Quantum Chemistry Can Locally Improve Protein Crystal Structures

Ryde, Ulf; Nilsson, Kristina

Published in:
Journal of the American Chemical Society

DOI:
10.1021/ja0365328

2003

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

Quantum chemistry can locally improve protein crystal structures

Ulf Ryde* and Kristina Nilsson

Department of Theoretical Chemistry, Lund University, Chemical Center, P. O. Box 124, S-221 00 Lund, Sweden

RECEIVED DATE (automatically inserted by publisher) 2017-04-02; Ulf.Ryde@teokem.lu.se

X-ray crystallography is the major source of structural information, especially for proteins. The solution of a protein structure involves repeated cycles of model building and refinement.[1] Because of the limited resolution obtained for biomolecules, the experimental data are normally supplemented by some sort of chemical information, typically in the form of a molecular-mechanics (MM) force field. In all except the most accurate structures, this force field determines the details of the structure (e.g. bond lengths and angles), whereas the experimental data determine the general fold and the dihedral angles.

From a theoretical point of view, this is somewhat unfortunate, because MM is the least accurate (but fastest) theoretical method, especially as the potential traditionally used in crystallography is of a simple form (a diagonal and harmonic force field without electrostatics). Moreover, MM involves a large number of parameters and the results depend on the quality of these. For the normal amino acids, accurate parameters exist, based on a statistical survey of small-molecule data.[2] However, for unusual molecules, such as metal centers, substrates, and inhibitors, experimental data are often partly lacking. In particular, force constants are normally not available, so the crystallographer has to construct them, a complicated and error-prone procedure.[2]

This raises the possibility that crystal structures could be improved if a more accurate level of theory was used in the refinement, e.g. quantum mechanics (QM). We have recently developed such a method, *quantum refinement*, implemented in the software ComQum-X.[3] In that and later papers[4], we have illustrated the excellent performance of this method and tried to show that it may improve crystal structures locally. This is not a trivial task, because most crystallographic quality criteria (e.g. *R* factors and electron-density maps) compare how well the model fits the experimental data. Thus, the optimum fit is obtained if *no* empirical constraints are used. However, 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 force field in standard refinement is used to remedy this problem at the expense of giving a slightly worse fit. Thus, improved *R* factors and maps not necessarily flag an improved structure.

A way out of this dilemma is to find a protein that has been solved at both a low and atomic resolution (where geometric restraints have a small influence on the structure), but otherwise at as similar conditions as possible. An improved method would then bring the low-resolution structure closer to the high-resolution structure. In this study, we employ such a pair of structures, cytochrome c_{553} from *Bacillus pasteurii*, which has been solved at 97 pm resolution with ab initio phasing and independently by the same group at 170 pm resolution in a multiple anomalous dispersion experiment.[5]

ComQum-X[3] is a combination of the QM software Turbomole[6] and the refinement software Crystallography and NMR system (CNS).[7] In essence, ComQum-X replaces the MM potential in CNS by a QM calculation for a small part of the protein. Alternatively, it can be seen as a standard combined QM and MM (QM/MM) program,[8] where the structure is restrained to be close to the experimental electron density. Chemical bonds between the quantum system and the surroundings is treated by the hydrogen link-atom method.[3,8] Thus, the total energy is calculated as:

$$E_{tot} = E_{QM1} - E_{MM1} + E_{MM12} + w_A E_{Xref} \quad (1)$$

where E_{QM1} is the QM energy of the quantum system, truncated with hydrogen atoms, E_{MM1} is the MM energy of the quantum system, still with hydrogen link atoms, E_{MM12} is the MM energy of the whole protein, E_{Xref} is the crystallographic penalty function (we have used the default maximum-likelihood refinement target with amplitudes[9]), and w_A is a weight function that takes into account that E_{Xref} is in arbitrary units, whereas the other terms are in energy units.

As the QM method, we have used the density functional Becke–Perdew method (BP86),[10] and we have employed the 6–31G* basis set[11] for all atoms, except iron.[12] Density functional methods have been shown to give excellent structures for transition-metal complexes in general[14] and in cytochrome models in particular.[13] Thus, Fe–N_{Por}, Fe–N_{His}, and Fe–S_{Met} bonds to the equatorial and axial bonds to the porphyrin, and His or Met ligands are reproduced within 2–3, 4–5, and 6 pm, respectively, with the B3LYP density functional method.[13] If the latter calculations are repeated with the present method, the errors are reduced to 1–3, 0–2, and 1–3 pm.[15] This is the basis of the success of the ComQum-X method: The theoretical uncertainties are appreciably smaller than those in standard medium-resolution crystal structures (an *average* error of ~10 pm[18]). Moreover, the theoretical errors are systematic (too long bonds to the metals) and can therefore be compensated for.[19]

All ComQum-X calculations were performed with the 170-pm structure of cytochrome c_{553}[5]; the high-resolution structure was only used to judge the result of these re-refinements. The CNS calculations involved the standard CNS force field. Coordinates, occupancies, *B* factors, structure factors, space group, unit cell parameters, and resolution limits were obtained from the PDB files.[5] For other entries, default CNS values were used. The full protein was used in all calculations, including all crystal water molecules. The quantum system was Fe^{III}(porphyrin)(imidazole)-($S(CH_3)_2$), where the porphyrin was modeled either with or without all side chains.[20]

The results of the calculations are shown in Table 1 and Figure 1. It can be seen that the Fe–ligand distances in the two crystal structures are quite different, with errors of up to 32 pm in the low-resolution structure. When the heme site is re-refined with ComQum-X, the distances change appreciably and become much closer to the high-resolution structure: The error in the Fe–N_{His} bond length is reduced from 32 to 0–3 pm, that of the Fe–S_{Met} bond length is reduced from 12 to 1–5 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; errors in the vacuum structures are less than 3 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 deposited files. Therefore, we can only measure how much the *R* factor (but not R_{free}) is reduced by COMQUM-X compared to the low-resolution structure (ΔR_{low} in Table 1). It can be seen that it is improved by 0.006–0.011. We can calculate a similar *R* factor based on the high-resolution reflections. These are also given in Table 1 (ΔR_{high}) and show a similar improvement.

The amelioration is even clearer when we compare the COMQUM-X structure with the low-resolution crystal structure in Figure 1. The largest movements are seen for one of the atoms in the substituted vinyl side chains (214 pm),[21] the iron ion (18 pm), and the $N^{\delta 1}$ atom (21 pm; the average movement of all atoms is 10 pm). In all cases, the COMQUM-X structure is closer to the high-resolution structure. The high-resolution electron-density map also confirms this.

Thus, we see that the inclusion of quantum chemistry in the refinement procedure of a protein may locally improve a medium-resolution protein structure. The result is relatively insensitive to the w_A factor, but it can be used to bias the structure towards the crystal or QM structure (a low value gives a structure closer to the theoretical structure). The optimum value of w_A can be selected by optimizing the R_{free} factor. For example, in the present case, such a procedure indicates that $w_A = 0.3$ (without) or 3 (with side chains) is best, which is in reasonable agreement with the optimum in ΔR_{high}, found for $w_A = 1$ (without) and 10 (with side chains).

At present, quantum refinement is quite costly, compared to standard refinement. Thus, the first refinement took ~2 weeks on a standard PC, whereas calculations with other values of w_A converged within a few days.[22] However, we predict that as the computers become faster, quantum refinement may become a standard tool in the refinement of non-standard sites in proteins. Moreover, the effort is not prohibitive if the re-refined site is of major interest, e.g. before a theoretical investigation of the active site of a protein. In particular, COMQUM-X provides an optimum compromise between quantum chemistry and crystallography, giving a structure that can be directly compared to QM data.

Acknowledgment. This investigation has been supported by grants from the Crafoord foundation and the Swedish research council, and by computer resources of Lunarc at Lund University

Supporting Information Available: COMQUM-X structure of the quantum system ($w_A = 0.16$, with side chains). This material is available free of charge via the Internet at http://pubs.acs.org.

(1) Kleywegt, G. J.; Jones, T. A., *Meth. Enzymol.* **1997**, *227*, 208.
(2) Kleywegt, G. J.; Jones, T. A. *Acta Cryst. D* **1998**, *54*, 1119.
(3) Ryde, U.; Olsen, L.; Nilsson, K. *J. Comp. Chem.* **2002**, *23*, 1058.
(4) Ryde, U.; Nilsson, K. *J. Mol. Struct. (Theochem)* **2003**, ASAP review.
(5) Benini, S.; González, A.; Rypniewski, W. R.; Wilson, K. S.; Van Beeumen, J. J.; Ciurli, S.; *Biochemistry*, **2000**, *39*, 13115; PDB entries 1b7v and 1c75.
(6) Treutler, D.; Ahlrichs, R. *J. Chem. Phys.* **1995**, *102*, 346.
(7) Brünger, A. T.; Adams, P. D.; Clore, G. M.; Delano, W. L.; Gros, P.; Grosse-Kunstleve, R. W.; Jiang, J.-S.; Kuszewsk, J. I.; Nilges, M.; Pannu, N. S.; Read, R. J.; Rice, L. M.; Simonson; T.; Warren, G. L. Crystallography & NMR System, CNS, Version 1.0, Yale University, **2000**.
(8) Mulholland, A. J. In Theoretical biochemistry – processes and properties of biological systems (Theoretical and computational chemistry, Vol. 9). Eriksson, L. A., ed., Elsevier, Amsterdam; **2001**, 597.
(9) Pannu, N. S.; Read, R. J.; Brünger, A. T. *Proc. Natnl. Acad. Sci. US*, **1997**, *94*, 5018.
(10) (a) Becke, A. D. *Phys. Rev. A* **1988**, *38*, 3098. (b) Perdew, J. P. *Phys. Rev. B* **1986**, *33*, 8822.
(11) Hehre, W. J; Radom L.; Schleyer P. v. R.; Pople J. A. In Ab initio molecular orbital theory, Wiley-Interscience, New York, **1986**.
(12) Schäfer, A.; Horn, H.; Alrichs, R. *J. Chem. Phys.* **1992**, *97*, 2571, enhanced with one *f*, one *d*, and two *p* functions.[13]
(13) Sigfridsson, E.; Olsson, M. H. M.; Ryde, U. *J. Phys. Chem. B* **2001**, *105*, 5546.
(14) Siegbahn, P. E. M.; Blomberg, M. R. A. *Chem. Rev.* **2000**, *100*, 421.
(15) The calculated Fe–N_{Por} and Fe–N_{His} bond in $FePorIm_2$ are 201 and 199 pm both in the reduced and the oxidized complex, compared to the experimental estimates[16] of 200 (reduced) or 199–200 (oxidized), and 201 or 196–199 pm. Likewise, the calculated Fe–N and Fe–S bond lengths of $FePorS(CH_3)_2$ are 201 and 232 (reduced) or 236 pm, compared the 200 or 198 and 234 or 233–235 pm in the experimental models.[17]
(16) (a) Scheidt, W. R. *Acc. Chem. Res.* **1977**, *10*, 339. (b) Collins, D. M.; Countryman, R.; Hoard, J. L. *J. Am. Chem. Soc.* **1972**, *94*, 2066.
(17) Mahiko, T.; Marchon, J.-C.; Musser, D. T.; Reed, C. A.; Kastner, M. E.; Scheidt, W. R. *J. Am. Chem. Soc.* **1979**, *101*, 3653.
(18) Fields, B. A.; Bartsch, H. H.; Bartunik, H. D.; Cordes, F.; Guss, J. M.; Freeman, H. C. *Acta Cryst.* D **1994**, *50*, 709.
(19) Sigfridsson, E.; Ryde, U. *J. Inorg. Biochem.* **2002**, *91*, 116.
(20) It should be noted that both heme vinyl side chains are covalently attached to a Cys residue and that one of the two propionate groups (on the *A* ring) is protonated.[5] Thus, there were junctions in the His and Met ligands and either two junctions in the Cys residues or eight junctions in the porphyrin side chains. In the latter case, the side chains were described by the standard CNS force field.
(21) Although the force field used in the original refinement is not available, it is likely that the reason for this large change is that the force field was not appropriately modified for the Cys substitution and desaturation.
(22) An alternative would be to extract a MM force field from the QM calculations and use it in a standard refinement. We have also tested such a method,[23] but it gave slightly worse result and was not significantly faster.
(23) Nilsson, K; Lecerof, D.; Sigfridsson, E.; Ryde, U. *Acta Cryst. D* **2003**, *59*, 274.

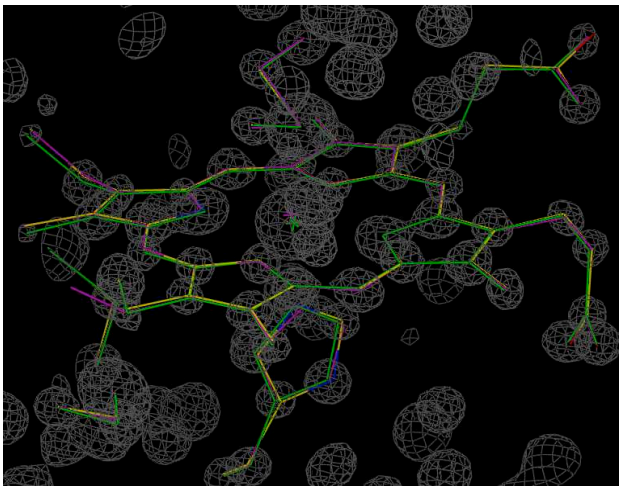
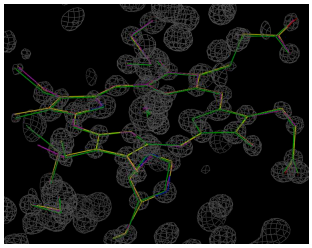


Figure 1. The low- (magenta) and high-resolution (green) crystal structure of heme cytochrome c_{553} compared to the COMQUM-X structure ($w_A = 0.16$, with side chains) and the electron density ($2f_o - f_c$ omit map at the 2.5 σ level) from the high-resolution data.

Table 1. Fe–ligand distances, strain energies[3] (ΔE_1, kJ/mole), and *R* factors for the heme group in cytochrome c_{553} calculated with COMQUM-X using the low-resolution data.[5]

Side chain	w_A	Distance to Fe (pm) N_{His}	S_{Met}	N_{Por}	ΔE_1	ΔR_{low}	ΔR_{high}
Crystal	Low[5]	231	221	202–208		0.0000	0.0000
	High[5]	199	233	197–200			
Yes	3	201	229	197–199	107.5	–0.0112	–0.0063
	1	201	229	197–199	66.4	–0.0107	–0.0063
	0.3	199	231	199	53.9	–0.0101	–0.0057
	0.16	199	231	199	51.2	–0.0055	–0.0057
	0.03	199	232	200–201	47.9	–0.0056	–0.0041
	Vacuum	199	233	200–201	0.0		
No	1	202	228	197–200	37.9	–0.0064	–0.0176
	0.3	200	230	198–201	33.6	–0.0059	–0.0171
	0.15	200	231	199–201	35.5	–0.0055	–0.0164
	Vacuum	200	235	200–201	0.0		

Authors are required to submit a graphic entry for the Table of Contents (TOC) that, in conjunction with the manuscript title, should give the reader a representative idea of one of the following: A key structure, reaction, equation, concept, or theorem, etc., that is discussed in the manuscript. The TOC graphic should be no wider than 4.72 in. (12 cm) and no taller than 1.81 in. (4.6 cm).



We have re-refined the X-ray structure of the heme site in cytochrome c_{553}, supplementing the crystallographic data with quantum chemical geometry optimizations, instead of the molecular-mechanics force field used in standard crystallographic refinement. By comparing the resulting structure, obtained using medium-resolution data (170 pm), with an atomic-resolution structure (95 pm) of the same protein, we show that the inclusion of quantum chemical information into the refinement procedure improves the structure significantly. Thus, errors in the Fe–ligand distances are reduced from 3–32 pm in the low-resolution structure to 0–5 pm in the re-refined structure, one side-chain atom changes its conformation (a movement by 214 pm towards its position in the high-resolution structure), and the *R* factors are improved by up to 0.018. Thus, quantum refinement may be a powerful method to obtain an accurate structure for interesting parts of a protein.