<http://www.diva-portal.org>

Postprint

This is the accepted version of a paper published in *Journal of Biological Inorganic Chemistry*. This paper has been peer-reviewed but does not include the final publisher proof-corrections or journal pagination.

Citation for the original published paper (version of record):

Vedin, N P., Lundberg, M. (2016)

Protein effects in non-heme iron enzyme catalysis: insights from multiscale models.

Journal of Biological Inorganic Chemistry, 21(5-6): 645-657

<http://dx.doi.org/10.1007/s00775-016-1374-7>

Access to the published version may require subscription.

N.B. When citing this work, cite the original published paper.

Permanent link to this version:

<http://urn.kb.se/resolve?urn=urn:nbn:se:uu:diva-304673>

Protein effects in non-heme iron enzyme catalysis: Insights from multiscale models

Nathalie Proos Vedin¹ and Marcus Lundberg¹✉

¹ Department of Chemistry - Ångström Laboratory, Uppsala University, Box 518, SE-751 20 Uppsala, Sweden. Fax: +46-18-471 36 33. E-mail: marcus.lundberg@kemi.uu.se

Abbreviations

CASSCF: Complete active space self-consistent field
CD: circular dichroism
CDO: Cysteine dioxygenase
DFT: Density-functional theory
DMRG: Density-matrix renormalization group
EE: Electronic embedding
FEP: Free-energy perturbation
FIH-1: Hypoxia-inducible factor asparaginyl hydroxylase
HF: Hartree-Fock
HG: Homogentisate
HGDO: Homogentisate 1,2-dioxygenase
HMS: Hydroxymandelate synthase
HPA: Hydroxyphenylacetate
HPPD: 4-Hydroxyphenylpyruvate dioxygenase
IPNS: Isopenicillin N synthase
MD: Molecular dynamics
ME: Mechanical embedding
MECP: Minimum energy crossing point
MIOX: *Myo*-inositol oxygenase
MM: Molecular mechanics
MMO: Methane monooxygenase
MMOH: Methane monooxygenase hydroxylase
RMS: Root-mean-square
RNR: Ribonucleotide reductase
PT2: Second-order perturbation theory
ToMOH: Toluene monooxygenase hydroxylase
 Δ^9 D: Δ^9 desaturase

Abstract

Many non-heme iron enzymes have similar sets of ligands but still catalyze widely different reactions. A key question is therefore the role of the protein in controlling reactivity and selectivity. Examples from multiscale simulations, primarily QM/MM, of both mono- and binuclear non-heme iron enzymes are used to analyze the stability of these models and what they reveal about the protein effects. Consistent results from QM/MM modeling are the importance of the hydrogen bond network to control reactivity and electrostatic stabilization of electron transfer from second-sphere residues. The long-range electrostatic effects on reaction barriers are small for many systems. In the systems where large electrostatic effects have been reported, these lead to higher barriers. There is thus no evidence of any significant long-range electrostatic effects contributing to the catalytic efficiency of non-heme iron enzymes. However, the correct evaluation of electrostatic contributions is challenging, and the correlation between calculated residue contributions and the effects of mutation experiments is not very strong. The largest benefits of QM/MM models are thus the improved active-site geometries, rather than the calculation of accurate energies. Reported differences in mechanistic predictions between QM and QM/MM models can be explained by differences in hydrogen bonding patterns in and around the active site. Correctly constructed cluster models can give results with similar accuracy as those from multiscale models, but the latter reduces the risk of drawing the wrong mechanistic conclusions based on incorrect geometries and are preferable for all types of modeling, even when using very large QM parts.

Keywords

computational chemistry, density functional theory, enzyme catalysis, ligand binding, QM/MM, transition state, transition metal

1. Introduction

The non-heme iron enzyme family consists of both mononuclear and binuclear iron proteins that catalyze a wide range of different reactions.[1-3] Their catalytic activity depends on an interplay between the metal center, loosely defined as the metal and its direct ligands, and the surrounding protein. Many enzymes in this class have similar sets of iron ligands, like the 2-His-1-carboxylate facial triad, but still catalyze widely different reactions.[4] Some structurally similar enzymes even catalyze different reactions using the same substrate.[5] A key question is therefore the importance of the protein matrix in controlling reactivity and selectivity.

Theoretical models have been used extensively to understand enzymatic activity. For systems with transition metals, these models often use quantum mechanical (QM) methods, and then almost exclusively density-functional theory (DFT). The results are energies and electronic structures that can be interpreted in terms of chemically intuitive molecular orbitals. When looking at the role of the protein, the computationally demanding molecular orbital description can be combined with one or more additional layer in a multiscale model that treats the full system, see Figure 1.

Figure 1. Electronic structure, exemplified by the highest occupied molecular orbital, of the iron(III)-superoxo species in isopenicillin N synthase (IPNS) from a QM/MM calculation.

Multiscale models come in several flavors, differing in both the number of layers and the methods used to describe these layers. The most common low-layer methods are different flavors of molecular mechanics (MM), and the most common multiscale models are the two-layer

QM/MM approaches.[6-9] In this mini-review no difference is made between additive QM/MM models and extrapolation models like ONIOM QM:MM,[10] because there are no fundamental differences in how they describe protein effects. A large number of non-heme iron studies have been performed with ONIOM, but to avoid focusing on technical aspects, the QM/MM label has been used throughout.

For transition metal enzymes, the main alternative to QM/MM modeling is to use a QM-only cluster model where the environmental effects are described by a continuum dielectric medium.[11] Proponents of the cluster models argue that these environmental effects converge relatively quickly with respect to the size of the QM model, after approximately 200 atoms.[12] However, other studies have show significant dependence of relative energies on QM system size up to 500-1000 atoms, with QM/MM models converging faster than cluster models.[13, 14] These effects remain even after geometry optimization.[15] A possible explanation for the apparent differences between these convergence tests is that they simply do not apply to the same properties. The first test probes changes in solvation energy of a dipole, which in the ideal case depends on inverse cube of the model radius, while the others include the energies of charge-dipole interactions that decrease slower, as the inverse square of the distance. As discussed in more detail below, the description of long-range electrostatic effects are challenging and must be carefully considered in any type of enzyme modeling.

For non-heme iron enzymes, a related observation is that in several cases apparently contradictory mechanistic conclusions have been drawn from cluster and QM/MM models, sometimes from the same groups using the same QM methods,[5, 16, 17] and sometimes in studies by different authors.[18-21] The origins of these discrepancies are considerably easier to analyze in the latter cases because they minimize the differences in how the reaction is modeled. In the methane monooxygenase hydroxylase (MMOH) subunit of methane monooxygenase (MMO) the relative stability of two intermediates, a peroxo species (H_{peroxo}) and the bis- μ -oxo diiron intermediate Q (H_Q) differed by 12 kcal/mol in the QM study, but were almost equally stable in a later QM/MM model.[16, 22] In isopenicillin N synthase (IPNS), two different mechanisms for O-O bond cleavage were predicted. An, arguably small, QM model proposed a route where the required proton came from a water ligand bound to iron, while the QM/MM model favored a route where the substrate donates the proton, concerted with formation of a four-membered β -lactam ring.[17, 23] In 4-Hydroxyphenylpyruvate dioxygenase (HPPD), a

radical path for the formation of the hydroxyphenylacetate (HPA) product was proposed from an early cluster model, while a subsequent QM/MM study favored a heterolytic path.[5, 24]

These differences could be interpreted as that there are large effects of the surrounding protein, and that these effects are critical for a correct description of any enzymatic reaction mechanism. This mini-review will use these examples, as well as others, to discuss what multiscale models reveal about the role of the protein in non-heme iron enzymes. Here the protein effect will be loosely defined as any effect in reaction rates originating from protein residues outside the first coordination sphere, including water molecules. When considering whether the results are significant or not, the size of the effect must be compared to the stability of the method. Therefore, the first part of the mini-review is devoted to model parameters and how these affect the results of non-heme iron enzyme modeling. The second part of the review analyzes two major factors suggested to enhance catalysis, proper arrangement of reactants and long-range electrostatic interactions. When possible, the level of agreement between theoretical predictions and experiments will be discussed, e.g., when it comes to the roles of individual amino acids.

2. Theoretical treatments

2.1 Model setup

Descriptions for setting up QM/MM models are described in detail elsewhere.[6, 25] Here a brief summary is given just to outline the most important choices. Starting from the protein X-ray structure, missing side chains and hydrogens are added. Protonation states of basic and acidic residues, other than those ligated to iron, are assigned based on pKa predictors. In the active site, starting structures for the substrate can be obtained by docking, or in many cases by simply replacing an inhibitor with the real substrate. Water molecules are added inside the protein, based on estimations of binding energies, and as a solvent shell. The protein structure is then equilibrated using a molecular dynamics (MD) simulation, and one or several snapshots are used as starting structures for the QM/MM calculations, see Figure 2.

Figure 2. QM/MM model of 4-Hydroxyphenylpyruvate dioxygenase (HPPD) with the hydroxyphenylacetate (HPA) substrate. The high-level selection is shown to the left. Solvent waters within 15 Å of the iron center have been kept in the model.

Models typically show a moderate sensitivity to the choice of snapshot. In cysteine dioxygenase (CDO), which is a mononuclear site with a 3His metal binding motif, the energy difference between the stable singlet and the triplet excited states differ by 1 kcal/mol between snapshots, and slightly more, 4 kcal/mol, for the quintet state.[26] When it comes to transition state energies, these for heterolytic O–O bond cleavage in nitrobenzene 1,2-dioxygenase range from 23 to 26 kcal/mol.[27] Similar variability in barrier heights, 2-3 kcal/mol, could be found in lipoxygenase by a systematic variation in the starting position of O₂. [28] Monte Carlo sampling has also been used to generate possible starting structures, e.g., for the peroxo diiron(III) structures within the active site of toluene monooxygenase hydroxylase (ToMOH).[29]

Larger effects appear if the fluctuations involve significant changes in the relative positions of groups directly participating in the reaction, with differences in barrier heights of up to 18 kcal/mol in studies of 15-lipoxygenases.[30, 31] From this perspective, special attention has to be paid to water as it is a highly mobile molecule, not necessarily resolved in the X-ray structures. It was early recognized that discrepancies between different QM/MM models could be due to differences in the amount, and position, of internal water molecules around the active

site, with effects in the order of the strength of a hydrogen bond, 5 kcal/mol.[32] Water can also be directly involved in the catalytic reactions.[33, 34] In IPNS, barriers differed by up to 20 kcal/mol between different snapshots for a proton transfer reaction involving an active site water.[17] The possibility of water entering the active site should always be considered, possibly by a molecular dynamics sampling, prior to the design of the model.[35]

2.2 High-level treatment

The large majority of non-heme iron studies have used hybrid DFT. As is well known, results are sensitive to the choice of DFT functional, especially the amount of Hartree-Fock (HF) exchange. In general, more HF exchange favors high-spin states, decreases the strength of metal-ligand bonding, and gives higher reaction barriers.[36] The accuracy of the electronic structure calculations do not influence the protein effects calculated using the low-level method, as long as they give similar electrostatic potentials. However, the errors in the underlying QM methods make it more difficult to compare with experimental observables, like reaction rates, to determine the accuracy of the calculated protein effects.

For metal-ligand distances, most combinations of functionals and basis sets show mean average deviations compared to experiment of less than 0.02 Å.[37] For an iron-oxo model system, standard functionals show deviations in the Fe-O distance of up to 0.02 Å. This value increases to 0.07 Å for the longer weaker iron-ligand distances.[38] Large differences between functionals, up to 0.13 Å, can be observed also for strong metal-oxo bonds if the functionals predict different electronic structures.[39]

Differences in relative spin-state energies of more than 10 kcal/mol have been observed between B3LYP and M06 in AlkB and 8R-Lipoxygenase,[20, 40] possibly due to problems in geometry optimization using the M06 family of functionals.[40] There are also large differences between functionals when it comes to reaction barriers, more than 10 kcal/mol between B3LYP and M06 in both AlkB and Δ^9 desaturase (Δ^9 D).[41, 42]

The alternatives to DFT are the multiconfigurational wavefunction methods. The complete active space self-consistent field (CASSCF) method has been used to calculate structures for oxygen binding in hypoxia-inducible factor asparaginyl hydroxylase (FIH-1) and

to get accurate excitation energies, meaning errors of approximately 4 kcal/mol, for a catechol dioxygenase with the inclusion of second-order perturbation theory (PT2).[43, 44] For reaction mechanisms, the outstanding example is the modeling of a large number of reaction steps in Δ^9 D using the density-matrix renormalization group (DMRG), which offers an approximate solution to a system with a large number of active orbitals.[41] However, accurate energies in transition metal systems are highly challenging also for wavefunction methods.[45] This is partly due to the need for a large basis sets. Effects of 5 kcal/mol on the relative spin-state energies of a heme iron system have been observed when going from a triple to a quadruple-zeta basis set on iron,[46] while DFT results are expected to converge at the triple-zeta level. The multiconfigurational methods have yet to find widespread use for non-heme iron enzymes, but can be expected to increase in frequency.

The use of wavefunction methods would lead to smaller QM models, 60-70 atoms in the DMRG study, than what currently can be afforded with DFT (100-800 atoms). As mentioned in the Introduction, QM/MM calculations converge faster than cluster models with respect to the size of the QM model, although they can still require very large high-level partitions.[13, 14] For non-heme systems, the size of the QM model had little effect on the reaction energies in nitrobenzene 1,2-dioxygenase while increasing the size of the QM part in CDO from 42 to 81 atoms gave singlet-triplet splittings within a few kcal/mol, while the singlet-quintet splitting is again more sensitive with changes of 5 kcal/mol.[26]

2.3 Low-level treatment and interactions between layers

When selecting a multiscale approach, the description of interactions between the layers is the most important aspect, together with the selection of the QM method. The difficult part is to correctly describe the electrostatic interactions. A detailed comparison of different schemes to treat these interactions have been made for the reaction profile of IPNS.[47]

For non-heme iron enzymes, the low-level methods have mainly been different versions of classical force fields like Amber, CHARMM and OPLS. In standard QM/MM the major issues are lack of charge transfer in the boundary, which can lead to an underestimation of polarization, and lack of dielectric effects, which can lead to an overestimation of the long-range

Coulomb interactions. In DFT studies, the most straightforward solution to the charge-transfer problem is to increase the size of the QM system. Another option is to combine two molecular orbital methods that both include charge transfer.[48] Tests of density-functional tight-binding method combined with B3LYP in an ONIOM model gave good results for IPNS, but the requirement is that both molecular orbital methods give the same description of the electronic structure of the iron center.[49, 50]

QM/MM models typically include polarization of the QM part by the MM charges, called electronic embedding (EE). The alternative is a fully classical description of the electrostatic interactions, called mechanical embedding (ME). These two descriptions show large differences if the ME charges are not reparameterized during the reaction, or if the polarizability of the QM system changes significantly during the reaction.[17, 51] If the model is carefully designed, ME can give results of similar, or even better, quality than EE.[14]

What is lacking in most non-heme QM/MM studies is electronic and geometric polarization of the MM part to describe dielectric effects. Due to the significant cost of the high-level calculation, most calculations use a static picture where the surrounding protein is kept in the same local minimum. To avoid structural changes unrelated to the reaction coordinate, a possible source of large errors in QM/MM calculations, it is important to use a well-behaving optimization procedure.[52] It is also common to freeze all atoms at a certain distance from the QM region (10-15 Å) in the geometry optimization. However, with this approach the solvent layer included in many models, see Figure 2, has little effect on the energies.[32, 53] The reason is that they can neither change their dipole moment through electronic polarization nor reorient when the electric field changes. The importance of the solvent layer of course changes if the active site is solvent exposed and the direct interactions become important.[54]

A solution to the problem with electronic polarization is to use a polarizable force field, but this has yet to be tested for non-heme iron systems. Geometric polarization can be included through the dynamics of a free-energy perturbation (FEP) calculation. FEP using umbrella sampling has been extensively performed for several lipxygenase enzymes. For 15-Lipxygenase-2 the barrier on the free-energy surface was 2 kcal/mol higher compared to the potential energy surface.[55] This relatively small effect is similar to what could be seen in other lipxygenases.[31, 56] A study of hydrogen abstraction by the Fe(IV)-oxo species in AlkB gave

rather similar free-energy and potential energy surface, see Figure 3, which could be rationalized by the small electrostatic effects already in the static model.[42]

Figure 3. Calculated minimum potential and free energy paths for hydrogen abstraction by the Fe(IV)-oxo species in AlkB. The minimum energy crossing point (MECP) is between triplet and quintet spin states. (Adapted from reference [42]. Copyright © 2013 American Chemical Society.)

In contrast, for IPNS the electrostatic effects from a static QM/MM model was found to increase the relative energy of a transition state for electron transfer from iron to a peroxo species by as much as 17 kcal/mol, which incorrectly predicted this to be the rate-limiting step. However, the effect decreased by 6 eV in the FEP description, leading to a correct prediction.[57]

2.4 Exploring reaction pathways

Multiscale models of course have the same principal problems as other modeling approaches, one of them the manually guided explorations of the potential energy surface. A more satisfying alternative is the use of an unbiased potential energy surface mapping that can find both reaction pathways and alternative reactant minima. A step in this direction is the artificial force induced

reaction scheme. When used for the iron(III)-superoxide species in IPNS, see Figure 1, the algorithm located 21 stationary structures with reasonable energies without an initial guess of the reaction mechanism. Although the lowest barrier was the same as for the manual exploration, C-H bond activation, a possible minor channel was identified where the superoxide instead attacks the sulfur coordinating to iron.[17, 58]

Atom-centered density matrix propagation has been used to facilitate the sampling of rare events in the dynamics of Soybean Lipoxygenase-1. Unlike the FEP studies described above, the QM part was here included in the dynamics. Physical constraints on the dynamics lead to decrease in the electrostatic fluctuations, which could in turn affect the catalytic reaction.[59]

3. Protein effects on structures and energies

Protein effects on energies should be viewed in relation to the stability of the model. The choice of electronic structure method is very important for the final result, but it does not necessarily affect the evaluation of the effects from the surrounding. The first important consideration is then the hydrogen bond network in and around the active site, with special attention paid to water molecules. As shown above, the effects on barrier heights can be more than 10 kcal/mol, large enough to change the proposed reaction mechanism in a theoretical study. The second important consideration is then the long-range electrostatic effects, potentially affecting reaction energies by more than 10 kcal/mol. Here these two effects are analyzed in more detail for a number of enzymes.

3.1. Improved geometries with QM/MM models

For metalloenzymes the most straightforward comparisons between theory and experiment can be made for structures from X-ray crystallography. Early studies of MMOH and the R2 unit of ribonucleotide reductase (RNR), with arguably small QM models, showed distinct improvements in geometries with the use of QM/MM, see Figure 4, primarily in the orientation of the iron ligands.[60] A separate study of only the reduced form of MMOH, MMOH_{red} , showed similar

results, and a significant reduction in the deviation of the Fe–Fe distance, from 0.3 Å to ± 0.02 Å. The geometric structures of these sites are strongly dependent on the surrounding hydrogen-bonding network, which involves a large number of residues outside the first coordination square.[16] Other examples are hemerythrin, where the orientation of the histidine ligands are better reproduced in the QM/MM structure, and *myo*-inositol oxygenase (MIOX) where the crystallographic water molecules play key roles for the active site structure.[53, 61] In all cases, the protein effects reduce the geometric errors by more than 0.2 Å, much more than the possible errors from the choice of density functional.[37] It is therefore clear that explicit inclusion of the protein leads to a significant improvement in the description of these active sites.

Figure 4. a) Root-mean-square (RMS) difference between optimized and X-ray structures with QM and QM/MM models for the active sites of the R2 unit in ribonucleotide reductase (RNR),[60] methane monooxygenase hydroxylase (MMOH),[60] its reduced form MMOH_{red}, [16] hemerythrin,[53] and *myo*-Inositol Oxygenase (MIOX).[61] b) Superposition of the X-ray (blue) and QM/MM optimized (orange) geometries for all heavy atoms in MIOX (ribbon representation). c) Same for the heavy atoms in the active site (stick representation). (Structures reproduced from reference [61]. Copyright © 2009 American Chemical Society.)

3.2 Reactant structures and correlation to reactivity

The capability of QM/MM models to predict structures better than QM models of the same size gives increased confidence to the use of structural arguments when rationalizing reactivity and spectroscopic results.[62-65] One interesting example is the suggestion of two different

conformations in ferrous soybean Lipxygenase-1.[62] The X-ray structure shows a water ligand bound to iron, but as hydrogens are missing, its orientation is not clear. The two alternatives, differing in the description of the hydrogen bond between water and Asn694, leads to distinct differences in coordination geometries, see Figure 5. None of the QM/MM structures matches the X-ray structure, and the suggestion is that the crystal structure is an average of the two alternatives. The two conformations also give significantly different circular dichroism (CD) spectra that can be matched to experimental observations.[66]

Figure 5. Multiscale structures of two different configurations of the soybean Lipxygenase-1 active site. (Adapted from reference [62]. Copyright © 2010 American Chemical Society.)

Another example is the ferroxidase site in ferritin, where the hydration, protonation, and coordination states could be assigned from comparisons with CD and Mössbauer spectra. The explicit effects of the MM layer are small, only 1 kcal/mol on the CD transitions. For Mössbauer quadrupole splittings, ΔE_Q , protein effects are only 1/7 of the difference between structural alternatives, but the use of a QM/MM model is anyway beneficial because it allows for structural flexibility while still maintaining reasonable conformations.[65]

Most iron enzymes are oxygen activated and QM/MM models have been used to rationalize reactivity from the structure of the O_2 bound state. Early studies showed significant shortening of the Fe–O bond distance in the protein models, compared to active-site models.[43,

67] In AlkB a molecular oxygen binding site that does not directly point to the substrate has been identified,[20] which introduces an isomerization step for the subsequent iron(IV)–oxo species not included in previous QM/MM studies.[42] This could be a way to control the reactivity of the iron-superoxo species, retain regioselectivity and prevent the formation of by-products. Several other studies highlight the role of a correct hydrogen bond network in controlling orientation and reactivity.[19, 68-70] As an example, in homoprotocatechuate 2,3-dioxygenase (HPCD), the active site residue His200 changes its orientation during O₂ binding to interact with the proximal oxygen of dioxygen, thus stabilizing a partial electron transfer from the substrate to the iron center after binding.[19]

3.3 Reaction energy profiles – effects of hydrogen bond networks

Reaction energy profiles, including barrier heights, are among the most important criteria when discriminating between proposed reaction mechanisms. The studies discussed above, as well as others, have highlighted the important effects of the hydrogen-bonding network on geometric and electronic structure.[71-73] An illustrative example is the difference in stability of the bis-μ-oxo diiron intermediate Q compared to a peroxo species in the related enzymes MMOH and ToMOH. In ToMOH a Thr201 residue can form a hydrogen bond with the peroxo species H_{peroxo}, which makes it more stable than H_Q, see Figure 6. Oxidation is therefore performed by the peroxo species. In MMOH QM/MM calculations predict that the corresponding residue points away from the active site, which leads to similar stability between peroxo and Q, and oxidations are instead performed by Q.[29]

Figure 6. The two lowest-energy toluene monooxygenase hydroxylase (ToMOH) peroxo structures and the intermediate Q structure optimized with QM/MM. The only unlabeled residue is Glu231. (Reproduced from reference [29]. Copyright © 2013 American Chemical Society.)

In the Introduction, three cases were mentioned where QM and QM/MM models gave apparently contradicting results, despite using the same QM method. The first of these cases was the above-mentioned energy difference between H_{peroxo} and H_Q in MMOH. The original QM model predicted Q to be stable by 12 kcal/mol, while the QM/MM model gave almost equal energies for the two species.[16, 22] The major difference between the models is an artificial hydrogen bond in the QM model of Q, formed between one of the Glu residues coordinating to Fe1 and the protein backbone, see Figure 7. This hydrogen bond, which leads to an apparent stabilization of this intermediate, is not present in either X-ray or QM/MM structure. In this case a truncation of the QM model far away from the active site leads to problems describing the hydrogen bond network and this has significant impact on the reaction energies.

Figure 7. a) QM/MM structure (colors) and QM model of intermediate Q in MMOH. The artificial hydrogen bond in the QM model is shown with a red dashed line. b) QM model of IPNS optimized with atoms frozen in their position in X-ray structure (marked with x). c) QM/MM model of IPNS with MM residues directly interacting with the substrate carboxylate shown in gray. (Reproduced from references [16] and [17]. Copyright © 2009 American Chemical Society.)

The second example of differences between cluster and QM/MM models was the identity of the second proton donor for O-O bond cleavage in IPNS. The cluster model proposed a water ligand bound to iron, while the multiscale model favored a route where the proton came from the substrate itself, concerted with formation of a β -lactam ring.[17, 23] Here the QM model failed to properly stabilize a substrate carboxylate, which lead to an overestimation of the substrate's proton affinity, thus incorrectly predicting it to be a poor proton donor, see Figure 7. In this case the problem was that the selected QM model was too small and it could have been corrected by either QM/MM or by using a larger QM model that can now be routinely afforded.

The third example was the reactivity of the substrate hydroxyphenylacetate (HPA), in the active sites of HPPD, see Figure 8. For this reaction, a radical path was proposed from QM modeling, while the QM/MM model favored a heterolytic path.[5, 24] In this case, the important difference is that the QM/MM model includes second-sphere ligands that stabilize iron-substrate electron transfer by forming stronger hydrogen bonds with the polarized metal-ligands. These ligands were not included in the original QM model, designed prior to the presence of any X-ray

structure. As was the case for IPNS, enlarging the cluster model with a full set of second-sphere ligands should be able to describe this effect.

Figure 8. Reaction pathways for HPA in hydroxymandelate synthase (HMS), forming (S)-hydroxymandelate (HMA), and HPPD, forming 2,5-dihydroxyphenylacetate (homogentisate, HG) (top). Representative snapshot from the molecular dynamics simulation of the HPPD-Fe(IV)-O-HPA complex with the starting structure from HMS in gray (bottom). (Adapted from reference [5]. Copyright © 2014 American Chemical Society.)

As mentioned in the Introduction, structurally similar enzymes can catalyze different reactions using the same substrate. A key example is the difference in reactivity of the substrate HPA in the active sites of HPPD and hydroxymandelate synthase (HMS), see Figure 8. This was studied using structures obtained from MD simulations.[5] The QM/MM calculations highlight the contributions of the second-sphere residues to reactivity, with Thr163 stabilizing the formation of the HPPD product, while interestingly the equivalent residue in HMS stabilizes its native

reaction. Although the neighboring Asn216 has no direct effect on the reaction barriers, it plays an important role in arranging the local structure of the active site. Specificity is a complicated interplay between residues, and correct predictions require accurate structures.

These three examples show that there are important effects of residues in the second shell. All these effects can be captured in large QM cluster models. However, the selection of a cluster model always leads to truncation of the hydrogen bond network. The electrostatic interactions at the border can be reasonably modeled using a continuum dielectric medium, but geometry optimization may lead to formation of artificial hydrogen bonds that disrupt the network. Instead of taking that risk, including the protein environment in a QM/MM description gives a stable hydrogen-bond pattern at a low computational cost.

3.4 Reaction energy profiles – electrostatic and van der Waals interactions

Previous examples showed the importance of correctly describing interactions around the active site and HPPD modeling also showed the catalytic effect of electrostatic polarization from second-shell residues. Similar effects, as well as other long-range interactions, have been consistently studied for O₂ binding, which is a common step for many non-heme iron enzymes. Early QM/MM results showed large protein effects, 10 kcal/mol, on the binding energy of O₂ in the oxygen transport protein hemerythrin.[53] Later the protein contributions to binding have been quantified also for IPNS, MIOX, and JMJD2A.[74-76] These results show a consistent stabilization, see Figure 9. There are two major contributions: 1) the van der Waals interactions between the bound dioxygen and the protein atoms and (2) electrostatic effects on a charge transfer from iron(II) to oxygen to form iron(III)-superoxide. In hemerythrin the van der Waals interactions stabilized binding by 6 kcal/mol, while the electrostatic contributions were 3 kcal/mol. The latter effect came primarily from an increase in the hydrogen-bond strength between iron ligands, which become positively polarized upon generation of the ferric ion, and negative side chains in the second coordination sphere.[53] This is similar to the effects of second-sphere ligands in the electron transfer step in HPPD described above.[5]

Figure 9. Coulomb and van der Waals contributions to the stabilization of O₂ binding from the MM part of QM/MM models for hemerythrin,[53] IPNS,[74] MIOX,[75] and JMJD2A.[76]

The IPNS study benefitted from the presence of two different X-ray structures, one representing a state prior to O₂ binding (five-coordinate iron) and another with NO bound as an analogue for O₂ (six-coordinate iron). Both structures gave almost identical energies, stabilizing binding by 8-10 kcal/mol, slightly depending on binding mode and spin state.[74] This shows that the static optimization approach method is flexible enough to handle changes in metal coordination. On average the electrostatic stabilization was 2 kcal/mol and the van der Waals contributions 3 kcal/mol. The remaining stabilization comes from a correction to the orientation of the histidine ligands that, like in other systems, adopted an incorrect orientation without the explicit protein environment. In MIOX the non-bonded MM stabilization of oxygen binding is similar to that of IPNS, although slightly smaller.[75] In both enzymes, the relative energies of different binding modes and spin states are not strongly affected by the MM layer, less than 2 kcal/mol. In JMJD2A, the surrounding increases the binding energy of the end-on quintet by 8 kcal/mol, with the electrostatic interactions three times larger than the van der Waals contributions.[76]

Comparisons with experiment are complicated by the large sensitivity of the binding energy to the choice of functional. Decreasing the amount of HF exchange in B3LYP from 20 to 15% increases the binding energy by 3-4 kcal/mol.[36, 74, 75] The effect of van der Waals interactions in the QM/MM calculations does not show that the protein “use” these interactions to preferentially stabilize the bound state, it only shows they cannot be neglected in calculations of binding energies. The situation is similar to the use of empirical dispersion in DFT, which

stabilizes O₂ binding by 4-7 kcal/mol.[75-77] A larger QM model decreases the van der Waals contributions from the MM part, but with a dispersion-corrected functional, these interactions are instead captured by the high-level method.

For barrier heights, a systematic study of electrostatic effects was done for hydrogen abstraction of substituted methanes, CH₃X, by intermediate Q in MMOH.[78] The mean absolute deviation compared to experiment is 2 kcal/mol. The trends in reactivity are correct, except for CH₃OH, although with larger differences than observed experimentally, see Figure 10. There are also no significant improvements compared to QM-only energies. An analysis of individual contributions shows large electrostatic effects, but also large differences between substrates, from -5 to +10 kcal/mol. Strangely enough, the two extremes were for reactions with the non-polar substrates ethane and methane respectively, which is difficult to rationalize. The total QM/MM contributions to these two transition states are much smaller, 1 and -2 kcal/mol, and one possibility is that it is the breakdown into individual contributions that is problematic. Relatively small changes in a hydrogen bond distance, that does not significantly affect the total MM contribution, can give apparently large individual terms working in different directions.

Figure 10. Reaction barriers for CH₃X substrates in MMOH.[78]

Other studies have shown relatively small electrostatic effects on hydrogen abstraction barriers. In MIOX and nitrobenzene 1,2-dioxygenase the protein surrounding lowers the hydrogen abstraction barrier by 2-3 kcal/mol, and effects are similar also for other types of barriers.[27, 61]. The corresponding reaction catalyzed by the Fe(IV)-oxo species in AlkB shows

minimal electrostatic stabilization, similar to the case for taurine/ α -ketoglutarate dioxygenase (TauD).[42, 79]

The small protein contribution in AlkB is due to the cancellation of larger individual effects, with most of the important residues close to the active site, see Figure 11.[42] The effects of three of these amino acids, T51, Y76, and R161 can be compared to mutation studies. Experimentally the mutation of T51 to alanine increases the reaction barrier by 0.5 kcal/mol while the Y76A mutation leads to an increase of 1.0 kcal/mol. These effects are reproduced in the theoretical calculations, although slightly overestimated, where T51 and Y76 reduce the barrier by -1.0 and -2.6 kcal/mol respectively. However, the success is not universal. R161 is theoretically predicted to decrease the reaction barriers significantly, but experimentally the mutation of R161 to alanine does not affect the reaction rate. Other reaction steps in AlkB have also been modeled with QM/MM, although without detailed analyses of residue effects.[80, 81]

Figure 11. a) Non-bonded interactions on the hydrogen abstraction step in AlkB. b) Structure of the AlkB active site including residues with large contributions to the barrier height. (Reproduced from reference [42]. Copyright © 2013 American Chemical Society.)

In general, large electrostatic effects can be expected for reactions that involve electron and proton transfer. As discussed above, electrostatic interactions evaluated using FEP increased the relative energy of a transition state for O–O bond cleavage in IPNS by 10 kcal/mol.[47, 57] These electrostatic effects, which again mainly originated from residues just outside the first

shell, increased the barrier and thus did not make any contribution to catalysis. Also $\Delta^9\text{D}$ showed large static MM effects increasing two of the barriers by almost 10 kcal/mol, and even if the origin is not analyzed in detail, the effect would be to slow catalysis.[41] A calculation with extremely large electrostatic effects is for a proton transfer reaction in nitrile hydratase where the reaction barrier decreases by 40 kcal/mol,[82] but as other reaction steps are stabilized by an unreasonable 200 kcal/mol, without any clear protein origin, it is possible that these effects were not correctly evaluated. Other studies of QM/MM of nitrile hydratase also give more reasonable energy profiles.[21, 83]

There can of course be huge effects on calculated pKa and redox potentials,[84] where there is a change in the total charge of the system, but the difficulty in modeling these properties are discussed in more detail elsewhere, and are not considered further here.[85, 86] The difficulty to predict the protonation state in the active site adds another important dimension to the modeling. A strategy for modeling reactions where the protonation state is unknown is to include both alternatives and compare with available X-ray structures, spectroscopic data and/or reaction barrier heights. [65, 87]

These examples highlight the considerable challenges in correctly calculating and evaluating electrostatic contributions, which makes it difficult to determine their role in catalysis. Any large electrostatic effects in a QM/MM model must be carefully analyzed to understand whether it originates from a real protein effect, rather than an artifact of the computational scheme. The most consistent effect is stabilization of electron transfer from iron, which is seen both in O_2 binding and the heterolytic mechanism in HPPD. However, for many enzymes the effects are small, and for a few enzymes with apparently large effects, these have led to an increase in the barrier heights. There is thus no consistent evidence that long-range electrostatic effects contribute to the catalytic efficiency of non-heme iron enzymes. The largest benefits of QM/MM models are thus the improved active-site geometries, rather than the calculation of accurate energies.

4. Conclusions

The modeling of non-heme iron enzymes, as well as other transition metal-containing enzymes, remains a challenging task. There are several factors that affect the accuracy of the modeling, but

the important ones are; the choice of high-level (QM) method, the presence of structural water, the hydrogen bond network, and the description of electrostatic effects. Multiscale models, most commonly different flavors of QM/MM, are cost-effective ways of including and evaluating these different factors.

The use of QM/MM models consistently leads to better predictions of geometries compared to QM cluster models. The improvements were mainly due to a better orientation of ligand side chains, especially histidines, and correct descriptions of hydrogen bond networks. Accurate descriptions of active-site conformations also lead to an increased capability to correlate structure and reactivity. The hydrogen bond networks are also important in the calculations of reaction energy profiles, and can in many cases explain differences in reactivity between similar enzyme active sites.

The reported differences in mechanistic predictions between QM and QM/MM models can be explained by differences in hydrogen bonding pattern in and around the active site. *Correctly* constructed cluster models that include also second-shell ligands can give results with similar accuracy as those from QM/MM models, but there are no convincing arguments why they should be used instead of multiscale models. The latter reduces the risk of faulty conclusions drawn from the use of incorrect geometries. QM/MM models also lead to faster convergence of relative energies with respect to the size of the QM part. The extra effort for the protein preparation is anyway required to properly explore different binding patterns and to evaluate potential positions of water molecules. The increased complexity of the model can lead to spurious changes in protein geometry that give artificial effects on the reaction energies, but these errors can be controlled with the use of a good optimization scheme.

The most consistent electrostatic effect is stabilization of the charge transfer from iron(II) to dioxygen by residues in the second sphere. For most reactions the long-range electrostatic effects on barrier heights appear to be small. In the few systems where large effects are reported, these led to increased reaction barriers. There is thus no evidence of any large long-range electrostatic effects contributing to the catalytic efficiency of non-heme iron enzymes. However, these effects are difficult to evaluate correctly, and results from different multiscale approaches often varies significantly. The correlation between calculated residue contributions in QM/MM models and results from mutation experiments is not very strong. FEP on classical potentials has not yet found widespread use for calculating reaction energies in non-heme systems, but is likely

to become important when analyzing the effects of residue mutations. QM/MM MD simulations are likely to remain challenging due to problems to accurately describe the electronic structure of the iron site with fast QM methods.

Considering that the accuracy of the QM method remains a very important parameter in non-heme iron modeling, another likely trend is further exploration of multiconfigurational wavefunction methods. They are now able to treat the first coordination shell of non-heme iron systems with reasonable timing and accuracy. These relatively small QM systems will suffer from the same limitations as early DFT/MM models, but most calculations will anyway be performed with standard QM/MM methodology, as the focus of these studies will be to control the performance of the QM part of the calculation.

The dominant trend in modeling is expected to be a significant increase in the size of the high-level part of DFT/MM models from around 100 to 300-800 atoms, similar to what is currently seen for cluster models. This would move the layer boundary away from the active site, which decreases the problems associated with this interface. This trend benefits from the mainstream DFT development, which will be more generally available than schemes to improve the boundary itself, especially considering the large number of different QM/MM flavors that are already in use. The big benefits of QM/MM models are the improvements in active-site geometries, rather than in the calculation of the energies. With larger QM models, it will be critical to correctly describe the hydrogen bond networks, and multiscale models will become even more useful.

Acknowledgements

We thank Tomasz Borowski for helpful discussions and Anna Wójcik for assistance in analyzing the QM/MM contributions in HPPD.

References

- 1 E. I. Solomon, T. C. Brunold, M. I. Davis, J. N. Kemsley, S.-K. Lee, N. Lehnert, F. Neese, A. J. Skulan, Y.-S. Yang and J. Zhou (2000) *Chem Rev* 100:235-350
- 2 E. I. Solomon, K. M. Light, L. V. Liu, M. Srncic and S. D. Wong (2013) *Acc Chem Res* 46:2725-2739
- 3 L. Que and R. Y. Ho (1996) *Chem Rev* 96:2607-2624
- 4 K. D. Koehntop, J. P. Emerson and L. Que Jr (2005) *J Biol Inorg Chem* 10:87-93
- 5 A. Wójcik, E. Broclawik, M. Lundberg, P. E. M. Siegbahn, M. Lundberg, G. Moran and T. Borowski (2014) *J Am Chem Soc* 136:14472-14485
- 6 H. M. Senn and W. Thiel (2009) *Angew Chem Int Ed* 48:1198-1229
- 7 M. W. van der Kamp and A. J. Mulholland (2013) *Biochemistry* 52:2708-2728
- 8 T. Borowski, M. Quesne and M. Szaleniec (2015) In: C. T. Karabeneva (ed) *Combined quantum mechanical and molecular mechanical modelling of biomolecular interactions*. Elsevier, San Diego, pp. 187-224
- 9 E. A. C. Bushnell, V. E. J. Berryman, J. W. Gault and R. J. Boyd (2015) In: C. T. Karabeneva (ed) *Combined quantum mechanical and molecular mechanical modelling of biomolecular interactions*. Elsevier, San Diego, pp. 153-185
- 10 L. W. Chung, W. Sameera, R. Ramozzi, A. J. Page, M. Hatanaka, G. P. Petrova, T. V. Harris, X. Li, Z. Ke and F. Liu (2015) *Chem Rev* 115:5678-5796
- 11 M. R. A. Blomberg, T. Borowski, F. Himo, R.-Z. Liao, M. Lundberg and P. E. M. Siegbahn (2014) *Chem Rev* 114:3601-3658
- 12 P. E. M. Siegbahn and F. Himo (2011) *Wiley Interdiscip Rev Comput Mol Sci* 1:323-336
- 13 C. V. Sumowski, B. B. Schmitt, S. Schweizer and C. Ochsenfeld (2010) *Angew Chem* 122:10147-10151
- 14 L. Hu, P. Söderhjelm and U. Ryde (2011) *J Chem Theory Comput* 7:761-777
- 15 S. Sumner, P. Söderhjelm and U. Ryde (2013) *J Chem Theory Comput* 9:4205-4214
- 16 D. Rinaldo, D. M. Philipp, S. J. Lippard and R. A. Friesner (2007) *J Am Chem Soc* 129:3135-3147
- 17 M. Lundberg, T. Kawatsu, T. Vreven, M. J. Frisch and K. Morokuma (2009) *J Chem Theory Comput* 5:222-234
- 18 Y. Shiota and K. Yoshizawa (2004) *J Phys Chem B* 108:17226-17237
- 19 G. Dong, S. Shaik and W. Lai (2013) *Chem Sci* 4:3624-3635
- 20 M. G. Quesne, R. Latifi, L. E. Gonzalez-Ovalle, D. Kumar and S. P. de Visser (2014) *Chem-Eur J* 20:435-446
- 21 M. Kayanuma, K. Hanaoka, M. Shoji and Y. Shigeta (2015) *Chem Phys Lett* 623:8-13
- 22 B. F. Gherman, M.-H. Baik, S. J. Lippard and R. A. Friesner (2004) *J Am Chem Soc* 126:2978-2990
- 23 M. Lundberg, P. E. M. Siegbahn and K. Morokuma (2008) *Biochemistry* 47:1031-1042
- 24 T. Borowski, A. Bassan and P. E. M. Siegbahn (2004) *Biochemistry* 43:12331-12342
- 25 M. G. Quesne, T. Borowski and S. P. de Visser (2016) *Chem-Eur J* 22:2562-2581
- 26 D. Kumar, W. Thiel and S. P. de Visser (2011) *J Am Chem Soc* 133:3869-3882
- 27 I. Geronimo and P. Paneth (2014) *Phys Chem Chem Phys* 16:13889-13899
- 28 R. Suardiaz, L. Masgrau, J. M. Lluch and À. González-Lafont (2014) *ChemPhysChem* 15:2303-2310

- 29 A. D. Bochevarov, J. Li, W. J. Song, R. A. Friesner and S. J. Lippard (2011) *J Am Chem Soc* 133:7384-7397
- 30 P. Saura, R. Suardiaz, L. Masgrau, J. M. Lluch and À. González-Lafont (2014) *ACS Catal* 4:4351-4363
- 31 J. Soler, P. Saura, D. García-López, L. Masgrau, J. M. Lluch and À. González-Lafont (2016) *J Phys Chem B* 120:1950-1960
- 32 A. Altun, S. Shaik and W. Thiel (2006) *J Comput Chem* 27:1324-1337
- 33 L. Du, J. Gao, Y. Liu and C. Liu (2012) *J Phys Chem B* 116:11837-11844
- 34 W. Huang, R. Gherib and J. W. Gault (2012) *J Phys Chem B* 116:8916-8929
- 35 B. Wang, Z. Cao, D. A. Sharon and S. Shaik (2015) *ACS Catal* 5:7077-7090
- 36 P. E. M. Siegbahn (2006) *J Biol Inorg Chem* 11:695-701
- 37 M. Bühl and H. Kabrede (2006) *J Chem Theory Comput* 2:1282-1290
- 38 F. Neese (2006) *J Biol Inor Chem* 11:702-711
- 39 M. Lundberg, M. R. A. Blomberg and P. E. M. Siegbahn (2004) *Inorg Chem* 43:264-274
- 40 E. A. Bushnell and J. W. Gault (2013) *J Comput Chem* 34:141-148
- 41 J. Chalupský, T. A. Rokob, Y. Kurashige, T. Yanai, E. I. Solomon, L. Rulíšek and M. Srnc (2014) *J Am Chem Soc* 136:15977-15991
- 42 D. Fang, R. L. Lord and G. A. Cisneros (2013) *J Phys Chem B* 117:6410-6420
- 43 A. Nemukhin, B. Grigorenko, I. Topol and S. Burt (2006) *Int J Quantum Chem* 106:2184-2190
- 44 N. Nakatani, Y. Hitomi and S. Sakaki (2011) *J Phys Chem B* 115:4781-4789
- 45 K. Pierloot (2011) *Int J Quantum Chem* 111:3291-3301
- 46 S. Vancoillie, H. Zhao, M. Radon and K. Pierloot (2010) *J Chem Theory Comput* 6:576-582
- 47 M. Lundberg and T. Borowski (2013) *Coord Chem Rev* 257:277-289
- 48 G. S. Zheng, M. Lundberg, J. Jakowski, T. Vreven, M. J. Frisch and K. Morokuma (2009) *Int J Quantum Chem* 109:1841-1854
- 49 G. S. Zheng, H. A. Witek, P. Bobadova-Parvanova, S. Irle, D. G. Musaev, R. Prabhakar, K. Morokuma, M. Lundberg, M. Elstner, C. Köhler and T. Frauenheim (2007) *J Chem Theory Comput* 3:1349-1367
- 50 M. Lundberg, Y. Sasakura, G. Zheng and K. Morokuma (2010) *J Chem Theory Comput* 6:1413-1427
- 51 H. Brkić, B. Kovačević and S. Tomić (2015) *Mol BioSyst* 11:898-907
- 52 T. Vreven, M. Frisch, K. Kudin, H. Schlegel and K. Morokuma (2006) *Mol Phys* 104:701-714
- 53 M. Wirstam, S. J. Lippard and R. A. Friesner (2003) *J Am Chem Soc* 125:3980-3987
- 54 T. Vasilevskaya, M. G. Khrenova, A. V. Nemukhin and W. Thiel (2016) *J Comput Chem*,
- 55 R. Suardiaz, P. G. Jambrina, L. Masgrau, À. González-Lafont, E. Rosta and J. M. Lluch (2016) *J Chem Theory Comput* 12:2079-2090
- 56 I. Tejero, M. Garcia-Viloca, À. González-Lafont, J. M. Lluch and D. M. York (2006) *J Phys Chem B* 110:24708-24719
- 57 T. Kawatsu, M. Lundberg and K. Morokuma (2011) *J Chem Theory Comput* 7:390-401
- 58 S. Maeda, E. Abe, M. Hatanaka, T. Taketsugu and K. Morokuma (2012) *J Chem Theory Comput* 8:5058-5063
- 59 P. Phatak, I. Sumner and S. S. Iyengar (2012) *J Phys Chem B* 116:10145-10164

- 60 M. Torrent, T. Vreven, D. G. Musaev, K. Morokuma, Ö. Farkas and H. B. Schlegel (2002) *J Am Chem Soc* 124:192-193
- 61 H. Hirao and K. Morokuma (2009) *J Am Chem Soc* 131:17206-17214
- 62 H. Hirao and K. Morokuma (2010) *J Phys Chem Lett* 1:901-906
- 63 C. Raspail, M. Graindorge, Y. Moreau, S. Crouzy, B. Lefèbvre, A. Y. Robin, R. Dumas and M. Matringe (2011) *J Biol Chem* 286:26061-26070
- 64 M. Srnec, T. A. Rokob, J. K. Schwartz, Y. Kwak, L. Rulíšek and E. I. Solomon (2012) *Inorg Chem* 51:2806-2820
- 65 T. V. Harris and K. Morokuma (2013) *Inorg Chem* 52:8551-8563
- 66 M. A. Pavlosky, Y. Zhang, T. E. Westre, Q.-F. Gan, E. G. Pavel, C. Campochiaro, B. Hedman, K. O. Hodgson and E. I. Solomon (1995) *J Am Chem Soc* 117:4316-4327
- 67 M. Hoffmann, I. V. Khavrutskii, D. G. Musaev and K. Morokuma (2004) *Int J Quantum Chem* 99:972-980
- 68 P. H.-L. Sit, A. Migliore, M.-H. Ho and M. L. Klein (2010) *J Chem Theory Comput* 6:2896-2909
- 69 G. Dong and W. Lai (2014) *J Phys Chem B* 118:1791-1798
- 70 S. Roy and J. Kästner (2016) *Angew Chem Int Ed* 55:1168-1172
- 71 H. Hirao and K. Morokuma (2011) *J Am Chem Soc* 133:14550-14553
- 72 J. Huang, C. Li, B. Wang, D. A. Sharon, W. Wu and S. Shaik (2016) *ACS Catal* 6:2694-2704
- 73 G. Ma, W. Zhu, H. Su, N. Cheng and Y. Liu (2015) *ACS Catal* 5:5556-5566
- 74 M. Lundberg and K. Morokuma (2007) *J Phys Chem B* 111:9380-9389
- 75 H. Hirao (2011) *J Phys Chem B* 115:11278-11285
- 76 W. A. Cortopassi, R. Simion, C. E. Honsby, T. C. França and R. S. Paton (2015) *Chem-Eur J* 21:18983-18992
- 77 P. E. M. Siegbahn, M. R. A. Blomberg and S.-L. Chen (2010) *J Chem Theory Comput* 6:2040-2044
- 78 B. F. Gherman, S. J. Lippard and R. A. Friesner (2005) *J Am Chem Soc* 127:1025-1037
- 79 E. Godfrey, C. S. Porro and S. P. de Visser (2008) *J Phys Chem A* 112:2464-2468
- 80 D. Fang and G. A. Cisneros (2014) *J Chem Theory Comput* 10:5136-5148
- 81 B. Wang, D. Usharani, C. Li and S. Shaik (2014) *J Am Chem Soc* 136:13895-13901
- 82 C. A. MacDonald and R. J. Boyd (2015) *Comput Theor Chem* 1070:48-54
- 83 M. Kayanuma, M. Shoji, M. Yohda, M. Odaka and Y. Shigeta (2016) *J Phys Chem B* 120:3259-3266
- 84 L. E. Grove, J. Xie, E. Yikilmaz, A.-F. Miller and T. C. Brunold (2008) *Inorg Chem* 47:3978-3992
- 85 J. Li, M. Farrokhnia, L. Rulíšek and U. Ryde (2015) *J Phys Chem B* 119:8268-8284
- 86 J. Heimdal, M. Kaukonen, M. Srnec, L. Rulíšek and U. Ryde (2011) *ChemPhysChem* 12:3337-3347
- 87 Y. Qi, J. Lu and W. Lai (2016) *J Phys Chem B* 120:4579-4590