Quantum chemical geometry optimizations in proteins using crystallographic raw data.

Ryde, Ulf; Olsen, Lars; Nilsson, Kristina

Published in:
Journal of Computational Chemistry

DOI:
10.1002/jcc.10093

2002

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

Quantum chemical geometry optimisations in proteins using crystallographic raw data

Ulf Ryde[1], Lars Olsen[2] & Kristina Nilsson[1]

[1]Department of Theoretical Chemistry

Lund University

Chemical Centre, P. O. Box 124

S-221 00 Lund, Sweden

[2]Department of Mathematics and Physics

Royal Veterinary and Agricultural University

Thorvaldsensvej 40

DK-1871 Frederiksberg C

Denmark

Correspondence to U. Ryde

Tel: +46-46-222 45 02

Fax: +46-46-222 45 43

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

2017-04-02

Abstract

A method is developed for the combination of quantum chemical geometry optimisations and crystallographic structure refinement. The method is implemented by integrating the quantum chemical software Turbomole with the crystallographic software Crystallography and NMR System (CNS), using three small procedures transferring information between the two programs. The program (COMQUM-X) is used to study the binding of the inhibitor *N*-methylmesoporphyrin to ferrochelatase and we show that the method behaves properly and leads to an improvement of the structure of the inhibitor. It allows us to directly quantify in energy terms how much the protein distort the structure of the bound inhibitor compared to the optimum vacuum structure (4-6 kJ/mole). The approach improves the standard combined quantum chemical and molecular mechanics (QC/MM) approach by guaranteeing that the final structure is in accordance with experimental data (the reflections) and avoiding the risk of propagating errors in the crystal coordinates. The program can also be seen as an improvement of standard crystallographic refinement, providing an accurate empirical potential function for any group of interest. The results can be directly interpreted in standard crystallographic terms (e.g. *R* factors or electron density maps). The method can be used to interpret crystal structures (e.g. the protonation status of metal-bound water molecules) and even to locally improve them.

Key words: crystallographic refinement, ferrochelatase, *N*-methylmesoporphyrin, QC/MM methods, strain energy.

Abbreviations: QC, quantum chemistry; QC/MM, combined quantum chemical and molecular mechanics; MM, molecular mechanics; MMP, *N*-methylmesoporphyrin

Introduction

During the last years, the explosive development of computers have allowed quantum chemical methods to establish themselves as a promising technique for the study of structure and function of proteins [1-3]. Various levels of sophistication have been used for modelling the influence of the protein on the properties of the quantum system. One of the most popular type of methods is the combination of quantum chemical and molecular mechanics (QC/MM) methods [4,5]. In these, a small part of the protein (e.g. the active site) is studied by quantum chemical methods, whereas the rest of the protein and the surrounding solvent are treated by molecular mechanics. Thereby, they combine the accuracy of the quantum mechanical calculations with the speed of the molecular mechanics method. Several variants of QC/MM methods have been suggested, e.g. Quest, IMOMM, Oniom, and COMQUM [6-9]. They differ mainly in the treatment of electrostatic interactions and in the interface between the quantum and molecular mechanics systems (the link region).

In standard QC/MM methods, the geometry of a part of the protein is optimised within the protein. However, a major problem with such methods is the restricted accuracy of molecular mechanics potentials used, which may distort the final structure so that it no longer is compatible with the crystal data. This problem could be partly solved by keeping the protein coordinates at or close to the positions observed in the crystal structure of the protein. Yet, crystallographic data for proteins have also a limited accuracy, so this would propagate errors in the crystal coordinates into the calculated structures. A natural solution to both these problems is to include the crystallographic raw data in the calculations, i.e. the reflections. These are normally deposited in the data banks together with the coordinates, so they are readily available.

Such an approach could also improve standard crystallographic refinement. In normal protein crystallography, the reflections are supplemented by a molecular mechanics force field, which ensures that the bond lengths and angles are realistic. For the amino acids, such force field exist and are accurate. However, for unusual molecules or groups (heterocompounds), such a force field is normally not available and has to be constructed by the crystallographers, a time-consuming and error-prone procedure. The use of quantum chemical methods could avoid this problem.

In this article, we develop a program which combines quantum chemical geometry optimisations with crystallographic refinement. We have modified our QC/MM program (COMQUM) to use crystallographic raw data instead of (or together with) the molecular mechanics potential, using the

freely available crystallographic refinement program CNS (Crystallography and NMR system) [10]. We discuss the method and its possible developments, and apply it to the structure of a transition-state analogue in the enzyme ferrochelatase, which inserts iron into the porphyrin ring, yielding haem. We show that the program behaves properly and that it allows us to calculate strain energies directly from the crystal data and discuss the accuracy of crystal structures. Possible applications of the method are also suggested.

Methods

The modular ComQum-01

The new geometry optimisation method is based on our local QC/MM program ComQum-00 [9,11]. We will therefore briefly describe this program, before we discuss how crystallographic data can be introduced into it. For a more thorough discussion of the method, we refer to the original articles [9,11].

ComQum divides the protein (including solvent) into three subsystems. The central system 1 is optimised by a quantum chemical method. System 2 consists of all atoms in all amino acids (and solvent molecules) within a radius r_1 of any atom in the quantum system. It is optimised with molecular mechanics methods. Similarly, system 3 comprises all atoms in all amino acids and solvent molecules within a radius r_2 of any atom in system 2. Typically, the rest of the protein and a sphere of water molecules are included in system 3. It is considered in all calculations, but it is kept fixed at the crystal geometry.

In the quantum chemical 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. Therefore, the polarisation of the quantum system by the protein is considered in a self-consistent way. In the classical 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). The program flow of ComQum-00 is shown in Scheme 1. Special action is taken when there is a bond between the classical and quantum chemical systems (a junction) [9]. The quantum chemical system is truncated by hydrogen atoms at the junctions, the positions of which are linearly related to the corresponding heavy (typically carbon) atoms in the full system.

The total energy is calculated as:

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

Here, E_{QC} is the quantum chemical 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. w_{MM} is a weight factor between the quantum chemical and molecular mechanics energy, which we introduce for future convenience. In ComQum-00 and -01, it is always 1. This approach is similar to the one used in the Oniom method [8]. The calculated forces are the gradient of this energy, taking into account the variation in the junction atoms [11].

In the original version, ComQum-00 is a combination of the quantum chemical software Turbomole 5.3 [12], and the classical simulation package Amber 5.0 [13] (although the approach is independent of the programs used in the actual implementation). ComQum-00 is an interface consisting of six small procedures which transfer information between the two programs or constructs the files needed. The philosophy behind ComQum is that the QC and MM programs shall be used without any modifications, so that they can be easily changed. This ensures that we always can use state-of-the-art programs, without needing to develop them.

Since we wanted to change the molecular mechanics program to include crystallographic raw data, we started by rewriting all the programs in a modular and program-independent fashion. This means that each of the six interface programs were divided into three or four new programs (cf. Figure 1). One of these programs is the core ComQum procedure, which reads QC and MM data from temporary text files in a specified format and writes QC and MM output to other text files in the same format. These six programs defines the ComQum core and they need not to be changed if the QC or MM software is changed (but they will change if ComQum is modified).

In addition, for each of the six interface programs, two or three input and output programs were also written. These programs either read or write the temporary text files, using data from or transferring data to the QC or MM software. These programs are specific for the QC and MM software and need to be changed if you want to switch software. A detailed description of the various programs and file formats is available on our home page (http://www.teokem.lu.se/~ulf/comqum.html). Together, they define ComQum-01, implemented for Turbomole and Amber.

ComQum-X

We now describe how crystal data is incorporated into the ComQum algorithm. The first step was to shift the classical program from Amber to Crystallography & NMR System (CNS) [10]. It is a free and widely used software for the determination of structures using crystallographic and nuclear magnetic resonance data. The program contains procedures for molecular mechanics minimisations, so the first step can be seen only as a change of the MM program from Amber to CNS.

CNS consists of a symbolic language so the existing task files could be used with small modifications (reading in or writing out coordinates, energies, and forces). A complication is that CNS reads in coordinates only in the form of Brookhaven protein data bank (PDB) files. Thus, the accuracy is restricted to only three decimal places. This led to convergence problems in some cases, so we modified the scripts to read in coordinates as two PDB files, one with the first three decimals, and one with the following six. Otherwise, the switch from Amber to CNS was straightforward using the modular implementation of ComQum-01.

Next, we wanted to include the crystallographic raw data into the calculations. This was done in the molecular mechanics calculations of systems 1-3 only (i.e. in the E_{MM123} term in Eqn. 1), because it should be included only in the calculations with the whole protein with C junction atoms. The energy comes from a normal CNS calculation, where the crystallographic data were included. Thus, the E_{MM123} term in Eqn. 1 was replaced by E_{XMM123}, defined by:

$$E_{XMM123} = E_{MM123} + w_A E_{xref} \qquad (2)$$

Here, E_{MM123} is the normal MM (CNS) energy for systems 1-3 with C junction atoms, as above, but with the addition of symmetry-related Van der Waals interactions. E_{xref} is the crystallographic penalty function. Since it is in arbitrary units, a weight factor (w_A) has to be included. By default, this term is determined so that the gradients of the two terms have a similar magnitude [14-16]. By changing this weight, the importance of the crystallographic data and the molecular mechanics and quantum chemical data can be varied, as will be discussed below.

In practice, we used the CNS sample input file minimize.inp (crystallographic conjugate gradient minimisation refinement) for this calculation (so that all the normal crystallographic manipulations were included, such as calculation of the *R* factors, the w_A weight, bulk solvent correction, etc.), with only two modifications: two sets of PDB files were read, giving a numerical accuracy of 10^{-8} (as was described above), and a number of lines were added for the output of energies and forces. The

number of minimisation steps was set to 0 (no change in the coordinates) and we used the default maximum likelihood refinement target using amplitudes (*mlf*) [17,18]. For the other entries we used the default values or choices. Naturally, we have to include the appropriate coordinate, structure, parameter, and reflection file names, and input the appropriate crystallographic data (resolution range, space group, and unit cell parameters).

In the molecular mechanics minimisations of system 2, we also included the crystallographic data. In fact, we used the same CNS script for this calculation (minimize.inp), with the only modification that atoms of systems 1 and 3 were kept fixed and that the number of minimisation steps was one (it will be discussed why below). In addition, we also run a restrained refinement of individual *B* factors for all atoms after each molecular mechanics minimisation. This was done using the CNS sample input file bindividual.inp, using default choices (but only one step of minimisation). This file was only modified to read in the second PDB file. Optionally, other crystallographic manipulations can also be run during this step.

Crystallographic refinement traditionally ignores electrostatic interactions in the molecular mechanics force field and hydrogen atoms are not considered, except for the most accurate structures. For practical convenience (we thereby avoid the problem of determining the protonation status of the histidine residues and the position of hydrogen atoms) and to get results that are directly comparable to the original crystal structures, we decided to follow this practice, at least in this first version of the program. This forced us to a few modifications of the ComQum algorithm. First, electrostatics were not included in the CNS calculations and no point-charge model of the surrounding protein was included in the quantum chemical calculations. Second, hydrogen atoms were included in the quantum chemical calculations but not in any of the MM calculations. The only exception to this is that H junction atoms must be included in the MM system 1, so that energies and forces are corrected for the conversion of carbon to hydrogen atoms in system 1. Third, all programs were modified to allow hydrogen atoms to be missing in the MM systems. Thus, the positions of the QC (non-junction) hydrogen atoms were determined from the QC forces only, because the corresponding MM forces are zero.

The final algorithm of ComQum-X is shown in Scheme 2. It can be seen that the exclusion of electrostatics reduce the number of ComQum specific procedures from five to three (one for adding the forces, one for adding the energies, and one for moving the coordinates of system 1 to the CNS

representation). In the present implementation of ComQum-X, three calculations are performed by Turbomole (evaluation of wavefunction, gradient calculations, and relaxation of the quantum system), and four calculations are performed by CNS (calculation of the crystallographic energy function and the forces, minimisation of system 2, and refinement of the individual *B* factors). The whole optimisation procedure is driven by a simple UNIX shell script.

Applications on ferrochelatase

The performance of ComQum-X was evaluated by optimising the inhibitor *N*-methylmesoporphyrin (MMP) inside the enzyme ferrochelatase. Ferrochelatase is the terminal enzyme in haem synthesis, which inserts an iron ion into the porphyrin ring. It is believed that the protein distorts the porphyrin ring so that one of the pyrrole nitrogen atoms becomes exposed and may bind the ion [19]. MMP is closely similar to the porphyrin substrate (cf. Figure 2), but the two vinyl groups of the substrate have been replaced by ethyl groups (increasing the solubility) and one of the pyrrole nitrogen atoms has been methylated, thereby forcing it out of the porphyrin plane. The structure of MMP bound to the enzyme has recently been solved at 190 pm resolution [20]. Our calculations are based on this structure.

The porphyrin ring of MMP shows interesting distortions, besides the out-of-plane tilt of the methylated pyrrole ring [20]. The goal of the present investigation was, in addition to test the ComQum-X method and evaluate its potential, to examine whether these additional distortions are real or caused by an improper force field in the refinement. We also wanted to calculate how much MMP is distorted by the protein in energy terms. This is not possible directly in the protein, because small errors in the crystal structure (especially in bond lengths) and systematic errors in the B3LYP method give rise to an unrealistically high strain energy. Thus, if hydrogen atoms are added to the MMP ring at standard positions using the crystal structure, side chains are replaced by hydrogen atoms, and the energy is calculated at the B3LYP/6-31G* level, the resulting structure is 156 kJ/mole less stable than the same molecule optimised in vacuum (if the side chains are included in the calculation, the difference is even higher, 264 kJ/mole).

As a test of the new ComQum-X program, we optimised the geometry of MMP inside the enzyme ferrochelatase. The coordinates (2848 atoms) of the protein were downloaded from the Brookhaven protein databank, access code 1c1h [20]. Space group ($P2_12_12_1$), unit cell symmetry (48.51, 49.97,

and 119.24 Å), and resolution (20.0-1.90 Å) were retrieved from this file. The structure factors (21 958 unique reflections, 93.5 % completeness), the test set for the structure factors (for the evaluation of the R_{free} factor; 4.9 % of the reflections, i.e. 1042 reflections), as well as the topology and parameter files for MMP and $Mg(H_2O)_6$ were obtained directly from the authors. This is not necessary, but it makes the comparison with the published results easier.

The full protein was used in all calculations, including all atoms in the structure. For simplicity however, we excluded the alternative configuration of residues 33 and 120-122 (we used the A conformation). This omission did not change the R_{free} factor (but the standard R factor increased from 0.181 to 0.183), so it is questionable if these alternative configurations are justified by the crystal data. The quantum system (system 1) consisted either of the MMP ring system with the side groups replaced by hydrogen atoms or of the whole MMP ring including the side chains. In the former case, there were eight hydrogen junction atoms in the quantum chemical calculations; in the latter case, no junction atoms were needed.

The quantum chemical calculations were performed at the density functional B3LYP level of theory for calculations without side chains, and at the Becke-Perdew86 level for the calculations with side chains [21-23]. In the latter calculations, the Coulomb operators were treated with the RI (resolution of identity) approximation [24,25]. In all calculation, we used the 6-31G* basis set [26] and the Turbomole software [12,21]. Only the five pure *d*-type functions were used. Becke-Perdew86 gives excellent geometries, whereas B3LYP has been shown to give the most reliable energetic results among the widely available density functional methods [27]. Test calculations showed that for MMP without side chains, the two methods give closely similar results. For example, the MMP tilt angle differs by less than 0.2° for the optimum vacuum geometries. Similarly, for the same COMQUM-X calculation, the results of the B3LYP and Becke-Perdew86 methods differed by less than 0.1 kJ/mole in strain energy and 0.00003 in the R_{free} factor.

For the MM calculations on MMP we used the force fields developed by the crystallographers and employed when they solved the structure of the MMP:ferrochelatase complex [20]. In some calculations, we corrected a few details in this force field, as is described below. The force field for the quantum system was the same as for MMP, except for the junction atoms, where the ideal bond lengths were taken from the vacuum calculation of the quantum system, the force constant for the bond was determined from the force constant of the same bond with C junction atoms multiplied

with the square of the quotient of the corresponding ideal bond lengths (this is necessary to not introduce an unphysical force), whereas the other parameters were identical to the parameters of the corresponding interactions with the C junction atoms.

The full geometry of all models was optimised until the change in energy between two iterations was below 10^{-6} Hartree (2.6 J/mole) and the norm of the internal gradients was below 10^{-3} a.u. (0.053 pm or 0.057°). Unless otherwise stated, the whole enzyme was included in system 2 (i.e. there was no system 3), and it was kept fixed during the geometry optimisation of MMP.

The strain energy (E_1) of the MMP ring in the ComQum-X calculations is calculated as the quantum chemical energy difference of the quantum system at the optimum vacuum geometry and at the ComQum-X geometry. It should be noted that this energy includes terms that normally are not considered as strain, in particular electrostatic effects [11,28]. We also study the tilt angle of the *A* pyrrole ring out of the plane of the other pyrrole rings. This angle is called simply the tilt angle below and it is defined as the angle between the two planes NA-CA2-CA3 and NB-NC-ND (the names of the atoms are defined in Figure 2).

Results and Discussion

The w_A *weight factor*

The performance of ComQum-X was tested by optimising *N*-methylmesoporphyrin (MMP) inside the protein ferrochelatase. We first run a series of calculations where the whole MMP molecule, including the side chains, was in the quantum system, because we then avoid the use of junction atoms, which may make the results harder to interpret. Similarly, the surrounding protein was kept fixed at the crystal structure during these calculations.

The present version of ComQum-X involves two new parameters, the weight factors w_{MM} and w_A, defined in Eqns. 1 and 2, respectively. Our first goal was to determine appropriate values for these two factors. The w_{MM} factor determines the relative weight between the quantum chemical and molecular mechanics energy (cf. Eqn. 1). For a normal energy-based force field, it should always be 1 (this is the value used in ComQum-00/01). However, the MM force field in CNS is not based on energies, but rather on a statistical analysis of crystal structures [29]. Therefore, the QC and MM energies are not comparable. In practice, it has turned out that CNS MM forces are typically 3 times larger than energy-based forces (like the QC forces) [29]. Therefore, w_{MM} should normally be 1/3 in

all ComQum-X calculations. We always use this value, if not otherwise stated.

The factor w_A determines the relative importance of the reflections (the crystallographic data) and the empirical energy function in the forces used for the geometry optimisation (Eqn. 2). In CNS, it is determined so that the MM and crystallographic forces have a similar magnitude [14-16]. In Table 1, we have studied how the tilt angle, the strain energy (E_1), and the *R* factors change when w_A is varied.

The results show that E_1 is sensitive to the value of this factor. For low values of w_A, E_1 approaches 24 kJ/mole. The reason why it does not go towards zero (E_1 is the difference in QC energy of the ComQum-X structure and the optimum structure in vacuum) is that MMP makes Van der Waals interactions with the surrounding protein. In the absence of junction atoms, E_{MM123} and E_{MM1} in Eqn. 1 includes exactly the same terms for MMP, except for the Van der Waals interactions between MMP and surrounding amino acids. Apparently, these interactions are quite strong and important, as we will also see below.

Similarly, E_1 increases when w_A is increased. Interestingly, E_1 can become larger than the value observed in the crystal structure (264 kJ/mole). This illustrates that also the original crystal structure involves a compromise between the crystallographic raw data and an empirical (MM) force field (determined by the w_A factor in CNS); if only the crystallographic data were used, the structure would become unrealistic with strange bond lengths and angles.

The tilt angle is much less sensitive to the w_A factor, even if it varies from slightly larger than the crystal value (38°) down to ~35° as w_A is decreased. This parameter does not either approach the vacuum value (30°; the tilt angle in crystals of free MMP is 28° [30]) when w_A is decreased, again owing to the Van der Waals interactions.

The *R* factors show an even smaller variation. R_{free} varies from 0.2314 for low values of w_A to 0.2308 for the structure optimised with $w_A = 0.1$. Encouragingly, the lowest value is clearly lower than in the original crystal structure 0.2312, showing that ComQum-X actually improves the crystal structure locally. At first, the small improvement, 0.0004, may seem a bit disappointing (we will see below that this can be slightly improved). However, it must be remembered that R_{free} is a global property of the whole protein (with 308 residues, 338 water molecules, and 2848 atoms). Even the vacuum structure of MMP, with completely different orientation of the the side chains, gives only an increase in R_{free} of 0.0049.

Interestingly, the normal R factor does not improve in the same way as R_{free}. On the contrary, the R factor tends to *increase* as R_{free} decreases. This illustrates that the original crystal structure is strongly optimised (refined) with respect to the normal R factor. Ideally (without overfitting), the two R factor 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.

This is the reason why we cannot use a residue R factor [31] (which is almost 100 times more sensitive to the variations in the MMP structure than the R factors) as the quality criterion: It is strongly correlated to the normal R factor, but not to the R_{free} factor, and therefore tends to increase for the ComQum-X structures. Moreover, the residue R factor (as implemented in the CNS and O [32] programs) is sensitive to details in the calculation (e.g. the type of density map, resolution range, and parameters in the equation used when the factor is calculated) [32].

Consequently, the choice of the w_A factor is crucial for the results. Apparently, the default choice of w_A by CNS (0.8707) does not give the best structure of MMP, in terms of the R_{free} factor. This is also manifested by the fact that the QC energy converges appreciably faster than the total ComQum-X energies. Thus, the default choice of w_A, determined automatically by CNS for the whole protein is not optimal for ComQum-X. A natural choice, given the data in Table 1 is to select the value of w_A that gives the lowest value of the R_{free} factor. This is found for $w_A = 0.1$ ($R_{free} = 0.2308$). At this value, we obtain a strain energy of 35 kJ/mole and a tilt angle of ~36°.

However, this procedure is quite time consuming. A simpler way would be to use the same criterion as CNS, i.e. to select w_A so that the QC and crystallographic forces have equal magnitude. CNS runs a short molecular dynamics simulation to determine w_A. Such a procedure would be too expensive with ComQum-X, because a new QC energy and force calculation would be needed for each dynamics step. Instead, we decided to use the data already available during the geometry optimisation. Thus, we compared the change in the various energy terms during each step in the geometry optimisation. It turned out that the quotient of the change in E_{QC} and in $w_{MM}(E_{MM123} - E_{MM1})$, averaged over the latest geometry optimisation steps, indicated that the two forces would be similar for w_A around 0.02, which is slightly too low (it gives an R_{free} factor of 0.2309). It seems to be reasonably general that this quotient suggests a w_A factor that is about ten times too small. Therefore, such a procedure to determine w_A can be used for crude calculations, whereas an investigation of the relation between R_{free} and w_A is best for accurate results.

Calculations with junction atoms

These results show that ComQum-X behaves well and that a reasonable choice can be made for the w_A factor with a moderate effort. However, the quantum system was large, making these calculations quite expensive, and there were no junction atoms, i.e. no chemical bonds between the quantum system and the surroundings. Therefore, we next performed a series of calculations with only the porphyrin ring of MMP in the quantum system (the methyl, ethyl, and propionate side chains on the periphery of the MMP ring were in the MM system, giving a total of eight junction atoms). The molecular mechanics force field for MMP was the one used during the original refinement of the crystal structure of the MMP:ferrochelatase complex [20].

The results of a series of ComQum-X optimisations of MMP in ferrochelatase with different values of w_A is presented in Table 2. The trends are similar those observed in Table 1. In particular, neither the strain energy nor the tilt angle converge towards the QC values as w_A is decreased; instead they stabilise at 9 kJ/mole and 37°, respectively. With junction atoms, there are two additional reasons for this, besides the Van der Waals interactions. First, the junction atoms are connected to the surrounding MM system (the side chains), which in these calculations are kept fixed at the crystal positions. Therefore, the junction atoms cannot move freely.

Second, the bonded interactions for MMP in E_{MM123} and E_{MM1} no longer cancel, because E_{MM1} contains terms from the H junction atoms, whereas E_{MM123} contains terms with C junction atoms. Ideally, these differences should correct for the truncation of the quantum system. This can be tested by varying the w_{MM} factor.

The w_{MM} *weight factor*

As discussed above, the w_{MM} factor determines the relative weight between the quantum chemical and molecular mechanics energy (Eqn. 1) and should normally be 1/3 in ComQum-X calculations. However, this factor can also be used to increase the weight of the QC calculations, independent of the MM forces, i.e. to bias the results toward the QC structure (which could be more accurate than the crystal data for low-resolution structures). It can also be used to test the quality of the MM force field and the treatment of the junction atoms.

This was investigated in Table 3, where w_{MM} has been varied from 3 to 1/300, using the default CNS value of w_A (~0.84). As expected, the strain energy decreases smoothly toward 0 as w_{MM} is decreased, and the tilt angle also converges towards the vacuum value of 34°. We also note that the strain energy and the tilt angle becomes larger than in the crystal structure for large values of w_{MM}, like in the other calculations.

Interestingly, R_{free} attains its lowest value at w_{MM} = 0.001, and the value is lower (0.2310) than in the calculations in Table 2. This indicates that there are some problems with the MM force field or the positions of the side chains. Therefore, we performed another series of calculations where the MM force field was omitted in all calculations. This was done by simply turning off all MM terms in the CNS calculations. This is possible because the electron density alone can determine the position of all the atoms (but the MM force field normally improves the local structure, i.e. the bond lengths and angles).

The results of such calculations with different values of w_{MM} are shown in Table 4. They show that the problem comes from the MM force field. When it is removed, the strain energy decreases by a factor of 2-3 for every value of w_{MM}. Already at the normal value w_{MM}=1/3, the strain energy is only 7 kJ/mole. Moreover, R_{free} has slightly decreased to 0.2309 and it attains its minimum value close to the normal value of w_{MM}, ~1/6. At this value, the tilt angle is 38° and the strain energy is 4 kJ/mole. This gives an indication of the actual strain energy of the MMP ring in the crystal.

Thus, CNS-X can be used to identify problems in the MM force field for the molecule of interest. The problem can be localised by systematically removing terms from the force field. However, such a procedure is quite tedious and very time consuming. A direct inspection of the force field of MMP identifies two inconsistencies in the definition of improper dihedrals which keep the ring system flat (the asymmetric CMA, CAA, CMD, and CAD atoms seem to have been mixed up), and several suboptimal choices. A correction of these improper torsions leads to an improvement of the R_{free} factor by 0.00015.

A more general procedure, if the accuracy of the MM force field for the quantum system is in doubt, is to remove all bonded interactions (bond, angle, and dihedral terms) of it (or, equivalently, only those interactions involving the junction atoms). Such a procedure has the advantage over the results in Table 4 of including Van der Waals interactions with the surrounding protein, but on the other hand it gives up the opportunity to correct for the truncation of the quantum system.

Encouragingly, such a procedure gave the most accurate structure (in terms of the R_{free} factor, 0.2308), as can be seen in Table 5, row 13. Therefore, we recommend such a procedure, unless you are certain that the MM force field for the junction atoms will lead to an improvement. Thereby, you also avoid the considerable work of developing and testing the force field.

Relaxation of the surrounding protein

In all calculations up to now, the surrounding protein has been kept fixed at the crystal structure. This is not necessary. On the contrary, it is probable that changes in the quantum system will affect the structure of the surroundings also. Conversely, a fixed protein will restrict the possible changes of the quantum system, especially if there are any junction atoms (as was discussed above). Therefore, we have tested various methods to allow the surrounding protein to relax during the geometry optimisation.

First, we observed that a reoptimisation of the *B* factors of the quantum-system atoms in general improved the R_{free} factor for all calculations by ~0.0001. Strictly speaking, this is not a relaxation of the surroundings, but rather an optimisation of other crystallographic variables for the quantum system than the coordinates. Such an optimisation could either be done in each step in the geometry optimisation or as a final improvement at the end of the optimisation. These two possibilities gave closely similar results for the present calculations of MMP in ferrochelatase. Therefore, it seems to be sufficient to optimise them only after the geometry optimisation.

In standard QC/MM calculations, the whole protein is normally not allowed to relax, but only the atoms closest to the quantum system (typically atoms within 0.8-2.0 nm of the quantum system). This restriction is used to ensure that the protein does not move too far from the crystal structure. However, in ComQum-X, such restrictions are unnecessary because the crystallographic raw data ensure that the final structure will always be close to (and normally even better than) the crystal structure. Therefore, no system 3 is necessary in the ComQum-X and all non-quantum atoms are included in system 2, which is optimised by standard crystallographic methods.

In our initial calculations, we optimised system 2 by a few (5-50) steps of crystallographic conjugate gradient minimisation and restrained individual *B*-factor refinement in each cycle of the ComQum-X optimisation. However, it was invariably observed that both the *R* and R_{free} factors increased steadily by such a treatment. The reason for this is that the original structure is strongly

optimised with respect to the reflections and this optimisation was always stopped as soon as the *R* factors started to increase.

After several tests, we decided to follow this practice also in the ComQum-X calculations. Thus, we do only one step of crystallographic conjugate gradient minimisation refinement and one step of restrained individual *B* factor refinement in each cycle of the ComQum-X geometry optimisation. However, the new coordinates and *B* factors are accepted only if the R_{free} factor decreases; otherwise, the old variables are kept unchanged. Since a typical geometry optimisation runs for 50-200 iterations, there can still be an appreciable improvement in the structure of the surrounding protein. Moreover, it is normally observed that during the end of the optimisation, few new coordinates and *B* factors are accepted, indicating that also the surroundings has converged. In particular, we have invariably observed that the R_{free} factor does not improve if the surroundings are optimised by several steps of optimisation after convergence of the ComQum-X geometry optimisation.

Relaxation of the surrounding protein, typically leads to improved structures in terms of the R_{free} factor, but only by 0.0001-0.0002. The effect is largest for the calculations with the whole MMP molecule in the quantum system. Some typical examples (the best results in Tables 1-4) are collected in Table 5. It is notable that the lowest value for the R_{free} factor is not always found for the same value of the weight factors if the protein was allowed to relax or not. However, the difference in strain energy and the tilt angle for the best calculations is not very large.

Even if the effect is not very large, the cost in computer time to relax the protein is quite small, less than a tenth of the total computer-time consumption for the current calculations. Moreover, if there are larger changes in the quantum system, relaxation of the surroundings may be important. Therefore, we recommend all ComQum-X calculations be run with the surroundings free to relax.

Comparison with standard QC/MM methods

Finally, it could be very interesting to compare the ComQum-X results with those obtained with standard QC/MM methods. Therefore, we have also optimised MMP in ferrochelatase with two different QC/MM protocols. The first is ComQum-01 in its original version, i.e. Turbomole combined with Amber. The other is based on ComQum-01 in the modified version, i.e. Turbomole combined with CNS, but without any crystallographic data (i.e. not ComQum-X). For the Amber calculations we run one optimisation with the protein fixed at the crystal positions and one

calculation where all amino acids within 0.8 nm of the quantum system were allowed to relax by a full MM energy minimisation in each geometry optimisation cycle. Apart from the change in the MM program, the two sets of calculations differ also in the force field for the protein (Amber or CNS) and in that electrostatics are included in the Amber calculations, but not in those with CNS. The results of the calculations are collected in Table 6 and two structures are shown in Figure 3.

Apparently, the various calculations gave quite differing results for the strain energy: 10-21 kJ/mole. E_1 is higher for the Amber calculations than for those with CNS, which is most likely an effect of the inclusion of electrostatics in these calculations, which may significantly distort the structure. The strain energy increases slightly when the surroundings are allowed to relax, which is somewhat unexpected (the opposite is normally observed). This indicates that the surrounding protein structure is strained in the crystal structure (at least according to the Amber force field) and this strain is released by slightly distorting MMP. However, the increase in E_1 is small, 2 kJ/mole.

The tilt angles are more similar, 37-39° and also close to the crystallographic value (37°). Likewise, the crystallographic R_{free} factors are quite similar; it is 0.2312 for both calculations with fixed surroundings. Thus, the ComQum-01 calculations fit the experimental electron density about as well as the crystal structure (but slightly worse than the ComQum-X result). The reason for this is that the coordinates are very similar.

However, if the surroundings are allowed to relax, the ComQum-01 structure changes more. In particular, the MMP ring becomes more planar. This effect is largest if the protein is allowed to equilibrate before the ComQum-01 calculation. Then, the MMP ring becomes almost completely planar except for the tilt of the methylated pyrrole ring, as can be seen in Figure 3b. This indicates that the force field for MMP is not fully satisfactorily (the ring is too stiff) and it nicely illustrates the risk of using QC/MM calculations (the structure may diverge from the experimental structure).

Quite naturally, the *R* factors increase strongly for the calculations where the protein is allowed to relax, especially in the equilibrated structure. This is caused by the change in the coordinates of the surrounding protein. It is partly an effect of the fact that the simulation is run in solvent and not in the crystal, partly an effect of shortcomings of the force field.

The pure QC/MM calculations can be used to identify the interactions that give rise to the distortions of the MMP ring. The fact that similar distortions are seen in the ComQum-01 calculations both with Amber and CNS (i.e. both with and without electrostatics) show that

electrostatics if of less importance. Instead, Van der Waals interactions turned out to give the whole effect: If the Van der Waals radii of the MMP atoms were decreased by 0.1 nm in the ComQum-CNS calculations, the MMP ring became completely planar. This was accompanied with a minimal strain energy but large *R* factors (R_{free} = 0.2334).

Thus, Van der Waals interactions are important for the MMP structure. We therefore tried to modify these parameters also in the ComQum-X calculations. However, this did not lead to any improvement of the structure - the lowest value of R_{free} was obtained for the original parameters.

In conclusion, we see that the standard QC/MM methods give reasonable structures if the surroundings are fixed at the crystal structure. However, if the surroundings are allowed to relax, the structure may start to diverge from the crystal structure. Consequently, it is strongly advisable to keep the surroundings fixed during QC/MM calculations, if the aim of the calculations is only to study the crystal conformation. Even if another conformation is studied, it is probably wise to relax as little of the surroundings as possible. Moreover, we see that the strain energies are 4-15 kJ/mole too high compared to the best ComQum-X estimate. Thus, ComQum-X seems to give appreciably more accurate strain energies than QC/MM methods, provided that the crystal structure is known.

Implications for ferrochelatase

We have shown that ComQum-X behaves properly and can give rise to improved structures. It is now time to collect the results and discuss their implications for the function of the enzyme ferrochelatase.

The best structure of MMP in ferrochelatase seems to be obtained in the calculations of MMP with side chains, with a w_A factor of 0.1 (cf. Table 5). This structure fits excellently into the electron density, as can be seen in Figure 4. The density is well-defined for the porphyrin ring, whereas the side groups, especially the propionate side chains, are not so easy to position. Consequently, the largest differences between this ComQum-X structure and the original crystal structure are seen for the side chains (the propionate carboxylic atoms have moved by up to 155 pm), as can be seen in Figure 5.

However, also the ring atoms have moved; the average movement of all the heavy atoms is 48 pm. The differences are most pronounced around the *A* ring, which is affected by the erroneous improper dihedral parameters. This is most clearly seen for the position of the CHA atom relative to

the A ring; in the original structure, CHA shows a strange and sharp kink. In the ComQum-X structure, there is a more gradual transition between the *A* ring and the rest of the porphyrin ring. This shows that crystal structures are sensitive to the MM force field used in the refinement and that possible errors in this will propagate to the final coordinates. ComQum-X can be used to identify and correct such errors.

The tilt angle in the ComQum-X structure (36°) is similar to the that in the crystal structure (37°). Encouragingly, the best ComQum-X calculations without side chains gives the same result, showing that the tilt angle is insensitive to the theoretical treatment. This tilt angle is significantly larger than in the optimised structure of free MMP (30°) or in the crystal structure of free MMP (28°) [30].

Interestingly, in the vacuum structure of MMP, the porphyrin ring is completely planar except for the tilt of the *A* ring. However, this is not the case in the crystal, where the ring is strongly ruffled (it looks like the figure eight seen from the edge; cf. Figure 3a) [33]. This ruffling is also seen in the ComQum-X structure. Thus, this ruffling is inherent in the crystal structure and clearly caused by Van der Waals interactions in the protein. However, this distortion is a low-energy mode. This can be seen from the strain energies.

The strain energies are very different in the ComQum-X calculations with or without side chains. In the former, the strain energy is quite high 34 kJ/mole, whereas in the best of the latter calculations, it is very low 1-10 kJ/mole. This large difference illustrates the problem of directly interpreting E_1 as a strain energy.

E_1 is defined as the difference in vacuum energy of the ComQum-X structure and the optimum vacuum structure. If the quantum system contains polar groups, these will form strong hydrogen bonds or ion pairs with the surrounding protein or solvent, leading to an extended conformation of these groups. However, in vacuum, these interactions are not present, and the polar groups instead have to curl back to the molecule itself and try to form as favourable interactions as possible. When E_1 is calculated, we compare the energy of the extended conformation in vacuum, i.e. without any compensation hydrogen bonds, with the compact vacuum structure with internal hydrogen bonds. Therefore, E_1 can become quite large. However, these differences in electrostatic interactions are not strain in the normal sense of the word [34]. This has been thoroughly discussed before [11,28].

Thus, the large values of E_1 for the calculations of MMP with side chains are caused by electrostatic interactions between the propionate side chains and the other side chains. In MMP

without side chains, there are no flexible polar groups in the molecule. Therefore, there are no dubious electrostatic terms in E_1 for this molecule. Consequently, E_1 is low and it can directly be interpreted as a strain energy in the normal sense of the word (mechanical distortions of the molecule when it is bound to a protein) [34].

Our best estimate of the strain energy of MMP when bound to ferrochelatase is 6 kJ/mole. This estimate is obtained from the calculations where the bonding interactions in the MM force field for MMP are ignored (rows 13-14 in Table 5). It is probably advantageous to calculate strain energies in calculations without any bonded MM interactions for the quantum system, because you then avoid the risk of distorting the results by errors in the force field. This strain energy is much lower than E_1 values observed in other QC/MM calculations 20-70 kJ/mole [28], but the latter calculations have invariably involved polar groups (note that the values are similar to E_1 for MMP with side chains, 34 kJ/mole). 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 [35].

Thus, we can conclude that MMP is very little strained by the protein in energy terms, although there is a clear distortion (ruffling) of the porphyrin ring and a change in the tilt angle, which undoubtedly are caused by the protein. Apparently, these distortions are low-energy modes. However, it should be noted that MMP is not the true substrate of the protein, but rather a strong inhibitor. Therefore, the observed distortions may give rise to larger energy terms for the true substrate and in the transition state of the chelatase reaction.

Concluding remarks

In this paper, we have developed a method, *quantum refinement*, to incorporate crystallographic raw data into a quantum chemical geometry optimisation. The method is implemented in the program ComQum-X and we have illustrated its possibilities. We show that the program behaves well and that reasonable structures are obtained. These structures provides an optimum compromise between quantum chemistry and crystallography. The method provides a clear improvement of normal QC/MM calculations, because the resulting structures are guaranteed to be consistent with the experimental data. Moreover, there is no risk that the structure is biased by errors in the crystal coordinates, as there would be if you fix some atoms at their positions in the crystal structure.

Thus, the method could be used to dock quantum mechanical vacuum structures into a protein.

The study of enzymatic reaction mechanisms with quantum chemical methods is a rapidly growing area of research [1-3]. Normally, the structures are obtained in vacuum and there is no guarantee that they may fit into the protein. ComQum-X can be used to test if the structure may apply also in the protein.

We have also seen that the method allow us to compare structures and energies obtained with and without the protein. The method avoids any bias from systematic errors in the quantum chemical method or errors in the experimental coordinates. Naturally, strain energies become much more realistic than those obtained directly from the crystal structure (for example 6 kJ/mole compared to 156 kJ/mole for MMP in ferrochelatase). Therefore, ComQum-X may become a valuable tool in the study of how a protein may distort a bound group.

We have thoroughly discussed and tested how the two new weight factors in ComQum-X (which are needed to make the three types of energies, quantum chemical, molecular mechanics, and refinement penalty function, comparable) affect the results. For w_{MM}, the only realistic choice is around 1/3 when the CNS force fields are used, and 1 when an energy-based force field is used. The w_A factor, on the other hand, can be determined by running several calculations and select the one with the lowest value for the R_{free} factor. Alternatively, but less accurately, it could be determined as the value that gives QC and crystallographic forces of the same magnitude.

Even more interestingly, the w_{MM} and w_A weights can be used to bias the structure towards the quantum chemical structure. A medium-resolution crystal structure of a heterocompound, e.g. a metal site, in a protein typically has an uncertainty in the bond lengths of at least 10 pm. A quantum chemical calculation of the same site has an uncertainty of 3-7 pm [36,37]. Moreover, the latter errors are systematic, whereas the crystallographic errors are random. Therefore, the quantum chemical errors can be compensated for, e.g. by the method of offset forces [38]. We could even use other experimental data, e.g. EXAFS data, to improve the quantum chemical structure. Therefore, we can expect that the quantum chemical calculations are more accurate than the crystal structures. By giving a higher weight to the quantum chemical calculations, we could therefore actually improve the crystal structure locally. We are currently investigating this possibility.

ComQum-X can also be regarded as an improved crystallographic refinement procedure. The refinement of protein structures already involves the use of theoretical methods, so the change is not very large. We have replaced (or supplemented) the normal molecular mechanics potential by a

quantum chemical potential. Naturally, this can be expected to give improved results. For normal amino acids, the molecular mechanics potential used in standard refinement programs is quite accurate. However, for other groups and molecules, such as metal centres, substrates, inhibitors, etc. (i.e. heterocompounds), experimental data is often partly lacking and the potential is much worse (and normally constructed by the crystallographers themselves). Naturally, this makes these parts of the crystal structure less well-determined. ComQum-X may solve this problem in an unbiased way. In fact, we can remove the molecular mechanics potential fully, giving a purely quantum chemical potential for the heterocompound. This gives an ultimate test of the potential used in the crystal structure. In the studied example, we have seen that the force field for MMP contains errors, which propagates to the final crystallographic coordinates.

An interesting and probably the most important application of ComQum-X is to *interpret* crystal structures. Quite often, it is not clear exactly what atoms are seen in the electron density or in what oxidation states the atoms are. By performing several ComQum-X calculations on different alternative structures, we may identify the most probable structure with an appropriate quality criterion, e.g. the R_{free} factor. The advantage with quantum chemical methods is that we can calculate the optimum structure for each alternative and see how well it fits into the electron density. Numerous examples of promising projects of this type can be found.

One example would be to study the protonation status of metal-bound water molecules in the active site of proteins. The protonation status (i.e. if it is a water molecule or a hydroxide ion) is most important for the reaction mechanism of metalloenzymes, but it is not available by normal crystallographic techniques. The reason for this is mainly that the ideal bond lengths of water and a hydroxide ion to the metal ion are not known. By quantum chemical methods, the two ligands can easily be compared, and their bond lengths typically differ by 30 pm. Moreover, they often give rise to further changes in the surrounding structure. Therefore, it should be possible to decide which structure fits the electron density best using ComQum-X. There is a large number of enzymes that can be investigated with such techniques. This possibility will be thoroughly explored in a future publication.

In the present version of ComQum-X, we have followed the crystallographic practice of not including hydrogen atoms and electrostatics in the force field used in the refinement. The reason for this decision is mainly that the hydrogen atoms are not discernible in the crystal structure, so we

would have to speculate about their positions. For many hydrogen atoms, this can be done with a satisfactorily accuracy. However, for other atoms, e.g. methyl groups or hydroxyl groups there is a rotational degree of freedom which is not easily settled. For water molecules, the situation is even worse and for histidine residues, it is not even clear to what atom the hydrogen atom should be added. Therefore, there is a large risk of making an erroneous choice when the positions of the hydrogen atoms are determined and this may bias the structure in an unwanted way.

On the other hand, the lack of electrostatics in the quantum chemical calculations will distort the structure. For example, it is known from earlier QC/MM calculations that the structure of metal sites change significantly when hydrogen bonds and solvation effects are included in the calculations. For example, the Fe-S distances in reduced rubredoxin decrease by 6 pm when the site is moved from vacuum into the protein [39]. Therefore, we need to include electrostatics (and at least polar hydrogens also) in the calculations in order to obtain accurate results for such systems. This can be attained by minor changes in the ComQum-X and we currently investigate this possibility.

Acknowledgements

This investigation has been supported by grants from the Crafoord foundation and by computer resources of the Swedish Council for Planning and Coordination of Research (FRN), Parallelldatorcentrum (PDC) at the Royal Institute of Technology, Stockholm, the High Performance Computing Centre North (HPC2N) at the University of Umeå, and Lunarc at Lund university. We thank David Lecerof, Lund, for giving us access to the ferrochelatase files and for help with the crystallographic work. We also thank Gerard Kleywegt, Uppsala, for fruitful discussions.

References

1. Siegbahn, P. E. M.; Blomberg, M. R. A. Annu. Rev. Phys. Chem. 1999, 50, 221–249.
2. Siegbahn, P. E. M.; Blomberg, M. R. A. Chem. Rev. 2000, 100, 421–437.
3. Tantillo, D. J.; Chen, J. G.; Houk, K. N. Curr. Opin. Chem. Biol. 1998, 2, 743-750.
4. Eurenius, K.P., Chatfield, D.C., Brooks, B.R. Int J Quant Chem 1996, 60, 1189.
5. Field, M.L. in The encyclypaedia of computational chemistry, P.v.R. Schleyer, N.L. Allinger, T. Clark, J. Gasteiger, P.A. Kollman, H.F. Schaefer III & P.R. Schreiner eds., John Wiley & Sons, Chichester, 1998, p. 2255.
6. Singh, U. C.; Kollman, P. A. J Comput Chem 1986, 7, 718.
7. Maseras, F.; Morokuma, K. J Comput Chem 1995, 16, 1170-1179.
8. Svensson, M., Humbel, S., Froese, R. D. J., Matsubara, T., Sieber, S., Morokuma, K. J Phys Chem 1996, 100, 19357.
9. Ryde, U. J. Comp.-Aided Mol. Design. 1996, 10, 153–164.
10. A.T. Brunger, P.D. Adams, G.M. Clore, W.L. Delano, P. Gros, R.W. Grosse-Kunstleve, J.-S. Jiang, J. Kuszewski, M. Nilges, N.S. Pannu, R.J. Read, L.M. Rice, T. Simonson, G.L. Warren, Crystallography & NMR System CNS, Version 1.0, Yale University, 2000.
11. U. Ryde & M.H.M. Olsson Intern. J. Quant. Chem 2001, 81, 335-347
12. Treutler D, Ahlrichs R J Chem Phys 1995,102, 346
13. D.A. Case, D.A. Pearlman, J.W. Caldwell, T.E. Cheatham III, W.S. Ross, C.L. Simmerling, T.A. Darden, K.M. Merz, R.V. Stanton. A.L. Cheng, J.J. Vincent, M. Crowley, D.M. Ferguson, R.J. Radmer, G.L. Seibel, U.C. Singh, P.K. Weiner & P.A. Kollman, Amber 5, University of California, San Francisco, 1997.
14. Jack, A; Levitt, M. Acta Cryst 1978, A34, 931.
15. Brünger, A. T.; Karplus, M.; Petsko, G. A. Acta Cryst 1989, A45, 50.
16. Brünger, A. T. Methods Enzymol 1997, 277, 243-269.
17. Pannu, N. S.; Read, R. J. Acta Cryst 1996, A52, 659-668.
18. Adams, P. D.; Pannu, N. S.; Read R. J.; Brünger, A. T. Proc. Natl. Acad. Sci. USA 1997, 94, 5018-5023.
19. Lavallee, D. K. Mol. Struct. Energ. 1988, 9, 279.
20. D. Lecerof, M. Fodje, A. Hansson, M. Hansson and S. Al-Karadaghi, J. Mol. Biol. 2000,

297, 221

21. Hertwig RH, Koch W Chem Phys Lett 1997, 268, 345-351
22. Becke AD Phys Rev A 1988, 38, 3098-3102.
23. Perdew JP Phys Rev B 1986, 33, 8822-8824.
24. Eichkorn K,Treutler O,Öhm H, Häser M, Ahlrichs R Chem Phys Lett 1995, 240, 283-290
25. Eichkorn K,Weigend F, Ahlrichs R Theor Chem Acc 1997, 97, 119-124
26. Hehre, W. J., Radom, L., Schleyer, P. v. R.; Pople, J. A. In Ab initio molecular orbital theory, Wiley-Interscience, New York1986,. .
27. Bauschlicher CW Chem Phys Lett 1995, 246, 40.
28. U. Ryde in, Recent Research Developments in Protein Engineering, 2001, in press.
29. Engh, R. A.; Huber, R. Acta Cryst 1991, A47, 392-400.
30. Lavalle, D. K.; Anderson, O. P. J. Am. Chem. Soc. 1982, 104, 4707-4708.
31. Jones, T. A.; Zou, J.-Y.; Cowan, S. W.; Kjeldgaard, M. Acta Cryst 1991, A47, 110-119.
32. Jones, T. A.; Kjeldgaard, M. O - the manual, Uppsala, 1995.
33. Jentzen, W.; Song, X.-Z-; Shelnutt, J. A. J Phys Chem B 1997, 101, 1684-1699.
34. Warshel, A. Computer modelling of chemical reactions in enzymes and solution. Wiley & Sons, New York, 1991, pp. 209-211.
35. Boström, J.; Norrby, P.-O.; Liljefors, T. J. Comp.-Aided Mol. Design. 1998, 12, 383-396.
36. E. Sigfridsson, M. H. M. Olsson, and U. Ryde J Phys Chem B, 2001, 105, 5546-5552.
37. Olsson, M.H. M.; Ryde U. J. Am. Chem. Soc., 2001, 123, 7866-7876.
38. Fogarasi, G, Zhou, X, Taylor, P W & Pulay, P J Am Chem Soc 1992, 114, 8191-8201
39. E. Sigfridsson, M. H. M. Olsson & U. Ryde Inorg Chem 2001, 40, 2509-2519

Table 1. Variation of the strain energy (E_1, kJ/mole), tilt angle (°), and *R* factors with the w_A factor when MMP (with side chains) is optimised in ferrochelatase with COMQUM-X.

The surrounding protein was not allowed to relax and w_{MM} = 1/3. For comparison, data for the crystal structure and a geometry optimisation in vacuum with the same QC method are also included.

w_A	E_1	Tilt angle	R_{free}	R
Crystal	264.5	37.2	0.23120	0.18271
30	333.6	42.2	0.23143	0.18256
10	131.7	40.7	0.23129	0.18263
3	74.2	39.4	0.23119	0.18272
1	53.4	38.3	0.23111	0.18286
0.8707	52.7	38.3	0.23109	0.18288
0.3	41.8	37.0	0.23106	0.18319
0.1	34.6	36.3	0.23084	0.18390
0.03	28.1	35.6	0.23093	0.18577
0.01	24.9	35.5	0.23129	0.18655
0.003	24.3	35.3	0.23136	0.18670
0.001	24.3	35.3	0.23138	0.18671
QC	0.0	29.9	0.23606	0.18859

Table 2. Variation of the strain energy (E_1, kJ/mole), tilt angle (°), and R factors with the w_A factor when MMP (without side chains) is optimised in ferrochelatase with ComQum-X.

The surrounding protein was not allowed to relax and w_{MM} = 1/3. For comparison, data for the crystal structure and a geometry optimisation in vacuum with the same QC method are also included.

w_A	E_1	Tilt	R_{free}	R
Crystal	156.3	37.2	0.23120	0.18271
30	188.2	42.3	0.23123	0.18262
10	65.8	40.5	0.23117	0.18264
3	23.6	38.8	0.23117	0.18269
1	12.9	37.7	0.23114	0.18272
0.84	12.2	37.6	0.23114	0.18273
0.3	9.6	37.1	0.231117	0.18277
0.1	8.8	36.9	0.231116	0.18282
0.03	9.0	37.0	0.23119	0.18296
0.01	9.3	37.2	0.23124	0.18308
QC	0.0	34.1	0.23373	0.18490

Table 3. Variation of the strain energy (E_1, kJ/mole), tilt angle (°), and *R* factors with the w_{MM} factor when MMP (without side chains) is optimised in ferrochelatase with COMQUM-X.

The surrounding protein was not allowed to relax and w_A was calculated by CNS (w_A = 0.841-0.848). For comparison, data for the crystal structure and a geometry optimisation in vacuum with the same QC method are also included.

w_{MM}	E_1	Tilt	R_{free}	R
Crystal	156.6	37.2	0.23120	0.18271
3	188.4	41.1	0.23149	0.18267
1	37.2	39.3	0.23128	0.18269
1/3	12.2	37.6	0.23114	0.18273
1/10	4.9	36.2	0.23100	0.18279
1/30	2.2	35.7	0.230982	0.18292
1/100	0.7	35.0	0.230978	0.18313
1/300	0.1	34.1	0.23119	0.18331
Pure QC	0.0	34.1	0.23373	0.18490

Table 4. Variation of the strain energy (E_1, kJ/mole), tilt angle (°), and *R* factors with the w_{MM} factor when MMP (without side chains) is optimised in ferrochelatase with COMQUM-X but without any molecular mechanics force field.

The surrounding protein was not allowed to relax and w_A was set to 1 (it cannot be determined by CNS without a MM potential). For comparison, data for the crystal structure and a geometry optimisation in vacuum with the same QC method are also included.

w_{MM}	E_1	Tilt	R_{free}	R
Crystal	156.6	37.2	0.23120	0.18271
10	180.5	42.3	0.23102	0.18263
3	50.7	40.7	0.23098	0.18270
1	16.3	39.6	0.23096	0.18276
1/3	7.3	38.5	0.23092	0.18279
1/6	4.4	37.9	0.23086	0.18281
1/10	2.9	37.3	0.23091	0.18285
1/15	2.0	36.4	0.23098	0.18288
1/30	1.0	35.8	0.23109	0.18296
1/100	0.1	34.1	0.23147	0.18311
Pure QC	0.0	34.1	0.23373	0.18490

Table 5. Change in the strain energy (E_1, kJ/mole), tilt angle (°), and *R* factors when the protein was allowed to relax during the optimisation of MMP in ferrochelatase with ComQum-X.

Relaxed?	Side chains?	w_{MM}	w_A	E_1	Tilt	R_{free}	R
No	Yes	1/3	0.1	34.6	36.3	0.23084	0.18390
Yes	Yes	1/3	0.1	34.3	36.1	0.23067	0.18388
No	No	1/3	0.3	9.6	37.1	0.23112	0.18277
Yes	No	1/3	0.3	9.6	37.2	0.23103	0.18277
No	No	1/3	0.1	8.8	36.9	0.23112	0.18282
Yes	No	1/3	0.1	8.8	36.9	0.23107	0.18284
No	No	1/10	0.8423	4.9	36.2	0.23100	0.18279
Yes	No	1/10	0.8423	4.9	36.2	0.23095	0.18280
No	No	1/100	0.8422	0.7	35.0	0.23098	0.18313
Yes	No	1/100	0.8423	0.8	35.0	0.23096	0.18311
No	No	1/6	No MM[a]	4.4	37.9	0.23086	0.18281
Yes	No	1/6	No MM[a]	4.4	37.9	0.23076	0.18281
No[b]	No	1/3	0.03	6.3	35.7	0.23085	0.18300
Yes[b]	No	1/3	0.03	6.2	35.2	0.23078	0.18321
Crystal				156.6	37.2	0.23120	0.18271

[a] These calculations were run without any MM interactions.

[b] These calculations were run without any bonded MM interactions for MMP.

Table 6. Strain energy (E_1, kJ/mole), tilt angle (°), and *R* factors for three standard QC/MM (ComQum-01) optimisations of MMP (without side chains) in ferrochelatase.

The calculations differ in the MM program and force field used (Amber or CNS) and whether the MM system (system 2) is allowed to relax or not. The *R* factors are calculated after a refinement of the individual *B* factors of the quantum system or the whole protein, if it was relaxed.

MM program	MM relaxed?	E_1	Tilt	R_{free}	*R*
Amber	No	19.1	38.7	0.23115	0.18276
Amber	Yes	21.4	37.7	0.24834	0.20833
Amber	Yes[a]	13.1	29.8	0.45926	0.45450
CNS	No	9.6	37.4	0.23122	0.18327
CNS[b]	No	0.0	33.6	0.23336	0.18440

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

[b] In this calculation, the Van der Waals radii of the MMP atoms were decreased by 1.0 Å

Figure 1. The structure and logic of the modular ComQum-01 program, illustrated by the division of one of the six interface programs (fixforce, the program which adds the QC and MM forces) into four independent programs (one of which there are two variants in this figure, depending on the MM program). Programs are shown in bold face, MM or QC program files in italics, and intermediate files in normal face.

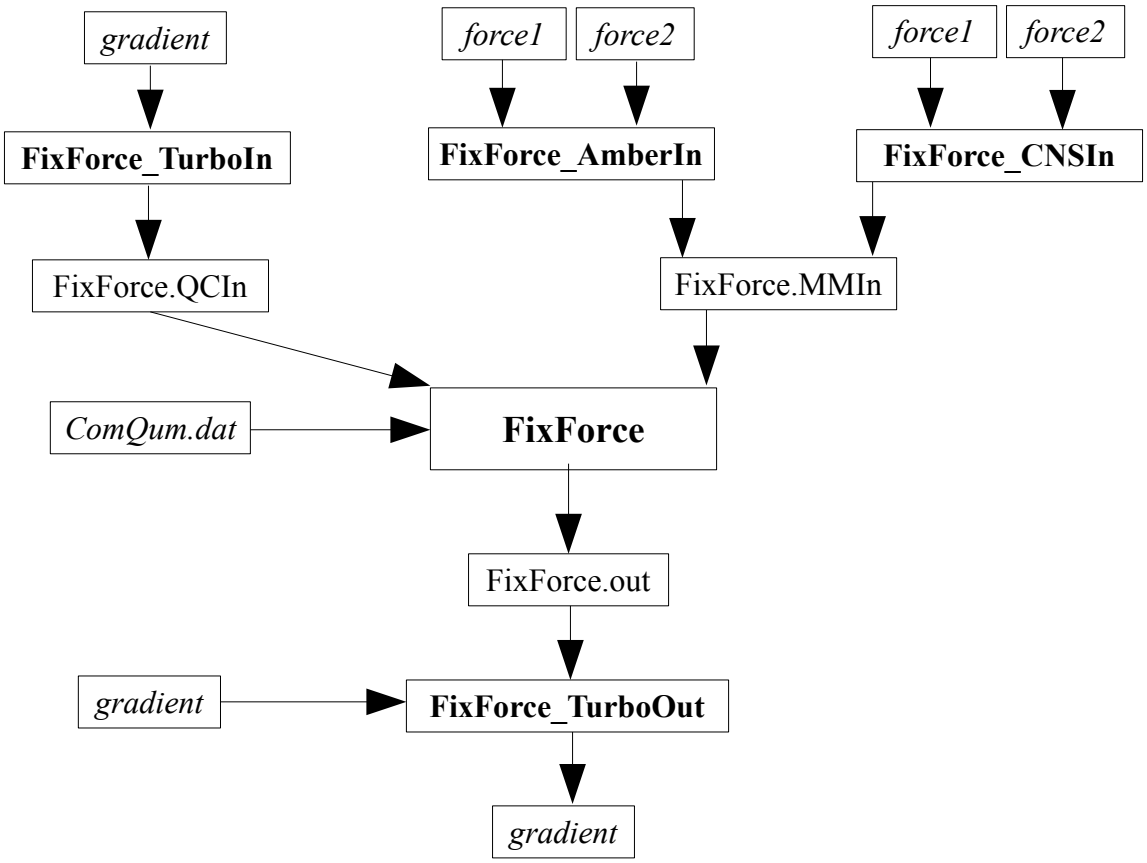


Figure 2. A comparison between MMP (a; optimised vacuum structure) and a haem group (b).

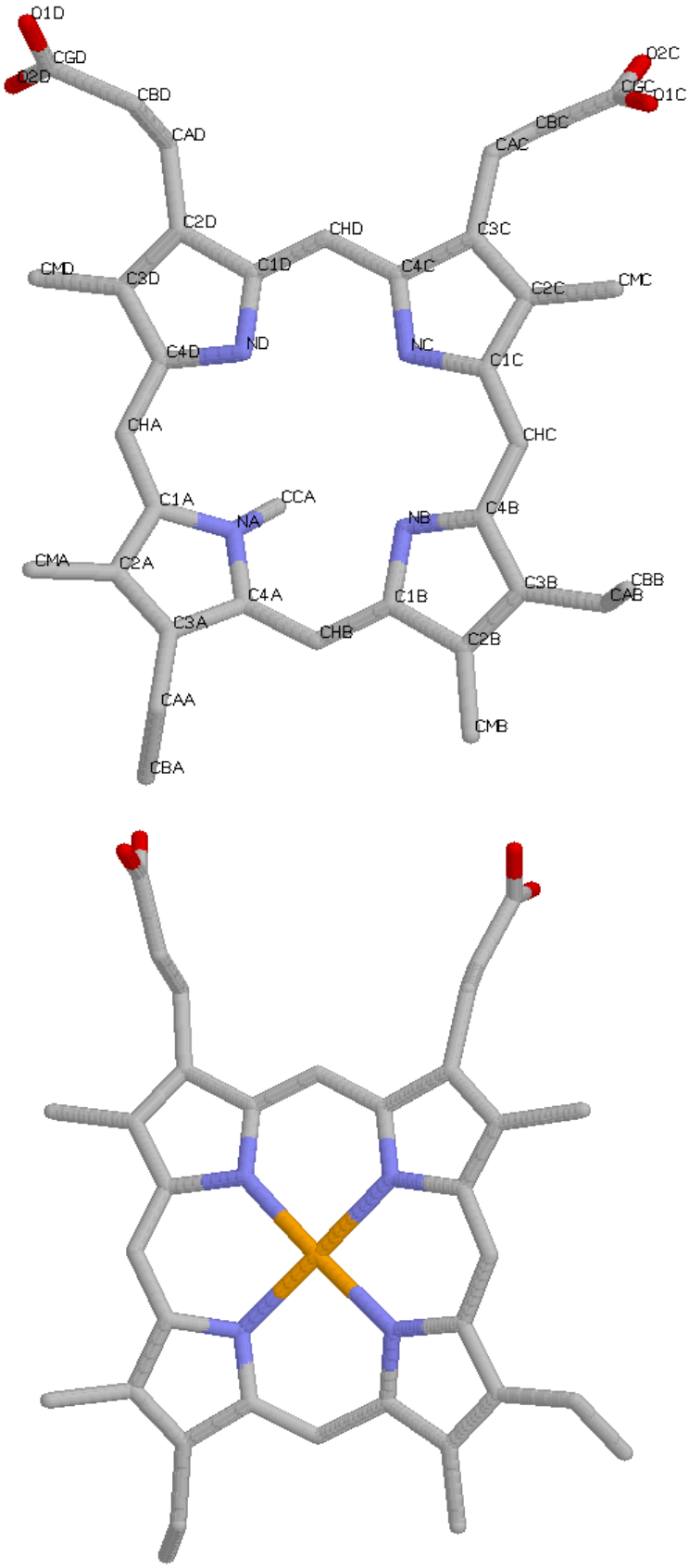


Figure 3. A comparison of MMP optimised (not the side chains) with ComQum-01/Amber without (a) or with (b) a relaxation of the surrounding protein.

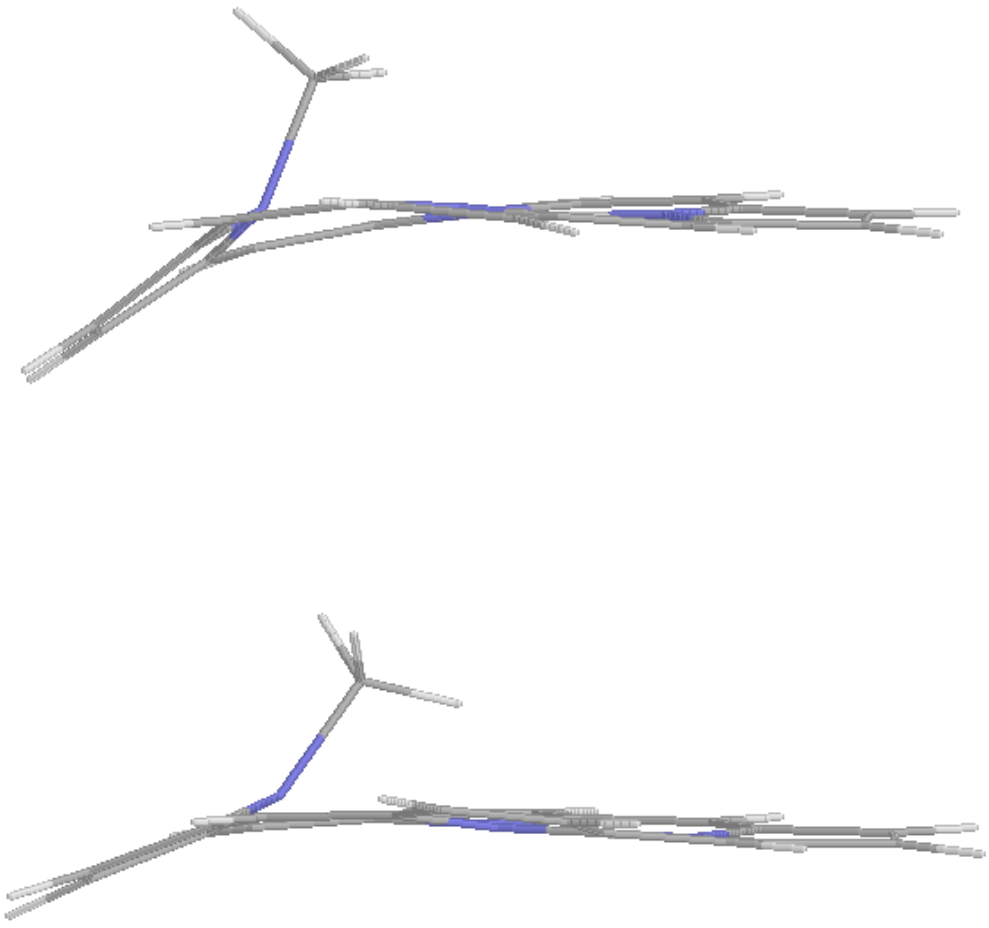


Figure 4. The ComQum-X structure of MMP in ferrochelatase, compared to the experimental electron density ($2f_o - f_c$ map, 1.4 level) and the crystal structure of MMP (blue).

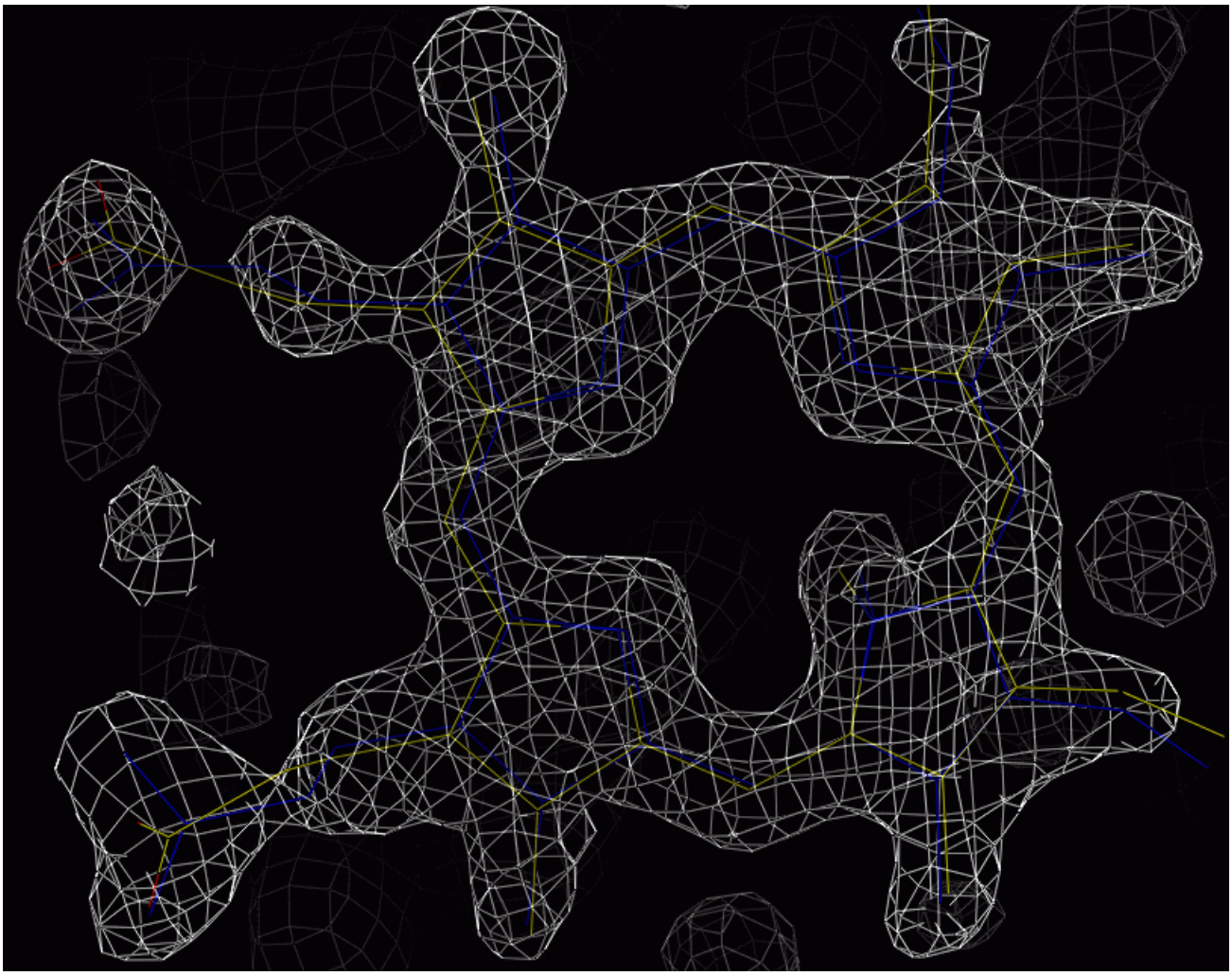
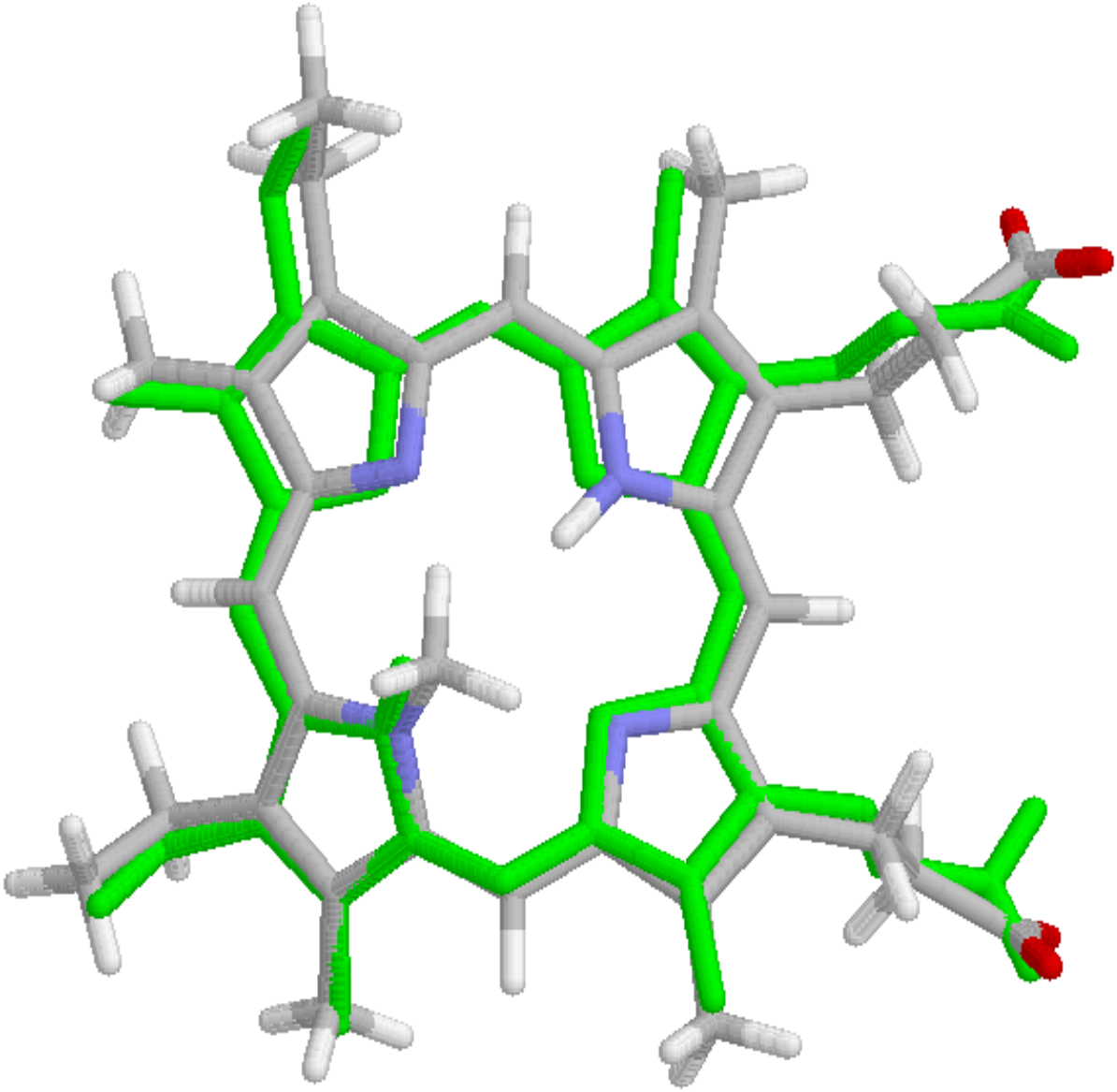


Figure 5. A comparison between the ComQum-X structure of MMP (with side chains) in ferrochelatase (w_A = 0.03, w_{MM} = 1/3) and the crystal structure (green and no hydrogen atoms) of the MMP:ferrochelatase complex [20].



Scheme 1. A flow scheme of the original ComQum-01 program. Steps in bold face constitute the actual ComQum interface (five small programs, which in the modular ComQum-01 version are divided into four programs each, two for input from the QC and MM programs, one for the actual ComQum procedure, and one for moving data to either the QC or MM program). The other steps are performed either by the QC or the MM program, whereas the whole program is a simple UNIX shell script. S1, S2, and S3 denotes systems 1 to 3.

```
Evaluate QC wavefunction
Repeat
    Evaluate the QC 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 to be relaxed
        Calculate the QC charges of S1
        Insert them into the MM representation
        Relax S2 by MM minimisation with S1 & S3 fixed
        Change the coordinates of S2 in QC representation
    Evaluate the QC wavefunction (and energy) of S1
    Evaluate the MM potential energy
    Add the energies
until convergence
```

Scheme 2. A flow scheme of the ComQum-X program. S1 and S2 denotes systems 1 and 2. Note that there is no system 3 in ComQum-X. Steps in bold face constitute the actual ComQum-X interface (three sets of four small programs). Steps in italics are performed by the crystallographic refinement program (CNS), whereas those underlined are run by the quantum chemistry (QC) program (Turbomole). The relaxation step could in principle be run with any relaxation program. In the current version, it is run by the Relax program in the Turbomole suite. The whole procedure is driven by a simple UNIX shell script.

Evaluate QC wavefunction
Repeat
    Evaluate the QC forces (within S1)
    Evaluate the crystallographic forces (from S1 & S2 onto S1)
    Add the forces
    Relax the geometry of S1 using these forces
    Change the coordinates of S1 in CNS representation
    Relax S2 by crystallographic refinement with S1 fixed
    Perform an individual B *factor refinement of S1 & S2*
    Evaluate the QC wavefunction and energy of S1
    Evaluate the crystallographic energy function
    Add the energies
until convergence