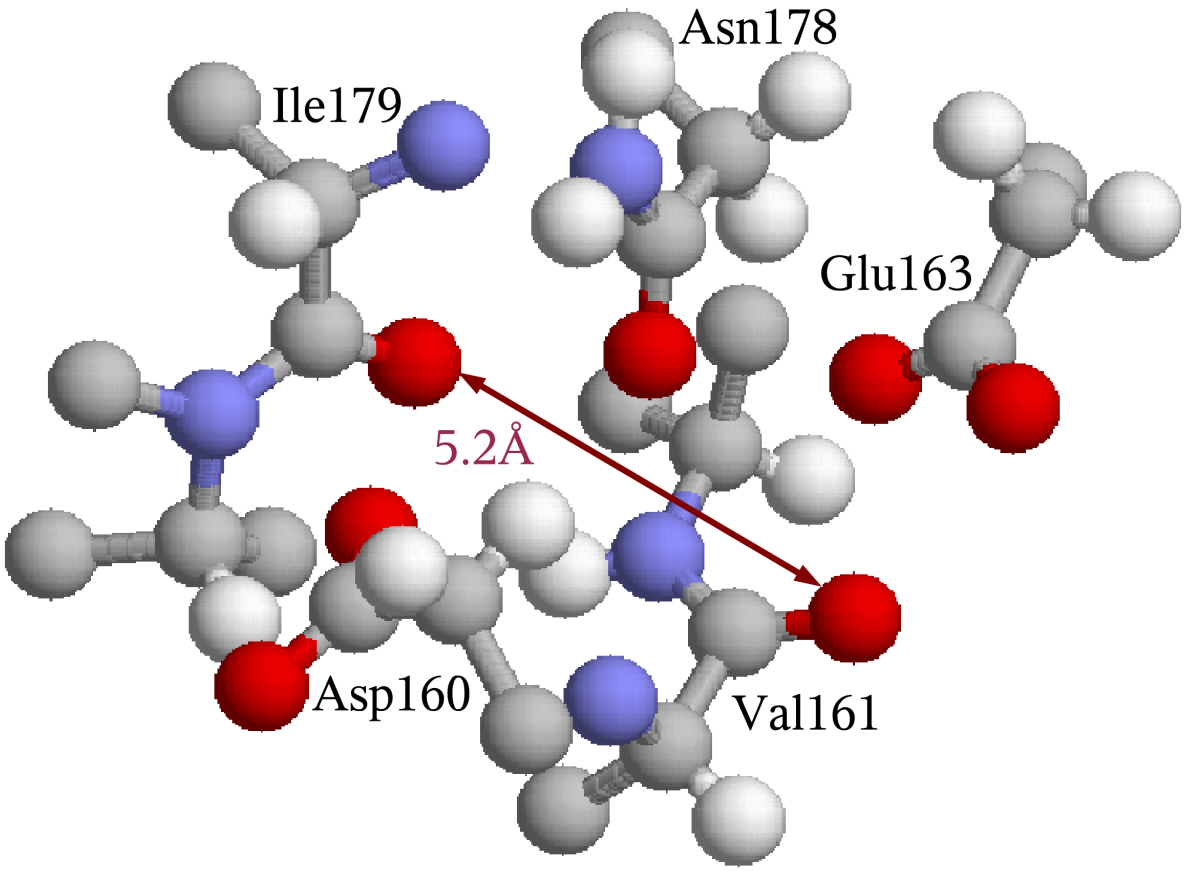


Figure 12. The ComQum-N structures of EFG34. Colouring as in Figure 10, but with 60, rather than 20, superimposed structures. The Ca ions are shown as black balls.

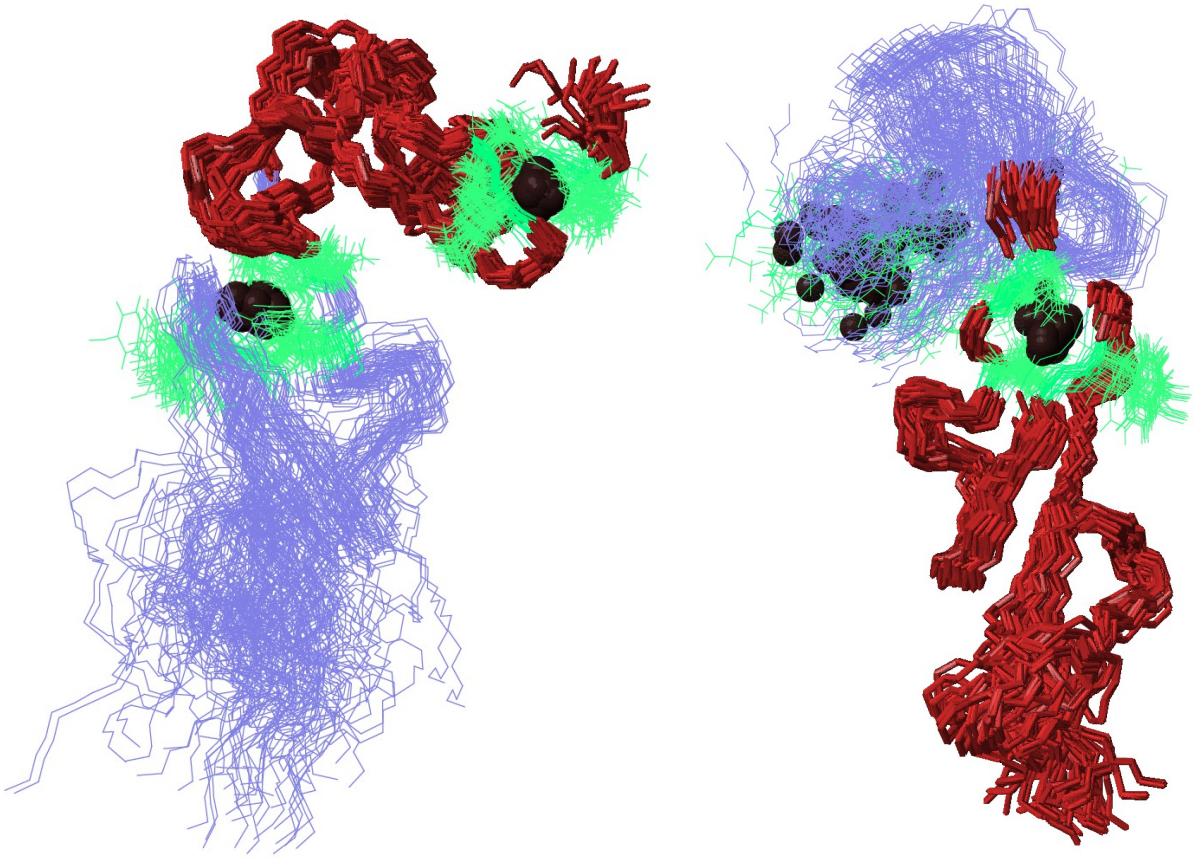
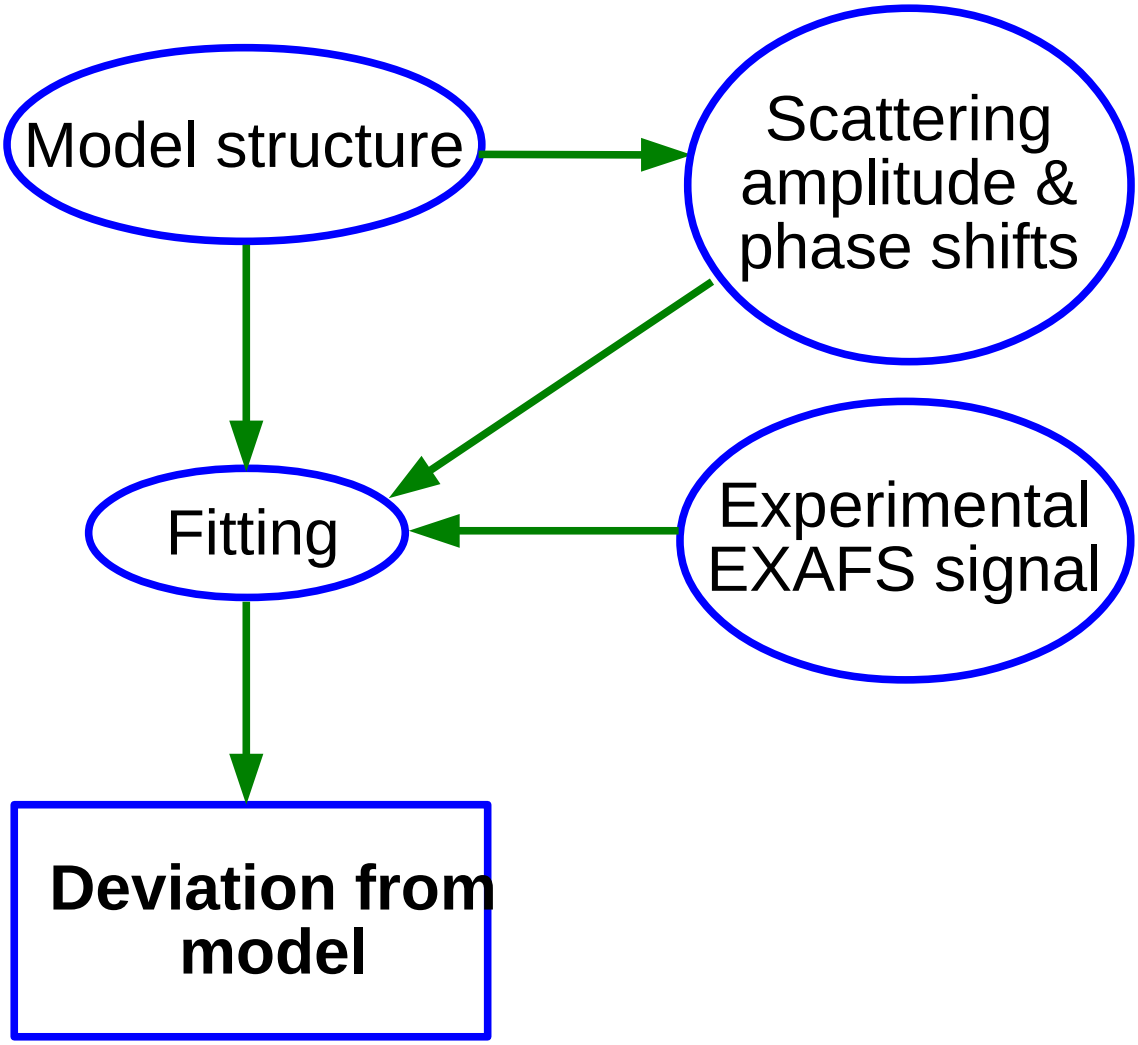


Figure 13. Flow scheme of the standard EXAFS procedure (a) and the EXAFS/CC approach (b).



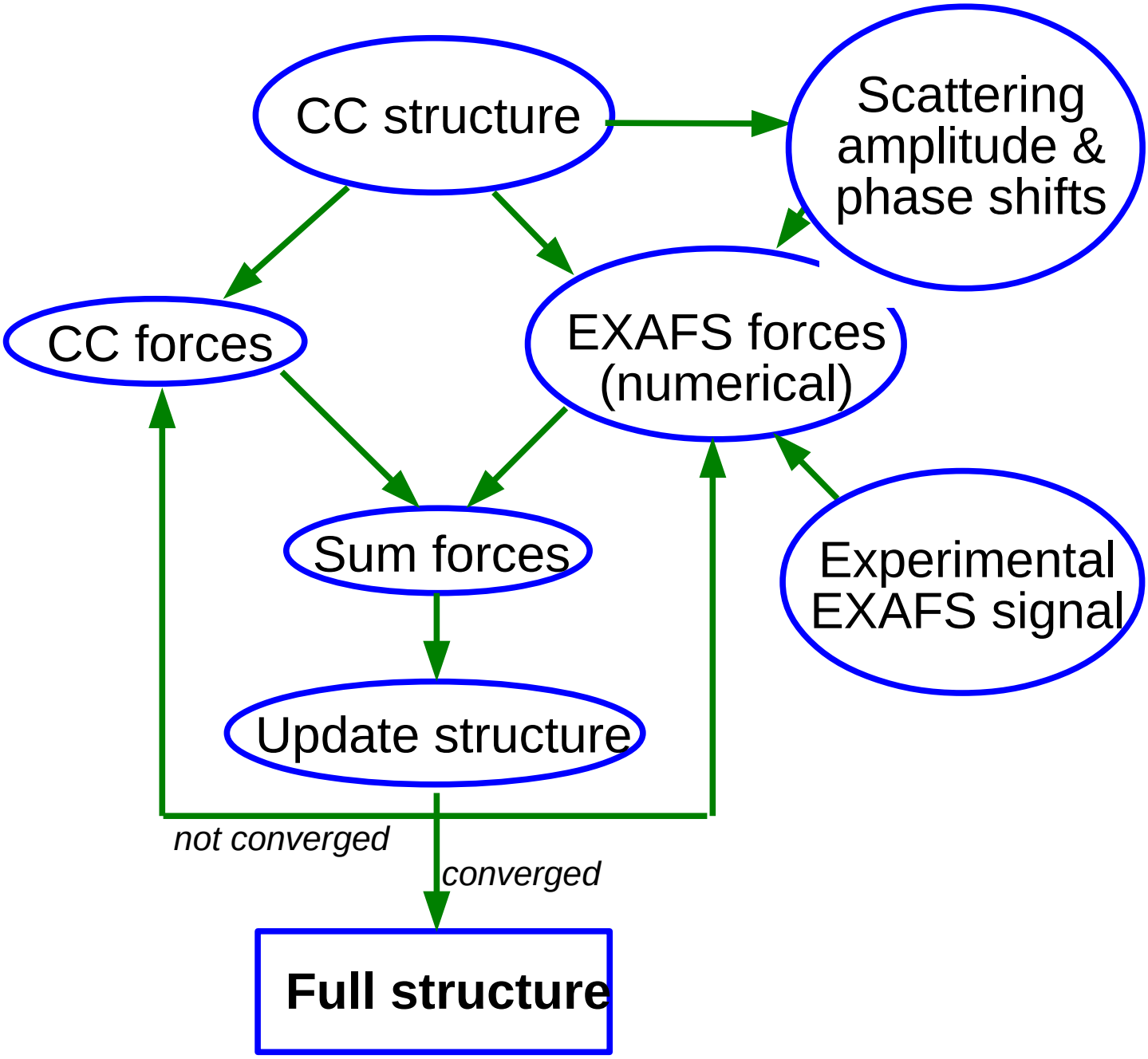

CC structure
Scattering amplitude & phase shifts
CC forces
EXAFS forces (numerical)
Sum forces
Experimental EXAFS signal
Update structure
not converged
converged
Full structure

Figure 14. The $Ni(NH_2CH_2CH_2NH_2)_3^{2+}$ structure used to calibrate of the EXAFS/CC method.

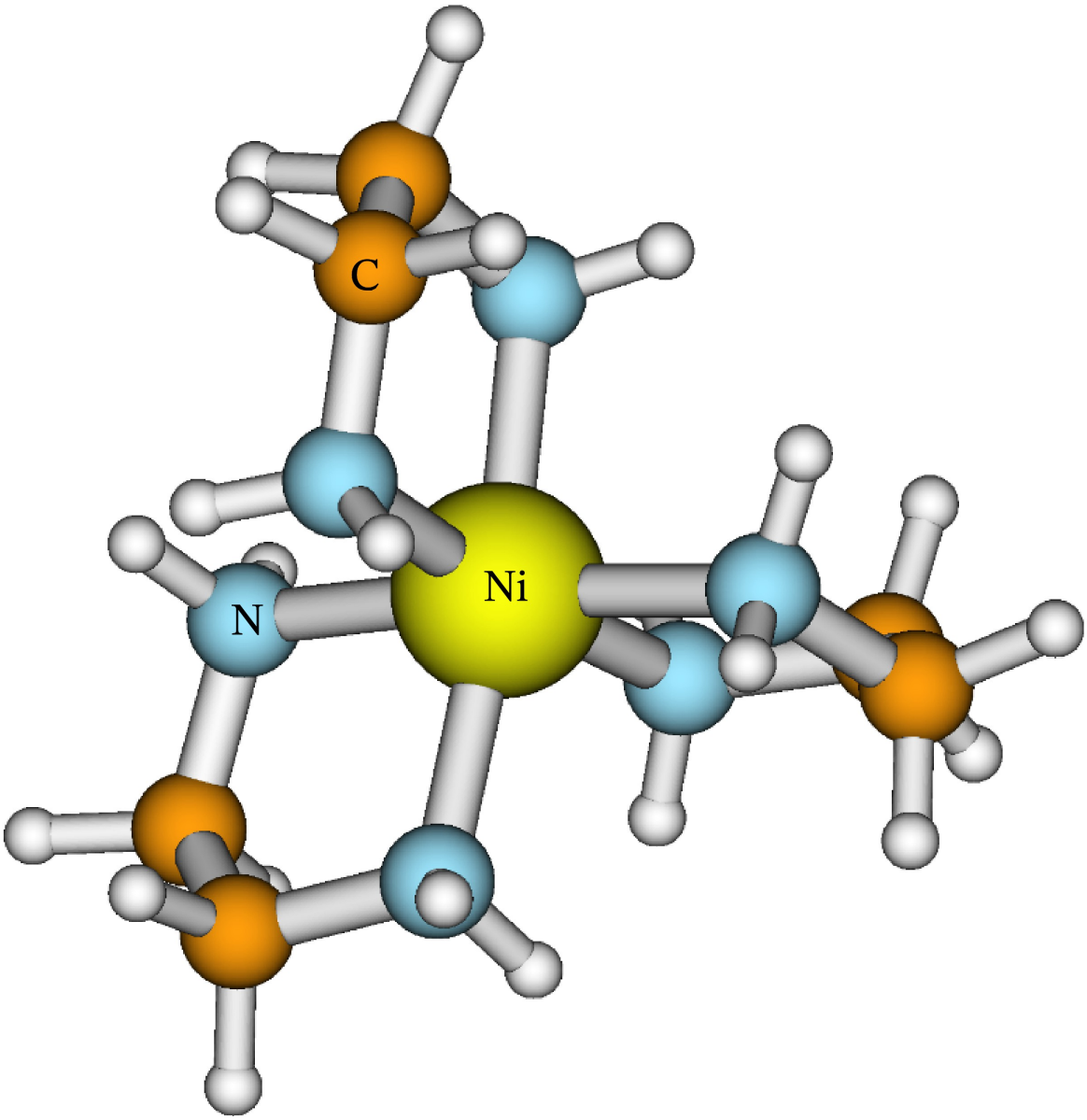
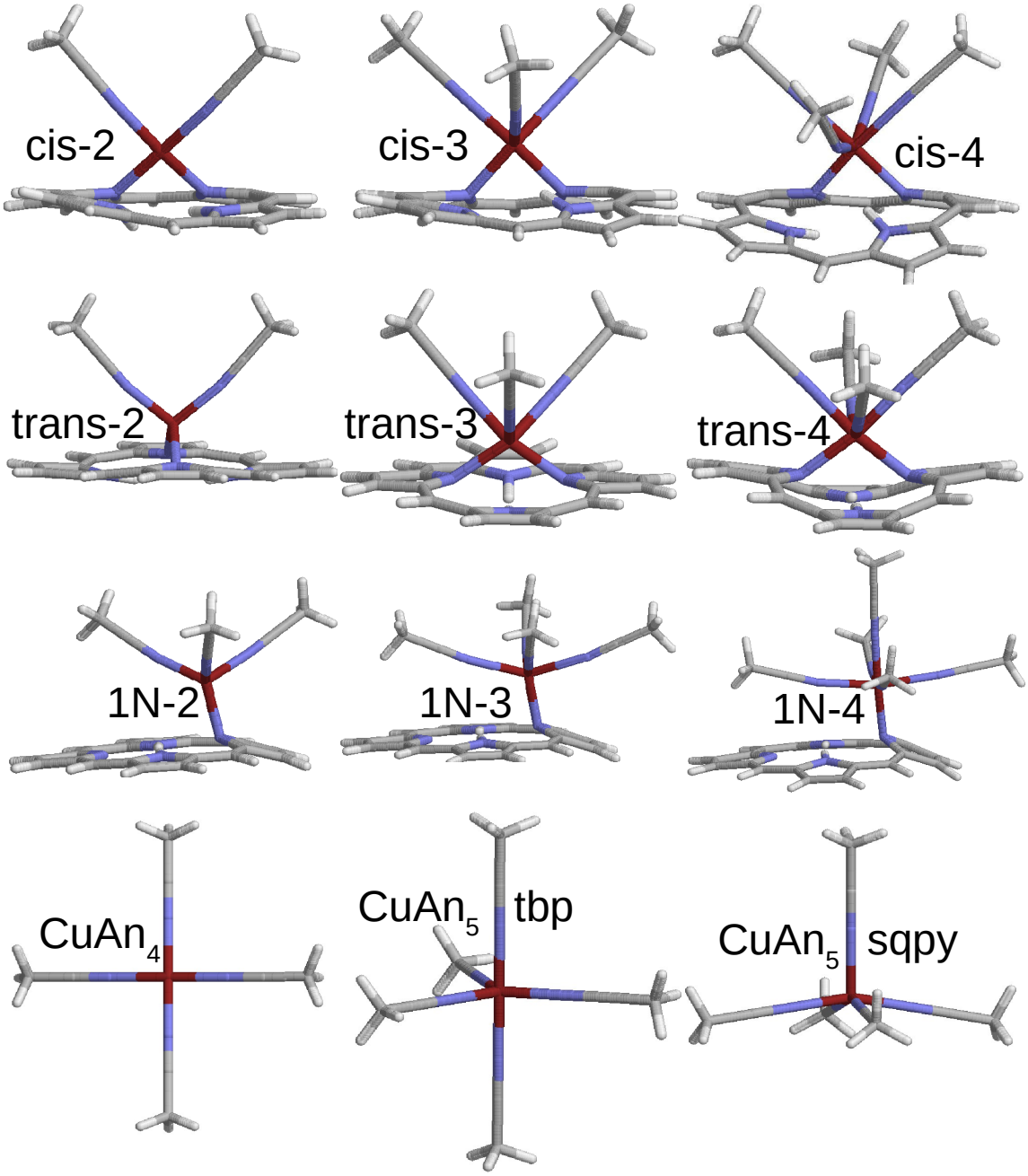


Figure 15. A set of nine putative SAT complexes of copper and porphine in acetonitrile solution, together with four possible reactant and two product states. All these complexes were considered in our EXAFS/QM investigation.[79] The structures were optimised with the Becke–Perdew86/6-31G* method.



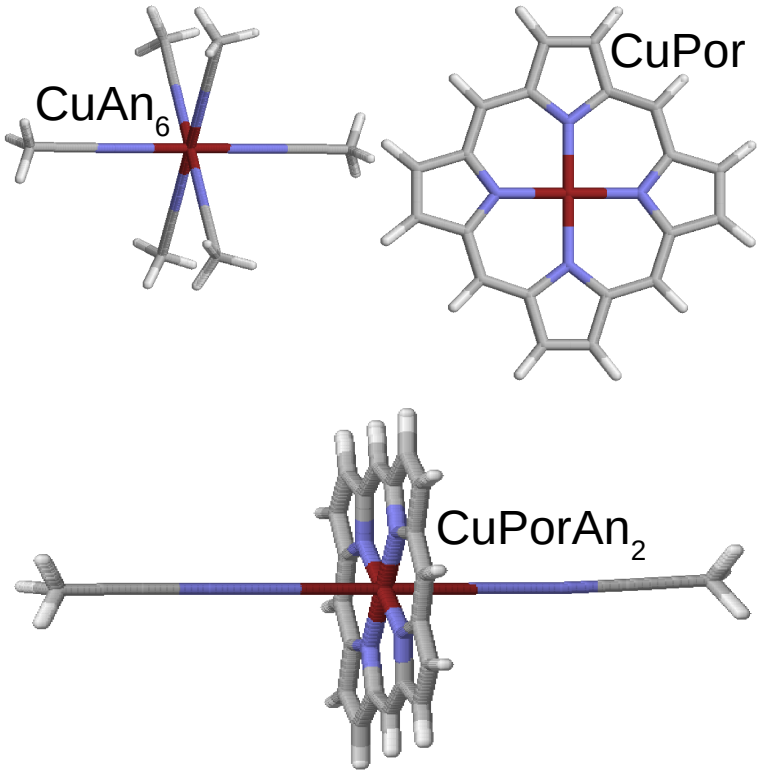

CuAn6
CuPor
CuPorAn2

Figure 16. A comparison between the EXAFS raw data (blue) and the fit to the cis-2 complex (a) and the 0.5 $CuAn_4$ + 0.5 $CuPorAn_2$ mixture (b), involving E_0, Cu–N distances, and Debye–Waller factors of the first-sphere paths (2 in (a) and 3 in (b)), as well as a single Debye–Waller factor for all the other paths (red), displayed as Fourier-transformed magnitude (k^3-weighted, no phase correction) versus the distance from the Ni atom (R; left), and k^3 (k) versus k (right).

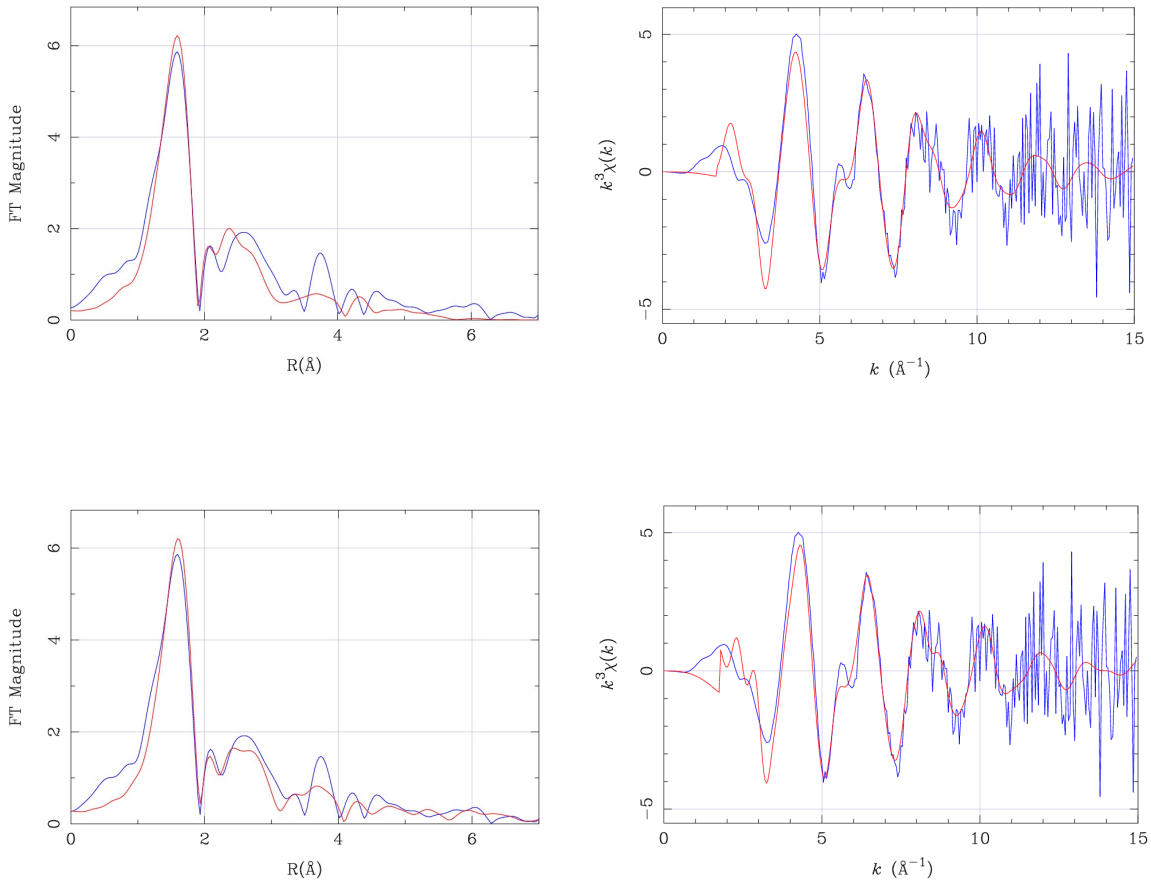
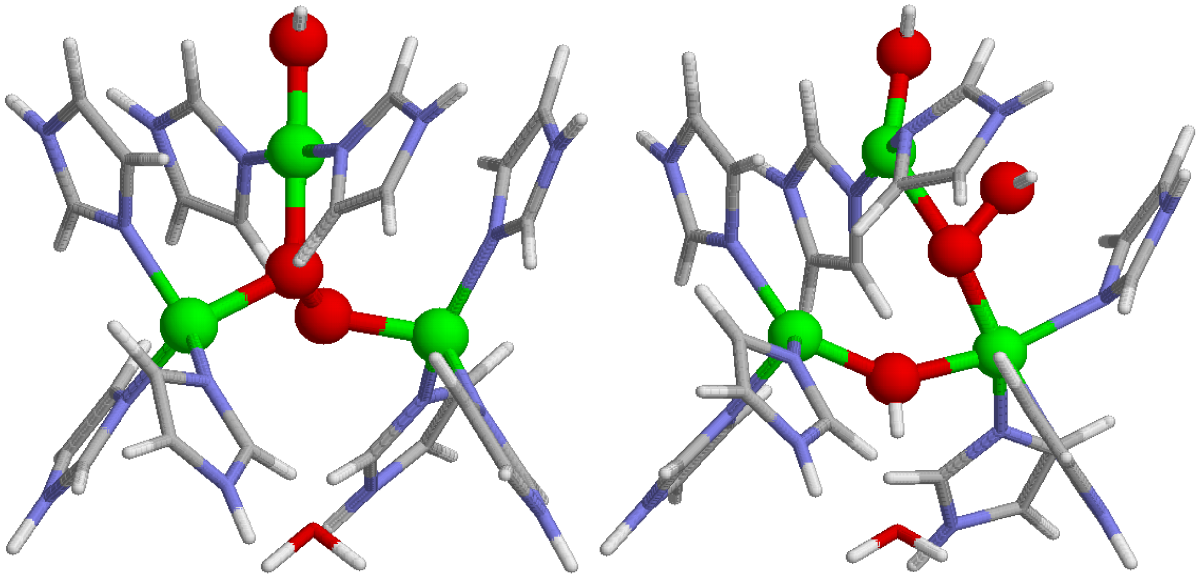


Figure 17. The QM/MM structures of the C_3 (left) and S_{23} (right) states of the peroxy adduct of the multi-copper oxidases.[86]



Graphical Abstract

The use of quantum chemical calculations in X-ray crystallography, NMR, and EXAFS refinement is reviewed. Such combinations give improved structures of metal sites in proteins and allow determination of protonation states of metal-bound solvent molecules.