

# Modeling Aspects of Mechanisms for Reactions Catalyzed by Metalloenzymes

P. E. M. SIEGBAHN

*Department of Physics, Stockholm University, Box 6730, S-113 85 Stockholm, Sweden*

*Received 31 October 2000; accepted 26 February 2001*

**ABSTRACT:** Different models to treat metal-catalyzed enzyme reactions are investigated. The test case chosen is the recently suggested full catalytic cycle of manganese catalase including eight different steps. This cycle contains O—O and O—H activations, as well as electron transfer steps and redox active reactions, and is therefore believed to be representative of many similar systems. Questions concerning modeling of ligands and the accuracy of the computational model are studied. Imidazole modeling of histidines are compared to ammonia modeling, and formate modeling compared to acetate modeling of glutamates. The basis set size required for the geometry optimization and for the final energy evaluation is also investigated. The adequacy of the model is judged in relation to the inherent accuracy achievable with the hybrid DFT method B3LYP. © 2001 John Wiley & Sons, Inc. *J Comput Chem* 22: 1634–1645, 2001

## Introduction

Developments the past decade, in particular of density functional theory (DFT),<sup>1</sup> have made it possible to study mechanisms of reactions occurring at active sites in metalloenzymes at a reasonable level of accuracy. Several studies of this type have thus appeared during recent years.<sup>2</sup> In these studies, the metal active site including its first coordination shell ligands have, in general, been included in the model fully quantum mechanically, which commonly leads to a model system in the range of 30–50 atoms. When models of these active sites are constructed, a key parameter is the inherent accuracy of the DFT method itself. For first and second row molecules, benchmark tests show

that all DFT methods perform quite well for geometries, but for energetics, the hybrid methods, which include Hartree–Fock exchange, are substantially superior.<sup>3</sup> The average error on the atomization energies for 55 small first- and second-row molecules is 2.2 kcal/mol using the B3LYP method. For transition metals there are no benchmarks due to the lack of accurate experimental numbers but indications from normal metal–ligand bond strengths are that the accuracy is slightly lower with errors of 3–5 kcal/mol.<sup>2</sup> Also, for transition states the accuracy can be expected to be somewhat lower than for equilibrium structures. Overall, a result that is within 3–5 kcal/mol for a barrier of a transition metal-catalyzed reaction for a model system consisting of 30–50 atoms, must be regarded as quite satisfactory. With this background on the inherent accuracy, the questions addressed in the present study concern the design of a balanced model, both chemical and

Correspondence to: P. E. M. Siegbahn; e-mail: ps@physto.se

computational, which does not severely destroy this accuracy. This model should obviously not be too small to recover the main effects in the reaction. The model should not be unnecessarily large either, because there is no gain to be expected by modeling effects, which only modify the final results by 1–2 kcal/mol. Rather than using an unnecessarily large model, time is then better spent on increasing the number of investigations. Clearly, the limitations on the inherent accuracy also put limits on the questions that can be meaningfully addressed by quantum chemical techniques, as further discussed below.

The next step in complexity compared to the model described above, where only first shell ligands are included, is to include also second shell ligands quantum mechanically, which would require on the order of at least 100 atoms in the model. Although such calculations are possible,<sup>4</sup> models of this size require an order of magnitude more computer time, and will therefore severely limit the number of mechanistic possibilities that can be tested. It is, therefore, a big advantage if they can be avoided unless it is absolutely necessary for the problem under consideration. An interesting alternative is to include the protein by a simple molecular mechanical modeling which now starts to become possible even for redox-active metalloenzymes.<sup>5,6</sup> However, this introduces other questions concerning the accuracy that have not yet been answered. For example, if the inclusion of the protein in this way requires the use of a less accurate DFT method, the overall improvement in accuracy is quite questionable in general and certainly problem dependent. Questions concerning the accuracy using different treatments of the interface between the different parts with relation to covalency and charge transfer are not resolved either. A very simple way to approximately evaluate the effects of the protein surrounding the metal complex is available through dielectric cavity methods. However, the use of such simple methods requires that these effects are rather small and largely independent of the dielectric constant used, because this constant usually has to be rather arbitrarily chosen. Another question in this context is if direct hydrogen bonding from the second to the first coordination shell ligands needs to be included in the model to answer the relevant question.

Even for a model of 50 atoms it is important to know the minimum level of accuracy needed to solve the chemical question posed. If large basis sets can be avoided in the geometry optimization this is a major advantage and allows more detailed in-

vestigations of possible mechanisms. Also, avoiding large basis sets for the final energy evaluation is an advantage, although this part of a study normally represents a more limited problem. For metal complexes with more than one metal center, the question whether ferromagnetic or antiferromagnetic coupling should be used is also a technically important one. With standard programs<sup>7,8</sup> it is much easier to converge wave functions with ferromagnetic coupling, and to model the metal complex with this coupling is therefore a major advantage. In fact, it may even decide whether the study is at all possible to perform or not. With this background it is clear that underlying investigations concerning these model aspects have been a normal part of the studies already published, and many quite definite conclusions have already been drawn. However, no systematic comparisons of different levels of modeling have so far been made for an entire catalytic cycle for a metalloenzyme.

In the present article, the recently suggested eight steps of the catalytic cycle for manganese catalase<sup>9</sup> are investigated to try to answer some of the above questions. For the modeling aspects discussed here, the question whether the suggested catalytic cycle is indeed the correct one or not should not matter. It is more important that the manganese catalase cycle is representative for reaction mechanisms encountered in metalloenzymes. Of the eight steps in this cycle there are both O—O and O—H bond cleavages, O<sub>2</sub> formation, electron transfer steps, and redox active reactions, which together should represent a quite typical mixture of reactions. The conclusions drawn should therefore have a reasonable degree of generality. Finally, it should be noted that even if a large model can be afforded it is still one of the major tasks in a quantum chemical investigation to analyze what the important effects are. Because an energy criterion can be provided for the importance of each effect, quantum chemistry is particularly suited to address this question and quantify this importance by its energy contribution. Such information is much more difficult, or often even impossible, to obtain from experiments.

---

## Computational Details

The normal procedure used in recent DFT applications on metalloenzymes is that the calculations are performed in three different steps.<sup>2,10</sup> First, an optimization of the geometry is performed using the B3LYP method.<sup>1,11</sup> In most applications performed so far this has been done using rather small double zeta basis sets. In the second step the energy is

evaluated for the optimized geometry using a larger basis set, also at the B3LYP level. In the final step, polarization effects from the surrounding protein are evaluated using a dielectric cavity model. The calculations discussed here were carried out using the GAUSSIAN-94<sup>7</sup> and GAUSSIAN-98<sup>8</sup> programs.

The double zeta basis used in the B3LYP geometry optimizations is the LANL2DZ set of the GAUSSIAN program. For manganese, this means that a nonrelativistic ECP<sup>12</sup> is used. In one of the investigations discussed below this basis was extended for the geometry optimization by adding a diffuse d- and p-function on manganese and a d-function on all first row atoms. The exponents were taken from the 6-311+G(1d,1p) basis set. For the most critical step in the catalase cycle an even larger basis was used in the geometry optimization step, where all polarization and diffuse functions from the 6-311+G(1d,1p) set were included except the f-function on manganese.

The standard large basis used in the B3LYP energy calculations starts with the LANL2DZ basis and adds diffuse and polarization functions from the 6-311+G(1d,1p) basis set. This basis set has a single set of polarization functions on all atoms including an f-set on manganese, and also diffuse functions. In one of the model studies discussed below, an even larger basis was used in this step with two sets of polarization functions instead of one for manganese and oxygen taken from the 6-311+G(2d,2p) set. An additional diffuse d-function was also added on manganese.

For all energies discussed below, zero-point vibrational effects were added. These were obtained by calculating molecular Hessians at the B3LYP level using the same double zeta basis set as used for the geometry optimization, and were in most cases only performed on the smallest model where ammonia was used instead of imidazole as model for the actual histidine ligands. The same zero-point effects were used for all the different models discussed below. The molecular Hessians were also used to optimize saddle point structures, for the ammonia and imidazole model structures, and define these as true transition states. For the other models discussed below, transition state structures were obtained by an energy minimization, where the key geometrical parameters (one or two) were kept frozen from the imidazole model. Entropy and other temperature effects were found to be small, and are not included in the energies in the figures and in the text.

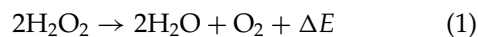
Dielectric effects from the surrounding protein were obtained using the Self-Consistent Isodensity

Polarized Continuum Model (SCI-PCM),<sup>13</sup> using a dielectric constant of 4. These effects were computed at the B3LYP level using the LANL2DZ basis. Because the dielectric effects were generally found to be small, the same dielectric energy contributions were used for all the models discussed below.

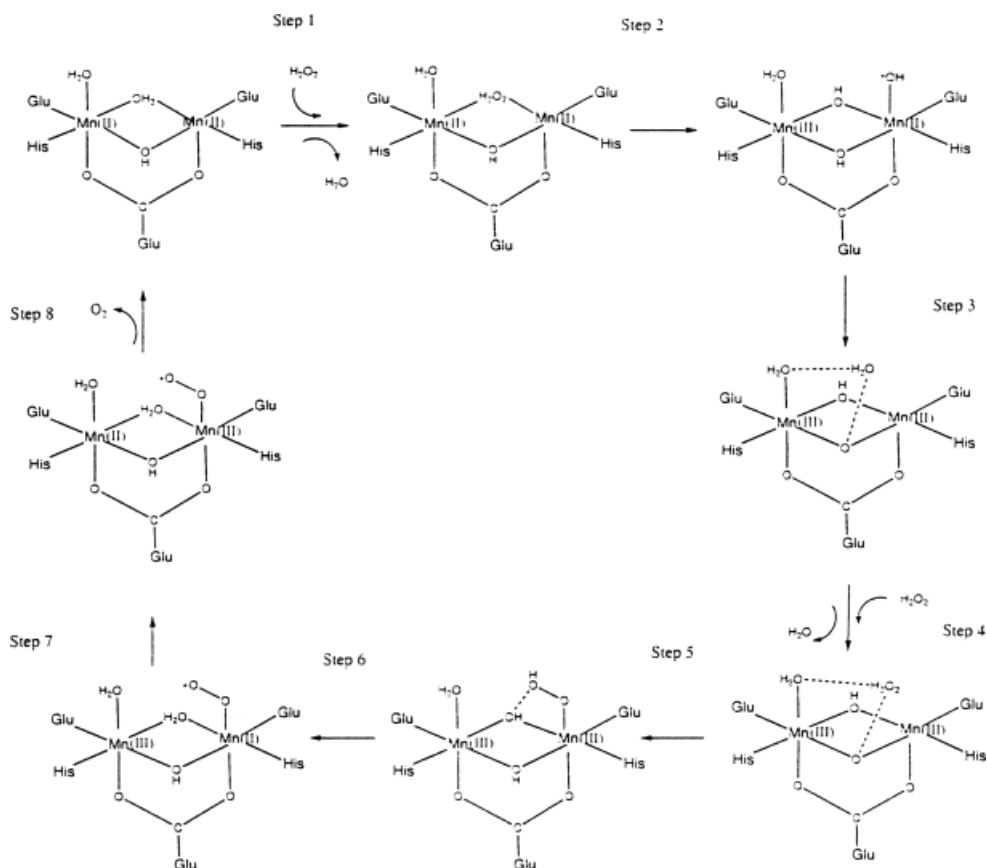
For the discussion below where the accuracy of different models are discussed, it is furthermore important to know the convergence criteria used in the calculations. For practical reasons, not to spend unreasonable amounts of unnecessary computer time, rather loose convergence criteria were used. A geometry is considered sufficiently converged when the rms force has been less than 0.0003 for a few geometry steps and when the energy changes are simultaneously less than 0.1 kcal/mol per step. This procedure normally leads to an energy converged to 0.5 kcal/mol and distances converged to 0.01 Å. However, exceptions can obviously occur where the deviations to the fully converged case is larger than this, as will be discussed below.

## Results and Discussion

The enzyme used here to investigate the models and their accuracy is manganese catalase. Catalases are metalloenzymes that protect the cell from oxidative damage by excess hydrogen peroxide produced during O<sub>2</sub> metabolism.<sup>14</sup> Hydrogen peroxide is destroyed, forming oxygen and water in a disproportionation reaction:



There are two major classes of catalases. The most abundant one of these has an Fe protoporphyrin IX cofactor with a proximal tyrosinate ligand *trans* to the position where the substrate binds. In the second class of catalases there is instead an active dimanganese complex, and this is the type of catalase considered in the present study. This type of catalase has been isolated from the thermophilic eubacteria, for example, *Thermus thermophilus*<sup>15,16</sup> and *Thermomoleophilum album*,<sup>17</sup> and from lactic acid bacteria, for example, *Lactobacillus plantarum*.<sup>18</sup> Crystal structures have been determined for the cases *Thermus thermophilus*<sup>19</sup> and *Lactobacillus plantarum*.<sup>20</sup> These structures show that both enzymes have a bridged binuclear manganese cluster. During turnover the binuclear cluster cycles between two different sets of oxidation states, the reduced form, Mn<sub>2</sub>(II,II), and the oxidized form, Mn<sub>2</sub>(III,III), which are both in principle indefinitely stable. Several mechanisms have been suggested previously based on



**FIGURE 1.** Schematic picture of the eight suggested steps of the catalytic cycle of manganese catalase.

experiments,<sup>21–24</sup> but the mechanism considered here is the one suggested recently based on a DFT study.<sup>9</sup> As emphasized in the introduction, for the present purposes it is not necessary that this mechanism will turn out to be the correct one, but more important, that it is representative for reactions occurring in a metalloenzyme.

The eight suggested steps of the catalytic cycle for manganese catalase are shown schematically in Figure 1. These steps can be briefly described as follows: *Step 1*: a water molecule bridging the two manganese centers in the  $\text{Mn}_2(\text{II},\text{II})$  complex is being substituted by a hydrogen peroxide molecule. *Step 2*: the O—O bond of hydrogen peroxide is cleaved forming a hydroxyl radical and a bridging hydroxide. The manganese complex has the oxidation states  $\text{Mn}_2(\text{III},\text{II})$  at the end of this step. The transition state is one of electron transfer from one of the manganese centers to the hydrogen peroxide. *Step 3*: a spin-transition and a proton transfer occur leading to a  $\text{Mn}_2(\text{III},\text{III})$  complex with a bound water molecule. *Step 4*: the water molecule produced in Step 3 is being substituted by a hydrogen peroxide

molecule. The manganese complex is  $\text{Mn}_2(\text{III},\text{III})$ . *Step 5*: an O—H bond of the hydrogen peroxide that entered in Step 4 is being heterolytically cleaved. The manganese complex remains in  $\text{Mn}_2(\text{III},\text{III})$ . *Step 6*: the second O—H bond of hydrogen peroxide is being cleaved leading to a terminally bound  $\text{O}_2$  ligand and a  $\text{Mn}_2(\text{III},\text{II})$  complex. The transition state is one of electron transfer from the peroxide to one of the manganese centers. *Step 7*: an electron is transferred between the manganese centers leading to a transfer from  $\text{Mn}_2(\text{III},\text{II})$  to  $\text{Mn}_2(\text{II},\text{III})$ . No transition state was obtained for this step. *Step 8*: a triplet oxygen molecule is being released.

Seven different models of varying size and accuracy have been used in the present study, and each one of these is discussed below. All energetic results are summarized in Table I, where the energy is set to zero for the reactant of step 1 for each model. Unfortunately, for the reactant of Step 7 only one model (the  $\text{NH}_3$  model) converged, so for this step the results shown for the other models in the figures were taken from this model. The comparison in the figures discussed below is always made to

**TABLE I.**  
**The Energetics (kcal/mol) for the Different Steps of Manganese Catalase (see Text), Modeled at Different Levels Described in the Text.**

	Imid.	NH <sub>3</sub>	Acet.	–Diel.	+d(geom)	LanI	+2d(energy)
Step 1	0	0	0	0	0	0	0
Step 2	–0.3	1.2	–1.5	2.5	–0.3	3.1	0.5
TS	15.1	16.4	13.5	17.5	15.9	16.0	15.2
Barrier	15.4	15.2	15.0	15.0	16.2	12.9	14.7
Reac.en.	–3.6	–5.7	–3.4	–4.7	–3.1	–6.8	–4.8
Step 3	–3.9	–4.5	–4.9	–2.2	–3.4	–3.7	–4.3
Reac.en.	–39.3	–39.1	–40.8	–38.7	–41.5	–32.0	–40.3
Step 4	–43.2	–43.6	–45.7	–40.9	–44.9	–35.7	–44.6
Reac.en.	–4.8	–9.0	–4.3	–4.9	–3.5	–3.8	–4.0
Step 5	–48.0	–52.6	–50.0	–45.8	–48.4	–39.5	–48.6
TS	–39.1	–39.9	–42.2	–37.7	–39.3	–32.7	–40.2
Barrier	8.9	12.7	7.8	8.1	9.1	6.8	8.4
Reac.en.	–2.8	+1.3	–3.7	–0.8	–2.5	–5.2	–3.0
Step 6	–50.8	–51.3	–53.7	–46.6	–50.9	–44.7	–51.6
TS	–44.4	–43.2	–45.9	–40.9	–44.6	–42.0	–44.5
Barrier	6.4	8.1	7.8	5.7	6.3	2.7	7.1
Reac.en.	+3.6	+4.3	+3.8	+2.1	+3.0	–1.1	+4.8
Step 7	–47.2	–47.0	–49.9	–44.5	–47.9	–45.8	–46.8
Reac.en.	+3.6	+4.3	+3.8	+2.1	+3.0	–1.1	+4.8
Step 8 <sup>a</sup>	—	–51.4	—	—	—	—	—
Product	–51.0	–49.4	–51.7	–49.3	–50.2	–41.6	–49.6

The energy is set to zero for the reactant of step 1 for all models.  
<sup>a</sup> Convergence only obtained for the NH<sub>3</sub> model.

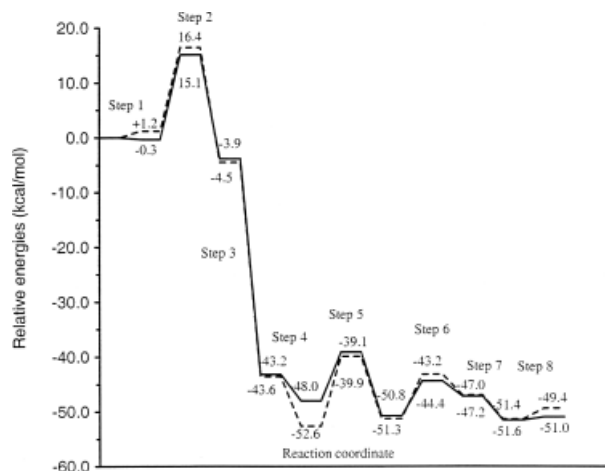
the “Imid.” model, which is typical for the level of modeling and accuracy used in previous studies on metalloenzymes,<sup>2,25–29</sup> including the one on manganese catalase.<sup>9</sup> As models for the ligands, it uses imidazole to model histidines, and is therefore termed “Imid.” in Table I, and the glutamates are modeled by formates. The geometry optimization is performed using a double zeta basis set and for the final energy evaluation a single set of polarization and diffuse functions are added.

When comparisons between the models are made it should be kept in mind that the overall accuracy of the B3LYP method is 3–5 kcal/mol. Deviations between the models that are much smaller than this are, therefore, unlikely to be significant. Exceptions to this rule are possibly cases where very similar mechanisms and models are compared, for which the trend of the variation may be significant even if it is small. In general, larger differences are required to discriminate between mechanisms. Fortunately, distinctly different mechanisms com-

monly differ by much more than 3–5 kcal/mol. For example, the different activation mechanisms of O<sub>2</sub> in cytochrome oxidase differed by more than 15 kcal/mol, as did some of the different mechanisms for isopenicillin N synthase, while the different mechanisms tried in photosystem II commonly differed by at least 5–10 kcal/mol. Before the comparisons between the models are discussed below, it is furthermore important to note that energy deviations less than 0.5 kcal/mol and distance variations less than 0.01 Å are numerically insignificant. For some cases, explicitly mentioned below, even larger deviations found between the models may be numerically insignificant.

**THE AMMONIA MODEL**

The first chemical model for which a comparison is made to the Imidazole model, described above, is one that is the same except that the histidines are modeled by ammonia, and this model is therefore



**FIGURE 2.** Energy diagram for the suggested mechanism of manganese catalase. The thick lines are the results for the imidazole model of histidine, and the dashed lines are those for the ammonia model.

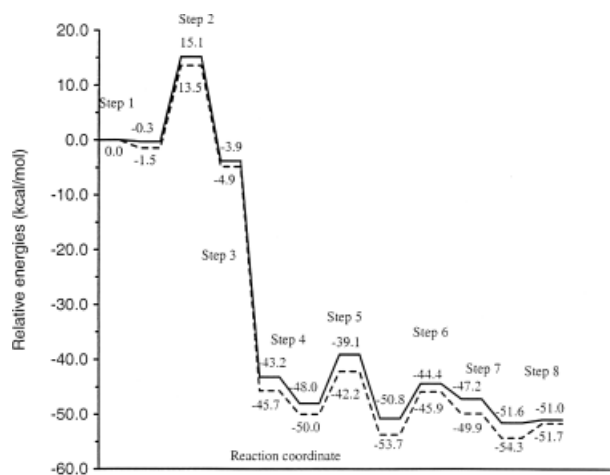
termed “NH<sub>3</sub>” in Table I. The computational level is identical to the one used for the Imidazole model. The energies for the two models are compared schematically in Figure 2. The main conclusion that can be drawn from these energy comparisons is that the two models give strikingly similar results. Most energies are within 1 kcal/mol, which is close to the numerically significant limit using the present convergence thresholds (see above). For example, the barrier for the critical rate-limiting step of O—O bond cleavage is very similar between the models, 15.4 and 15.2 kcal/mol, respectively. There are two additional, somewhat larger, deviations that can be noted in the figure that may still not be numerically significant. These concern the first and the final step where the deviations are 1.5 and 1.6 kcal/mol. For these points there are weakly bound ligands, water and O<sub>2</sub>, whose positions are difficult to converge precisely and where very close local minima exist. Minor deviations for these steps therefore have to be regarded with caution throughout the model comparisons of the present study. The large similarity between the models as shown in the figure may be considered quite surprising, because ammonia lacks the conjugation present in imidazole. However, based on detailed comparisons of ligand effects performed over the past decade for different catalytic reactions,<sup>30</sup> the present results are actually expected. From those studies, it can even be expected that water would probably be a qualitatively reasonable model of histidine in this context. This experience was the background for the quite simple model used in the first B3LYP study of methane

monooxygenase, where water and hydroxide ligands were used.<sup>31</sup> It is clearly a major advantage to use these smaller ligands, at least in the large number of exploratory investigations needed to find the best mechanism. This reduces the number of atoms in the models significantly, from about 40 to 30 in manganese catalase, which means about a factor of 5 on the computation time. Ammonia ligands were, therefore, used in the previous catalase study except at the very end, where imidazole ligands were used.

From the comparison of the energies discussed above, the conclusion drawn is that the electronic effects on the mechanism from using ammonia and imidazole ligands are very similar. However, there is one exception to the good agreement between the models, and this occurs for the product of Step 4. The energy deviation is here as large as 4.6 kcal/mol, 52.6 kcal/mol compared to 48.0 kcal/mol, which is significant. The reason for this deviation is easy to find. When the second hydrogen peroxide binds to the complex, an artificial hydrogen bond is formed to one of the N—H bonds of ammonia, which could not form to histidine (or imidazole). When ammonia is used as a model ligand, this situation must therefore be avoided. In the present case, it is easy to avoid this except in this step, but it is, in fact, possible also for this step. When a local minimum without hydrogen bonding is located, the deviation to the imidazole model is reduced from 4.6 to 1.5 kcal/mol. However, for some systems like tyrosinase where there are six histidines, it turned out to be impossible to avoid artificial hydrogen bonding when ammonia was used,<sup>32</sup> and there seems to be no alternative but to use imidazole models in that case.

### THE ACETATE MODEL

The second chemical model compared to the Imidazole model is the same as this one except that the glutamates are modeled by acetates, and it is, therefore, termed “Acet.,” in Table I. The histidines are modeled by imidazoles, and the computational level is identical to the one used for the Imidazole model. Because acetate must be considered much more electronically similar to formate than ammonia is to imidazole, even smaller deviations between the models could be expected in this case. On the other hand, the glutamates are more active in the suggested mechanism than the histidines are. For example, the glutamates significantly change their hydrogen bonding in a mechanistically important way, both in Steps 2 and 5. The energies of the two models are compared in Figure 3. The main conclu-

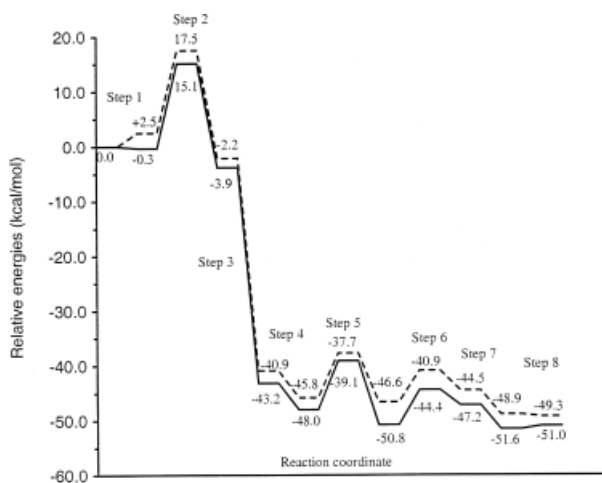


**FIGURE 3.** Energy diagram for the suggested mechanism of manganese catalase. The thick lines are the results for the formate model of glutamate (the "Imid." model in Table I), and the dashed lines are those for the acetate model.

sion when the energies are compared must again be that the results are very similar. From the figure it may appear as if the deviations are slightly larger than for the ammonia model, but this is actually not the case but rather a result of an initial deviation for Step 1 of 1.2 kcal/mol, that remains throughout the cycle and that is probably not numerically significant (see above). If the individual reaction energies and barriers of the different steps are compared in Table I, it can be concluded that they are, in general, more similar than the Ammonia model to the reference Imidazole model. The energetic deviations for the Acetate model is almost within the convergence thresholds in most cases, even for the steps where the glutamates are most involved. Even more important is that, in contrast to using ammonia ligands where artificial hydrogen bonding is always a potential danger, no such problem can occur for the formate model. Also, the geometrical changes compared to the formate model are essentially within the convergence thresholds. From these results and similar comparisons for methane monooxygenase, it appears that larger models for glutamates than formates should almost never be motivated when the B3LYP method is used. Clearly, substantial amounts of computer time are saved by avoiding acetates in the modeling.

### THE -DIELECTRIC MODEL

The third model compared to the standard Imidazole model is chemically identical, with the same modeling of all ligands, but is different computa-



**FIGURE 4.** Energy diagram for the suggested mechanism of manganese catalase. The thick lines are the results with dielectric corrections (the "Imid." model in Table I) and the dashed lines are those without this correction (the "-Diel." model).

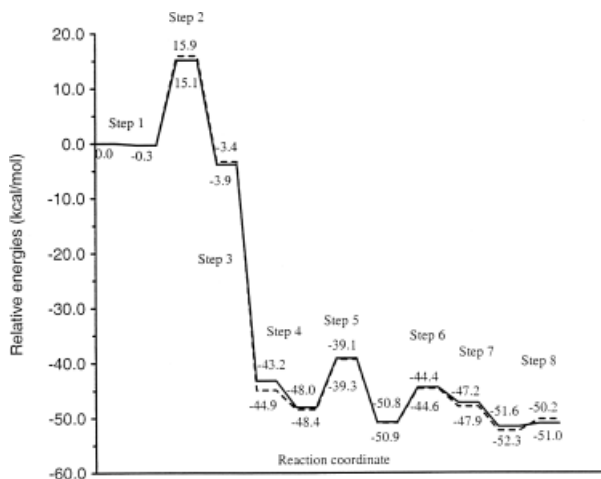
tionally by not including the dielectric effects. From the energy difference of this model and the standard model, the dielectric effects can, thus, be studied. As seen on the results shown in Figure 4 and in the table, the dielectric effects are somewhat larger than the effects noted in the previous figures. For the reaction energy of Step 5 the effect is a decrease by 2.0 kcal/mol, for example. An advantage with these effects, as opposed to the previous effects discussed above, is that they take very little computation time and can, therefore, always be evaluated. This means that there is by now a quite large number of cases where the effects have been studied. The present results are quite typical for the effects found when the system does not change its charge. For these systems, a dielectric effect of more than 3 kcal/mol is often a sign of a badly designed model where, for example, an essential hydrogen bond is left out. In cases with electron or proton transfers over long distances, the situation is clearly different. For these situations the dielectric cavity method should be regarded as qualitative, although still with surprisingly accurate results in some cases.<sup>33</sup> The dielectric cavity method is very poor at reproducing effects from hydrogen bonding left out of the model. If a hydrogen bond to the second shell is expected to have a large effect on the property under investigation, it should, therefore, be included in the quantum chemical model. As a general rule, hydrogen bonds to uncharged groups can be left out of the model unless this bond is expected to be broken in the reaction studied. The electronic and geomet-

ric structure effects of hydrogen bonds from inactive uncharged groups are generally found to be very small. As an example, in a recent model of compound **Q** in MMO,<sup>4</sup> one of the carboxylates, Glu243, bound to one of the Fe(IV) atoms of the dimer, forms a hydrogen bond to a second shell uncharged ligand. Because the carboxylate is charged it might be expected that this hydrogen bond should be important to include in the model. However, including this ligand does not change the metal–ligand bond distances beyond the convergence thresholds used (those mentioned above), and the same is true for the critical spin-populations, which change by less than 0.01.<sup>34</sup> Any chemical effects of this hydrogen bond on the property of compound **Q** is, therefore, very unlikely if it is not broken during the reaction studied.

### THE +d(GEOMETRY) MODEL

The fourth model compared to the standard Imidazole model is also purely a test of the computational model. In this case the same chemical model is thus used, but the computational scheme is changed by optimizing the geometries using a larger polarized basis set (for details see the second section). The final energy evaluation is done at the same standard level as before.

When polarization functions are added on the first row atoms, this is expected to have noticeable effects on the geometry, which is also found to be the case. For the O—O activation TS, for example, the Mn—Mn bond distance is decreased by as much as 0.11 Å while the Mn—His distances are increased by 0.04–0.05 Å. For the hydrogen peroxide, the distance from the lower oxygen to one of the manganese atoms decreases by 0.07 Å, while the proton on the same oxygen moves away by 0.14 Å towards the glutamate. The other distances change by 0.02 Å and less. Even though these distance changes may appear mechanistically significant, they turn out to have only very small effects on the energetics. There are two reasons for this. First, the largest bond distance changes occur for the bonds with the most shallow energy minima. This meant that the total energy obtained in the reoptimized geometries, using the same large basis as in the standard procedure, decreased by only 6.7 kcal/mol. Second, the distance changes are often essentially the same for all structures along the reaction pathway. As seen in Figure 5 and in Table I, the relative energies using the two different sets of geometries are extremely similar. The only numerically significant differences occur for the product of Step 3, where the difference



**FIGURE 5.** Energy diagram for the suggested mechanism of manganese catalase. The thick lines are the results with geometries optimized using the lanl2dz basis (the “Imid.” model in Table I) and the dashed lines are those with geometries obtained using a polarized basis (the “+d(geom)” model).

is 1.7 kcal/mol, and possibly for the TS of Step 2 with 0.8 kcal/mol and the product of Step 8, also with 0.8 kcal/mol. The very high stability of the energies with respect to geometric changes is in line with the experience obtained for similar systems the past years. In several cases where notable effects were obtained on the final energies when the basis set was increased, the geometries were reoptimized at a higher level, and the effects were always found to be small. This is quite fortunate, because the geometry optimizations using the smaller basis are substantially faster, for the present systems by about a factor of 3. This is in line with an  $n^3$  dependence on the basis set size, because the basis set increases from 283 in the small to 414 basis functions in the large basis.

For the most critical rate-limiting step of the catalase cycle, the O—O bond cleavage in Step 2, the geometry was also optimized using a still larger basis set. The first set of polarization functions, discussed above, increased the barrier by 0.8 kcal/mol from 15.1 to 15.9 kcal/mol. Even though this does not indicate any particular sensitivity to the basis set size, there could be hidden problems in the basis sets chosen. The extended basis set used includes all the polarization and diffuse functions from the 6-311G+(1d,1p) basis except the manganese f-functions, leading to a total of 500 basis functions and an expected computation time increase by a factor of 5 compared to the double zeta basis. Most notably, diffuse functions on oxygen,

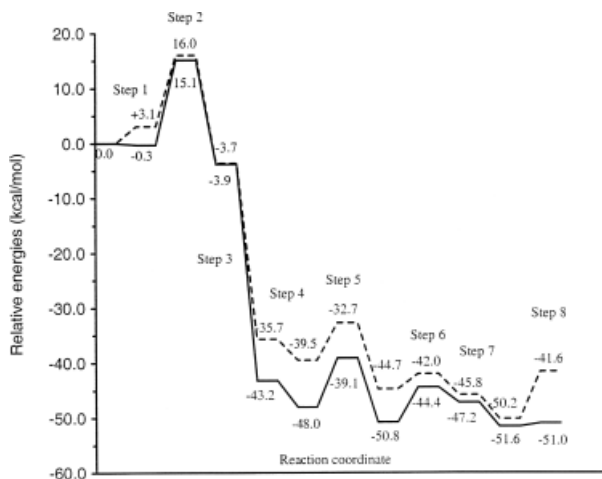


p-polarization on hydrogen and diffuse d-functions on manganese were added. The geometries, also for the transition state, were fully optimized using the Hessian obtained from the small basis set. The geometric changes due to the large extension of the basis set are quite small, on the order of 0.02 Å and less, hardly above the convergence thresholds. The total energy obtained in the reoptimized geometries, using the same large basis as in the standard procedure, decreased by only 0.5 kcal/mol for the TS, and remained the same for the reactant. The barrier decreased by 0.4 kcal/mol. Because there are negatively charged ligands, these truly negligible effects from adding diffuse functions might be considered somewhat surprising. However, it must be noted that the present comparisons for the “+d(geom)” model only concern the indirect effect of the geometries on the energies, not the direct effect that will be discussed below.

### THE LANL MODEL

One of the major findings in quantum chemical modeling of mechanisms the past decades is that geometries require a much less rigorous treatment in terms of methods and basis sets than the final energies do. This is particularly true for *ab initio* methods, where often a simple Hartree–Fock treatment of the geometry using a small basis set is enough to balance a final energy evaluation using a coupled cluster method with a huge basis set.<sup>30</sup> The same is true, although to a lesser extent, when DFT methods are used.<sup>2,3,10</sup> In this and the next section, the basis set requirement on the final energy will be evaluated, again in the light of the internal limitations on the accuracy of the B3LYP method. As discussed above, a double zeta basis (here LANL2DZ) appears quite adequate in most cases for the geometries. It is not expected that this will be true for the energies. A comparison between the LANL2DZ energies and the ones obtained using the large standard basis set described in the second section and used in all the above comparisons are shown in Figure 6.

As expected, the comparisons in the figure and table for the Lanl model show much larger deviations compared to the standard treatment than for any of the other comparisons. For example, the reaction energy of Step 3, which is a spin-transition where one of the manganese atoms changes oxidation state, is 32.0 kcal/mol compared to 39.3 kcal/mol in the standard treatment. Also, the step where O<sub>2</sub> is released, going from the reactant of Step 7 to the product of Step 8, is exothermic

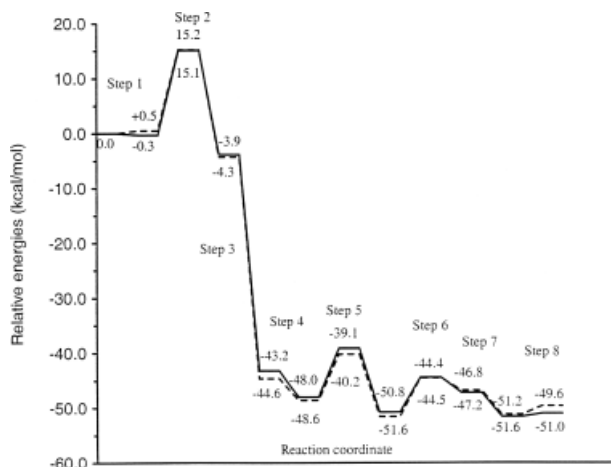


**FIGURE 6.** Energy diagram for the suggested mechanism of manganese catalase. The thick lines are the results obtained using the standard polarized basis (the “lmid.” model in Table I) and the dashed lines are those obtained using an unpolarized basis (the “Lanl” model).

by 2.8 kcal/mol in the standard treatment and endothermic by 4.2 kcal/mol in the Lanl model. These two cases, where a change of oxidation state occurs and where a  $\pi$ -bond is formed or broken, are the ones where the largest deviations are expected. In Step 2, where both these effects are present, they largely cancel. The barrier for the second O—H bond cleavage in Step 6 is furthermore notably underestimated at the Lanl level. Even though the deviations to the standard treatment are substantial, and definitely motivate the use of the larger basis for the energies, the overall picture given for the catalase cycle using the small unpolarized basis set is qualitatively correct. This would probably not have been the case for a corresponding *ab initio* treatment, where the effects of increasing the basis set in the correlation treatment are generally much more dramatic. In fact, at the DFT level most qualitative conclusions concerning the mechanism can be already drawn after the geometry optimization before the final energies have been evaluated using the big basis, and this is often done in practice.

### THE +2d(ENERGY) MODEL

With the rather large effects on the final energies found by adding the first set of polarization functions, the effects of adding a second set might not be negligible. To test this, the energies for the catalase cycle were evaluated also for such a basis set. The results are shown in Figure 7 and in Table I. The effects are, as expected, smaller than when the first set



**FIGURE 7.** Energy diagram for the suggested mechanism of manganese catalase. The thick lines are the results obtained using the standard polarized basis (the “Imid.” model in Table I), and the dashed lines are those obtained using double polarization (the “+2d(energy)” model).

of polarization functions were added. The largest effects are found for the steps where also the first set had the largest effects, in Step 3 and Step 8. However, the effect on the reaction energy of Step 3 is only 1.0 kcal/mol and of Step 8 also 1.0 kcal/mol, and hardly motivate the extension of the basis set as a general procedure. This is in line with previous experience where the energies in several studies were evaluated both for one and two sets of polarization functions.<sup>10, 31</sup> This procedure was abolished after a few investigations, because the energies hardly ever were affected by the basis set extension.

### ANTIFERROMAGNETIC COUPLING

Most of the manganese complexes involved in the catalase cycle have experimentally been determined to have antiferromagnetically coupled manganese spins. However, due to severe convergence problems for antiferromagnetic coupling, all the calculations discussed above have used ferromagnetic coupling between the spins. Experimentally,  $\text{Mn}_2(\text{II},\text{II})$  and  $\text{Mn}_2(\text{III},\text{III})$  complexes of the present type do not show strong exchange couplings,<sup>24</sup> and the use of ferromagnetic coupling should therefore not be a severe problem when mechanistic effects are investigated. Nevertheless, this still represents an approximation and an estimate of the effect would be useful. The only manganese systems for which convergence to antiferromagnetic coupling was achieved so far with the present programs was for an  $\text{Mn}_2(\text{IV},\text{IV})$  and an  $\text{Mn}_2(\text{V},\text{IV})$  complex with

hydroxyl and water ligands. The two manganese centers are coupled by two  $\mu$ -oxo bridges for both systems and they are neutral. The energy splitting between parallel and antiparallel couplings of the manganese spins was found to be 2.2 kcal/mol for the  $\text{Mn}_2(\text{IV},\text{IV})$  complex and 1.8 kcal/mol for the  $\text{Mn}_2(\text{V},\text{IV})$  complex. This leads to predicted  $J$ -values of 0.49 and 0.60 kcal/mol, respectively.<sup>35</sup> Using these  $J$ -values the spin-corrected splittings between ferro- and antiferromagnetic coupling can be estimated to be 2.9 and 2.4 kcal/mol, respectively. From the calculations on the two complexes the antiferromagnetic effect on an O—H bond strength of an hydroxyl ligand can also be obtained, and was found to be a lowering of the bond strength by 0.5 kcal/mol. This latter value should give some idea of the size of the relative energetic effects on mechanisms from antiferromagnetic couplings. The splittings of 2.4 to 2.9 kcal/mol gives an estimate of the total energy effects, which thus largely cancel for the relative energies. The effects on the geometries and on the spinpopulations, which are good chemical indicators, were found to be within the convergence thresholds for both systems.

Convergence to antiferromagnetic coupling has also recently been achieved for a similar  $\text{Fe}_2$  complex.<sup>34</sup> This was for a model of compound **Q** of MMO, which has two  $\text{Fe}(\text{IV})$  centers bridged by two  $\mu$ -oxo bonds. The spins on the individual irons are low-spin coupled to  $S = 1$ . Also, in this case, the effect of antiferromagnetic coupling was found to be very small, actually smaller than for the manganese complexes. The conclusion from these few examples is that the effect of antiferromagnetic coupling is not expected to have significant effects on the chemistry and on the mechanisms in these types of complexes. Recent calculations on systems involving iron–sulfur bonds and on tightly bound iron-dimer complexes with  $S = 2$ , indicate that the antiferromagnetic effects are somewhat larger, and may not be negligible in those cases. Clearly, more examples are needed to draw more general conclusions.

### Conclusions

A recently suggested catalytic cycle for manganese catalase has been used to test the adequacy of different models, both chemical and computational, for the treatment of mechanisms of metal-catalyzed enzyme reactions. Previous experience has shown that the hybrid B3LYP method can usually provide relative energies for reactions of this

type with typical errors of 3 kcal/mol, even though larger errors can not be excluded. It is, therefore, argued that a balanced model should aim at reproducing effects of this size. A model that has a higher accuracy is obviously no drawback unless it will be so expensive to use that it will severely hamper the exploration of all the mechanistic possibilities.

In previous investigations of metal-enzyme mechanisms, a model has been arrived at where normally 30–50 atoms are treated using the B3LYP method. Geometries have been obtained using small double zeta basis sets while the final energies have been evaluated with the addition of polarization and diffuse functions. Effects of the protein surrounding have been estimated using dielectric cavity methods. In the present study it has been shown that it is normally not a worthwhile improvement of the accuracy to extend the basis set, neither by adding polarization for the geometry optimization nor by adding additional polarization for the final energy evaluation. Dielectric effects have furthermore been shown to usually be rather small when the charge of the system does not change, and can therefore be treated with rather crude dielectric cavity methods even though the precise dielectric constant is not well defined. To go in the other direction and make the treatment still more approximate does not appear to be recommendable either. For example, removing polarization functions for the final energy reduces the accuracy significantly with only a small gain of efficiency. The improvement from extensions of the size of the model significantly beyond 50 atoms is a more complicated issue, and must to some extent be model dependent. However, the stability of the results on sometimes rather large variations of the models demonstrated in the present study, will at least give some support for models that only take the most important part of the enzyme active site into account. The main problem when only the first sphere ligands around the metals are included in the model will appear when charged groups are present in the immediate surrounding in the second sphere, and special care therefore has to be devoted to these situations. Ideally, if it can be afforded, these charged groups should be included in the model as well.

## References

- Becke, A. D. *Phys Rev* 1988, A38, 3098; Becke, A. D. *J Chem Phys* 1993, 98, 1372; Becke, A. D. *J Chem Phys* 1993, 98, 5648.
- Siegbahn, P. E. M.; Blomberg, M. R. A. *Chem Rev* 2000, 100, 421.
- Bauschlicher, C. W., Jr.; Ricca, A.; Partridge, H.; Langhoff, S. R. In *Recent Advances in Density Functional Methods, Part II*; Chong, D. P., Ed.; World Scientific Publishing Company: Singapore, 1997; p. 165.
- Dunietz, B. D.; Beachy, M. D.; Cao, Y.; Whittington, D. A.; Lippard, S. J.; Friesner, R. A. *J Am Chem Soc* 2000, 122, 2828.
- Amara, P.; Volbeda, A.; Fontecilla-Camps, J. C.; Field, M. J. *J Am Chem Soc* 1999, 121, 4468.
- Rothlisberger, U.; Carloni, P.; Doclo, K.; Parinello, M. *J Biol Inorg Chem* 2000, 5, 236.
- Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Gill, P. M. W.; Johnson, B. G.; Robb, M. A.; Cheeseman, J. R.; Keith, T.; Petersson, G. A.; Montgomery, J. A.; Raghavachari, K.; Al-Laham, M. A.; Zakrzewski, V. G.; Ortiz, J. V.; Foresman, J. B.; Cioslowski, J.; Stefanov, B. B.; Nanayakkara, A.; Challacombe, M.; Peng, C. Y.; Ayala, P. Y.; Chen, W.; Wong, M. W.; Andres, J. L.; Replogle, E. S.; Gomperts, R.; Martin, R. L.; Fox, D. J.; Binkley, J. S.; Defrees, D. J.; Baker, J.; Stewart, J. P.; Head-Gordon, M.; Gonzalez, C.; Pople, J. A. *Gaussian 94, Revision B.2*; Gaussian, Inc.: Pittsburgh, PA, 1995.
- Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Zakrzewski, V. G.; Montgomery, J. A., Jr.; Stratmann, R. E.; Burant, J. C.; Dapprich, S.; Millan, J. M.; Daniels, A. D.; Kudin, K. N.; Strain, M. C.; Farkas, O.; Tomasi, J.; Barone, V.; Cossi, M.; Cammi, R.; Mennucci, B.; Pomelli, C.; Adamo, C.; Clifford, S.; Ochterski, J.; Petersson, G. A.; Ayala, P. Y.; Cui, Q.; Morokuma, K.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Cioslowski, J.; Ortiz, J. V.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Gomperts, R.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Gonzalez, C.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Andres, J. L.; Head-Gordon, M.; Replogle, E. S.; Pople, J. A.; *Gaussian 98*; Gaussian, Inc.: Pittsburgh, PA, 1998.
- Siegbahn, P. E. M. *Theoret Chem Acc* 2001, 105, 197.
- Siegbahn, P. E. M.; Blomberg, M. R. A. *Annu Rev Phys Chem* 1999, 50, 221.
- Stevens, P. J.; Devlin, F. J.; Chabowski, C. F.; Frisch, M. J. *J Phys Chem* 1994, 98, 11623.
- Hay, P. J.; Wadt, W. R. *J Chem Phys* 1985, 82, 299.
- Wiberg, K. B.; Rablen, P. R.; Rush, D. J.; Keith, T. A. *J Am Chem Soc* 1995, 117, 4261; Wiberg, K. B.; Keith, T. A.; Frisch, M. J.; Murcko, M. *J Phys Chem* 1995, 99, 9072.
- Voet, D.; Voet, J. G. *Biochemistry*; J. Wiley & Sons, Inc.: New York, 1995.
- Barynin, V. V.; Grebenko, A. I. *Dokl Akad Nauk SSSR* 1986, 286, 461.
- Barynin, V. V.; Vagin, A. A.; Melik-Adamyan, W. R.; Grebenko, A. I.; Khangulov, S. V.; Popov, A. N.; Andrianova, M. E.; Vainshtein, B. K. *Sov Phys Dokl* 1986, 31, 457.
- Allgood, G. S.; Perry, J. J. *J Bacteriol* 1986, 168, 563.
- Kono, Y.; Fridovich, I. *J Biol Chem* 1983, 258, 6015.
- Barynin, V. V.; Hempstead, P. D.; Vagin, A. A.; Antonyuk, S. V.; Melik-Adamyan, W. R.; Lamzin, V. S.; Harrison, P. M.; Artymiuk, P. J. *J Inorg Biochem* 1997, 67, 196.
- Barynin, V. V.; Harrison, P. M.; Artymiuk, P. J.; Antonyuk, S. V.; Lamzin, V. S.; Whittaker, M. M.; Whittaker, J. W., to be published.

21. Dismukes, G. C. *Chem Rev* 1996, 96, 2909.
22. Whittaker, M. M.; Barynin, V. V.; Antonyuk, S. V.; Whittaker, J. W. *Biochemistry* 1999, 38, 9126.
23. Meier, A. E.; Whittaker, M. M.; Whittaker, J. W. *Biochemistry* 1996, 35, 348.
24. Boelrijk, A. E. M.; Khangulov, S. V.; Dismukes, G. C. *Inorg Chem* 2000, 39, 3009; Abe, K.; Boelrijk, A. E. M.; Abe, Y.; Belle, C.; Pierre, J.-L.; Dismukes, G. C. *Inorg Chem* 2000, to appear.
25. Siegbahn, P. E. M. *J Inorg Chem* 1999, 38, 2880.
26. Siegbahn, P. E. M. *J Inorg Chem* 2000, 39, 2923.
27. Himo, F.; Eriksson, L. A.; Maseras, F.; Siegbahn, P. E. M. *J Am Chem Soc* 2000, 122, 8031.
28. Wirstam, M.; Siegbahn, P. E. M. *J Am Chem Soc* 2000, 122, 8539.
29. Blomberg, M. R. A.; Siegbahn, P. E. M.; Babcock, G. T.; Wikström, M. *J Inorg Biochem* 2000, 80, 261.
30. Siegbahn, P. E. M. In *Adv. Chem. Phys.*; Prigogine, I.; Rice, S. A., Eds.; J. Wiley: New York, 1996; p. 333, Vol. XCIII.
31. Siegbahn, P. E. M.; Crabtree, R. H. *J Am Chem Soc* 1997, 119, 3103.
32. Lind, T.; Siegbahn, P. E. M.; Crabtree, R. H. *J Phys Chem* 1999, 103, 1193.
33. Blomberg, M. R. A.; Siegbahn, P. E. M.; Babcock, G. T. *J Am Chem Soc* 1998, 120, 8812.
34. Siegbahn, P. E. M. *J Inorg Biochem* 2001, 6, 27.
35. Noodleman, L.; Case, D. A. *Adv Inorg Chem* 1992, 38, 423; Noodleman, L.; Li, J.; Zhao, X.-G.; Richardson, W. H. In *Methods in Chemistry and Materials Science*; Springborg, M., Ed.; John Wiley & Sons: New York, 1997; p. 149.