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

Postprint

This is the accepted version of a paper published in *Coordination chemistry reviews*. 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):

Lundberg, M., Borowski, T. (2013)

Oxoferryl species in mononuclear non-heme iron enzymes: biosynthesis, properties and reactivity from a theoretical perspective.

Coordination chemistry reviews, 257(1): 277-289

<http://dx.doi.org/10.1016/j.ccr.2012.03.047>

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-181244>

Oxoferryl species in mononuclear non-heme iron enzymes: biosynthesis, properties and reactivity from a theoretical perspective

Marcus Lundberg^{a,*}, Tomasz Borowski^b

^aDepartment of Chemistry - Ångström laboratory, Uppsala University, Box 518, SE-751 20 Uppsala, Sweden

^bJerzy Haber Institute of Catalysis and Surface Chemistry, Polish Academy of Sciences, ul. Niezapominajek 8, 30-239, Cracow, Poland

Abstract

Mononuclear non-heme iron enzymes perform a wide range of chemical reactions. Still, the catalytic mechanisms are usually remarkably similar, with formation of a key oxoferryl (Fe(IV)=O) intermediate through two well-defined steps. First, two-electron reduction of dioxygen occurs to form a peroxo species, followed by O-O bond cleavage. Even though the peroxo species have different chemical character in various enzyme families, the analogies between different enzymes in the group make it an excellent base for investigating factors that control metal-enzyme catalysis. We have used density-functional theory to model the complete chemical reaction mechanisms of several enzymes, e.g., for aromatic and aliphatic hydroxylation, chlorination, and oxidative ring-closure. Reactivity of the Fe(IV)=O species is discussed with focus on electronic and steric factors determining the preferred reaction path. Various spin states are compared, as well as the two reaction channels that stem from involvement of different frontier molecular orbitals of Fe(IV)=O. Further, the two distinctive species of Fe(IV)=O, revealed by Mössbauer spectroscopy, and possibly relevant for specificity of aliphatic chlorination, can be identified. The stability of the modeling results have been analyzed using a range of approaches, from active-site models to multi-scale models that include classical free-energy contributions. Large effects from an explicit treatment of the protein matrix (~ 10 kcal/mol) can be observed for O₂ binding, electron-transfer and product release.

Keywords:

enzyme catalysis, non-heme iron, O-O bond cleavage, oxoferryl, density-functional theory, multi-scale modeling

Contents

1 Introduction	2	6.3 Multi-scale (QM/MM) models II - Semi-classical effects	11
2 Computational approach	2	6.4 Multi-scale (QM/MM) models III - Dynamical effects	12
3 Biosynthesis of Fe(IV)=O	3	7 Discussion	13
3.1 Reaction energy diagrams	3	8 Summary	15
3.2 Details of O-O bond cleavage	4	Abbreviations	
4 Properties of Fe(IV)=O	6	α KAD: α -ketoacid dependent dioxygenases DFT: density-functional theory DFTB: density-functional tight-binding	
4.1 Electronic structure of Fe(IV)=O	6	EE: electronic embedding FEP: free-energy perturbation	
4.2 Isoforms important for selectivity	7	HPPD 4-Hydroxyphenylpyruvate dioxygenase IPNS: isopenicillin N synthase ME: mechanical embedding MNIE: mononuclear non-heme iron enzymes NOSD: natural orbitals for the spin density ONIOM: our own N-layered integrated molecular orbital + molecular mechanics QM/MM: quantum mechanics/molecular mechanics THBH: tetrahydrobiopterin-dependent hydroxylases TS: transition state TST: transition-state theory	
5 Reactivity of Fe(IV)=O	7		
5.1 σ - and π channels in oxoferryl reactions . .	7		
6 Stability of protein modeling	9		
6.1 Active-site model	9		
6.2 Multi-scale (QM/MM) models I - Classical effects	9		

*Corresponding author. Telephone: +46-18-4713708. Fax: +46-18-4715830

Email addresses: marcus.lundberg@kemi.uu.se (Marcus Lundberg), ncborows@cyf-kr.edu.pl (Tomasz Borowski)

1. Introduction

Iron proteins are involved in a wide variety of biological processes. One reason behind the important role of iron in biology is the possibility to finely tune its chemical properties with different coordination environments. Several classes of iron enzymes exist, e.g., iron-sulfur, heme iron, and non-heme iron. In the latter, iron is coordinated by several amino acid side chains, e.g., histidine, tyrosine, aspartate, and glutamate[1]. Among the mononuclear non-heme iron enzymes, there exist several apparently unrelated proteins, with similar active sites that still catalyze a wide range chemical reactions, such as desaturations, oxidative cyclizations, mono-oxygenations and di-oxygenations, hydro-peroxidations and epoxidations [2–4]. Despite the difference in the outcome of the chemical reactions, the catalytic mechanisms for many of them are remarkably similar, with formation of an oxoferryl (Fe(IV)=O) species as a key intermediate, see Figure 1[5, 6]. The large diversity in reactivity between enzymes with similar active sites makes this class an excellent target for investigations of enzyme catalysis.

Figure 1: Fe(IV)=O intermediate in the mononuclear non-heme iron enzyme isopenicillin N synthase. The structure is obtained using a QM/MM model (QM atoms shown in ball-and-stick representation).

The present review outlines how theoretical modeling, mainly using density-functional theory (DFT), can be used to clarify the important similarities and differences in this class of enzymes. Detailed results are presented e.g., for different α -ketoacid dependent dioxygenases (α KAD) [7], tetrahydrobiopterin-dependent hydroxylases (THBH) [8], and isopenicillin N synthase (IPNS) [9]. Modeling results for non-heme enzymes that are not discussed in the present article can be found in references [10–13]. For computational studies of mononuclear non-heme iron enzymes by other groups, see e.g., references [14–17].

In this article, the different roles of theoretical modeling, and their connection to experimental data, are discussed. Further, it is shown how important conclusions can be drawn, even for systems where the calculations are not very accurate and there is lack of experimental data to compare with. This requires an understanding of the accuracy of both, for the quantum chemistry modeling of the chemical reaction, and for the treatment of the surrounding protein matrix. To better understand the latter factor, the stability of the results for IPNS have been evaluated when going from an active-site model to a quantum mechanics/molecular mechanics (QM/MM) model that also includes free-energy contributions.

2. Computational approach

The enzymatic reactions have been analyzed based on calculations of the full reaction energy diagram using hybrid density functionals, typically B3LYP[18]. The general idea is to discriminate between alternative mechanisms, by showing that only one alternative has reaction barriers consistent with the experimentally observed rate. This approach is widely used for the study of enzymatic mechanisms, but it has two major drawbacks: relatively weak connections to experiment, discussed in section 7, and large uncertainties in the computational results.

To draw relevant conclusions from these mechanistic studies requires that the difference in energy between two alternatives, i.e., the barrier height between two mechanistic proposals, is larger than the error in the computational results. All density functionals have large mean absolute deviations for transition-metal test sets, with errors for B3LYP of ~ 12 kcal/mol[19, 20]. However, these benchmarks are direct metal-metal or metal-ligand bonds, while reactions in transition metal enzymes mainly concern bonds in organic molecules bound to metals. An optimistic view of the *average* errors for B3LYP in transition-metal catalyzed reactions is an estimate of ~ 5 kcal/mol[21, 22]. Further, most DFT functionals underestimate the reaction barriers of organic reactions, with B3LYP having a mean absolute error of ~ 3 kcal/mol[20]. It has been argued that the same relation does not hold for transition states in redox reactions because they have a higher degree of multi-reference character[22], but no proper benchmark tests have been performed so far.

To overcome the clear limitations in the computational accuracy, one possibility is to use a semi-empirical approach, e.g., to start from an experimental observation, and then adapt one or more empirical parameters to fit the experimental data. In DFT calculations of transition-metal complexes, the important parameter is the amount of Hartree-Fock exchange in the functional, with values of ~ 10 – 15 % suggested for some iron complexes[23, 24]. The problem with a more empirical approach is that it is not clear that the parameters adapted to match one set of data also gives the best description of reaction energies or transition-state barriers. There are functionals that perform better

than B3LYP in benchmark tests[20], but experience with transition-metal systems have shown that the accuracy of a certain functional depends strongly on the type of reaction. The widespread use of B3LYP has provided a better understanding of its successes and failures, and has also made it possible to compare results for related enzymes from different studies.

For transition-metal enzymes, additional uncertainty comes from the modeling of the protein environment. In the present review most calculations have been performed using active-site QM models that include 50-200 atoms. To better understand the potential effects of an explicit inclusion of the entire protein, the reaction energy diagram of IPNS is studied using a range of methods, from an active-site model to a QM/MM model that includes classical free-energy contributions[25-28]. Detailed results will be discussed separately in section 6, but protein effects can be significant (~ 10 kcal/mol) for electron-transfer reactions (large change in dipole moment) or when binding/releasing molecules to/from the active site is calculated.

As a rule of thumb, energy differences smaller than 5 kcal/mol are inconclusive, while values larger than 10 kcal/mol can be considered relatively certain, unless a large protein effect is expected. Energy differences between 5-10 kcal/mol must be treated with care, and the sensitivity of the modeling should be tested by changing e.g., the functional or the treatment of environmental effects.

3. Biosynthesis of Fe(IV)=O

The most common purpose for which Nature uses oxoferryl species is oxidation of chemically inert substrates. Since reactions involving such compounds and Fe(IV)=O typically involve energy barriers of around 15 kcal/mol[29, 30], the oxoferryl species itself needs to be a stable intermediate, preferentially with an energy lower than the preceding stable structure, so that the concentration of Fe(IV)=O is sufficient for the catalytic reaction to advance. This "low energy postulate" is realized in catalytic reactions of mononuclear non-heme iron enzymes (MNIE) firstly, by involvement of powerful reductants which provide electrons necessary for initial two-electron reduction of O_2 , and second, by cleavage of the O-O bond in a manner that guarantees that the leaving oxygen atom exits the reaction with two covalent bonds[31]. This general scheme of Fe(IV)=O biosynthesis is illustrated in the Figure 2 for three representative systems: α -ketoacid dependent dioxygenases (α KAD), tetrahydrobiopterin-dependent hydroxylases (THBH), and isopenicillin N synthase (IPNS).

3.1. Reaction energy diagrams

As emphasized in Figure 2, in the first stage of the synthesis molecular dioxygen is reduced to peroxo species with two electrons provided by organic co-substrate, or

Figure 2: General biosynthesis route leading to Fe(IV)=O in selected mononuclear non-heme iron enzymes: a) α -ketoacid dependent dioxygenases [32], b) tetrahydrobiopterin -dependent hydroxylases [33], c) isopenicillin N synthase [26].

substrate,[34] which is either bound directly to the metal (α KAD, IPNS), or in its immediate vicinity (THBH). More specifically, in α KAD, see Figure 2a, α -ketoacid is decarboxylated to form an Fe(II)-peracid intermediate[32]; in THBH oxygen is reduced by tetrahydrobiopterin and forms a peroxo bridge between Fe(II) and an oxidized cofactor[33], whereas in IPNS the Cys- β -C hydrogen of the substrate is transferred to the distal oxygen atom of O_2 , which is concomitant with oxidation of thiolate to thioaldehyde[26]. It should be noted here that for α KAD sometimes a different mechanism is presented, where instead of the Fe(II)-peracid intermediate a Fe(IV)-peroxohemiketal bicyclic species is proposed to lie on the reaction path between the enzyme-substrates complex and the oxoferryl species (see e.g. [35]). However, to the best of our knowledge, the experimental evidence supporting existence of such species is missing, whereas in computational studies it has never been observed [32, 36].

Despite the difference in chemistry, these reactions have very similar barriers (~ 15 kcal/mol) in the modeling, see Figure 3.

Values of reaction energies calculated with cluster models for these initial two-electron steps, amount to: -50.8, +8.0, and -10.8 kcal/mol, respectively. This indicates that the first stage of Fe(IV)=O biosynthesis is, in a least favorable case, modestly endothermic, and in other cases exothermic.

The second stage of the Fe(IV)=O biosynthesis is, without exception, an exothermic process, with oxoferryl species

Figure 3: General reaction energy diagrams for biosynthesis of Fe(IV)=O in selected non-heme iron enzymes: solid line - α -ketoacid dependent dioxxygenases [32]; dotted line - tetrahydrobiopterin-dependent hydroxylases [33]; dashed line - isopenicillin N synthase [26].

considerably more stable than the reactant of the (total) reaction, i.e. Fe(II)-O₂, see Figure 3. Reaction energies of these O-O cleavage steps, calculated with cluster models, are in all instances substantial and negative, i.e.: -19.8, -14.3, and -20.2, which makes the overall biosynthesis of Fe(IV)=O exothermic by: -70.6, -6.3, and -31.0 for α KAD, THBH, and IPNS, respectively[26, 32, 33].

3.2. Details of O-O bond cleavage

Cleavage of the single O-O bond in the Fe(II)-peroxo intermediate is formally a heterolytic process, whereby the two bonding electrons (of the O-O bond) are moved to the oxygen atom distal with respect to iron. The proximal oxygen accepts two electrons from Fe(II) and hence oxidizes it to Fe(IV). Heterolysis usually follows a generic two-step mechanism, see Figure 4, where in the first, more difficult step, Fe(II) provides one electron which goes to the O-O σ^* orbital. In the second, low-barrier step, the unpaired electron of the O-O group localizes on the proximal oxygen atom.

In the peroxo intermediate (peroxo_INT in Figure 4), a doubly occupied iron orbital locates in the Fe-O-O plane, where it has a maximum overlap with the O-O σ^* orbital. The significance of such arrangement stems from the fact, that it allows for efficient electron transfer from Fe(II) to O-O, which is realized as mixing of the two critical orbitals at the TS1 geometry. As an example, the contour of the β HOMO-1 orbital for TS1 in IPNS, which is depicted in Figure 5a, shows mixing of Fe 3d and O-O σ^* orbitals (an animation of the evolution of this orbital in the vicinity of TS1 is available in the on-line version of the paper).

Changes of gross spin populations on iron and O-O group, for both IPNS and α KAD, support this interpretation of electronic structure changes, e.g. the iron spin population changes from 3.8, which is a typical value for a high-spin Fe(II), to 4.1, which indicates a high-spin Fe(III).

Figure 4: Geometrical and electronic structure changes along the reaction coordinate for O-O bond heterolysis: a) isopenicillin N synthase (alternative with protonation from *aqua* ligand) [26], b) α -ketoacid dependent dioxxygenases [32]. The sign '-' indicates the unpaired electron localized in the O-O σ^* orbital. Gross atomic spin populations in italics, interatomic distances (in Å) in bold, relative energies (E) in kcal/mol.

Concerning the geometry changes, at the TS1 the O-O bond is stretched to ca. 1.8 Å, while the Fe-O distance is shortened to approximately the same value.

The product of the first step is a one-electron intermediate (1-e_INT in Figure 4) with a O-O bond stretched to ca. 2.1 Å, whose formal bond order is 0.5. In this structure the unpaired β electron is shared unevenly by the two oxygen atoms, with, as evidenced by the gross spin populations, the main share on the distal oxygen. Note, however, that the oxygen atom proximal to iron is spin polarized by the high-spin Fe(III), which effect masks the contribution from the β electron, and hence, the gross spin population of this atom can be even positive.

A recurrent feature of the O-O bond heterolysis is development of a second covalent bond by the distal oxygen atom as the O-O bond finally cleaves. Typically, for this purpose a proton is provided either by an *aqua* iron ligand, as in THBH (Fig. 2b), or by some acidic group from the protein surrounding or the substrate, as in IPNS (Fig. 2c). For IPNS, the proton could potentially come also from the

IPNS TS2 NOSD(-0.9) animation

Figure 5: Orbitals showing key electronic structure changes in the two transition states for O-O bond heterolysis. Linked animations: IPNS TS1 β HOMO-1, IPNS TS2 NOSD(-0.9)

the carbonyl carbon, as the peracid moiety transforms to carboxylic group. This second transition state for O-O bond heterolysis (TS2 in Figure 4) is coupled with a tiny energy barrier, sometimes vanishing completely.

Focusing on the electronic structure, at TS2 the unpaired β electron from the O-O σ^* orbital localizes on the proximal oxygen atom, whereas the distal one becomes O^{-2} and part of a closed-shell group, i.e. carboxylate or water. This flow of unpaired β electron density can be most readily visualized with the aid of natural orbitals for the spin density (NOSD). At TS2 there is only one NOSD with large negative eigenvalue, i.e. describing the unpaired β electron, see Figure 5b. For TS2 the lobes on the two oxygen atoms are approximately of the same size, yet, as the animation available in the on-line version of the paper shows, the lobe on the proximal oxygen increases at the expense of the distal one as the system moves along the transition vector.

Once TS2 is passed, the unpaired β electron is located on the proximal oxygen atom on the 2p orbital which is perpendicular to the Fe-O axis. Thus, in order to form a regular σ bond between iron and oxygen, the singly occupied 2p orbital of oxygen has to rotate so that it has a good overlap with the iron's singly occupied d_{z^2} orbital, as schematically drawn in brackets in Figure 4. Such a rotation takes place once the system has passed TS2, and when it is completed the singly occupied 2p orbital of the oxyl radical overlaps with Fe $3d_{z^2}$.

The remaining two 2p orbitals of the oxyl radical are doubly occupied, and as they are perpendicular to the Fe-O axis, they form two π bonds, of formal bond order 0.5 each, by overlapping with two singly occupied 3d orbitals of iron. Short Fe-O bond length, of ca. 1.6 Å as typically

IPNS TS1 β HOMO-1 animation.

aqua ligand,[26] but direct protonation from the substrate is favored by a few kcal/mol in the large protein model.[27] Protonation of the leaving oxygen is not, however, the only option as evidenced for example by α KAD (Fig. 2a). In this case, the distal oxygen atom forms a second bond with

observed for Fe(IV)=O [37], is consistent with this description of the electronic structure of the oxoferryl species. A more detailed molecular orbital picture of the Fe(IV)=O species can be found in the following section.

4. Properties of Fe(IV)=O

Oxoferryl species in MNIE can attain various coordination geometries depending on the strength and number of ligands other than the oxo group. Thus, when the total coordination number is 6, the geometry around iron can be described as close to octahedral, as exemplified by Fe(IV)=O species in THBH and IPNS (Fig. 2), although the strength of the iron-oxo bond distorts these complexes away from centrosymmetry. When one of the six ligands is very weak, or absent, the geometry of the oxoferryl species is either square pyramidal with the oxo ligand in the base of the pyramid, or trigonal bipyramidal with the oxo ligand in an axial position. Examples of the former are provided by αKAD in cases when the carboxylate formed in O-O cleavage is a monodentate ligand (Fig. 2a), whereas trigonal bipyramidal coordination is realized, for example, in the active site of 4-hydroxyphenylpyruvate dioxygenase (HPPD) [38], which is also one of αKAD (Fig. 8).

4.1. Electronic structure of Fe(IV)=O

Concerning the electronic structure of the oxoferryl species, a sigma component of the Fe=O bond, corresponds to doubly occupied σ orbital and an empty σ^* , as shown in Figure 6a. Two other 3d orbitals of iron overlap with the 2p orbitals of oxygen forming two (π , π^*) pairs, each accommodating three electrons, which correspond to two half-bonds. In a single-determinant wave function picture the π 's are doubly occupied, whereas the π^* are occupied by unpaired electrons. The remaining two 3d orbitals, labeled as $d_{x^2-y^2}$ and d_{xy} , do not have bonding partners on oxygen and become non-bonding (nb) orbitals in the Fe(IV)=O species.

The ground spin state of Fe(IV)=O is in a quintet, as for example shown with the use of spectroscopy methods for one of αKAD [39], which means that the two non-bonding and two π^* orbitals are all singly occupied with electrons of the same spin polarization. The lowest energy excited state is the spin triplet, which is obtained by pairing the two electrons from the two non-bonding orbitals and localizing the pair on the one with lower energy (d_{xy}). Since such excitation has negligible impact on the Fe-O bond, the reactivity of $^3[\text{Fe(IV)=O}]$ is not enhanced compared to $^5[\text{Fe(IV)=O}]$, and the triplet state is not catalytically relevant [40].

Consequences of formal excitation of one of the electrons from a π orbital to the empty σ^* are far more significant for the reactivity of the system. Such excitation can be viewed as ligand-to-metal charge transfer process because the π orbital has a predominant oxygen character whereas σ^* is primarily Fe $3d_{z^2}$, see Figure 6a. The

Figure 6: Schematic molecular orbital diagram for: a) Fe(IV)=O in the spin quintet ground state, b) excited quintet state ($\text{Fe(III)-O}^\cdot$).

result of this electron transfer is that the Fe-O bond elongates substantially, from ca. 1.65 to ca. 1.90 Å, and the orbitals lose bonding/antibonding character and become almost pure iron or oxygen orbitals, see Figure 6b.

Thus, the electronic structure of such excited state species is best summarized by the formula $\text{Fe(III)-O}^\cdot$, where iron is in a high-spin configuration and the electron hole is localized on an oxygen's 2p orbital perpendicular to the Fe-O bond.

In the excited quintet spin state, the unpaired O 2p electron has minority spin polarization, whereas a corresponding septet state, of very similar energy, is achieved by ferromagnetic coupling of the electrons on Fe(III) and the oxyl radical. $\text{Fe(III)-O}^\cdot$, irrespective of the spin state, is very reactive due to the presence of the oxyl radical, and inherent barriers to its reactions, for example C-H cleavage, are significantly lower than for ground state $^5[\text{Fe(IV)=O}]$. However, the $\text{Fe(III)-O}^\cdot$ electromer was computed to lie at least 16 kcal/mol above the ground Fe(IV)=O state [40], which rules out $\text{Fe(III)-O}^\cdot$ as species participating in the catalytic cycle. The possible exception is the, yet hypothetical, case where $\text{Fe(III)-O}^\cdot$ is preferentially formed in the O-O cleavage step and its subsequent reaction with the substrate is faster than its de-excitation to $^5[\text{Fe(IV)=O}]$.

4.2. Isoforms important for selectivity

SyrB2 is an α -ketoacid dependent halogenase catalyzing chlorination of an aliphatic group of its substrate [41]. A unique feature of α KAD halogenases is that in the first coordination shell of iron, a halogen ligand is bound instead of the Glu or Asp residues that are usual for α KAD (see Figure 2) [42]. The catalytic steps of SyrB2 following the synthesis of the oxoferryl species encompasses C-H bond cleavage, yielding a Fe(III)-OH/substrate radical species, and subsequent recombination of Cl and the radical. Thus, in the final rebound step Cl is the ligand transferred to the radical, and not the usual OH. For a more thorough discussion of oxoferryl reactivity, see section 5.

Interestingly, Mössbauer spectra for freeze-quench trapped Fe(IV)=O in SyrB2 revealed that it exists in two forms characterized by similar (free) energies and distinctive isomer shifts (δ) and quadrupole splittings (ΔE_Q) [43]; values for exp "1" and exp "2" are reported in Figure 7. A similar equilibrium has also been reported for the halogenase CytC3 [44].

Figure 7: Ligand-exchange reaction for the Fe(IV)=O species in the active site of SyrB2 halogenase. Exp "1" and exp "2" are experimental values for the two forms observed for SyrB2. Energies in kcal/mol, Mössbauer spectra parameters in mm/s [40].

In a computational study seeking the plausible origins of SyrB2 preference for chlorination over hydroxylation, a mechanistic hypothesis, which links the existence of the two different forms of oxoferryl species with the selectivity for chlorination, could be put forth [40]. It was proposed that the two species could simply differ in the placement of the oxo and chloro ligands (**a** and **a'** in Figure 7). The Mössbauer spectra parameters computed for molecular models of **a** and **a'** agree reasonably well with the experimental data, and the agreement improves if one compares relative values computed for **a** - **a'** pair with those measured for exp "1" - exp "2".

Swapping of the oxo and chloro ligands proceeds with a barrier of around 13 kcal/mol, which is ca. 5 kcal/mol lower than the barrier for C-H bond cleavage, and hence, the equilibrium between **a** and **a'** can be established. Species **a** is produced in oxidative decarboxylation and O-O cleavage, yet it is species **a'** which is proposed to react with the organic substrate with the lowest barrier. Since in **a'** Cl is exposed towards the substrate, this ligand is preferentially used in the rebound step. Alternative explanation

of SyrB2 preference for chlorination over hydroxylation, which however does not conflict with the above described equilibrium between the two forms of the oxoferryl species, rests on the precise positioning of the substrate, as inferred from the results of studies with substrate analogues [45].

5. Reactivity of Fe(IV)=O

Probably the most frequently encountered reaction of oxoferryl species is oxygenation of an organic substrate, which leads to insertion of an oxygen atom into the skeleton of the oxidized compound [29]. As a representative example we take here electrophilic attack on the aromatic ring of 4-hydroxyphenylacetate, which is one of the catalytic cycle steps of HPPD [46]. The most important geometrical and electronic structure features of this reaction are summarized in Figure 8.

Figure 8: Geometrical and electronic structure changes along the Fe(IV)=O reaction coordinates for electrophilic attack on the aromatic ring (HPPD). Values in parenthesis for excited state species [38, 47]. Gross atomic spin populations in italics, interatomic distances (in Å) in bold, relative energies (E) in kcal/mol. Linked animations: TS α HOMO, TS* β HOMO

5.1. σ - and π channels in oxoferryl reactions

Notably, for this system it was possible to optimize two different transition states: TS which leads to the product (so-called σ -complex) in a ground electronic state (with a high spin Fe(III)), and TS* which lies on a path to an excited state product (with an intermediate spin Fe(III)) [38, 47]. The electronic structure differences can be discerned from the spin populations reported in Figure 8 for TS:s and σ -complexes, i.e. on the ground-state path, the population on iron changes to the value typical for high-spin Fe(III) and the ring-based radical is antiferromagnetically coupled to iron. On the other hand, in the higher energy path, the unpaired electron on the organic radical has the same spin orientation as three unpaired electrons on iron (spin-intermediate Fe(III)). As shown in Figure 8, the activation energy computed for TS (12.0 kcal/mol) is almost 5 kcal/mol lower than that connected with TS* (16.9 kcal/mol). An important geometric difference between TS and TS* is the value of an Fe-O-C angle, where oxygen and carbon atoms are those forming the new bond. Its values are 132 and 122 degrees, respectively.

TS α HOMO animation

Figure 9: Cartoon showing two reaction channels for electrophilic attack by Fe(IV)=O. S stands for substrate.

TS* β HOMO animation

Such a dual behavior of the oxoferryl oxidant was first explained based on molecular orbital symmetry by Solomon and co-workers [48], and it can be rationalized taking into account the molecular orbital picture of $^5[\text{Fe(IV)=O}]$ (see Figure 6a). More specifically, when the substrate approaches the Fe=O group at a wide Fe-O-S angle (180 - ca. 130 degrees; Figure 9a) the substrate's frontier orbital has a good overlap with the σ and σ^* of the Fe=O.

When the substrate approaches the Fe=O group at the sharper angle, in our example 122 degrees, the π and π^* of the Fe=O are the best partners for interaction with the substrate's frontier orbital. Then, since in the transition states the Fe-O bond is significantly stretched, typically to ca. 1.8 Å, the Fe-O group gains significant amount of

Fe(III)-O $\cdot$ character, with the electron hole localized in one of the oxygen's 2p orbitals. At the transition states the hole is stabilized by delocalization between oxygen's 2p orbital and the substrate's orbital, and hence the angle of substrate's approach determines on which 2p orbital the hole develops. In the wide-angle reaction (σ -channel) it is the $2p_z$, whereas in the sharp angle (π -channel) it is e.g. $2p_x$. As can be easily noticed on molecular orbital correlation diagrams in Figure 9, the σ -channel leads to high-spin Fe(III) and a β spin on the substrate radical, whereas the π -channel yields intermediate spin Fe(III) and α spin on the substrate radical.

The fact that different iron 3d orbitals become populated in TS and TS* is visualized in animations showing evolution of α and β HOMO orbitals in the vicinity of TS:s. The animations are to be inserted in the online version of the article.

Finally, it should be clarified here that for the reaction

shown in Figure 8 the π -channel is a higher energy path, and thus most probably it is not catalytically relevant. However, in cases when the substrate has limited freedom to move and it can approach the Fe=O group only at a sharp angle, close to 90 degrees, the π -channel is used for the reaction. See also reference [49] for a discussion of the relative importance of σ and π -channels in oxoferryl reactivity.

6. Stability of protein modeling

As mentioned in section 2, uncertainties in the computational results come from both the quantum-mechanical description of the reaction, and from the treatment of the surrounding protein. In a series of papers, the stability of the IPNS reaction energy diagram have been evaluated for a series of different protein models[25–28].

Computational methods are normally evaluated by benchmarks of relative energies. Experiments provide information about intermediates and reaction paths in enzymes, but rarely about relative energies. The exception is when two species exist in equilibrium, and relative energies can be calculated from the equilibrium constant. Turnover rates can be used to calculate reaction barriers, but that gives only one data point per reaction, unless the buildup and decay of intermediates can be observed. Instead of looking at the accuracy compared to experiment, a more approachable path is to test the stability of the results when using different models[50, 51]. To get a clear picture of the consequences for modeling of a reaction mechanism, protein effects were not studied only for a single step, but for the full IPNS energy diagram, with close to twenty stationary points[27]. The present review includes a selection of reaction steps, namely O₂-binding, Cys- β -C-H bond activation (two-electron reduction), and O-O bond cleavage. Results are presented for an active-site model, different static QM/MM models, and finally for a QM/MM model that includes classical free-energy contributions.

6.1. Active-site model

In all applications of quantum chemistry to proteins the key concept is the active site, which makes it possible to scale down large enzymatic systems to much smaller models. Active-site models have been particularly useful for metal-containing enzymes because the electronic and geometric structures are dominated by the metal and its first coordination sphere[52–54]. These models include a number (50-200) of atoms close to the metal center(s), while effects of the surrounding protein are approximately included by the use of a homogenous dielectric medium, and by freezing one atom per residue to their respective positions in the X-ray structure.

One advantage of a limited-size model is the ability to use efficient geometry optimization algorithms, which makes it possible to optimize transition states almost routinely. The active-site model is often the first approach

when exploring complicated potential energy surfaces, and these results can be used to estimate whether a large protein effect is expected.

For IPNS possible reaction mechanisms were first explored using an active-site QM model that includes iron, iron ligands, and parts of the substrate, see Figure 1. The potential energy diagram for the active-site model is shown in grey in Figure 10. The active-site model gives a reasonable energy profile, although the barrier for Cys- β -C-H bond activation is too high (25 kcal/mol), mainly due to a high endothermicity of O₂-binding. This will change with the use of a QM/MM model.

Figure 10: Reaction energy diagram for biosynthesis of an Fe(IV)=O intermediate in IPNS calculated with active-site and ONIOM QM:MM-ME models. Two different QM:MM-ME models are shown, one where the assigned MM charges are kept constant (No update), and one where atoms are assigned new charges at each stationary point (RESP update). Each stationary point is given a label representing the order it appears in the reaction energy diagram.

6.2. Multi-scale (QM/MM) models I - Classical effects

The active-site model is not always sufficient to understand all aspects of enzyme catalysis. Important enzyme functions like regulation are controlled by residues far from the active site, and many times explicit consideration of the whole protein is necessary to understand reaction mechanisms, relative reaction rates and substrate selectivity. Arguments have been made that reaction energies converge very slowly with system size,[51] and as it is not possible to know *a priori*, which residues that are important, it is unlikely that models with even up to 200 atoms would lead to converged relative energies.

A convenient way to take into account the effects of the whole protein, while keeping the accuracy of the quantum treatment is to use multi-scale models, e.g., QM/MM. The same atoms as in the active-site model are treated by quantum methods, while the rest of the enzyme is treated by a classical force field[55, 56]. The main advantage of multi-scale models is that they highlight potential effects

of the surrounding protein. The main disadvantage is that potential errors in the QM/MM interactions, or from artificial changes in the protein geometry, may lead to reasonable barriers even if the real catalytic effect is not correctly described.

The choice of method often depends on the capabilities of the program package used to perform the calculations. Results in the present review are obtained using the ONIOM QM:MM method. The QM:MM label highlights that it is an extrapolation scheme, while most QM/MM methods are additive schemes, for details see reference [55]. One major reason to use this method for IPNS was the possibility to efficiently locate transition states using an advanced optimizer [57, 58], implemented in the Gaussian program package [59].

Interactions between QM and MM layers can be described by either mechanical (ME) or electronic embedding (EE). In the first case, the interactions are fully described by the low-level (MM) method, i.e., a fully classical description. In the second case, electrostatic interactions are evaluated semi-classically by including protein point charges in the QM calculation. The main advantage of a classical description is the ease with which different contributions to the energy can be analyzed.

Figure 10 shows the reaction energy diagram calculated using both active-site and QM:MM-ME models. The multi-scale model includes all atoms in the X-ray structure, including crystal waters, but no added solvent shell. To easily compare the differences between active-site and QM:MM models, the QM parts of the two models are identical. Results for an extended active site are presented in reference [27].

To isolate the electrostatic interactions between the reacting core and the surrounding protein, Figure 10 contains two sets of QM:MM-ME data, one where the same charges have been used for all stationary points (No update) and one where charges are reparameterized at each stationary point (RESP update). Compared to the active-site model, the QM:MM-ME (No update) model includes geometric effects, van der Waals interactions, and protein-protein interactions. The difference between the two QM:MM curves shows the effects due to changes in the charge distribution during the reaction. To isolate the electrostatics, the comparison is made with the same protein geometry.

O₂ binding - Better active-site geometry. The first reaction step is O₂ binding, and as seen in Figure 10, there is a large stabilizing effect from the protein (8 kcal/mol for the side-on and 10 kcal/mol for the end-on septet) [25]. This is important, because in the active-site calculation the endothermicity of O₂ binding leads to very high barriers for the proceeding C-H bond activation step.

Comparing the QM energies of the active-site and QM:MM-ME models, which are only affected by the geometry, an effect of 3-6 kcal/mol is found, depending on the binding mode (side-on or end-on). A comparison of the iron ligand geometries shows that the active-site model overestimates the flexibility of the iron coordination sphere. The effect is

larger for the five-coordinate site before O₂ binding, compared to the six-coordinate site after binding, see Figure 11. This leads to an artificial destabilization of the structure where O₂ is bound [25].

Figure 11: Geometric effect of the protein environment on O₂ binding in IPNS. Reproduced with permission from reference [25]. Copyright 2007 American Chemical Society.

It is of course possible that the QM:MM model is too rigid, artificially stabilizing a state similar to the original X-ray structure (in the present case the product state where iron is six-coordinate). However, starting from an alternative five-coordinated X-ray structure, and then adding O₂ gave almost identical binding energies and QM geometry contributions. This indicates that optimization of a multi-scale model gives enough flexibility to describe relatively large changes in iron ligand geometry.

O₂ binding - Van der Waals interactions. An analysis of the classical contributions that stabilize O₂ binding in the QM:MM model shows that they mainly come from van der Waals interactions between the O₂ molecule and surrounding protein (3-4 kcal/mol). Since O₂ does not have any interactions with the protein in the deoxy state, any interactions in the oxy state directly affects the binding energy. Similar results have previously been found in a study of O₂ binding in hemerythrin [60]. These effects are found in addition to the effects of introducing van der Waals interaction terms to DFT [61], which are of approximately the same size (4 kcal/mol) [62]. If dispersion-corrected DFT is used, a larger QM model will catch a larger part of the van der Waals interactions, and MM contributions will decrease. Note that decreasing the amount of Hartree-Fock exchange in B3LYP from 20 to 15 % favors O₂ binding by 4.4 kcal/mol [25], which illustrates the challenges to get accurate energies for this process.

Electron transfer - Electrostatic effects The barrier for Cys-β-C-H activation (**3** TS in Figure 10) does not show any significant protein effects when compared to the previous end-on O₂-bound structure (**2** INT). This is expected as the reaction can be described as hydrogen-atom transfer (one proton and one electron). However, after passing the transition state a second electron is transferred from substrate to iron to form a ferrous peroxide intermediate, as discussed in section 3. The relative energies of this intermediate (**4** INT) differ by 17 kcal/mol, in the active-

site and QM:MM-ME (RESP update) models, and this large difference comes almost exclusively from unfavorable electrostatic effects (compare the two ME curves in Figure 10).

Such a large effect can potentially affect the conclusions of a mechanistic study, and it is important to understand how accurately different approaches describe these electrostatic effects. Results from mechanical embedding neither includes electronic polarization of the QM nor the MM part. The first contribution can be handled using electronic embedding, see section 6.3, while the second contribution would require a polarizable force field. In the present example geometric polarization is also completely ignored as the protein geometry is kept fixed after the charges are reparameterized. To properly include geometric polarization requires a dynamic description of the protein, with the energies evaluated using a free-energy technique, see section 6.4.

O-O bond cleavage - Electrostatic effects. O-O bond cleavage proceeds by initial electron transfer from iron to an antibonding O-O σ^* orbital, as described previously in section 3.2. As with the preceding electron transfer reaction, there is an electrostatic effect of the protein, although smaller (3 kcal/mol). Together with the increase in energy of the preceding intermediates, the barrier for O-O bond cleavage is now higher than the barrier for Cys- β -C-H bond activation, which is in disagreement with results from kinetic isotope experiments [63]. As argued above, the discrepancy is most likely due to the neglect of polarization effects.

Water release - Electrostatic effects. Another step with very large (~ 10 kcal/mol) protein effects is the formation of a water molecule after the O-O bond cleavage reaction, (9 INT) in Figure 10. In the QM:MM model the water makes hydrogen bonds with explicit MM waters, which leads to large effects on the reaction energy for this step. As a general observation, release of a product shows large protein effects due to explicit interactions with residues not included in the active-site model. Fortunately, these effects are not critical when determining the reaction mechanism, because product formation is often exothermic and the degree of exothermicity has no effect on the barrier of the next step.

6.3. Multi-scale (QM/MM) models II - Semi-classical effects

The use of a fully classical model makes it easy to analyze the protein contributions, e.g., the effects of individual residues. However, there has been significant disagreement as to whether the results are in line with those obtained from a semi-classical treatment (electronic embedding).

To compare mechanical with electronic embedding, the reaction energy diagram was recalculated with QM:MM-EE, see Figure 12. As before, all calculations were made on the same protein geometry to isolate the electrostatic effects.

Figure 12: Reaction energy diagram for biosynthesis of an Fe(IV)-oxo intermediate in IPNS calculated with QM:MM-ME models, with and without reparameterized QM charges, and QM:MM-EE.

The first observation is that the classical and semi-classical approximations are very similar over a large part of the potential energy diagram, even if the electrostatic effects are large. Only comparing the barrier heights of the three transition states, the mean absolute deviation is 0.7 kcal/mol, which does not in any way affect the conclusions of the mechanistic study.

Major deviations in relative energy appear for the Fe(II)-peroxide (4 INT in Figure 12), and for the release of water. In the first case, the origin of the large electrostatic effect was a charge transfer between substrate and iron. Including electronic polarization, i.e., using ONIOM-EE, stabilizes the peroxide intermediate by 7 kcal/mol (relative to the previous stationary point 3 TS). The large semi-classical effect is due to a highly polarizable electronic structure, originating from two resonance structures with very different charge distributions, see Figure 13. In the QM:MM-EE calculation, the charge transfer from substrate to iron is not complete, as judging by the spin population on the substrate carbon (-0.14).

Despite the large effect of electronic polarization on the energy of the Fe(II)-OOH intermediate, the barrier for O-O bond cleavage is not affected. The reason is that there are no viable resonance structures for the later stationary points, and the polarizability is therefore much smaller.

Figure 13: Resonance structures for the iron peroxide formed after Cys- β -C-H bond activation in IPNS. Reproduced from reference [28]. Copyright 2011 American Chemical Society.

There is also a large polarization effect on the release of water (7 kcal/mol). This illustrates the difficulty in modeling electrostatic effects when point charges are very close to a changing QM region. The semi-classical treatment still lacks mutual polarization (polarization also of the MM residues) and a description of charge transfer between layers. A possible improvement over QM:MM methods is the use of QM:QM' methods, where QM' is a fast molecular orbital method, e.g., a semi-empirical method or density-functional tight-binding (DFTB). The development of DFTB parameters for transition-metal elements, including iron, made it possible to apply the DFT:DFTB method to iron enzymes, including the reaction energy diagram of IPNS[64]. Benchmark tests for a series of enzymatic and biomimetic reactions, show that the critical factor for success is that both molecular orbital methods predict the same electronic structure[65].

6.4. Multi-scale (QM/MM) models III - Dynamical effects

Optimization methods cannot easily describe situations where the protein structure changes during the chemical reaction, e.g., new alignments of side chains or solvent water, thermal fluctuations, or large-scale protein motions. As an example, most QM/MM investigations use a solvent water shell, but keep it frozen to avoid artificial changes in water conformations during optimization. This procedure leads to the complete neglect of dipole reorientation, which is the main reason for the very high dielectric constant of water.

To investigate the role of dynamical fluctuations of the protein, the static interactions between protein and QM region were replaced by classical free-energy corrections from the dynamical sampling of a large number of protein configurations. QM/MM approaches with free-energy perturbation (FEP) have previously been used to describe reactions in both protein and solvent, see e.g., references [66–68]. To obtain reasonably converged free energies requires sampling over several millions of configurations, and the challenge is to accomplish this for an expensive QM/MM Hamiltonian. In addition, to cover the reaction energy diagram for biosynthesis of an oxoferryl species, with multiple stationary points, puts severe requirements on the efficiency of the calculations. For that reason, we developed the QM:[MM-FEP] protocol for complex reactions, e.g., multi-step redox reactions in a transition metal enzyme[28].

Transition-metal systems require relatively expensive QM methods, and the key is to minimize the number of QM calculations that are required. The most important approximations are; to perform the molecular dynamics simulations with a fixed QM geometry, and to use the fully classical (mechanical embedding) approximation. With these approximations, the QM energy is independent of the MM geometry, and only has to be evaluated once for each step in the free-energy perturbation scheme. The mechanical embedding approximation is critical because in electronic embedding, the QM Hamiltonian includes the

position of the MM atoms, which would lead to a recalculation of the QM energy for each protein geometry. As shown previously in Figure 12, the classical approximation is good when looking at the transition-state barriers, while larger deviations are expected for formation of Fe(II)-peroxo and the release of water.

The difference in geometric and electronic structure between two stationary states can be large, so in FEP calculations each reaction step is divided into several intermediate points, e.g., by following selected reaction coordinates (typically bond distances) or the intrinsic reaction coordinate. However, transition-metal systems have complicated multi-dimensional reaction coordinates and to avoid a detailed mapping of the reaction coordinate between all stationary points, the alchemical FEP technique was employed. Here intermediate points are generated by a virtual reaction coordinate that gradually mixes the initial and the final state, and the only information that is required comes from the previous static optimizations. For more details of the free-energy calculations, see reference [28].

The results of the free-energy calculations, QM:[MM-FEP], are compared to the static results in Figure 14. To get a fair comparison, the present comparison also includes static results where the MM geometry has been re-optimized after the charges were reparameterized. Note that the dynamical simulations are performed in a water box, using periodical boundary conditions, while the static calculations are performed without solvent water.

Figure 14: Reaction energy diagram for biosynthesis of an Fe(IV)-oxo intermediate in IPNS calculated with QM:MM-ME and QM:[MM-FEP] models. Note the difference in energy scale compared to Figures 11 and 12

First looking at the effects of MM relaxation, the relative energies of the two steps that had large electrostatic effects, formation of iron peroxide (4 INT) and O-O bond cleavage (6 TS), decreases by a minor amount (1 kcal/mol each). It should be noted that at no point during the reaction coordinate is the protein allowed to change its conformation, so these results only represent relaxation around the same minimum.

Understanding the free-energy results is more complicated. Compared to the stable Fe-O₂ state, the TS barriers for O-O bond formation follows the expected trend, decreasing by 6 kcal/mol compared to the static result, a larger effect than simply performing a geometry relaxation. However, behind this reasonable effect lies a slightly *higher* energy for the Fe(II)-OOH intermediate, and a big decrease in the energy (8 kcal/mol compared to the previous intermediate) of the first step of the O-O bond cleavage process.

The origin of the significant decrease in barrier for the O-O bond cleavage process was analyzed further. Only a minor stabilization was achieved due to a change in the average protein geometry in the dynamical simulation compared to the static optimization procedure. Instead, the low barrier is mainly due to fluctuations in the geometry, where some configurations strongly stabilize the transition state. The challenge is to access enough important conformations to get accurate statistics. In the present calculation, the error bars for this step are relatively large (± 1 kcal/mol) [28].

The FEP effects are very large also for the final step, release of water. This is not unexpected as the released water molecule makes several strong hydrogen bonds with MM waters, and other residues. However, the current method also gives large error bars (± 3 kcal/mol), which illustrates that convergence of the alchemical method is slow for reactions where a QM group moves out into the MM region.

7. Discussion

DFT energy diagrams have been criticized because they do not always agree with the current interpretation of experimental results, and it was argued that it must be a minimum requirement to first reproduce what is known, before making any predictions of what is not yet known. The present section outlines the reasons for a relatively weak connection between experiment and modeling, discusses when this criticism is valid, and when modeling results are important.

To make the discussion less abstract, some points are illustrated by looking at the formation of the Fe(II)-OOH intermediate in IPNS[26]. The B3LYP energy diagram for this reaction, calculated with an active-site model, is shown in Figure 15. As outlined above, O₂ binding results in the abstraction of one proton and two electrons, leading to the formation of the iron peroxide intermediate. The interesting questions are how triplet oxygen is activated to react with an organic substrate, and the role of the protein in catalyzing the reaction.

Focus on transition states. The modeling approach is based on transition-state theory (TST), i.e., that chemical reactivity is mainly determined by the relative energy of the transition state. Understanding enzyme catalysis thus requires an understanding of the differences in protein interactions between reactant and transition state. As computational methods are unique in their ability to describe

Figure 15: Transition-state barriers for Cys- β -C-H bond activation in isopenicillin N synthase. Reproduced with permission from reference [26]. Copyright 2008 American Chemical Society.

transition states (and short-lived intermediates), they can give unprecedented insight into reactivity and catalysis. However, the focus on transition states inherently gives a weak connection between modeling predictions and experiment. Theoretical insights can also come from ground-state theories like frontier molecular orbital theory, but extrapolating from the ground state becomes increasingly difficult as the complexity of the reaction coordinate grows.

The large uncertainty in the calculations of barrier heights routinely leads to errors in experimental rates >1000 times, which often makes direct comparisons with experimental rates meaningless. Transition-state theory also takes into account recrossing, tunneling and non-equilibrium events. Consideration of these events is important to understand experimental results, e.g., kinetic isotope effects, but as they rarely affect the barrier by more than a factor of 10[69], which is dwarfed by the uncertainties of the barrier height calculations, they have often been ignored when modeling transition-metal enzymes.

The general idea is to discriminate between alternative mechanisms, by showing that only one alternative is consistent with the experimentally observed rate. For IPNS, the lowest calculated barrier for C-H bond activation is 15 kcal/mol (compared to the O₂-bound state), see Figure 15. This is in apparent agreement with experiments, which show a barrier of 17 kcal/mol for this step [63, 70]. **Clarify which results that are significant.** A pedagogical problem is that results are presented with a level of detail that does not match the accuracy of the calcu-

lation, e.g., energies are reported in kcal/mol with one decimal, while errors may be 5-10 kcal/mol. The reason to present these seemingly exact results is not to make any claims about the preferred electronic structure, but to give enough details so the results can be tested and reproduced. The main problem is a lack of reliable error bars and confidence intervals for the calculations, which could help clarify whether the computational results are significant or not.

For the mononuclear non-heme iron enzymes, there exists a large number of alternatives for the electronic and geometric structure of O₂-bound reactant. First, O₂ can bind either end-on or side-on to iron. Second, there are two main alternatives for the electronic structure, either Fe(II)-O₂ or Fe(III)-O₂⁻ (superoxo radical). Third, the intermediate could exist in several different spin states. Even assuming that Fe(II) (d⁶) is high spin[71], coupling to the spin on triplet oxygen gives possible triplet, quintet, and septet states. Figure 15 shows the reactant to be a side-on septet, which is best described as an Fe(III)-O₂⁻ superoxo radical. What is not clear is that with a large number of states within 4 kcal/mol, not much can really be said about the ground state of the reactant.

However, when investigating reaction mechanisms, the barriers for two mechanistic proposals are very different, and only one of them is in agreement with experimental data, the computational results can be conclusive without being accurate. For many of the reaction pathways presented in section 3, barriers for alternative mechanisms can be up to 20 kcal/mol higher than for the "best" alternative. And despite the significant effects of the surrounding protein on the reaction energy diagram of IPNS, the general mechanism remained the same for all models as the effects on the transition-state barriers were smaller than 5 kcal/mol.

For IPNS, even if the calculations cannot identify the primary reactant with any certainty, the models can still be used to provide insight into C-H bond activation[26] or even to understand the preference for oxidase over oxygenase activity [16, 72]. Looking at the reaction barriers in Figure 15, the barrier on the quintet surface is lower by 6 kcal/mol compared to the triplet and septet surfaces. Judging by the geometry and spin population of the transition state, it corresponds to a hydrogen atom transfer, with no major changes in dipole moment or atomic positions (with the exception of hydrogen). The effects of the protein environment are therefore expected to be relatively small, see section 6.2. The only major modeling parameter that should affect the preference for reactivity on the quintet surface is the choice of DFT functional.

Assuming that the energy diagram in Figure 15 is correct, it illustrates a potential case when it is not necessarily the lowest-energy structure (septet) that is important when describing the reactivity. Instead, the real (quintet) reactant is "hidden", an excited state on the potential energy surface.

Direct coupling to experiments. An alternative strategy

is to mainly use computational models to aid in the interpretation of experimental data. For IPNS, the O₂-bound intermediate has not been characterized, so it is not possible to compare with experimental data, the closest thing is a thorough investigation of the Fe-NO system[72]. Based on the information from this analogue, and additional DFT calculations Brown *et al.* could propose that the reactant is a Fe(III)-superoxide species where the frontier molecular orbital, the O₂ π*-orbital, is oriented to form a σ-overlap with the hydrogen.

The present review shows how modeling can rationalize the findings from Mössbauer spectroscopy, that the oxoferyl species in halogenases has two distinctive isoforms, see section 4.2. Further, it could be proposed that this equilibrium is responsible for the product specificity of the SyrB2 enzyme. In other cases, a minor error in the prediction of the electronic structure of the reactant, or a failure to reproduce a certain spectrum, prevent a meaningful comparison to experiment. However, this does not necessarily mean that the same method cannot give a reasonable description of the reaction energy diagram, and be used to discriminate between mechanisms.

A mismatch between modeling results and the current interpretation of experiments, can, in addition to errors in the model, also depend on an incorrect, or too stringent, interpretation of the experiments. It can therefore be of great use to simulate experimental spectra to illustrate that a modeled structure indeed is compatible with all current experiments.

Problems with reaction modeling. Despite the generally positive perspective, there are major problems in reactivity modeling. The first problem is that the number of mechanistic alternatives is not known, and a manually guided exploration of the potential energy surface can miss any number of alternative mechanisms. This makes it impossible to conclusively prove that the proposed mechanism is correct. It would be more satisfying to use an unbiased global reaction route mapping. This can be done for small molecules, and in combination with a multi-scale method, also for larger systems[73]. However, it will probably take some more time before the full reaction mechanism for transition-metal enzymes can be explored using this method. A full exploration of all stationary points could still fail to describe all alternative structures, because distortions of the real energy surface can hide stationary structures that are important for the description of the reaction pathway.

Another problem is that with so many variables in the modeling, and very little experimental data, there is generous room for mistakes and misinterpretations of the computational results. In many studies, the only experimental information that is used is the starting structure and the rate of the enzymatic reaction. With a number of potential sources of error, from the density functional, the description of environmental effects, or even the execution of the calculations, it is possible to arrive at the right barrier height for a number of different (incorrect) reasons.

As an example, in a very early study of the IPNS reaction mechanism, an incorrectly converged electronic structure of the Fe-O₂ reactant lead to an incorrect barrier for the first reaction step, which led to the authors to discard the later confirmed mechanism[74]. Instead, an alternative mechanism was proposed based on protonation of the substrate prior to the reaction. These results were inconsistent with later studies[25, 26, 72], but the detail of the reported results made it easy to analyze the origin of the error, and to track down the apparent inconsistency between the studies. A second mechanistic study [26], still failed to explore an alternative mechanism for O-O bond cleavage, partly because a minimal selection of active-site residues artificially destabilized that potential mechanism and partly because the reaction pathways were not properly explored[27].

8. Summary

Theoretical models show the similarities and differences in biosynthesis, properties, and reactivity of the oxoferryl species in mononuclear iron enzymes. Two-electron reduction of dioxygen to form a Fe(II)-peroxo species, is followed by O-O bond cleavage, the latter reaction often occurring in two discrete steps. Reactions between the oxoferryl species and organic substrates can occur in σ and π -channels, and the preferred channel depends on the steric effects around the active site. Further, the two distinctive species of Fe(IV)=O, revealed by Mössbauer spectroscopy and whose identity was proposed based on the computational study, are possibly relevant for specificity of aliphatic chlorination.

Explicit consideration of the protein matrix has significant effects (~10 kcal/mol) on the energy for O₂ binding and the stability of the Fe(II)-OOH species in IPNS. The latter effect is an electrostatic response on an electron transfer between substrate and iron. Different multi-scale models give very different electrostatic contributions, by up to 9 kcal/mol. Fortunately, the conclusions from the modeling are still stable, as the barrier heights used to discriminate between mechanisms do not change by more than 5 kcal/mol.

Acknowledgments

We thank Dr. Tsutomu Kawatsu for designing and analyzing the free-energy perturbation calculations. M.L. acknowledges financial support from the Marcus and Amalia Wallenberg foundation and the Fukui Institute for Fundamental Chemistry. This research project was partly supported by grant No UMO-2011/01/B/ST4/02620 from the National Science Centre, Poland (T.B.).

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, J. Zhou, *Chem. Rev.* 100 (2000) 235–350.
- [2] M. Costas, M. P. Mehn, M. P. Jensen, L. Que Jr, *Chem. Rev.* 104 (2004) 939–986.
- [3] K. D. Koehnert, J. P. Emerson, L. Que Jr, *J. Biol. Inorg. Chem.* 10 (2005) 87–93.
- [4] P. C. A. Bruijninx, G. van Koten, R. J. M. K. Gebbink, *Chem. Soc. Rev.* 37 (2008) 2716–2744.
- [5] C. Krebs, D. Galonic Fujimori, C. T. Walsh, J. M. Bollinger Jr, *Accounts Chem. Res.* 40 (2007) 484–492.
- [6] E. I. Solomon, S. D. Wong, L. V. Liu, A. Decker, M. S. Chow, *Curr. Opin. Chem. Biol.* 13 (2009) 99–113.
- [7] R. P. Hausinger, *Crit. Rev. Biochem. Mol. Biol.* 39 (2004) 21–68.
- [8] T. Flatmark, R. C. Stevens, *Chem. Rev.* 99 (1999) 2137–2160.
- [9] P. L. Roach, I. J. Clifton, C. M. H. Hensgens, N. Shibata, C. J. Schofield, J. Hajdu, J. E. Baldwin, *Nature* 387 (1997) 827–829.
- [10] T. Borowski, P. E. M. Siegbahn, *J. Am. Chem. Soc.* 128 (2006) 12941–12953.
- [11] A. Bassan, T. Borowski, C. J. Schofield, P. E. M. Siegbahn, *Chem. Eur. J.* 12 (2006) 8835–8846.
- [12] T. Borowski, M. R. A. Blomberg, P. E. M. Siegbahn, *Chem. Eur. J.* 14 (2008) 2264–2276.
- [13] T. Borowski, S. de Maroth, E. Broclawik, C. J. Schofield, P. E. M. Siegbahn, *Biochemistry* 46 (2007) 3682–3691.
- [14] S. P. de Visser, *J. Am. Chem. Soc.* 128 (2006) 9813–9824.
- [15] S. Ye, F. Neese, *Curr. Opin. Chem. Biol.* 13 (2009) 89–98.
- [16] C. D. Brown-Marshall, A. R. Diebold, E. I. Solomon, *Biochemistry* 49 (2010) 1176–1182.
- [17] L. W. Chung, X. Li, H. Hirao, K. Morokuma, *J. Am. Chem. Soc.* 133 (2011) 20076–20079.
- [18] A. D. Becke, *J. Chem. Phys.* 98 (1993) 1372.
- [19] F. Furche, J. P. Perdew, *J. Chem. Phys.* 124 (2006) 044103.
- [20] Y. Zhao, D. G. Truhlar, *Theor. Chim. Acta* 120 (2008) 215–241.
- [21] M. Lundberg, P. E. M. Siegbahn, *J. Comput. Chem.* 26 (2005) 661–667.
- [22] P. E. M. Siegbahn, *J. Biol. Inorg. Chem.* 11 (2006) 695–701.
- [23] M. Reiher, O. Salomon, A. B. Hess, *Theor. Chim. Acta* 107 (2001) 48–55.
- [24] G. Schenk, M. Y. M. Pau, E. I. Solomon, *J. Am. Chem. Soc.* 126 (2004) 505–515.
- [25] M. Lundberg, K. Morokuma, *J. Phys. Chem. B* 111 (2007) 9380–9389.
- [26] M. Lundberg, P. E. M. Siegbahn, K. Morokuma, *Biochemistry* 47 (2008) 1031–1042.
- [27] M. Lundberg, T. Kawatsu, T. Vreven, M. J. Frisch, K. Morokuma, *J. Chem. Theory Comput.* 5 (2009) 222–234.
- [28] T. Kawatsu, M. Lundberg, K. Morokuma, *J. Chem. Theory Comput.* 7 (2011) 390–401.
- [29] A. Bassan, M. R. A. Blomberg, T. Borowski, P. E. M. Siegbahn, *J. Inorg. Biochem.* 100 (2006) 727–743.
- [30] P. E. M. Siegbahn, T. Borowski, *Accounts Chem. Res.* 39 (2006) 729–738.
- [31] A. Bassan, T. Borowski, P. E. M. Siegbahn, *Dalton T.* 20 (2004) 3153–3162.
- [32] T. Borowski, A. Bassan, P. E. M. Siegbahn, *Chem. Eur. J.* 10 (2004) 1031–1041.
- [33] A. Bassan, M. R. A. Blomberg, P. E. M. Siegbahn, *Chem. Eur. J.* 9 (2003) 106–115.
- [34] W. A. van der Donk, C. Krebs, J. M. Bollinger, *Curr. Opin. Struc. Biol.* 20 (2010) 673–683.
- [35] K. P. McCusker, J. P. Klinman, *J. Am. Chem. Soc.* 132 (2010) 5114–5120.
- [36] I. A. Topol, A. V. Nemukhin, K. Salnikow, R. E. Cachau, Y. G. Abashkin, K. S. Kasprzak, S. K. Burt, *J. Phys. Chem. A* 110 (2006) 4223–4228.
- [37] P. J. Riggs-Gelasco, J. C. Price, R. B. Guyer, J. H. Brehm,

- E. W. Barr, J. M. Bollinger, C. Krebs, *J. Am. Chem. Soc.* **126** (2004) 8108–8109.
- [38] T. Borowski, A. Bassan, P. E. M. Siegbahn, *Biochemistry* **43** (2004) 12331–12342.
- [39] J. C. Price, E. W. Barr, B. Tirupati, J. M. Bollinger, C. Krebs, *Biochemistry* **42** (2003) 7497–7508.
- [40] T. Borowski, H. Noack, M. Radoń, K. Zych, P. E. M. Siegbahn, *J. Am. Chem. Soc.* **132** (2010) 12887–12898.
- [41] F. H. Vaillancourt, J. Yin, C. T. Walsh, *P. Natl. Acad. Sci. USA* **102** (2005) 10111–10116.
- [42] L. C. Blasiak, F. H. Vaillancourt, C. T. Walsh, C. L. Drennan, *Nature* **440** (2006) 368–371.
- [43] M. L. Matthews, C. M. Krest, E. W. Barr, F. H. Vaillancourt, C. T. Walsh, M. T. Green, C. Krebs, J. M. Bollinger, *Biochemistry* **48** (2009) 4331–4343.
- [44] D. P. Galonić, E. W. Barr, C. T. Walsh, J. M. Bollinger, C. Krebs, *Nat. Chem. Biol.* **3** (2007) 113–116.
- [45] M. L. Matthews, C. S. Neumann, L. A. Miles, T. L. Grove, S. J. Booker, C. Krebs, C. T. Walsh, J. M. Bollinger, *P. Natl. Acad. Sci. U S A* **106** (2009) 17723–17728.
- [46] G. R. Moran, *Arch. Biochem. Biophys.* **433** (2005) 117–128.
- [47] A. Wójcik, E. Broclawik, T. Borowski, in preparation .
- [48] M. L. Neidig, A. Decker, O. W. Choroba, F. Huang, M. Kavana, G. R. Moran, J. B. Spencer, E. I. Solomon, *P. Natl. Acad. Sci. USA* **103** (2006) 12966–12973.
- [49] S. Ye, F. Neese, *P. Natl. Acad. Sci. USA* **108** (2011) 1228.
- [50] A. Altun, S. Shaik, W. Thiel, *J. Comput. Chem.* **27** (2006) 1324–1337.
- [51] L. H. Hu, P. Söderhjelm, U. Ryde, *J. Chem. Theory Comput.* **7** (2011) 761–777.
- [52] P. E. M. Siegbahn, M. R. A. Blomberg, *Chem. Rev.* **100** (2000) 421–438.
- [53] L. Noodleman, T. Lovell, W. G. Han, J. Li, F. Himo, *Chem. Rev.* **104** (2004) 459–508.
- [54] S. Shaik, D. Kumar, S. P. de Visser, A. Altun, W. Thiel, *Chem. Rev.* **105** (2005) 2279–2328.
- [55] T. Vreven, K. S. Byun, I. Komáromi, S. Dapprich, J. A. Montgomery Jr, K. Morokuma, M. J. Frisch, *J. Chem. Theory Comput.* **2** (2006) 815–826.
- [56] H. M. Senn, W. Thiel, *Angew. Chem. Int. Edit.* **48** (2009) 1198–1229.
- [57] T. Vreven, M. J. Frisch, K. N. Kudin, H. B. Schlegel, K. Morokuma, *Mol. Phys.* **104** (2006) 701–714.
- [58] M. Lundberg, K. Morokuma, in: *Encyclopedia of Inorganic Chemistry: Computational Inorganic and Bioinorganic Chemistry*, John Wiley & Sons, Ltd, Hoboken, 2009, pp. 17–31.
- [59] M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G. A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H. P. Hratchian, A. F. Izmaylov, J. Bloino, G. Zheng, J. L. Sonnenberg, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, J. A. Montgomery, J. E. Peralta, F. Ogliaro, M. Bearpark, J. J. Heyd, E. Brothers, K. N. Kudin, V. N. Staroverov, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, N. Rega, J. M. Millam, M. Klene, J. E. Knox, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, R. L. Martin, K. Morokuma, V. G. Zakrzewski, G. A. Voth, P. Salvador, J. J. Dannenberg, S. Dapprich, A. D. Daniels, Farkas, J. B. Foresman, J. V. Ortiz, J. Cioslowski, D. J. Fox, *Gaussian 09*, Revision B.01 (2009).
- [60] M. Wirstam, S. J. Lippard, R. A. Friesner, *J. Am. Chem. Soc.* **125** (2003) 3980–3987.
- [61] S. Grimme, *J. Comput. Chem.* **27** (2006) 1787–1799.
- [62] P. E. M. Siegbahn, M. R. A. Blomberg, S. L. Chen, *J. Chem. Theory Comput.* **6** (2010) 2040–2044.
- [63] J. E. Baldwin, E. Abraham, *Nat. Prod. Rep.* **5** (1988) 129–145.
- [64] G. Zheng, H. A. Witek, P. Bobadova-Parvanova, S. Irle, D. G. Musaev, R. Prabhakar, K. Morokuma, M. Lundberg, M. Elstner, C. Köhler, T. Frauenheim, *J. Chem. Theory Comput.* **3** (2007) 1349–1367.
- [65] M. Lundberg, Y. Sasakura, G. Zheng, K. Morokuma, *J. Chem. Theory Comput.* **6** (2010) 1413–1427.
- [66] E. Rosta, M. Klähn, A. Warshel, *J. Phys. Chem. B* **110** (2006) 2934–2941.
- [67] H. Hu, W. Yang, *Annu. Rev. Phys. Chem.* **59** (2008) 573–601.
- [68] M. Kaukonen, P. Söderhjelm, J. Heimdal, U. Ryde, *J. Phys. Chem. B* **112** (2008) 12537–12548.
- [69] M. Garcia-Viloca, J. Gao, M. Karplus, D. G. Truhlar, *Science* **303** (2004) 186.
- [70] A. Kriauciunas, C. A. Frolik, T. C. Hassell, P. L. Skatrud, M. G. Johnson, N. L. Holbrook, V. J. Chen, *J. Biol. Chem.* **266** (1991) 11779.
- [71] V. J. Chen, A. M. Orville, M. R. Harpel, C. A. Frolik, K. K. Surerus, E. Münck, J. D. Lipscomb, *J. Biol. Chem.* **264** (1989) 21677.
- [72] C. D. Brown, M. L. Neidig, M. B. Neibergall, J. D. Lipscomb, E. I. Solomon, *J. Am. Chem. Soc.* **129** (2007) 7427–7438.
- [73] S. Maeda, K. Ohno, K. Morokuma, *J. Chem. Theory Comput.* **5** (2009) 2734–2743.
- [74] M. Wirstam, P. E. M. Siegbahn, *J. Am. Chem. Soc.* **122** (2000) 8539–8547.