Quantum refinement - a combination of quantum chemistry and protein crystallography.

Ryde, Ulf; Nilsson, Kristina

Published in:
Journal of molecular structure. Theochem

DOI:
10.1016/S0166-1280(03)00304-X

2003

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

Quantum refinement – a combination of quantum chemistry and protein crystallography

Ulf Ryde and Kristina Nilsson

Department of Theoretical Chemistry

Lund University

Chemical Centre

P. O. Box 124

S-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

The combination of quantum mechanics and molecular mechanics (QM/MM) is one of the most promising approaches to study the structure, function, and properties of proteins. We here review our applications of QM/MM methods to alcohol dehydrogenase, blue copper proteins, iron–sulphur clusters, ferrochelatase, and myoglobin. We also describe our new quantum refinement method, which is a combination of quantum chemistry and protein crystallography. It is shown to work properly and it can be used to improve the structure of protein metal centres in terms of the crystallographic R_{free} factor and electron-density maps. It can be used to determine the protonation status of metal-bound solvent molecules in proteins by refining the various possible states and see which fits the crystallographic raw data best. Applications to ferrochelatase, cytochrome c_{553}, alcohol dehydrogenase, myoglobin, and methylmalonyl coenzyme A mutase are described.

Keywords

QM/MM, crystallographic refinement, metalloproteins, density functional theory, protonation status.

Introduction

During the last decade, quantum chemical methods have become an important complement to experiments for the study of structure and function of proteins, mainly owing to the increase in computer power and the introduction of accurate density functional methods [1,2]. However, an entire protein is still too large to study by such methods. Therefore, quantum chemical studies have traditionally isolated the interesting part of the protein (e.g. the active site), including only a restricted number of atoms (30–200) in the calculations, whereas the surrounding protein has been ignored or is represented as a homogenous continuum solvent [1,2].

Naturally, such a treatment is not fully satisfactorily, even if it often works surprisingly well [1-3]. A natural improvement is to include the surrounding protein by molecular mechanics, adding the quantum chemical and molecular mechanics energies and forces (ensuring that no interactions are double-counted). This gives the combined quantum mechanics and molecular mechanics methods (QM/MM), pioneered by Warshel, Levitt, and Kollman [4,5]. Many variants of QM/MM methods have been proposed and several reviews are available [6-13].

In this paper, we shortly review our developments and applications of QM/MM methods. In particularly, we will discuss a method to refine protein structures using quantum chemical methods for the most interesting parts of the system.

The QM/MM program ComQum

We will start with a short description of our local QM/MM program ComQum [14-16], comparing it with other available QM/MM codes. We started to develop this program in 1992, unaware of the already published QM/MM optimisation methods [5, 17]. Therefore, the algorithm of ComQum is similar but not identical to that of many

other QM/MM codes. The underlying philosophy is that ComQum should be an interface between a quantum mechanics (QM) and a molecular mechanics (MM) software, without any need to change the QM and MM codes. This makes it easy to change or update the QM and MM programs and therefore the best available software can always be used.

Like most other QM/MM methods, ComQum divides the protein (including solvent) into three subsystems. The central system 1 (the quantum system) is optimised by a QM method. System 2 consists of all amino acids (and solvent molecules) within a radius r_1 (0.4–2 nm) of the quantum system. It is optimised with MM methods. Similarly, system 3 comprises all atoms within a radius r_2 of system 2. Typically, it consists of the rest of the protein and a sphere of water molecules with a radius of 3–4 nm (we usually study a non-periodic spherical system). It is considered in all calculations, but is kept fixed at the crystal geometry (hydrogen atoms are optimised by a simulated annealing calculation).

In the QM calculations, system 1 is represented by a wavefunction, whereas systems 2 and 3 are modelled by an array of point charges, one for each atom, which is included in the one-electron Hamiltonian. Therefore, the polarisation of the quantum system by the protein is considered in a self-consistent way. No cut-off is used for the electrostatic interactions (the system is non-periodic and finite). In the MM energy and force calculations, systems 1–3 are represented by the molecular mechanics force field, but without any electrostatic interactions (which are already treated by quantum mechanics).

System 2 is optimised with system 1 fixed in every step of the geometry optimisation of system 1. Thereby, we exploit the speed of a MM optimisation compared to that of the QM calculations. In these calculations, systems 2 and 3 are represented by standard

MM parameters (including electrostatics), whereas system 1 is represented by MM parameters, but with charges from the QM calculation [15].

Special action is taken when there is a bond between the classical and quantum chemical systems (a junction) [14]. The quantum chemical system is truncated by hydrogen atoms, the positions of which are linearly related to the corresponding heavy (typically carbon) atoms in the full system. The position of the H junction atoms are optimised in the QM steps and the positions of the corresponding C junction atoms are updated from these positions. No constraints are imposed on any of the atoms in the boundary region, but the charges of the atoms bound to the C junction atoms are set to zero by spreading it out on the other atoms in the amino acid (a change of less than 0.05 *e*). More sophisticated methods exist for the treatment of the junction atoms [11,18], but they have not been shown to give any improved performance compared to the link-atom approach [18,19].

The total energy is calculated as

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

Here, E_{QM} is the QM energy of system 1 with H junction atoms, including all the electrostatic interactions. Similarly, E_{MM1} is the classical energy of system 1, still with H junction atoms, but without any electrostatic interactions. Finally, E_{MM123} is the classical energy of systems 1–3 with C junction atoms and no electrostatics. This approach is similar to the one used in the ONIOM method [20]. The forces are the negative gradient of this energy, taking into account the relation between the H and C junction atoms using the chain rule [21].

ComQum was originally [14] developed as a combination of the QM software Turbomole [22] and the local MM program Mumod [23,24]. Later [15], the MM software was changed to Amber [25]. The program now runs with version 5.5 of

Turbomole and version 7 of Amber. We also work on an implementation of ComQum with MOLCAS 5.4 [26]. This will allow us to run calculations with a continuum solvation model outside the water sphere, to include dipoles and polarisabilities in the force field, and effective group potentials [27] for the junctions.

ComQum consists of six small programs that transfer information between the two programs or constructs the files needed. To make the logic of the programs clear and facilitate changes of the QM and MM software, each of the interface programs are actually divided into four programs [28]. One of these programs is the core ComQum procedure, which reads QM and MM data from temporary text files in a standard format and writes the output to other text files in the same format. This program defines the ComQum core and they need not to be changed if the QM or MM software is changed. In addition, there are two input programs that extract data from the QM and MM output files and write them to the ComQum temporary text files, and one procedure that reads the output temporary text files and inserts the results into the QM or MM files. These programs are specific for the QM and MM software and need to be changed if you want to switch software. The program flow of ComQum is shown in Scheme 1 [16].

Applications to alcohol dehydrogenase

Our first application of ComQum was to the enzyme alcohol dehydrogenase, which oxidises alcohols to aldehydes [29]. At that time, there was a controversy whether the catalytic zinc ion is four-coordinate in all steps (as crystallography indicated) or if some complexes are five-coordinate (as some spectroscopic studies suggested) [29]. We had studied this issue with vacuum models and molecular dynamics simulations [24,30], both favouring four coordination (other early theoretical calculations had assumed a certain coordination number [31-33]), but we realised that QM/MM would give more

conclusive results. The ComQum calculations (Hartree–Fock calculations with basis sets of double quality) showed that alcohol dehydrogenase disfavours five-coordinate zinc complexes by ~80 kJ/mole [14]. Such an energy difference is so large that five-coordinate complexes can hardly play any significant role in the reaction mechanism of the enzyme. Encouragingly, this was later confirmed by experimental reinterpretations of the observations of five-coordinate complexes [34,35].

This study also resulted in two papers on related subjects. First, we showed by QM/MM calculations that Glu-68, a conserved amino acid located 0.5 nm from the catalytic zinc ion, may intermittently coordinate to the zinc ion [36]. This has later been confirmed by the crystal structure of human alcohol dehydrogenase, showing such a coordination [37].

Second, we considered the structural zinc ion and showed that all its four cysteine ligands are deprotonated and not only two of them, as had previously been suggested [38]. The Zn–S_{Cys} distances are very sensitive to the theoretical treatment: The experimental distances could only be reproduced if both electronic correlation (at the MP2 level) and a detailed picture of the surrounding enzyme was considered.

Finally, we developed a method to calculate electric field gradients of cadmium complexes (MP2 calculations with a large basis set and a point-charge model of the surroundings) [39] and used it to interpret perturbed angular -ray correlation measurements on alcohol dehydrogenase [40], using structures of the Cd-substituted enzyme with various ligands and coenzymes estimated by ComQum (Hartree–Fock calculations with core potentials and basis sets of svp quality). We reproduced seven experimental field gradients with an average error of less than 9 % and several of the experiments were reinterpreted. The results showed that the enzyme is four-coordinate in all examined complex and that Glu-68 coordinates to the cadmium ion in two of the

complexes. Recently, several QM/MM studies of the reaction mechanism, kinetic isotope effects, and hydrogen tunnelling in alcohol dehydrogenase have been published [41-44].

Electron transfer proteins, myoglobin, and ferrochelatase

Next, we turned to blue copper proteins. Our calculations on alcohol dehydrogenase had shown that proteins have a large flexibility around metal sites. Therefore, it was clear to us that the current hypotheses about the blue copper proteins could not be correct, viz. that the protein forces the Cu^{2+} ion to have a trigonal structure, although it prefers a tetragonal geometry [45,46]. By simply optimising a copper ion in vacuum with the same ligands as in the protein (B3LYP/6-31G* calculations), we showed that the optimum structure is close to that observed in crystal structures [47]. In a series of publications we then explained why copper prefers a trigonal structure with these ligands and how the various properties of the blue copper proteins can be explained without protein strain [48]. Several other groups have also contributed to the understanding of these proteins by theoretical investigations [49-55].

Most of these properties could be studied in vacuum or by a simple point-charge [56] or continuum-solvent model [57] of the protein. However, for the inner-sphere reorganisation energies (i.e. the change in geometry, in energy terms, of the active site upon oxidation or reduction), the results were quite poor [58]. This prompted us to update the ComQum program with density functional methods and calculate the reorganisation energies inside several types of blue copper proteins (still with the B3LYP/6-31G* method) [15]. Interestingly, the protein environment decreased the reorganisation energy by a factor of two or more for all proteins, giving much more reasonable results (~30 kJ/mole). Recently, two other QM/MM studies of blue copper

proteins have been published, which do not include electrostatic interactions between the metal site and the surrounding protein [59,60]. Interestingly, these give geometries closer to the crystal structures, indicating some problems in the use of undamped electrostatics in the QM/MM calculations.

Encouraged by these results, we extended the investigation to the other two types of electron carriers in nature, cytochromes and iron–sulphur clusters (also with the B3LYP/6-31G* method) [61,62]. For the cytochromes, the reorganisation energies were small already in vacuum (8 kJ/mole). However, for the iron–sulphur clusters, inclusion of the protein environment reduced the reorganisation energies by a factor of two (to 12–57 kJ/mole). Thus, the inclusion of the surrounding protein with its hydrogen bonds and solvation effect is important for the reorganisation energies of electron carriers.

For many other properties, the protein environment is less important, provided that crucial protein residues are included in the calculations [1,2]. An example is the hydrogen-bond energy in myoglobin. A free haem group binds CO much stronger than O_2, but myoglobin reduces this preference by a factor of ~800 (17 kJ/mole). Traditionally, this has been explained by the protein forcing CO to bind in a bent mode (the ideal binding mode of O_2) [63]. Recently, it has been realised, however, that electrostatic interactions are more important for this differential bonding, in particular a hydrogen bond to the distal histidine residue [64]. We have estimated the strength of hydrogen bonds from imidazole to CO and O_2, both in vacuum and by making the structure more similar to that in the protein by constraining two or three angles or dihedrals (with the B3LYP/6-31G* method). The difference in hydrogen-bond energy between CO and O_2 amounts to 21 and 24 kJ/mole, with and without the constraints, respectively [65]. Recently, we repeated these calculations inside the protein with QM/MM methods [3], stimulated by an article suggesting that the discrimination is

actually based on strain [66]. Interestingly, the QM/MM calculations gave virtually the same result for the differential hydrogen-bond strength as the vacuum calculations, 21–22 kJ/mole. However, they gave also an estimate of the strain energy and information about the protonation state of the distal histidine residue. Other vacuum and QM/MM studies of myoglobin have given similar results [67-69].

Finally, we have studied the enzyme ferrochelatase [70], which is the terminal enzyme in haem synthesis, i.e. the one that inserts Fe^{2+} into the porphyrin ring. It is generally believed that the enzyme distorts the substrate to expose the lone-pair orbitals on the pyrrole nitrogen atoms. In fact, N-methylmesoporphyrin (MMP), which has a one of the pyrrole rings tilted 30º out of the porphyrin plane, is a strong inhibitor of the enzyme and antibodies raised against MMP have ferrochelatase activity [71]. The crystal structure of ferrochelatase in complex with MMP is known [72]. However, it has not been possible to obtaine a structure of the substrate protoporphyrin IX bound to the enzyme. We have estimated this structure with QM/MM methods to see how much it is distorted by the enzyme (with the B3LYP/6-31G* method) [70]. It turns out that the porphyrin ring becomes saddled with all pyrrole rings tilted 2–11º out of the average ring plane. Thus, the tilt is much smaller than for MMP.

We currently study several other proteins with QM/MM methods, e.g. laccase, nitrite reductase, metallo- -lactamase, and coenzyme B_{12} dependent enzymes. In particular, we try to develop methods to obtain stable enthalpies and free energies from QM/MM structures using various methods, e.g. free energy perturbations, Langevin-dipole, or Poisson–Boltzmann methods [73]. We use the enzyme catechol O-methyltransferase as a test case, because it catalyses a simple S_N2 reaction with a well-defined reaction coordinate. It has also been thoroughly studied by theoretical methods [74-76].

Quantum refinement

Many QM/MM projects starts with reoptimising a crystal structure, in order to get structures and energies that are comparable with other quantum chemical calculations. Unfortunately, there is no guarantee that the reoptimised structure stays close to the crystal structure; inaccuracies in the molecular mechanics force field may distort the structure. On the other hand, normal crystal structures involve significant errors (e.g. up to 0.4 Å in the metal–ligand distances), which together with systematic errors in the theoretical method will lead to nonsense energies if the crystal structure is used directly [77-79].

A natural solution to this dilemma is to include the crystallographic raw data in the QM/MM calculations (the structure factors are available in the Brookhaven protein databank for most structures). Protein crystallography typically consists of crystallisation of the protein, data collection, phase determination, model building, refinement, and validation of the model. An initial model built into an electron-density map usually contains many errors. To produce an accurate model, one must carry out several cycles of crystallographic refinement and rebuilding [80]. Refinement programs change the model (coordinates, occupancies, *B* factors, etc.) to improve the fit of the observed and calculated structure-factor amplitudes (typically estimated as the residual disagreement, the *R* factor).

Because of the limited resolution normally obtained for biomolecules, the experimental data are supplemented by some sort of chemical information, usually in the form of a MM force field [80]. Therefore, the refinement takes the form of a minimisation or simulated annealing calculation by molecular dynamics using an energy function of the form

$$E_{cryst} = w_A E_{Xray} + E_{MM} \qquad (2)$$

Here, E_{Xray} is a penalty function, describing how well the model agrees with the experimental data, typically a maximum likelihood refinement target [81,82]. E_{MM} is a normal MM energy function with bond, angles, dihedral, and non-bonded terms. Finally, w_A is weight factor, which is necessary because E_{MM} is in energy units, whereas E_{Xray} is in arbitrary units.

Thus, we can include restraints to the crystallographic raw data in a QM/MM calculation by replacing E_{MM123} in Eqn. 1 by E_{cryst} from Eqn. 2. This gives a QM/MM method that is restrained to be close to the crystal data. Thereby, we obtain a structure that is an optimum compromise between quantum chemistry and crystallography, i.e. a structure where the general geometry and flexible dihedral angles are determined by the crystal data, whereas the details (bonds and angles) of the quantum system is determined mainly by quantum chemistry (and that of the rest of the enzyme is determined by the MM potential function, as in a normal crystal structure). In particular, the quantum system is compatible with similar QM vacuum calculations (i.e. it contains the same systematic errors), so we can directly compare differences in the structure and energies.

In principle, we could have obtained a similar effect by keeping as much as possible of the surrounding protein fixed at the crystal coordinates. However, such a procedure would not guarantee that the quantum system is compatible with the crystallographic data. Moreover, it would propagate errors in the original crystal coordinates (they are not the raw data but the result of an involved interpretation of the data). With our procedure, we may improve the structure also outside the quantum system, because the estimated phases of the whole protein change when the coordinates of the quantum system are modified.

From a crystallographic viewpoint, this can equivalently be seen as a standard

crystallographic refinement, where we have replaced the E_{MM} term in Eqn. 2 by a QM calculation for a small part of the protein (the quantum system). This solves a serious problem in crystallography. For the normal amino acids, accurate force fields exist, which are based on statistical analysis of small-molecule data [83]. However, for unusual molecules, such as metal centres, substrates, inhibitors, etc., i.e. *hetero-compounds*, experimental data are often lacking or are less accurate. In particular, force constants for hetero-compounds are typically not available, so the crystallographer has to determine them himself. This is a complicated and error-prone procedure, which may make parts of the crystal structure less well-determined [84]. In particular, it constitutes a serious bottleneck in high-throughput crystallography.

In practise, we implemented such a procedure by replacing the MM program in ComQum by a crystallographic refinement program. We have chosen the free and widely used program CNS (Crystallography & NMR system) [85]. Thereby, we automatically get methods to treat crystallographically related interactions, to calculate the w_A factor in Eqn. 2, to make corrections for the bulk solvent, and to calculate various crystallographic quality criteria, such as electron-density maps and R factors. This yielded the *quantum refinement* method, as implemented in the ComQum–X program [16].

Test calculations on ferrochelatase

We have tested the ComQum-X method by performing a re-refinement of the structure of MMP bound to ferrochelatase, mentioned above [72]. The results obtained with the Becke–Perdew86/6-31G* method are shown in Table 1. It can be seen that structure is improved in terms of the R_{free} factor, which is an R factor calculated for a set of reflections (typically 5–10%) that is not used in the refinement of the structure. It is

an objective quality criterion that is used to avoid overfitting of the model (inclusion of additional parameters will always reduce the standard *R* factor) [86,87]. However, the improvement is not very large – R_{free} decreases from 0.2312 to 0.2307, although there are substantial changes in the structure of MMP, as can be seen in Figure 1. This is because R_{free} is a global property of the whole protein (with 308 residues and 2848 atoms).

Interestingly, the standard *R* factor does not improve in the same way as R_{free}. On the contrary, it tends to *increase* as R_{free} decreases. This illustrates that the original crystal structure is strongly optimised with respect to the normal *R* factor. Ideally (without overfitting), the two *R* factors should be equal. Therefore, both the decrease in R_{free} and the decrease in the difference between R_{free} and *R* flag an improvement in the structure [87].

In Figures 1 and 2, we compare the ComQum-X and crystal structures of MMP in ferrochelatase [72]. We can see that there are appreciable changes. The ComQum-X structure fits excellently into the electron density. The density is well-defined for the porphyrin ring, whereas the side groups are harder to position. Consequently, the largest differences between the two structures are seen for the side chains (up to 155 pm). However, also the ring atoms have moved (48 pm on the average). The difference is most pronounced around the tilted *A* ring. In the original structure, there is a sharp kink, whereas in the ComQum-X structure, there is a more gradual transition between the *A* ring and the rest of the porphyrin. Many of the differences are caused by small inconsistencies in the force field used in the original refinement [16]. This shows that crystal structures are sensitive to the MM force field used in the refinement and that possible errors in it will propagate to the final coordinates.

Table 1 also shows how the tilt angle (the angle of the methylated pyrrole ring out of the porphyrin ring plane) and the strain energy (E_1; the energy difference between structures optimised in vacuum and with ComQum-X) vary with the w_A factor in Eqn. 2.

The tilt angle does not change much, whereas the strain energy is very sensitive to w_A. Neither of them approache the crystal value as w_A is increased. This illustrates that also the original crystal structure involves a compromise between the crystallographic raw data and a MM force field. Likewise, they do not go towards the vacuum value at low values of w_A, owing to van der Waals interactions between MMP and surrounding protein. Consequently, the choice of the w_A factor is crucial for the results. In CNS, it is determined so that the MM and crystallographic forces have a similar magnitude [88-90]. An alternative way is to select the value of w_A that gives the lowest value for the R_{free} factor; this is found for $w_A = 0.1$.

The calculations in Table 1 were performed with the whole MMP molecule in the quantum system (81 atoms). This way, we avoid to break any covalent bonds between the QM and MM systems. However, a more typical situation is to cut off the side chains of the MMP molecule in the QM calculations. We have also tried such calculations [16]. They give a similar tilt angle (37º), but a higher R_{free} factor (0.2310) and a smaller strain energy, 1–10 kJ/mole. The increase in R_{free} is caused by problems in the junctions between the QM and MM systems – the calculations are very sensitive to the MM force field around these junctions. If this force field was completely removed for MMP, the R_{free} factor decreased to 0.2308. Therefore, it seems to be advantageous to remove the force field of the quantum system and the junctions, unless it is very accurate.

Next, we compared the ComQum-X results with those obtained with standard QM/MM methods. First we ran a normal QM/MM calculation with ComQum, keeping the protein fixed at the crystal structure. As can be seen from Table 2, this gave similar results to those of ComQum-X, with a R_{free} factor of 0.2312. However, if the enzyme is allowed to relax, the R_{free} factor increases drastically (to 0.248), showing that the system starts to diverge from experimental data. If the protein is equilibrated with the MM force field before the QM/MM calculation, the R_{free} factor increases to 0.46.

This is also reflected in the structure of MMP. As can be seen in Figure 3, in the

calculation with a fixed enzyme, the structure of MMP is similar to that obtained with ComQum-X. However, if the protein is allowed to relax, the MMP structure becomes almost planar, like the optimum vacuum structure of MMP. Consequently, it is strongly advisable to keep the surroundings fixed (heavy atoms fixed at the crystal structure, with optimised positions of the hydrogen atoms) during QM/MM calculations if the aim of the calculation is to study the crystal conformation. Even if another conformation is studied, it is probably wise to relax as little of the surroundings as possible. Similar results have been obtained by comparing QM/MM and linear-scaling methods [91].

The ComQum-X calculations can also be used to estimate how much MMP is strained when bound to the protein. This is far from trivial [48,92]: E_1 in Table 1 includes terms that are not normally considered as strain, especially if the polar propionate side chains are included in the calculations [16]. Our best estimate of the strain energy is 6 kJ/mole, obtained without side chains and without any bonded MM interactions for the quantum system. This strain energy is lower than E_1 values observed in other QM/MM calculations 20–200 kJ/mole [3,14,15,36,38,40,62,70,93], but these calculations have invariably involved polar groups. However, it is fully in line with the suggestion that a molecule bound to a protein is in general strained by less than 10 kJ/mole if a proper reference state is used [94]. If we use instead the original crystal structure, the strain energy of MMP is very large, 265 kJ/mole with and 157 kJ/mole without the side chains. This illustrates the problem of using the crystal structures directly.

Test calculations on cytochrome c_{553}

The calculations on ferrochelatase were not fully satisfactorily because they gave so small changes in the R_{free} factor and that no other quality criteria seemed to be useful [16,79]. However, the problem is delicate, because the QM calculations are used to *improve* the crystal structure; if the goal was only to give the best fit to the experimental

data (i.e. to minimise the *R* factors), we should not use any QM calculation or MM force field at all. Yet, crystallographic experience shows that small errors in the raw data leads to chemically unreasonable structures with strange bond lengths and angles, at least for low- and medium-resolution protein structures. The MM force field is used in standard refinement to remedy this problem at the expense of giving a slightly worse fit to the experimental data. We claim that we can improve the result even more with QM calculations for parts of the system. Such a suggestion is reasonable, because density functional calculations with a medium-sized basis set typically reproduce experimental bond lengths within 2 pm for organic molecules and within 2–7 pm for bonds to metal ions [61,95,96], whereas low- and medium-resolution protein structures show *average* errors of ~10 pm [77,78], and much larger errors are frequently found, as we will see below.

A way to solve this dilemma is to find a protein that has been solved at both low and atomic resolution (where geometric restraints have a small influence on the structure), but otherwise at as similar conditions as possible. We could then investigate how close a ComQum-X structure refined with the low-resolution data is to the high-resolution structure and how well it fits to the high-resolution density map. We have found such a pair in the Brookhaven protein databank, which contains an appropriate hetero-compound: Cytochrome c_{553} from *Bacillus pasteurii* has been solved at 97 pm resolution with ab initio phasing and independently by the same group at 170 pm resolution in an multiple anomalous dispersion experiment [97]. The crystals were obtained at the same conditions. This protein contains a haem group, where the central Fe^{3+} ion binds also to a histidine and a methionine residue from the protein. The two crystal structures show an appreciable difference in the iron–ligand bond lengths, as can be seen in Table 3.

We have optimised the structure of this haem group with ComQum-X (Becke–

Perdew86/6-31G* method), using the low-resolution data [98]. The results are included in Table 3 and show that ComQum-X brings the structure much closer to the high-resolution structure: The error in the Fe–N_{His} bond length is reduced from 32 to 0–1 pm (for w_A = 0.1–0.3), that of the Fe–S_{Met} bond length is reduced from 12 to 1–3 pm, and those of the Fe–N_{Por} bond lengths are reduced from 3–9 to 0–3 pm. This is of course a manifestation of the excellent performance of density functional theory for this metal site; already the vacuum structure gives errors of only 1–4 pm.

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 3). It can be seen that it is improved by 0.0043 (this is ten times more than for ferrochelatase, because cytochrome c_{553} is a much smaller protein with 667 atoms). We can calculate a similar *R* factor based on the high-resolution reflections. These are also given in Table 3 (R_{high}) and show a similar improvement.

This is even clearer when we compare the ComQum-X structure with the low- and high resolution crystal structures in Figure 4. It is mainly the iron ion (19 pm), the N ^{1} atom (27 pm), and one of the ethyl side chains (210 pm) that have moved (the average movement of all atoms is 12 pm), and their positions in the ComQum-X structure are closer to those in the high- than in the low-resolution structure. The high-resolution electron-density map confirms that ComQum-X has improved the low-resolution structure.

Protonation of metal-bound solvent molecules

We have seen that ComQum-X works properly and that it can locally improve low- and medium-resolution protein crystal structures. Therefore, we could look for appropriate applications of the method. One of the most useful applications is probably

to *interpret* crystal structures, i.e. to decide exactly what chemical species are present in the structure.

Hydrogen atoms are not normally seen in protein structures. Therefore, ComQum-X could be used to decide where they are in the structure, e.g. to determine the protonation status of various molecules in the structure. There have been many attempts to calculate pK_a values in proteins using theoretical methods [99,100], but none gives very reliable results, especially not for metal-bound solvent molecules. Therefore, it would be highly interesting to test if we can determine the protonation status of metal-bound solvent molecules using ComQum-X, especially as it is 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 ~30 pm shorter than the corresponding bond to water (cf. Table 4). 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 test the method, we first need 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 [29]. We have studied a complex between alcohol dehydrogenase, NAD^+, and trifluoroethanol (at 200 pm resolution) [101], in which the alcohol should have a pK_a of ~6 [29]. 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 the ComQum-X structures of this complex with both an alkoxide and a protonated alcohol (Becke–Perdew86/6-31G* method) [28]. The results in Table 4

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). However, the difference is not very large and both possibilities give a lower R_{free} factor than the original crystal structure.

Second, the alkoxide gives a lower strain energy (E_1) than the alcohol for all values of w_A. This indicates that the alkoxide fits better into the electron density than the alcohol.

Third, the Zn–O distance in the alkoxide structure is close to that found in vacuum calculations at all values of w_A, whereas for the alcohol, the Zn–O distance is far from the vacuum value 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. The same appears to be true also for the other Zn–ligand distances.

Likewise, we have studied another crystal structure of alcohol dehydrogenase (also at 200 pm resolution) [102], for which the experimental data indicate that the zinc-bound water molecule is protonated. With the same three criteria as above, we show that the a zinc-bound water molecule fits the crystal data better than a hydroxide ion [28]. Thus, it is clear that ComQum-X can discern the two protonation forms of the zinc-bound solvent molecules in alcohol dehydrogenase.

This opens an important area of applications for ComQum-X. We have used it to study the protonation status of MMP in ferrochelatase [28], of the iron-bound water molecule in superoxide dismutase [28], and of compound II in myoglobin [103]. Finally, we have started to study available crystal structures of hydrogenases. These

enzymes perform the formally simple conversion of protons and electrons to a hydrogen molecule. However, because it involves protons and hydrogen, the substrate and product are not seen in crystal structures. Therefore, it is an ideal project for ComQum-X. Moreover, the *oxidation states* of the active-site metal ions (two irons or one iron and one nickel ion) are not known with certainty [104-107]. For example, it has recently been suggested that the reaction cycle of iron-only hydrogenase involves the Fe(I) oxidation state, unprecedented in biological systems [108].

Other applications of ComQum-X

Many other applications of ComQum-X are conceivable. We have studied CO–myoglobin to test if ComQum-X is useful also for a high-resolution structure [3]. There exist two recent structures of this complex, both at 115 pm resolution [66,109]. Still, they differ by 11 pm in the predicted Fe–CO bond length (cf. Table 5). We have re-refined one of the crystal structures using ComQum-X (with the B3LYP/6-31G* method) [3]. At this high resolution it is likely that the systematic overestimation of metal–ligand distances by density functional methods becomes significant [61,95]. Therefore, we investigated if such systematic errors can be corrected by the method of offset forces [110]. Table 5 shows that the ComQum-X results are unexpectedly insensitive to the correction. ComQum-X gives a Fe–C–O angle similar to that observed 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 (170 pm, compared to 173 and 182 pm) and a C–O distance that is longer (116 pm, compared to 109 and 113 pm) 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 (174 and 116 pm) [111]. The improvement of the crystal structure is flagged by a decrease in the R_{free} factor of 0.0013 [3]. Thus, appreciable errors (>10 pm)

in metal–ligand bond lengths can be expected also in high-resolution crystal structures.

Finally, we have studied the structure of 5'-deoxyadenosine in the crystal structure of methylmalonyl coenzyme A mutase (220 pm resolution) [112]. This enzyme binds adenosylcobalamin (coenzyme B_{12}). During catalysis, the Co–C bond breaks, giving rise to a 5'-deoxyadenosyl radical that extracts a hydrogen atom from the substrate, leading to a radical-based reorganisation. The structure is believed to contain a mixture of the substrate and product (methylmalonyl and succinyl coenzyme A) and a 5'-deoxyadenosine molecule (not a radical) but only with partial (0.5) occupancy.

However, the structure of 5'-deoxyadenosine in the crystal structure is strange, with a very short interaction between the C5' and C8 atoms (211 and 235 pm in the two subunits). This is intermediate between what is expected for a covalent bond (~150 pm) and a non-bonded interaction (>300 pm). We have therefore reoptimised the structure of this molecule with ComQum-X [113]. This gives a C5'–C8 distance of 336 pm. However, the distance is sensitive to the theoretical treatment, in particular to the van der Waals parameters of the two atoms. In fact, the structure can be re-refined with a normal MM force field also, giving C5'–C8 distances of 332–383 pm, depending on the parameters used [79]. All calculations give similar R_{free} factors (0.2681; the protein contains 22224 atoms) and it is hard to decide from electron-density maps which structure fits the experimental data best, although there are appreciable differences in the geometry [79]. The take-home message is that there may be errors of 170 pm for non-bonded interactions in medium-resolution crystal structures.

Concluding remarks

We have seen that QM/MM is a powerful approach to study geometry, properties, and energies of metalloproteins [114]. In particular, ComQum-X seems to be a promising method to interpret and improve crystal structures. Not the least, it is very informative for a theoretical biochemist to get an insight into the problems of building a

correct model into the experimental electron density and to learn how large errors actually can be encountered in crystal structures.

ComQum-X is far from fully developed. Many improvements can still be done. As mentioned above, hydrogen atoms are not visible in X-ray structures. Therefore, they are normally ignored in the refinement. Consequently, electrostatic interactions are also normally ignored. In the first version of ComQum-X, we have followed this custom. However, some of the results show that this is not optimal – it gives rise to spurious vacuum effects between polar groups in the QM calculations, which are not compensated for in the MM calculations. Moreover, hydrogen-bonds and solvation effects are ignored, although they significantly affects the QM structure of many systems [38,62].

On the other hand, inclusion of hydrogen atoms is also problematic. Since the hydrogens are not visible, we have to speculate about their positions in the structure. For some hydrogens, this is quite straight forward. However, for the side chains of cysteine, serine, and threonine, there is a rotational freedom of the hydrogen atom, which will strongly affect the electrostatics, and this freedom is even larger for water molecules. Moreover, for histidine, it is not even clear to what atoms the hydrogens should be connected. Thus, the addition of hydrogen atoms is not automatic and it depends on the pH. Therefore, there is a great risk to make erroneous assignments, thereby giving a suboptimal structure. We currently try to solve these problems.

Moreover, we want to develop appropriate parameter values for the quantum refinement. Up to now, we have selected the w_A weight by running several ComQum-X calculations, using the one that gives the lowest R_{free} factor. However, it is questionable if the accuracy of R_{free} justifies such a use (we look at changes in the fourth decimal) [79], even if it usually behaves well (i.e. it gives a well-defined minimum as a function

of the w_A weight and converges smoothly during the geometry optimisations). Perhaps, it is better to use the default CNS value in all calculations, especially if we want to look at the electron-density maps (they show a clear deterioration if we use too small values of w_A).

Likewise, it is not clear how to weight the MM and QM constraints. The reason for this is that the MM force field in CNS and other refinement programs is not based on energy considerations, but rather on a statistical analysis of available crystal structures [83]. Therefore, the force constants in the CNS force field are in arbitrary units and are not directly comparable to the QM energy function. Traditionally, it is assumed that the statistical force constants are approximately three times larger than energy-based force constants [83] and therefore, the QM energy function has been weighted up by a factor of three before it is combined with the CNS energy function. However, it remains to be shown that this is an optimal choice. We currently investigate this issue.

Finally, it should be mentioned that there are alternative ways of introducing QM data in crystallographic refinements. You could also calculate a Hessian matrix with any theoretical method and then extract a MM force field from this matrix, e.g. using the method suggested by Seminario [115]. We have implemented and tested such an approach to automatically obtain topology and parameter files for any hetero-compound [79]. Such a method also works properly, especially when based on a Hessian matrix obtained with a density functional method. It often gives results faster than ComQum-X. However, there is always the risk that the force field is a poor approximation of the potential around the geometry observed in the protein.

We see many possible applications of ComQum-X, besides obtaining improved structures, interpreting crystal structures, and calculating strain energies. For example, it can be used to test force fields for hetero-compounds in standard crystallographic

refinement. Moreover, it is a very powerful method to test various treatments of junctions in standard QM/MM methods. We have already seen that we could detect problems in the junctions for MMP in ferrochelatase, which could be solved by removing all MM parameters for the quantum system. However, it would be interesting to test also other methods to treat the junctions, based on pseudoatoms, core potentials, or localised hybrid orbitals [13].

Finally, the most important use of ComQum-X will probably be to obtain starting structures for other QM or QM/MM studies. A prerequisite for ComQum-X is that there exists an appropriate crystal structure with exactly the atoms of interest. Of course, this is not always the case, especially if we want to study the full reaction mechanism of a protein. Then, ComQum-X can be used to obtain a starting structure, involving the correct protonation status of all residues and molecules, the correct oxidation state of the metals, and an optimum compromise between quantum chemistry and the crystallographic raw data.

Acknowledgements

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

References

1. P. E. M. Siegbahn and M. R. A. Blomberg, Annu. Rev. Phys. Chem., 50 (1999) 221–249.
2. P. E. M. Siegbahn and M. R. A. Blomberg, Chem. Rev., 100, (2000) 421–437.
3. E. Sigfridsson and U. Ryde, J. Inorg. Biochem., 91 (2002) 116-124.
4. A. Warshel and M. Levitt, J. Mol. Biol., 103 (1976) 227-249.

5. U. C. Singh and P. A. Kollman, J. Comput. Chem., 7 (1986) 718-730.
6. G. Monard and K. M. Merz, Acc. Chem. Res., 32 (1999) 904-911.
7. M. Orozco and F. J. Luque, Chem. Rev., 100 (2000) 4187-4225.
8. F. Maseras, Chem. Commun. (2000) 1821-1827.
9. W. Wang, O. Donini, C. M. Reyes and P. A. Kollman, Annu. Rev. Biophys. Biomol. Struct., 30 (2001) 211-243.
10. V. Gogonea, D. Suárez, A. van der Vaart and K. M. Merz, Curr. Opin. Struct. Biol., 11 (2001) 217-223.
11. A. J. Mulholland In Theoretical biochemistry – processes and properties of biological systems (Theoretical and computational chemistry, Vol. 9). Edited by L. A. Eriksson. Amsterdam Elsevier Science; (2001) 597-653.
12. J. Gao and D. G. Truhlar, Annu. Rev. Phys. Chem., 53 (2002) 467-505.
13. M. J. Field, J. Comput. Chem., 23 (2002) 48-85.
14. U. Ryde, J. Comput.-Aided Mol Design, 10 (1996) 153-164.
15. U. Ryde and M. H. M. Olsson, Intern. J. Quant. Chem., 81 (2001) 335-347.
16. U. Ryde, L. Olsen and K. Nilsson, J. Comp. Chem., 23 (2002) 1058-1070.
17. M. J. Field, P. A. Bash and M. Karplus, J. Comput. Chem. 11 (1990) 700.
18. N. Reuter, A. Dejaegere, B. Maigret and M. Karplus, J. Phys. Chem. A, 104 (2000) 1720-1735.
19. R. M. Nicoll, S. A. Hindle, G. MacKenzie, I. H. Hillier and N. A. Burton, Theor. Chem. Acc., 106 (2001) 105-112.
20. Svensson, M., Humbel, S., Froese, R. D. J., Matsubara, T., Sieber, S., Morokuma, K. J Phys Chem 1996, 100, 19357.
21. F. Maseras and K. Morokuma, J. Comput. Chem., 16 (1995) 1170-1179.
22. D. Treutler and R. Ahlrichs, J. Chem. Phys., 102 (1995) 346.
23. O. Teleman and B. Jönsson, J. Comput. Chem., 7 (1986) 58-66.

24. U. Ryde, Proteins, Struct. Funct. Genet., 21 (1995) 40-56.

25. D. A. Case, D. A. Pearlman, J. W. Caldwell, T. E. Cheatham III, W. S. Ross, C. L. Simmerling, T. A. Darden, K. M. Merz, R. V. Stanton. A. L. Cheng, J. J. Vincent, M. Crowley, D. M. Ferguson, R. J. Radmer, G. L. Seibel, U. C. Singh, P. K. Weiner and P. A. Kollman, Amber 5, University of California, San Francisco, 1997.

26. K. Andersson, M. Barysz, A. Bernhardsson, M. R. A. Blomberg, D. L. Cooper, M. P. Fülscher, C. de Graaf, B. A. Hess, G. Karlström, R. Lindh, P.-Å. Malmqvist, T. Nakajima, P. Neogrády, J. Olsen, B. O. Roos, B. Schimmelpfenning, M. Schütz, L. Seijo, L. Serrano-Andrés, P. E. M. Siegbahn, J. Stålring, T. Thorsteinsson, V. Veryazov and P.-O. Widmark, MOLCAS version 5.4, Department of theoretical chemistry, Lund university, P.O.Box 124, S-221 00 Lund, Sweden (2002).

27. R. Poteau, I. Ortega, F. Alary, A. R. Solis, J.-C. Bathelat and J.-P. Daudey, J. Phys. Chem. A, 105 (2001) 198-205.

28. K. Nilsson and U. Ryde, Nature, Struct Biol, to be submitted.

29. G. Pettersson, CRC Crit. Rev. Biochem., 21 (1987) 349-389.

30. U. Ryde, Int. J. Quant. Chem., 52 (1994) 1229-1243.

31. I. Bertini and C. Luchinat, Metal Ions. Biol. Syst. 15 (1983) 101-156.

32. O. Tapia, R. Cardenas, J. Andres, J. Krechl, M. Campillo and F. Colonna-Cesari, Int. J. Quantum Chem. 39 (1991) 767-786.

33. A. R. von Onciul & T. Clark, J. Comput. Chem. 14 (1993) 392-400.

34. M. T. Werth, S.-F. Tang, G. Formicka, M. Zeppezauer and M. K. Johnson, Inorg. Chem. 34 (1995) 218-228.

35. L. Hemmingsen, R. Bauer, M. J. Bjerrum, M. Zeppezauer, H. W. Adolph, G. Formicka and E. Cedergren-Zeppezauer, Biochemistry 34 (1995) 7145-7153.

36. U. Ryde, Protein Science, 4 (1995) 1124-1132.

37. Z.-N. Yang, W. F. Bosron and T. D. Hurley, J. Mol. Biol. 265(97)330-343.

38. U. Ryde, Eur. J. Biophys., 24 (1996) 213-221.

39. L. Hemmingsen and U. Ryde, J. Phys. Chem., 100 (1996) 4803-4809.

40. U. Ryde and L. Hemmingsen, J. Biol. Inorg. Chem., 2 (1997) 567-579.

41. C. Alhambra, J. C. Corchado, M. L. Sánchez, J. Gao and D. G. Truhlar, J Am Chem Soc, 122 (2000) 8197-8203.

42. C. Alhambra, J. C. Corchado, M. L. Sánchez, J. Gao and D. G. Truhlar, J Phys Chem B, 105(2001) 11326-11340.

43. G. Tresandern, J. P. McNamara, M. Mohr, H. Wang, N. A. Burton and I. H. Hillier, Chem Phys Lett, 258 (2002) 489-494.

44. Q. Cui, M. Elstner and M. Karplus, J Phys Chem B 106 (2002) 2721-2740.

45. B. G. Malmström, Eur. J. Biochem., 223 (1994) 711-718.

46. R. J. P. Williams, Eur. J. Biochem., 224 (1995) 363-381.

47. U. Ryde, M. H. M. Olsson, K. Pierloot and B. O. Roos, J. Mol. Biol., 261 (1996) 586-596.

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

49. A. A. Gewirth and E. I. Solomon, J. Am. Chem. Soc, 110 (1988), 3811-3819.

50. J. A. Guckert, M. D. Lowery and E. I. Solomon, J. Am. Chem. Soc, 117 (1995), 2817-2844.

51. L. B. LaCroix, S. E. Shadle, Y. Wang, B. A. Averill, B. Hedman, K. O. Hodgson and E. I. Solomon, J. Am. Chem. Soc, 118 (1995), 7755-7768.

52. L. B. LaCroix, D. W. Randall, A. M. Nersissian, C. W. G. Hoitink, G. W. Canters, J. S. Valentine and E. I. Solomon, J. Am. Chem. Soc, 120 (1998), 9621-9631.

53. D. W. Randall, D. R. Gamelin, L. B. LaCroix and E. I. Solomon, J. Biol. Inorg.

Chem. 5 (2000) 16-29.

54. S. Larsson, A. Broo and L. Sjölin, J. Phys. Chem. 99 (1995) 4860-4865.

55. S. Larsson, J. Biol. Inorg. Chem. 5 (2000) 560-564.

56. K. Pierloot, J. O. A. De Kerpel, U. Ryde, M. H. M. Olsson and B. O. Roos, J. Am. Chem. Soc., 120 (1998) 13156-13166.

57. M. H. M. Olsson and U. Ryde, J. Biol. Inorg. Chem., 4 (1999) 654-663.

58. M. H. M. Olsson, U. Ryde and B. O. Roos, Prot. Sci., 7 (1998) 2659-2668.

59. P. Comba, A. Lledós, F. Maseras and R. Remenyi, Inorg Chim Acta, 324 (2001) 21-26.

60. M. Swart, M. van den Bosch, H. J. C. Berendsen, G. W. Canters and J. G. Snijders, to be published, described in M. Swart, Ph. D. Thesis, Rijksuniversität Groningen, 2002.

61. E. Sigfridsson, M. H. M. Olsson and U. Ryde, J. Phys. Chem. B, 105 (2001) 5546-5552.

62. E. Sigfridsson, M. H. M. Olsson and U. Ryde, Inorg. Chem., 40 (2001) 2509-2519.

63. L. Stryer, In, Biochemistry, 3rd edn. Freeman, New York, 1988, p 149.

64. J. S. Olson and G. N. Philips, J. Biol. Inorg. Chem. 2 (1997) 544-552.

65. E. Sigfridsson E and U. Ryde, J Biol Inorg Chem, 4 (1999) 99-110.

66. G. S. Kachalova, A. N. Popov and H. D. Bartunik, Science 284 (1999) 473-476.

67. C. Rovira, K. Kunc, J. Hutter, P. Ballone, M. Parrinello, J. Phys. Chem. A 101 (1997) 8914-8925.

68. T.G. Spiro, P.M. Kozlowski, Acc. Chem. Res. 34 (2001) 137-144.

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

70. E. Sigfridsson and U. Ryde, J. Biol. Inorg. Chem., (2002) in press; DOI: 10.1007/s00775-002-0413-8.

71. A. G. Cochran and P. G. Schultz, Science, 249 (1990) 781-783.

72. D. Lecerof, M. Fodje, A. Hansson, M. Hansson and S. Al-Karadaghi, J. Mol. Biol. 297 (2000) 221-232.

73. T. Rasmussen, K. Nilsson and U. Ryde, J. Phys. Chem. B, to be submitted.

74. Y.-J. Zheng and T. C. Bruice, J. Am. Chem. Soc. 122 (2000) 8137-8145.

75. K. Kahn and T. C. Bruice, J. Am. Chem. Soc. 122 (2000) 46-51.

76. B. Kuhn and P. A. Kollman, J. Am. Chem. Soc. 122 (2000) 2586-2596.

77. D. W. J. Cruickshank, Acta Cryst., D55 (1999) 583-601.

78. B. A. Fields, H. H. Bartsch, H. D. Bartunik, F. Cordes, J. M. Guss and H. C. Freeman, Acta Cryst., D50 (1994) 709-730.

79. K. Nilsson, D. Lecerof, E. Sigfridsson and U. Ryde, Acta Crystallogr. D, 59 (2003) 274-289.

80. G. J. Kleywegt and T. A. Jones, Meth. Enzymol., 227 (1997) 208-230.

81. N. S. Pannu and R. J. Read, Acta Cryst. A52 (1996) 659-668.

82. P. D. Adams, N. S. Pannu, R. J. Read and A. T. Brünger, Proc. Natnl. Acad. Sci. US, 94 (1997) 5018-5023.

83. R. A. Engh and R. Huber, Acta Cryst., A47 (1991) 392-400.

84. G. J. Kleywegt and T. A. Jones, Acta Cryst., D54 (1998) 1119-1131.

85. 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.

86. A. T. Brünger, Acta Cryst. D49 (1993) 24-36.

87. G. J. Kleywegt and T. A. Jones, Structure 3 (1995) 535-540.

88. Jack, A; Levitt, M. Acta Cryst 1978, A34, 931.

89. Brünger, A. T.; Karplus, M.; Petsko, G. A. Acta Cryst 1989, A45, 50.

90. Brünger, A. T. Methods Enzymol 1997, 277, 243-269.

91. S. J. Titmuss, P. L. Cummins, A. A. Bliznyuk, A. P. Rendell and J. E. Gready, Chem. Phys. Lett. 2000, 320169-176.

92. U. Ryde, In Recent Research Developments in Protein Engineering, 2 (2002) 65-91. Research Signpost, Trivandrum.

93. M. Torren, T. Vreven, D. G. Musaev and K. Morokuma, J. Am. Chem. Soc., 124 (2001) 192-193.

94. Boström, J.; Norrby, P.-O.; Liljefors, T. J. Comp.-Aided Mol. Design. 1998, 12, 383-396.

95. M. H. M. Olsson and U. Ryde, J. Am. Chem. Soc., 123 (2001) 7866-7876.

96. F. Jensen, Introduction to Computational Chemistry, J. Wiley & Sons, Chichester, 1999.

97. S. Benini, A. González, W. R. Rypniewski, K. S. Wilson, J. J. Van Beeumen and S. Ciurli, Biochemistry, 39 (2000) 13115-13126.

98. U. Ryde & K. Nilsson, J. Am. Chem. Soc., to be submitted.

99. I. A. Topol, S. K. Burt, A. A. Rashin and J. W. Erickson, J. Phys. Chem. A 104 (2000) 866-872.

100. G. M. Ullman, L. Noodlemann and D. A. Case, J. Biol. Inorg. Chem., 7 (2002) 632-639.

101. B. J. Bahnson, T. D. Colby, J. K. Chin, B. M. Goldstein and J. P. Klinman, Proc. Natl. Acad. Sci. USA, 94 (1997) 12797-12802.

102. J. K. Rubach, S. Ramaswamy and B. V. Plapp, Biochemistry, 40 (2001) 12686-12694.

103. K. Nilsson, U. Ryde and K. K. Andersson, J. Biol. Inorg. Chem., to be submitted.
104. M. J. Maroney and P. A. Bryngelson, J. Biol. Inorg. Chem., 6 (2001) 453-459
105. P. E. M. Siegbahn, M. R. A. Blomberg, M. Wirstam and R. H. Crabtree, J. Biol. Inorg. Chem., 6 (2001) 460-466.
106. H.-J. Fan and M. B. Hall, J. Biol. Inorg. Chem., 6 (2001) 467-473.
107. M. Stein and W. Lubitz, Curr. Opin. Chem. Biol., 6 (2002) 243-249.
108. M. B. Hall, Z. X. Cao, H.-J. Fan, S. H. Li, S. Q. Niu and L. Thomson L, J. Inorg. Biochem. 86 (2001) 245-245.
109. J. Vojtechovský, K. Chu, J. Berendzen, R. M. Sweet and I. Schlichting, Biopys. J., 77 (1999) 2153-2174.
110. G. Fogarasi, X. Zhou, P. W. Taylor and P. Pulay, J. Am. Chem. Soc. 114 (1992) 8191-8201.
111. R. Salzmann, M. T. McMahon, N. Godbout, L. K. Sanders, M. Wojdelski and E. Oldfield, J. Am. Chem. Soc. 121 (1999) 3818-3828.
112. F. Mancia, G. A. Smith and P. R. Evans, Biochemistry, 38 (1999) 7999-8005.
113. K. P. Jensen & U. Ryde, manuscript in preparation.
114. U. Ryde, Curr. Opin. Chem. Biol., in press.
115. J. M. Seminario, Intern. J. Quant. Chem., Quant. Chem. Symp., 30 (1996) 59-65.

Scheme 1. A flow scheme of ComQum. Steps in bold face constitute the actual ComQum interface. The other steps are performed either by the QM or the MM program, whereas the whole procedure is run by a simple UNIX shell script. S1, S2, and S3 denotes systems 1, 2, and 3.

```
Evaluate QM wavefunction
Repeat
    Evaluate the QM forces (from S1–S3 onto S1)
    Evaluate the MM forces (from S2–S3 onto S1)
    Add the forces
    Relax the geometry of S1 using these forces
    Change the coordinates of S1 in MM representation
    If S2 is relaxed
        Calculate the QM charges of S1
        Insert them into the MM representation
        Relax S2 by MM minimisation with S1 and S3 fixed
        Change the coordinates of S2 in QM representation
    Evaluate the QM wavefunction and energy of S1
    Evaluate the MM potential energy
    Add the energies
until convergence
```

Table 1. Variation of the tilt angle (°), strain energy (E_1, kJ/mole), and *R* factors with the w_A factor when MMP is optimised in ferrochelatase with ComQum-X (Becke–Perdew86/6-31G* method) [16]. For comparison, data for the crystal structure and a geometry optimisation in vacuum with the same QM method are also included.

w_A	*Tilt angle*	E_1	R_{free}	*R*
Crystal	37.2	264.5	0.2312	0.1827
30	42.2	333.6	0.2313	0.1826
3	39.4	74.2	0.2311	0.1827
0.87	38.3	52.7	0.2310	0.1829
0.3	37.0	41.8	0.2310	0.1832
0.1	36.3	34.6	0.2307	0.0000
0.01	35.5	24.9	0.2312	0.1865
0	35.3	24.3	0.2313	0.1867
QM	29.9	0.0	0.2360	0.1886

Table 2. Tilt angle (°), strain energy (E_1, kJ/mole), and *R* factors for three standard QM/MM (COMQUM) optimisations of MMP (without side chains) in ferrochelatase (B3LYP/6-31G* method) [16]. The calculations differ in whether the MM system (all amino acids within 0.8 nm of MMP) is allowed to relax or not.

MM relaxed?	Tilt	E_1	R_{free}	R
No	38.7	19.1	0.23115	0.18276
Yes	37.7	21.4	0.24834	0.20833
Yes[a]	29.8	13.1	0.45926	0.45450

[a] In this calculation, the protein was first equilibrated.

Table 3. Fe–ligand distances, strain energies (E_1, kJ/mole), and *R* factors for the haem group in cytochrome c_{553} calculated with COMQUM-X using the low-resolution data [98]. For comparison, the low- and high-resolution crystal structures [97] and the result of a QM vacuum calculation are also included. The quantum systems consisted of Fe^{III}(porphine)(imidazole)($S(CH_3)_2$) in the low-spin doublet state, and it was studied with the density functional Becke–Perdew86/6-31G* method.

w_A	Distance to Fe (pm)			E_1	R_{low}	R_{high}
	N_{His}	S_{Met}	N_{Por}			
Low	231	221	202–208		0.00000	0.00000
High	199	233	197–200			
1	202	228	197–200	37.9	–0.00430	–0.0176
0.3	200	230	198–201	33.6	–0.00433	–0.0171
0.15	200	231	199–201	35.5	–0.00421	–0.0164
0.1	199	232	199–201	36.7	–0.00396	–0.0160
0.01	199	236	200–201	42.2	–0.00257	–0.0117
Vacuum	200	235	200–201	0.0		

Table 4. Zn–ligand distances (the ligands are histidine, two cysteines, and the alcohol), strain energies (E_1, kJ/mole), and *R* factors for the catalytic zinc ion in alcohol dehydrogenase in complex with NAD^+ and trifluoroethanol calculated with COMQUM-X (Becke–Perdew86/6-31G* method) [28]. For comparison, the original crystal structure [101] and the result of QM vacuum optimisations are also included

Ligand	w_A	Distance to Zn (pm)			E_1	R_{free}
		N	S	O		
ROH	Vacuum	209	225–229	229	0	
RO^-	Vacuum	224	233	193	0	
ROH	10	218	233–237	201	119	0.2283
ROH	3	215	231–234	204	93	0.2283
ROH	1.76	214	230–232	207	87	0.2285
RO^-	3	228	236	190	62	0.2280
RO^-	1.85	228	234–236	190	54	0.2280
RO^-	0.3	225	231–233	192	54	0.2279
	Crystal	213–220	213–229	205–207		0.2303

Table 5. Geometric parameters (distances in pm, angle in degrees) of the ComQum-X structure of CO–myoglobin (B3LYP/6-31G* method), without or with a correction of the systematically too long Fe–ligand distances (and the C–O distance) with offset forces, and with or without a model of the distal histidine residue in the quantum system [3]. The results are compared with the two most accurate structures of this protein complex (both at 115 pm resolution) [66,109].

Method	$Fe–N_{His}$	$Fe–N_{Por}$	Fe–C	C–O	Fe–C–O
Uncorrected	207.0	199–204	169.0	117.8	170.3
Corrected	208.0	199–203	170.5	116.5	170.9
Uncorrected[a]	208.0	199–203	171.3	116.2	170.8
Corrected[a]	207.8	199–203	170.0	116.8	170.9
1bzr [66]	211.2	199–203	173.1	112.6	171.3
1a6g [109]	206.2	194–202	182.2	109.2	171.2

[a] In these calculations, a model of His-64 was included in the quantum system.

Figure 1. A comparison of the original crystal structure (green) [72] and the re-refined ComQum-X structure of MMP in ferrochelatase [16].

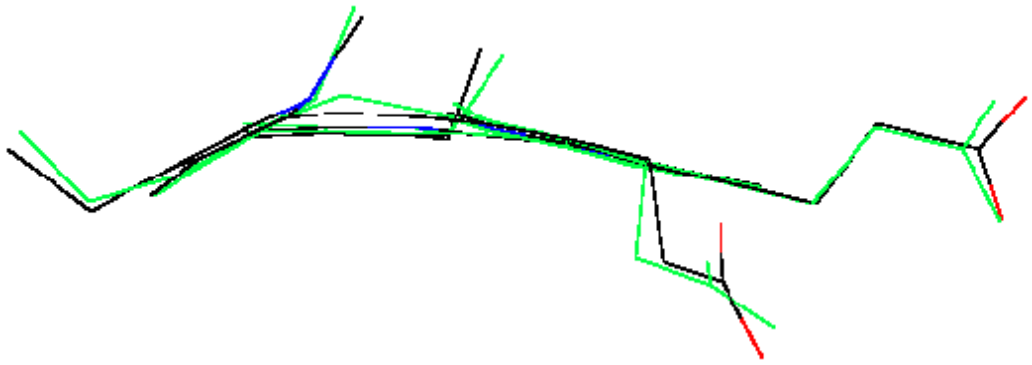

Figure 2. The ComQum-X structure of MMP in ferrochelatase, compared to the experimental electron density ($2f_o - f_c$ map, 1.4 level) and the crystal structure of MMP (blue) [16].

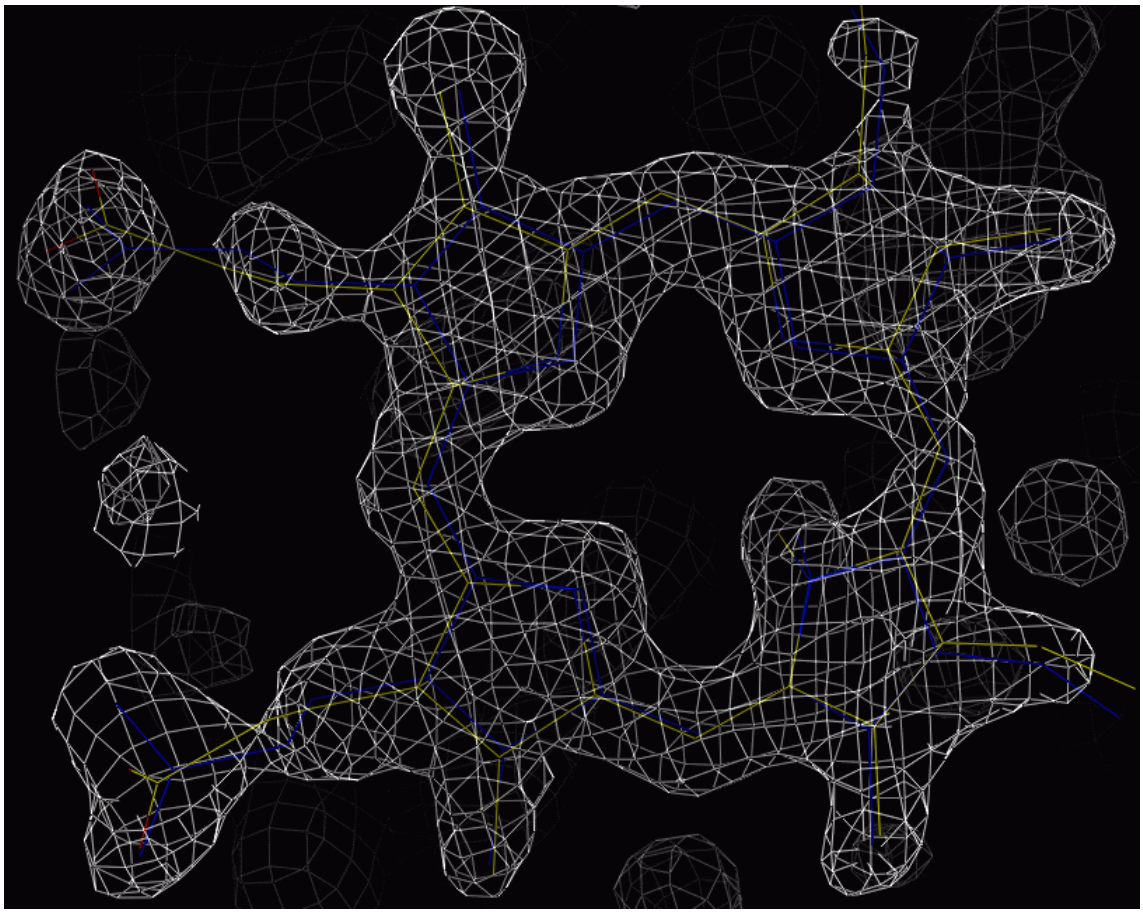

Figure 3. Two structures of MMP in ferrochelatase obtained by standard QM/MM methods (ComQum) with the protein fixed at the crystal structure (top) or with the protein free to move (bottom) [16].

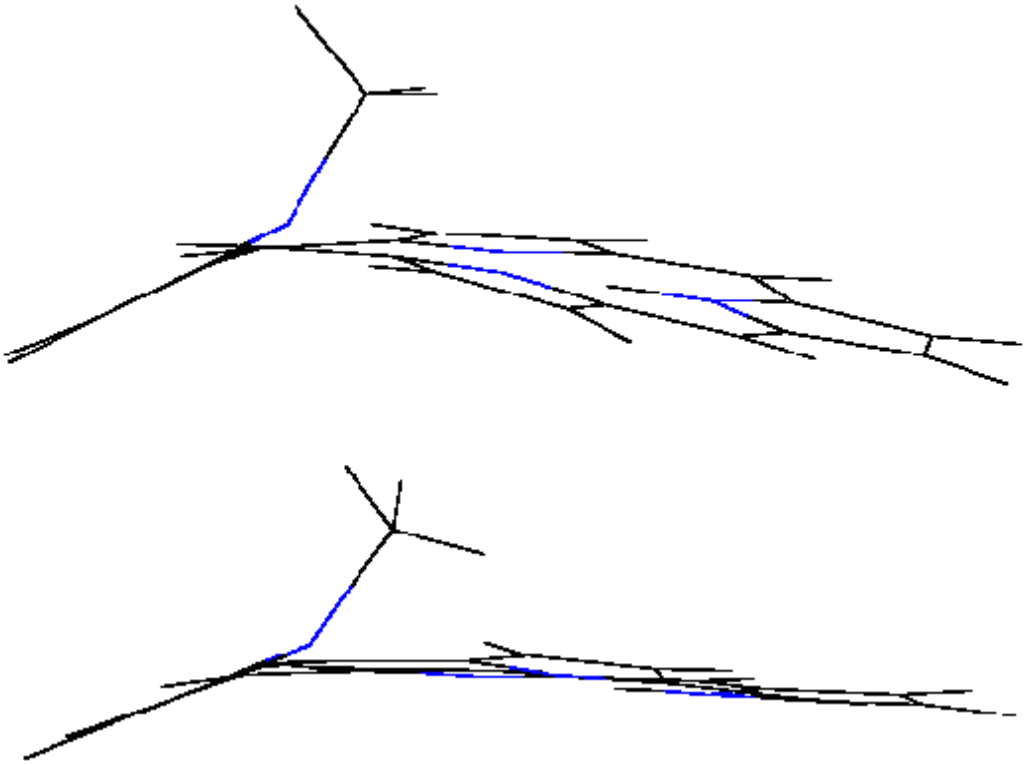

Figure 4. The low- (magenta) and high-resolution (orange) crystal structure of haem cytochrome c_{553} compared to the ComQum-X structure and the electron density ($2f_o - f_c$ omit map at the 2.5 level) from the high-resolution data.