Quantum mechanical/molecular mechanical (QM/MM) methods and applications in bioinorganic chemistry

Ryde, Ulf

Published in:
Computational inorganic and bioinorganic chemistry

2009

Document Version:
Early version, also known as pre-print

QM/MM Methods and Applications in Bioinorganic Chemistry

Ulf Ryde *

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

Correspondence to Ulf Ryde, E-mail: Ulf.Ryde@teokem.lu.se, Tel: +46 – 46 2224502, Fax: +46 – 46 2224543

2017-03-19

Abstract

The combination of quantum mechanics and molecular mechanics (QM/MM) methods is a promising approach to study the structure, function, and properties of proteins. The number of QM/MM applications on metalloproteins has been steadily increasing and it is the method of choice to study the influence of the protein onto the structure, spectroscopy, and energetics of an active site. In particular, it has been shown that QM/MM calculations can provide geometries of metal sites in proteins with an accuracy that is better than in low- and medium-resolution crystal structures. However, it is quite hard to obtain reliable and converged energies. Therefore, QM/MM is still an area of method development and no consensus of the best treatment of a protein has yet been reached. This review covers various aspects of QM/MM studies of metalloproteins, including methods to incorporate experimental data in the calculations, methods to obtain reliable energies, as well as some examples of typical applications.

Key Words: QM/MM, density functional theory, quantum mechanics, molecular mechanics, metalloproteins

1 Introduction

During the last decades, theoretical methods have become an important complement to experiments for the study of structure and function of metalloproteins, mainly owing to the increase in computer power and the introduction of accurate density functional methods.[1,2] Quantum mechanical (QM) methods have the advantage of being accurate and free of empirical parameters, and are therefore applicable to any types of systems, including those involving chemical reactions. However, even if QM calculations of whole proteins have started to appear,[3] the need of proper solvation and an account of dynamic and entropic effects make such calculations still too expensive for routine use.

The alternative is to use molecular mechanics (MM) methods, which ignore the electrons and therefore easily can be applied to systems of over 100 000 atoms, including sampling over many nanoseconds. However, MM methods need to be parameterized for all new types of molecules and typically have a lower accuracy than QM methods. In particular, it is challenging to obtain general parameters for metal-containing systems.

Therefore, QM/MM methods have been suggested as a way to combine the accuracy of QM methods with the speed of MM methods.[4] In these, a small but interesting system is studied by QM methods, whereas the surroundings are treated by MM methods. Many variants of QM/MM methods and programs have been proposed and QM/MM codes are now available in most quantum chemistry and molecular dynamics (MD) software.[5,6,7] This has led to a steadily increasing number of applications of QM/MM methods to proteins.[5,6,7]

In this chapter, I review development and applications of QM/MM methods relevant to bioinorganic chemistry. In particular, I will discuss the advantages and limitations of QM/MM methods compared to other approaches. The applications are restricted to metalloproteins; QM/MM applications to non-biological inorganic chemistry has recently been reviewed,[8] as well as applications to proteins in general.[5,6,7,9]

2 Methods

There are already many excellent reviews about QM/MM methods and methodology.[5,6,7] Therefore, I will only give a short and simple-minded introduction to the method. The philosophy behind the QM/MM methods is to divide the full system (protein + solvent) into two parts, as is shown in Figure 1. One part (called system 1 or the QM system), contains the most interesting part of the protein, typically the active site. It is treated by QM methods. The rest is called system 2 or the MM system and is treated by MM (in more sophisticated QM/MM methods, system 2 may be divided into additional layers).

In principle, the QM/MM approach is simple: The QM energy of system 1, E_{QM1}, is added to the MM energy of system 2, E_{MM2}, together with a term, describing the interaction between the two systems E_{int12}:

$$E_{QM/MM} = E_{QM1} + E_{MM2} + E_{int12} \quad (1)$$

Different QM/MM approaches differ in the details of the latter interaction energy:[5,6,7] The simplest approach is to calculate it at the MM level, leading to what is called mechanical embedding. The next step is to include a point-charge model of the surroundings in the QM calculations, giving a polarization of the QM system by the surroundings, electrostatic embedding. Finally, the QM may also polarize the surroundings in a self-consistent manner, requiring a polarisable MM force field and special QM programs.

Another important source of variations in the QM/MM methods is the treatment of covalent bonds between the QM and MM systems. QM calculations need fulfilled valencies everywhere. Therefore, the QM system is either truncated by hydrogen atoms (or other appropriate pseudoatoms) or are special localized orbitals used in the junctions.[5,6,7] Both methods have their advantages and shortcomings and they typically give similar results when used with care.[6,7] In fact, the actual treatment of the junction bond is not as crucial as the treatment of the electrostatics around it, i.e. whether charges close to the junction atoms

should be included, scaled down, or removed.[6,7,10] In fact, different treatments may give results that differ by up to 80 kJ/mol if the reaction involves a change in the total charge. Unfortunately, no single method seems to give good results for all calculations.[10] On the other hand, the effect of the junctions can be estimated by enlarging the QM system.

3 Structures

Undoubtedly, QM/MM is an excellent method to obtain structures of metalloproteins in various states. This is because structures are local properties, insensitive to details of the surroundings. In fact, already vacuum DFT structures typically reproduce covalent bonds with an error of less than 0.02 Å and metal–ligand distances to within 0.07 Å,[1,2] and this excellent performance applies also to the QM/MM methods.[11] This is actually better than low- and medium-resolution crystal structures, for which the average error is ~0.1 Å and errors of up to 0.3 Å are frequently encountered.[12] However, the accuracy of the theoretical bond lengths are directly related to the strength (and therefore to the length) of the bond. Strong and short bonds are most accurately reproduced. For example, metal–ligand double bond, such as the Fe=O bond in compound I of heme enzymes (*see* IA105-, IA106-) are typically reproduced within 0.03 Å, whereas normal metal–ligand bond lengths (1.8–2.2 Å) may have errors of up to 0.07 Å. Longer bonds are much weaker and often more determined by interactions with the surrounding protein than by the metal–ligand interaction. For such bonds, much larger differences between QM calculations and experiments may be encountered. A typical example is the weak axial Cu–S bond of the methionine ligand in the blue copper proteins[13] (*see* IA056-) and the axial Co–N bond in coenzyme B_{12} proteins (*see* IA045-).[14] Likewise metal–metal distances and hydrogen bonds can exhibit errors of over 0.1 Å.[15]

3.1 Combination with experimental data

Essentially all QM/MM studies of a protein start with a reoptimisation of a crystal structure. The reason for this is that the crystal structure itself cannot be used in a theoretical investigation, because there are small errors in both the crystal structure and in the QM/MM method that make energies calculated directly on a crystal structure useless (errors of >100 kJ/mol compared to a QM/MM structures).[16] On the other hand, there is no guarantee that a QM/MM structure remains close to a crystal structure, because the MM potential is often rather inaccurate and the crystal structure is an average structure.

A natural solution to this problem is to restrain the QM/MM calculations to the crystallographic raw data.[12] However, such a procedure should not use the crystal structure directly, because it is the result of an involved and partly subjective interpretation of the crystal data. Instead, the crystallographic raw data are the intensities of the reflections, called the structure factors (they are available in the Protein Data Bank for about half of the structures). These, together with some phase information give an electron-density map, in which an initial structure may be build. Then an involved cycle of refinement and model rebuilding is performed, during which errors in the initial structure are removed. The refinement takes the form of a geometry optimization, using an energy function of the form

$$E_{\text{X-ray}} = w_{\text{A}}\, E_{\text{Xref}} + E_{\text{MM}} \qquad (2)$$

where E_{Xref} is a measure of how close the structure reproduce the structure factors.[17] However, in all except the most accurate structures, this measure is supplemented by a standard MM energy function, which is used to ensure that the crystal structure is chemically reasonable (the crystal data are insensitive to small details in the structure, e.g. bond lengths and angles). Because E_{MM} is in energy units, whereas E_{Xref} is in arbitrary units, the two terms need to be weighted with the factor w_A, which determines the relative importance of the two terms. It is normally determined so that the X-ray and MM forces have equal root-mean-squared magnitudes in a short MD simulation.[18] This means that *most crystal structures actually are*

50 % theoretical.

Of course, this means that the final crystal structure will depend on what MM potential was used. This is a relatively small problem for standard amino-acid residues and nucleic acids, for which accurate force fields are available.[17] However, for metal sites, substrates, and inhibitors, no standard force field is available and the crystallographer has to construct it himself, an error-prone procedure. Even worse, such non-standard force fields are rarely specified, so the errors cannot be identified without re-refining the structure.

Metal sites provide a special problem in crystallography, because there are no accurate general force fields for metals. It is therefore often stated that the metal site was refined without any MM restraints. Unfortunately, this often means that the crystallographer has not defined any explicit bonds between the metal and its ligands (which has to be done by hand). However, this means that there will be van der Waals interactions between the metal and the ligands. With standard parameters, such an interaction has a minimum around 2.5 Å, so there is actually a quite strong restraint to increase the distances. Therefore, metal–ligand distances are often too long in crystal structures. The only way to have an unrestrained metal–ligand interaction is to define bonds between the metal and its ligands and set the force constants to zero.

A more accurate solution to these problems is to combine QM/MM and X-ray crystallography with the energy function[12]

$$E_{\text{ComQum-X}} = w_A E_{\text{Xref}} + E_{\text{QM/MM}} \quad (3).$$

This means that the QM/MM energy function is restrained to be close to the crystallographic structure factors, or equivalently, that the inaccurate MM energy function in a crystallographic refinement has been replaced by a more accurate QM potential for a small, but interesting part of the protein. This defines the method of quantum refinement. We have shown that such an approach may actually improve crystal structures of metalloproteins significantly. For example, quantum refinement reduces the errors in the Fe–N_{Por}, Fe–N_{His}, and Fe–S_{Met} distances in a 1.7-Å resolution structure of cytochrome c_{553} (*see* IA104-) from 0.03–0.09, 0.32, and 0.12 Å to 0.01–0.02, 0.00, and 0.02 Å, compared to an atomic-resolution structure of the same protein.[11]

Quantum refinement can also be used to interpret crystal structures, i.e. to determine what particles are actually present in the structure. This is especially important for electrons and protons, which are not directly visible in normal crystal structures. For example, the method can be used to determine the protonation state of metal-bound solvent molecules (water, OH^-, or O^{2-}) or the oxidation state of metal ions in a structure.[12] The latter is important, because most metals in crystal structures are reduced during data collection, making the final structure a mixture of two or more oxidation states. The quantum refinement method is sensitive to mixtures in the QM system, so it is a powerful method to identify such mixtures, showing that the original structure is not reliable.[12]

A second source of protein structural data is nuclear magnetic resonance (NMR; *see* IA319-). Although the method is very different from X-ray crystallography, the treatment is similar: The raw data (mainly atomic distance or dihedral restraints) need to be processed by theoretical methods to produce a final structure (or typically an ensemble of ~20 possible structures) and the details of the structure (bond lengths and angles) are to a major extent determined by a MM force field. Therefore, QM/MM calculations can be used to supplement the NMR data to obtain improved structures of important parts of the structure, e.g. metal sites, in the same way as for quantum refinement.[12] This has been done for two Ca^{2+} sites in the epidermal growth factor-like domains 3 and 4 of protein S (*see* IA032-).[12]

Local information of metal sites in proteins can also be obtained by extended X-ray absorption fine structure (EXAFS) measurements (*see* IA330-). This approach is quite different from X-ray crystallography and NMR structure determination. In particular, it does not give a full structure (i.e. coordinates), not even of the metal site, but only a set of metal–ligand distances. Therefore, it is harder to combine with theoretical methods. However, pure

QM calculations can be used to supplement the EXAFS data, in the same way as MM is used in NMR and X-ray refinements.[12] This has been used to identify possible sitting-atom complexes in the metallation of porphyrins (*see* IA111-).[12]

Moreover, QM/MM methods can be combined with EXAFS to obtain improved structures of metal sites in proteins.[12] This has been used to identify intermediates in the reaction cycle of the multi-copper oxidases (MCO). The MCOs are a group of enzymes that couple the four-electron reduction of molecular oxygen to water with four one-electron oxidations of various substrates,[19] e.g. laccase, ceruloplasmin, and ascorbate oxidase (*see* IA055-). The active site of the MCOs contains a trinuclear Cu cluster. Extensive kinetic and spectroscopic studies have been performed to deduce the mechanism of the MCOs.[19] In particular, the peroxy intermediate (PI) has been identified, arising after the binding of O_2 to the reduced protein. Unfortunately, it was not possible to deduce the structure of these two intermediates; instead two possible structures were suggested for it, one with O_2^{2-} bridged to the three coppers in the center of the cluster (C_3), and one with HO_2^- bridged to two coppers on the side of the cluster (S_{23}) and with an additional OH^- bridging two of the coppers. A QM/MM study refined these two structures and pinpointed C_3 as the more stable structure, although the electronic ground state was not in accordance with experimental data.[20] Both structures gave Cu–Cu distances in accordance with EXAFS data (~3.4 Å). However, combined EXAFS/QM/MM calculations showed that the C_3 structure fitted the experimental EXAFS raw data (i.e. not only the Cu–Cu distances) much better than the S_{23} structure.[12] This has later been confirmed by other calculations and experiments.[21,22]

4 Energies

The prime use of theoretical methods is not to give structures, but rather energies, which govern all chemical processes and are hard to obtain experimentally, especially for short-lived intermediates and transition states. Unfortunately, it is quite hard to obtain reliable energies also with QM/MM methods. The reason for this is that energies are global properties. For example, if a water molecule far from the active site forms an extra hydrogen bond, the energy may change by 20 kJ/mol, a highly significant amount, although this is most likely totally irrelevant for the process of interest. Moreover, many energy terms are long-range, in particular electrostatics and solvation effects. For example, the interaction between two charged groups is 17 kJ/mol at a distance of 20 Å, even if screened by a dielectric constant of 4, as is normally assumed for a protein. In fact, different methods to estimate relative energies between two states in a protein may differ by 100 kJ/mol.[23] Therefore, you must be extremely careful to ensure that the obtained energies are relevant and reproducible. These problems are much more severe for QM/MM calculations than for normal QM calculations in vacuum, where the studied system is so small that a visual inspection may directly show whether the structures of interest have the same peripheral interactions (i.e. whether they are in the same local minima). In a full protein, this is impossible and in standard MM studies, the problem is solved by studying a large ensemble of structures during a MD simulation, obtaining free energy. However, many QM/MM studies are still based on energies from a single minimization, something that would be impossible to publish in the MM community.

There are many ways to address these problems. First, you may thoroughly check the QM/MM energies, in particular, the energy components in Eqn. 1, i.e. the vacuum QM energy of the QM system E_{QM1} (both at the vacuum and QM/MM geometry), the MM energy of the MM system E_{MM2}, and the interaction energy between the QM and MM systems $E_{inter12}$. The latter term is normally dominated by electrostatics, which can be divided into contributions from each residue in the surrounding protein, if the QM system is described by a point-charge model. These energy contributions should be chemically reasonable. Moreover, E_{MM2} should be similar for related systems. It is in E_{MM2} the effect of local minima is expected to be seen; therefore, significant energy differences in E_{MM2} should only be accepted if they make

chemical sense (again, this term can be divided residue-wise). Finally, the E_{QM1} term should be similar to what is obtained in vacuum calculations; otherwise, it is an indication that you might study different electronic states, which is not uncommon for open-shell transition-metal complexes. Unfortunately, these energy components are rarely presented in QM/MM investigations, making it hard for a reader to judge the reliability of the results.

A second way to make QM/MM energies more reliable is to run the calculations several times. For example, you can ensure that two states are in the same local minimum by running the optimization forth and back between the two states several times until the QM/MM energy difference is stable.[24] Alternatively, you may start several QM/MM calculations from different structures (e.g. from snapshots of a MD simulation) and take the average of the resulting energies.[5,6,7,25]

However, the most accurate method is to perform free-energy perturbations (FEPs) between the various states of interest.[7,26,27] As the name indicates, this also provides free energies, rather than pure energies, giving a proper account of entropy effects. Unfortunately, FEPs require millions of energy and force calculations, which may be prohibitively expensive at an accurate QM/MM level of theory. Therefore, pure QM/MM FEPs can normally only be afforded at a semiempirical QM level. However, methods have been developed to perform QM/MM FEPs, using only a restricted number of QM calculations, e.g. the QM/MM-FE (free energy), QM/MM-PBSA (Poisson–Boltzmann solvation, combined with a surface-area method for non-polar solvation effects), or QTCP (QM/MM thermodynamic cycle perturbation) and related methods.[7,26,28,29,30] These methods perform the sampling only at the MM level and then estimate the QM/MM free energies either by single-point extrapolations, by simplified end-point perturbations, or by full MM → QM/MM FEPs. Comparisons between full FEPs and QTCP (at the semiempirical level) have shown that the methods give accurate estimates of the full QM/MM FEP free energy.[31]

As mentioned above, QM/MM energies are sensitive to long-range effects. Therefore, the set-up of the calculations is crucial. A common approach is to optimize only the part of the protein that is close to the active site (~10 Å).[7] The reason for this is that the MM force fields are not accurate enough to keep the structure close to the original crystal structure and that the optimization might become unstable for large systems.[7,16] Unfortunately, this typically means that the surrounding solvent is not properly relaxed for all states of the active site, meaning a systematic bias. Therefore, if only parts of the protein are optimized, it is mandatory to postprocess the QM/MM energies by a FEP or continuum solvation approach.[28,32]

Still another problem is caused by charged groups on the surface of the proteins. Experimentally, it is known that they have a small influence on properties of the active site.[33] However, even in a QM/MM FEP approach, they can have a strong influence on calculated energies, as is shown in Figure 2. This is related also to the common practice to truncate large proteins at a distance of 20–30 Å from the QM system (to reduce the computational load),[7,34] which can have a large impact on energies, as is also shown in Figure 2. This indicates that a MD sampling for several hundreds of ps is not enough to allow charged residues to fully relax. This may be solved by longer simulations (expensive) or continuum solvation methods, but a simpler practical solution is to scale down the interactions with solvent-exposed charges, as is done by several groups.[5,7,10,33,35] Fortunately, the results are insensitive to the actual scaling factor (the effective dielectric constant), provided that it is large enough.[32]

Probably the most important question is whether there is any reason to run QM/MM calculations on proteins, considering their problems. Do vacuum QM calculations with large enough systems give the same (and more stable) results?[1] The influence of the protein and solvent on protein reactions has been much discussed in literature[5,36,37,38] and the conclusion is that it depends on the system and properties studied. For some properties, e.g. many structures and some reaction energies, the surroundings have a rather limited influence. However, for other properties, the protein and solvent environment is crucial for the reactions, changing the energies by over 100 kJ/mol.

The question is then if this effect can be modeled in a simpler way than with explicit QM/MM calculations. Unfortunately, this does not seem to be the case: In Table 1, a comparison between the results obtained with various commonly used simplified QM methods for five simple proton-transfer reactions in two proteins are shown.[28,32] The results are compared to a state-of-the-art QM/MM FEP approach. It can be seen that only two approaches give reliable results with mean absolute errors of less than 20 kJ/mol compared to the QM/MM FEP results. These are (1) the QM/MM-FE approach,[26] which also use a FEP at the MM level to estimate the free energy change between the two states, but then simply add a QM energy estimate to this free energy, and (2) the QM/MM-PBSA approach,[28] which adds a continuum solvation and an entropy estimate to the raw QM/MM energies. The latter approach is especially interesting, because it can be used directly on the QM/MM structures, without any conformational sampling, which may save much computer time. This indicates that at least for these two systems, entropy effects outside the QM system are negligible.

On the other hand, it can be seen that many other popular approaches give large errors. This includes the original QM/MM data, vacuum calculations, with or without a continuum estimate of solvation effects[1] and the popular Poisson–Boltzmann approach.[39] Thus, a QM model with ~100 atoms cannot accurately model the effects from the surrounding protein, even if studied in a continuum solvent. This is also illustrated by the free energy components in Figure 2, which are divided into contributions from each residue and plotted against the distance between the residues and the QM system.[32] It can clearly be seen that the energy does not stabilize until after ~10 Å, which means that ~2000 atoms need to be included before the energy is converged.

Finally, two approaches should be mentioned that were not included in this comparison, viz. QM/MM FEP based on semiempirical methods[5,6,26] and pure MM FEP, typically based on the empirical valence-bond (EVB) approach.[37] Both methods have successfully been applied to many different (metallo)proteins. They have the advantage of being fast, thereby allowing a proper sampling of the conformational space and calculations of free energies using FEP. On the other hand they are based on inherently approximate methods, which need to be thoroughly parameterized to give reliable and accurate results, something that is not always trivial.

It should also be mentioned that recently methods have been developed to include high-level QM methods, such as local coupled-cluster methods in QM/MM calculations. Combined with free-energy calculations at a lower level of theory, it is claimed that they can obtain near chemical accuracy (~4 kJ/mol) for reaction and activation energies in proteins.[40] Although I doubt that this applies also for reactions that are strongly affected by the surroundings, this is a useful and important development. Another promising recent approach is to improve the description of the interaction between the QM system and the surrounding protein by a combination of very accurate MM force fields (with a multicentre–multipole description of electrostatics and anisotropic polarizabilities for induction), calculated directly on the relevant protein structure, combined with fragmentation techniques, allowing for pairwise calculations of the active site and all residues in the nearest surrounding (400–800 atoms), using high-level QM calculations with saturated basis sets.[41] This would provide as accurate energies as is possible with a partly MM-based approach.

5 Applications

The application of QM/MM methods to metalloproteins is a rather new field, with about 200 publications since 2002 (the time of the latest QM/MM metalloprotein review[36]). This is about a fifth of the QM/MM publications, the other being devoted to other proteins,[7] non-protein organic or inorganic reactions, solid-state applications, and method development. The number of publications is steady increasing from ~10 in 2002 to ~40 in 2006–2007.

Among the metals, applications to iron proteins are strongly dominating (almost 50%),

especially heme proteins (cytochrome P450, heme peroxidase, catalase, heme oxidase, chloroperoxidase, NO synthase, and various globins; *see* IA103-, IA104-, IA105-, IA106-), although many other iron proteins have also been studied, e.g. non-heme oxidases, methane monooxygenase, ribonucleotide reductase, superoxide dismutase, and hydrogenases (*see* IA112-, IA113-). There are also many applications on zinc (especially metallo-β-lactamase, carbonic anhydrase, zinc proteases, and alcohol dehydrogenase; *see* IA259-), copper (e.g. blue-copper proteins, multicopper oxidases, nitrite reductase, prion protein, Cu chaperones), and magnesium (catechol *O*-methyltransferase, kinases, and phosphatases; *see* IA039-) proteins. For the other metals, less than six applications each have been published since 2002, e.g. on oxygen-evolving complex (*see* IA192-), Mn superoxide dismutase (*see* IA0129), [Ni,Fe] hydrogenase (*see* IA149-), V-peroxidases (*see* IA252-), vitamin B_{12} enzymes (*see* IA045-), and Mo nitrite reductase (*see* IA146-).

As already discussed above, QM/MM is an excellent method to study various structural aspects of metalloproteins, e.g. to obtain accurate structures,[12] as well as electronic structures,[7] to predict the structure of a bound substrate or inhibitor to a metal site,[7] or to discuss the role of strain in protein function.[13,37] Other studies have considered enzymic reactions and mechanistic aspects of the proteins.[7] This includes studies of reduction potentials, and electron (*see* IA270-), as well as proton transfer reactions. These are especially complicated because they involve large changes in the electrostatics of the active site, making them sensitive to the treatment of the surroundings.[23,28,32] However, there are also studies of spectral properties, including electronic, NMR, EPR spectral properties, as well as Mössbauer parameters (*see* IA314-) and Heisenberg exchange-coupling constants.[7,9] In this section, I will discuss in detail only two selected applications, to illustrate some interesting aspects of the QM/MM modeling of metalloproteins.

The metalloprotein that has been most studied QM/MM methods is undoubtedly cytochrome P450 (*see* IA105-). In a series of almost 20 publications the groups of Shaik and Thiel have systematically studied all intermediate states in the reaction cycle of this interesting protein.[7,9,42] This has given a detailed understanding of all steps in this rather complicated reaction cycle. Particular interest has been devoted to the elusive compound I, which has been stipulated as the reactive intermediate in the cycle, although it has actually never been unambiguously observed for cytochrome P450. The electronic structure has been scrutinized and compared to the same intermediate in vacuum studies and in other heme proteins. The two-state reactivity of compound I, involving both the doublet and quartet states, has been confirmed, but other possible reactants have also been tested. Several other groups have also contributed to the understanding of the mechanism of this large group of related enzymes, and several human variants have also been studied.[42] The QM/MM studies have largely confirmed the results of small-model studies of this enzyme,[42,43] although they have pinpointed some modulation by the surrounding protein.

Of particular interest are two early studies of the hydrogen-abstraction from the substrate by compound I in the prototypical enzyme P450cam.[35] One of the investigations indicated that the activation energy is 49 kJ/mol and that one of the propionate side chains of the heme group is important for the catalysis, having significant spin density in some steps of the reaction. However, the other study gave an almost twice as high barrier, 92 kJ/mol, and saw no effect of or spin density on the propionate side chains. Of course, the two studies differed in many details of the calculations, but a joint study managed to pinpoint two main causes of the differences: In the first study, the two states had been in different local minima of the surroundings, which had artificially decreased the barrier by ~20 kJ/mol. On the other hand, in the second study, a crystal-water molecule was missing, which could decrease the barrier by ~20 kJ/mol. Moreover, the two studies differed in the assignments of the protonation state of some residues in the protein, which could change the energies by 8 kJ/mol. With a proper protonation of the surroundings, no spin density was found on the propionate groups. This illustrates how sensitive the QM/MM results are to the details of the calculations and the

importance of the initial setup of the calculations.

Likewise, three investigations have dealt with the cleavage of the unusual Co–C bond in coenzyme B_{12} enzymes (*see* IA045-), two with QM/MM and one with the EVB approach (i.e. with a reactive MM force field).[14] All three investigations obtained similar energetic results (viz. that the equilibrium constant for the cleavage is close to 1 in the proteins, although the reaction energy in solution is ~130 kJ/mol). However, the details of the results were quite different. For example, one investigation obtained a transition state and a reactive intermediate, whereas the other two showed monotonous reactions. Moreover, one of the investigations assigned all the catalytic effects to electrostatics, whereas another showed a more complicated picture with contributions from several factions, including direct and indirect (via the geometry) electrostatic, as well as steric effects. The discrepancy between the three investigations can be traced to the fact that no crystal structure is available for the reactive enzyme with an intact Co–C bond. Therefore, this structure has to be generated by the calculations, which might end up in incorrect states, even if one of the QM/MM investigations involved repeated runs forth and back between the two states and also cycles of simulated annealing MD simulations within the QM/MM steps, and the EVB investigation involves full dynamics and free-energy perturbations between the two states.

However, the latter two investigations nicely illustrated how the calculations can be used to analyze the results in detail and to provide contributions to the observed energies, both from various parts of the protein (residues) and substrates and coenzymes, as well as from various terms in the energy function (bonded terms, electrostatics, van der Waals interactions, solvation). In this respect, the QM/MM calculations have the advantage of giving a single, well-defined structure that can be shown and compared to experiments. On the other hand, the EVB calculations include dynamics and entropy effects that might be important. These two studies also illustrated the possibility to design computer experiments in the form of in silico mutations of both the protein and the coenzyme. They pinpointed the importance of the ribose moiety of the coenzyme as a polar handle for the protein, thereby explaining why the similar coenzyme methylcobalamin cannot be used in this type of reaction.[14]

6 Practical recommendations

Considering how sensitive the QM/MM results are to the setup of the calculations and that there is no consensus how this should be done, I give in this section, my recommendations how to do reliable QM/MM calculations for a metalloprotein.

- It is normally wise to first study the reaction thoroughly with a small QM model in vacuum. This gives the opportunity to test many different reaction mechanisms, compare different DFT methods and basis sets, and to get a first feeling of solvation effects, using continuum solvent models.
- The investigator should spend much time (several days) and effort for the setup of the QM/MM calculations, because errors at this stage cannot be cured at later stages. The investigator should consider previous (theoretical and experimental) investigations and use available computational tools[7] to obtain correct conformations and protonation states of all residues. The protein should be properly solvated by 2–3 layers of explicit water molecules, but it can probably be truncated outside ~25 Å of the active site.
- The QM system should be selected to include all metal ligands and other residues crucial to the reaction mechanism. Junction atoms should be at least four bonds away from the reactive center.
- The protein should first be equilibrated by a MD simulation. If the starting point is a relevant crystal structure, only added hydrogen atoms and water molecules need to be optimized.
- Next, QM/MM structures are optimized along the reaction path. The actual implementation of the QM/MM method and the treatment of the junction atoms are

probably not crucial for the structures. However, much care should be taken to avoid ending up in different local minima for different states by thoroughly examining the QM/MM energies and energy components, running each steps forward and backward several times, or using several starting conformations from a MD simulation. It is probably wise to optimize only residues within ~10 Å of the QM system. These calculations can be run by a pure DFT method and a medium-sized basis set (split-valence with one set of polarizing functions), or even at a semiempirical or EVB level.

- Then, energies should be recalculated by single-point calculations with a hybrid DFT method and larger basis set (triple zeta quality with two sets of polarizing functions). Moreover, entropy effects should be estimated, preferably by a QM/MM-FEP approach. If this cannot be afforded or is technically demanding, post-processed energies, including a continuum solvent model should be calculated. In these calculations, solvent-accessible charged residues should be neutralized. Moreover, entropy and zero-point energies of the QM system should be included, taken from the QM-only calculations.
- Finally, the results should be analyzed, using energy components and residue contributions. Residues with large energy contributions should be added into the QM system, especially if they are covalently attached. Likewise, the effects of the junctions should be tested by enlarging the QM system (single-point energy calculations). Computational mutations may be performed to get further understanding of the reaction.

7 Conclusions

In this review, I have tried to emphasize the strengths and limitations of the QM/MM methods for applications in metalloprotein chemistry. We have seen that QM/MM is a powerful approach to obtain structures of metal sites in proteins, providing data that are more accurate than low- and medium-resolution crystal structures. If combined by experimental data, it may be used to improve and interpret experimental structure, e.g. to deduce protonation states of metal-bound solvent molecules or identify photoreduction of metal sites.

On the other hand, QM/MM energies are more problematic, and there is still no consensus how such calculations should be performed. Yet, it is clear that only methods taking a full and detailed account of the surrounding protein are able to model the effects of the surroundings to an accuracy better than ~25 kJ/mol. Great care should be taken to avoid ending up in different local minima and fully reliable results are only obtained with free-energy methods. In addition, solvation and long-range interactions must be treated consistently. Still many interesting results concerning reactions and mechanisms have been obtained for many groups of metalloproteins, including heme enzymes, globins, blue-copper proteins, copper oxidases, coenzyme B_{12} enzymes, hydrogenases, as well as zinc and magnesium enzymes.

8 List of abbreviations and acronyms

EVB – empirical valence bond
EXAFS – extended X-ray absorption fine structure
FE – free energy
FEP – free-energy perturbation
MCO – multicopper oxidases
MD – molecular dynamics
MM – molecular mechanics
QM – quantum mechanics
QM/MM – combined quantum and molecular mechanics calculations

QTCP – QM/MM thermodynamic cycle perturbation
PBSA – Poisson–Boltzmann solvation, combined with a surface-area method
PI – peroxy intermediate

9 Acknowledgments

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

10 References

1 P. E. M. Siegbahn and T. Borowski, *Acc. Chem. Res.* 2006, **39**, 729-738.
2 F. Neese, *J. Biol. Inorg. Chem.*, 2006, **11**, 702-711.
3 M. Sulpizi, S. Raguel, J. VandeVondele, P. Carloni, M. Sprik, *J. Phys. Chem. B*, 2007, **111**, 3969-3976.
4 A. Warshel, M. and Levitt, *J. Mol. Biol.,* 1976, **103**, 227-249.
5 R. A. Friesner and V. Guallar, *Annu. Rev. Phys. Chem.*, 2005, **56**, 389-427.
6 H. Lin and D. G. Truhlar, *Theor. Chem. Acc.*, 2007, **117**, 185-199.
7 H. M. Senn and W. Thiel, *Top. Curr. Chem*. 2007, **268**, 173-290.
8 C. Bo and F. Maseras, *Dalton. Trans.*, 2008, 2911-2919.
9 H. M. Senn and W. Thiel, *Curr. Opin. Chem. Biol.*, 2007, **11**, 182-187.
10 P. H. König, M. Hoffmann, T. Frauenheim, Q. Cui, *J. Phys. Chem. B* 2005,**109**, 9082-9095.
11 U. Ryde and K. Nilsson, *J. Am. Chem. Soc.*, 2003, **125**, 14232-14233
12 U. Ryde, *Dalton Trans.*, 2007, 607-625.
13 U. Ryde, M. H. M. Olsson, B. O. Roos, K. Pierloot and J. O. A. De Kerpel, *J. Biol. Inorg. Chem.*, 2000, **5**, 565-574.
14 K. P. Jensen and U. Ryde, *Coord. Chem. Rev.*, 2008, in press.
15 M. H. M. Olsson and U. Ryde, *J. Am. Chem. Soc.*, 2001, **123**, 7866-7876.
16 U. Ryde and K. Nilsson, *J. Mol. Struct. (Theochem)*, 2003, **632**, 259-275
17 G. J. Kleywegt and T. A. Jones, *Methods Enzymol.*, 1997, **227**, 208-230.
18 A. T. Brünger, *Methods Enzymol.,* 1997, **277**, 243-269.
19 E. I. Solomon., Sundaram, U. M., Machonkin, T. E. *Chem. Rev.* 1996, **96**, 2563-2605.
20 L. Rulíšek, Solomon, E. I., Ryde, U. *Inorg. Chem.* **2005**, *44*, 5612-5628.
21 J. Chalupský, F. Neese, E. I. Solomon,U. Ryde, L. Rulíšek, *Inorg. Chem.*, 2006, **45**, 11051-11059
22 J. Yoon, E. I. Solomon, *J. Am. Chem. Soc.*, 2007, **129**, 13127-13136.
23 N. Källrot, K. Nilsson, T. Rasmussen & U. Ryde, *Intern. J. Quant. Chem.*, 2005, **102**, 520-541
24 K. P. Jensen and U. Ryde, *J. Am. Chem. Soc.*, 2005, **127**, 9117-9128
25 F. Claeyssens, K. E. Ranghan, F. R. Manby, J. N. Harvey, A. Mulholand, *Chem. Commun.* 2005, 5068-5070.
26 H. Hu and W. Yang, *Annu. Rev. Phys. Chem.*, 2008, **59**, 573-601.
27 J. L. Gao, D. G. Truhlar, *Annu. Rev. Phys. Chem.*, 2002, **53**, 467-505.
28 M. Kaukonen, P. Söderhjelm, J. Heimdal, U. Ryde, J. Phys. Chem. B, 2008, in press DOI: 10.1021/jp802648k.
29 R. P. Muller and A. Warshel, *J. Phys. Chem.,* 1995**, 99***,* 17516-17524.
30 T. H. Rod and U. Ryde, *J. Chem. Theory Comput.*, 2005, **1**, 1240-1251
31 J. Kästner, H. M. Senn, S. Thiel, N. Otte, W. Thiel, *J. Chem. Theory Comput.,* 2006, **2**, 452-461.

32 M. Kaukonen, P. Söderhjelm, J. Heimdal, U. Ryde, J. Chem. Theory Comput., 2008, **4**, 985-1001.
33 C. N. Schutz, and A. Warshel, *Proteins, Struct. Funct. Genet.* 2001**, 44**, 400-417.
34 A. J. Mulholland, *Biochem. Soc. Transact*, 2008, 22-26.
35 A. Altun, V. Guallar, R. A. Friesner, S. Shaik, W. Thiel. *J. Am. Chem. Soc.*, 2006, **128**, 3924-3925.
36 U. Ryde, *Curr. Opin. Chem. Biol.*, 2003, **7**, 136-142.
37 A. Shurki and A. Warshel, Adv. Prot. Chem., 2003, **66**, 249-313.
38 K. Morokuma, *Bull. Chem. Soc. Jpn.,* 2007, **80**, 2247-2261.
39 G. M. Ullmann and E. W. Knapp, *Eur. Biophys. J.*, 1999, **28**, 533-551
40 F. Claeyssens, J. N. Harvey, F. R. Manby, R. A. Maa, A. J. Mullholland, K. E. Ranghan, M. Schütz, S. Thiel, W. Thiel, H.-J. Werner, *Angew. Chem.,* 2006, **45**, 6856-6859.
41 P. Söderhjelm, U. Ryde, *J. Phys. Chem. B.*, 2008, accepted.
42 S. Shaik, D. Kumar, S. P. de Visser, A. Altun, W. Thiel. *Chem. Rev.,* 2005, **105**, 2279-2328.
43 P. Rydberg, E. Sigfridsson and U. Ryde, *J. Biol. Inorg. Chem.*, 2004, **9**, 203-223

Table 1. Mean absolute deviations of various energy estimates (kJ/mol) of proton-transfer reactions in two proteins, compared to the results of the QM/MM FEP method QTCP.[28,32] The tested methods are QM calculations on small models in vacuum (QM), the same calculations polarised by the surrounding protein (QM+pol), the QM energy with a point-charge model of the surroundings (QM+ptch), the QM model dissolved in a continuum solvent with a dielectric constant 4 (QM+ε=4), the QM system inserted in a Poisson–Boltzmann model of the protein and the surrounding solvent, using a point-charge model and a dielectric constant of 4 for the protein solvation energy (PB), QM/MM energy with the same (QM/MM) or different MM environments for each state (QM/MM free), as well as the QM/MM-FE[26] and QM/MM-PBSA[28] methods, the latter either using a single QM/MM structure, or structures sampled by a MD simulation.

Protein	[Ni,Fe] hydrogenase	Cu nitrite reductase
QM	26	28
QM+pol	25	36
QM+ptch	30	14
QM+ε=4	23	23
PB	32	39
QM/MM	32	18
QM/MM free	29	69
QM/MM-FE	5	9
QM/MM-PBSA	13	7
QM/MD-PBSA	10	8

Figure 1. The partitioning scheme in QM/MM, as exemplified by the protein catechol *O*-methyltransferase:[30] The active site and substrate (magenta) are included in the QM system, whereas the surrounding protein (yellow) and solvent (cyan) belong to the MM system.

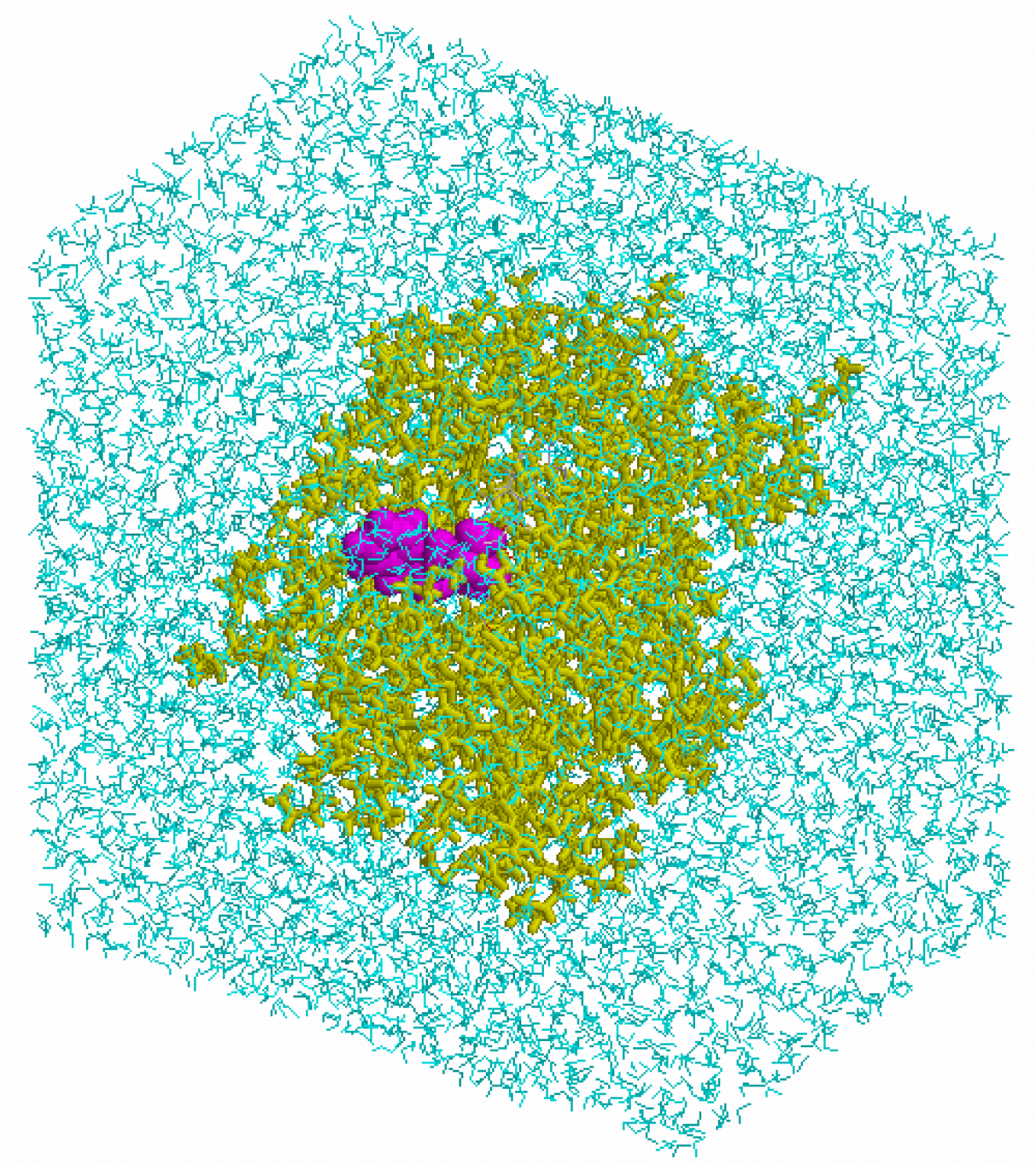

Figure 2. Cumulative residue components of the MM free energy for a proton-transfer reaction in [Ni,Fe] hydrogenase as a function of the distance of the residue from the QM system (which contained 91 atoms).[32] Four different simulations are shown, two with the full protein and two with a truncated protein, in which all residues more than 27 Å from the Ni ion were removed. For each system, one simulation used normal charges, whereas in the other, all the charged residues with a distance greater than 20 Å from the QM system were neutralised (0-ch; note that those curves are identical to the standard-charge curve up to approximately this distance).