Overview

The quantum chemical cluster approach for modeling enzyme reactions

Per E.M. Siegbahn[1]* and Fahmi Himo[2]*

This Overview describes the general concepts behind the quantum chemical cluster approach for modeling enzyme active sites and reaction mechanisms. First, the underlying density functional electronic structure method is briefly recapitulated. The cluster methodology is then discussed, including the important observation on the convergence of the solvation effects. The concepts are illustrated using examples from recent applications, such as the discrimination between different reaction mechanisms in phosphotriesterase, the elucidation of origins of regioselectivity in the epoxide-opening reaction of haloalcohol dehalogenase, and finally the use of the cluster methodology to establish the detailed structure of the oxygen-evolving complex in photosystem II. © 2011 John Wiley & Sons, Ltd. *WIREs Comput Mol Sci* 2011 1 323–336 DOI: 10.1002/wcms.13

INTRODUCTION

The basic idea of the cluster approach for modeling enzyme active sites and reaction mechanisms is to cut out a relatively small but well-chosen part of the enzyme and treat it with as accurate quantum chemical methods as possible.[1–8] To account for the excluded rest of the enzyme, two simple approximations are used, namely the polarizable continuum model and a coordinate-locking scheme (see *Methodology* section below). The model should be a good representation of the enzyme and behave and react like the real system.

The cluster approach for modeling enzymes is nearly three decades old. The present short Overview cannot give a full account of the history of this approach, but a few comments as background could be of interest. The early applications used mostly semiempirical or small basis set Hartree–Fock (HF) electronic structure methods.[9–14] The obtained accuracy was judged as insufficient and the general view was that a substantial part of the enzyme was needed to be included in the model in order to produce useful results. A review written in 1993 gives a representative picture of this approach seen at that time.[15] It was summarized that 'such models clearly neglect the effects of the surrounding protein and solvent and the results are difficult to apply to the catalytic process within the enzyme. The energies must be treated with caution'.

A different, more optimistic, view of the cluster approach was obtained at about the same time from the field of modeling homogenous catalysis including transition metals, for a review see Ref 16. At that time only, rather small models of less than 20 atoms could be afforded. Surprisingly, it was found that even these models reproduced rather well many of the experimental features of the potential surface, such as barriers and thermodynamics. Local structures including transition states were furthermore shown to be quite stable, independent of the choice of the model. A high level of geometry optimization was not found to be important either. Instead, it was found that the energetics depended sensitively on the accuracy of the electronic structure method (and basis set used). At an early stage, only highly correlated *ab initio* methods were therefore used,[17] which was in sharp contrast to the work done at the same time for enzymes, see above.

Only a few years after the development of hybrid density functional theory (DFT) in the beginning of the 1990s, it was surprisingly discovered that these methods were almost as accurate, even for transition metal complexes,[18] and allowed much faster

*Correspondence to: ps@physto.se; himo@organ.su.se

[1]Department of Physics, ALBANOVA, and Department of Biochemistry and Biophysics, Arrhenius Laboratory, Stockholm University, Stockholm, Sweden

[2]Department of Organic Chemistry, Arrhenius Laboratory, Stockholm University, Stockholm, Sweden

DOI: 10.1002/wcms.13

calculations and therefore larger models. The first application of the cluster model on an enzyme reaction mechanism using a high accuracy electronic structure method, was done for methane monooxygenase in 1997.[19] Those calculations included similar transition state determinations (for a transition metal complex) as described in the present Overview, but for a much smaller model.

In the past 15 years, the cluster approach for enzymes has gradually developed to become more robust and accurate, a work that is still continuing. For example, different models of the amino acids were tested,[20] and also how second-shell amino acids should be handled.[21] As a result of the accelerating computer power, active site models have become increasingly larger, which has enabled the modeling of evermore complex problems. Today, applications with models larger than 150 atoms are common. Many important mechanistic problems have thus been solved and a great wealth of insight into enzymatic reactions has been gained.[22–34]

This Overview first describes the electronic structure method usually used in the cluster approach. Then, the ideas behind the cluster methodology are reviewed and a systematic study of the size of the active site model is presented. Finally, a number of applications is discussed to illustrate the scope and capabilities of the methodology. The applications described are selected because of their systematic character, and/or as examples of the present limits of the approach.

METHODS

The most useful electronic structure method for treating large molecular systems during the past decades has been the hybrid DFT method with the B3LYP exchange correlation functional, given in Eq. (1)[35]:

$$F^{\mathrm{B3LYP}} = (1-A)\cdot F_{\mathrm{x}}^{\mathrm{Slater}} + A\cdot F_{\mathrm{x}}^{\mathrm{HF}} + B\cdot F_{\mathrm{x}}^{\mathrm{Becke}} + C\cdot F_{\mathrm{c}}^{\mathrm{LYP}} + (1-C)\cdot F_{\mathrm{c}}^{\mathrm{VWN}} \quad (1)$$

where F_x^{Slater} is the Slater exchange, $F_{\mathrm{x}}^{\mathrm{HF}}$ is the exact HF exchange, $F_{\mathrm{x}}^{\mathrm{Becke}}$ is the gradient part of the exchange functional, and $F_{\mathrm{c}}^{\mathrm{LYP}}$ and $F_{\mathrm{c}}^{\mathrm{VWN}}$ describe electron correlation. A, B, and C are coefficients determined using a fit to experimental heats of formation. The most important improvements compared with earlier DFT methods, which were not accurate enough for the present systems, are the inclusion of the gradient of the exchange and the exact exchange contributions. In spite of numerous attempts, it has been difficult to improve the accuracy beyond that of this functional.

There are three major sources of error in DFT methods. The first one is the self-interaction error, which gives an artificial nonzero contribution for the interaction of an electron with itself. In HF theory, the coulomb interaction of an electron by itself is exactly canceled by the exchange term for this electron. In DFT, the coulomb interaction is also described exactly, but exchange is only approximated. Therefore, these terms do not exactly cancel in DFT. The second error is due to the inherent single determinant description. The most severe limitation of the single determinant approximation is that it will dissociate a bond, like the one in H_2, not only into neutral fragments, as it should, but into a mixture of neutral and ionic fragments if there is a spin-restriction on the orbitals. In HF theory this can be corrected by using more than one determinant, which is not yet possible for DFT. Instead, a spin-unrestricted formalism is used, which can reasonably well describe dissociation of single bonds,[36] but not of double bonds. It has been found that the error due to the single determinant description and the error due to self-interaction to a large extent cancel in most cases. In fact, it can be argued that the functionals are optimized for this cancellation. Therefore, removing one of the errors normally leads to larger total errors. The third error in DFT is the lack of van der Waals interaction, which can lead to significant errors when large systems interact, such as for the binding between two porphyrin molecules.

One of the most important developments during the past years of DFT applications has been that approaches have been obtained by which one can get a reasonable idea of the accuracy of the results without having to go to a direct comparison to known experimental results. Such comparisons are often not possible due to a lack of accurate experiments. It has been realized that the most sensitive parameter in the hybrid DFT method is the amount of exact exchange in the functional. In B3LYP this amount is 20%. For transition metal-containing systems, it has quite generally been found that using 15% exact exchange (B3LYP*) instead represents an improvement for redox energies.[37] Other properties are not significantly modified by this change. Moreover, by varying the amount of exact exchange a general empirical rule has been found.[38] The finding is that if the result does not change when the amount of exact exchange is decreased from 20% to 15%, the method is reliable. No exceptions have been found so far. Of course, the chemical model also has to be reliable to produce accurate results.

The first significant improvement of the hybrid DFT method at a more fundamental level has

been presented rather recently.[39,40] In this approach, termed double hybrid DFT, the errors of DFT mentioned above, are directly addressed. The lack of van der Waals effects is corrected by adding an empirical attractive long-range dispersion correction to the functional. The more difficult problem, the lack of static correlation in DFT, is corrected by also performing a multideterminantal Moller-Plesset second order perturbation theory (MP2) calculation. By these changes, the double hybrid DFT is significantly better than B3LYP for the large general benchmark tests that contain first and second row atoms. For transition metal complexes, the results are unfortunately not yet very useful in most cases. The problem is the MP2 correction for static correlation, which is often largely overestimated.

On the contrary, the explicit correction for van der Waals effects appears to be quite useful also for transition metals and can be added without adding the MP2 correction. The prime example is the energy differences between the peroxo and bis-μ-oxo isomers of copper dimer complexes, which appear to show big B3LYP errors of 10–15 kcal/mol.[41,42] In contrast, B3LYP* combined with the van der Waals correction reproduces the experimental energy differences quite well. Another example is the binding of CO and NO to heme complexes, where the B3LYP binding energies have been demonstrated to be underestimated by as much as 10–15 kcal/mol.[43] The dispersion correction computed from the empirical formula is in the order of 10 kcal/mol and therefore improves the results significantly. An effect of the same size has been found for the binding of methyl to cobalamin, where B3LYP without this correction underestimates the binding by 10 kcal/mol.[44] It should be emphasized that for reaction profiles, apart from the binding step, van der Waals effects give only very small contributions.

METHODOLOGY

As mentioned above, in the cluster methodology, a relatively small but well-chosen part of the enzyme is cut out and treated with the quantum chemical methods described in the previous section. Today, the size of the active site models is typically in the order of 100–150 atoms. The protein surrounding that is not explicitly included, can affect the model in two main ways. First, the protein matrix can impose steric constraints on the various parts of the model. If not taken into account, this might lead to large artificial movements of the various groups, which can result in an incorrect description of the reactions, especially features like selectivity. Second, the surrounding can provide long-range polarization that can affect the computed energies. To account for these two effects, two basic approximations are used.

To model the steric effects, a rather simple method is to lock certain key coordinates at the periphery of the model, typically where the truncation is made. This way, large artificial movements of the active site groups are prevented from taking place during the geometry optimizations and the overall structure is kept close to the experimental one. If the model is too small and/or the coordinate-locking scheme is applied very close to the reacting parts, this method can lead to a too rigid model of the active site, and hence wrong energetics can be obtained. However, as the model size grows and more flexibility is granted to the groups of the model, this approximation works better and better.

To model the electrostatic effects, polarizable continuum techniques are usually used. This assumes that the surrounding is a homogenous polarizable medium with some assumed dielectric constant. The solvation energies are usually obtained from single-point calculations performed on the optimized geometries. The choice of this constant is somewhat arbitrary, but $\varepsilon = 4$ is usually considered to be a good representation of protein surrounding.[1–8] Although absolute solvation energies could be large, solvation effects, i.e., relative solvation between stationary points, are much smaller. Also here, it is easy to realize that as the model size grows, relative solvation effects will decrease because more groups that provide polarization are already explicitly included in the model. In fact, one of the important results in the last couple of years is that the solvation effects decrease very quickly with the model size. Systematic studies have shown that when the model reaches a size of ca 150–200 atoms, the solvation effects saturate.[45–47] This important observation means that at certain model size, addition of polarizable continuum makes no difference for relative energies, and as a result of this, the particular choice of dielectric constant becomes irrelevant.

Here, one example of these recent systematic investigations will be presented, namely the study of the methyl transfer reaction in histone lysine methyltransferase (HKMT) SET7/9 enzyme.[47] This enzyme catalyzes the methylation of the N-terminal histone tail of the chromatin structure and uses the S-adenosylmethionine (SAM) cofactor as a methylating agent in this reaction (Figure 1).[48,49] Models of increasing size were designed, as displayed in Figure 1. Model I consists of only truncated models of SAM and the substrate, whereas Model II contains additionally two important tyrosine residues, Tyr245 and

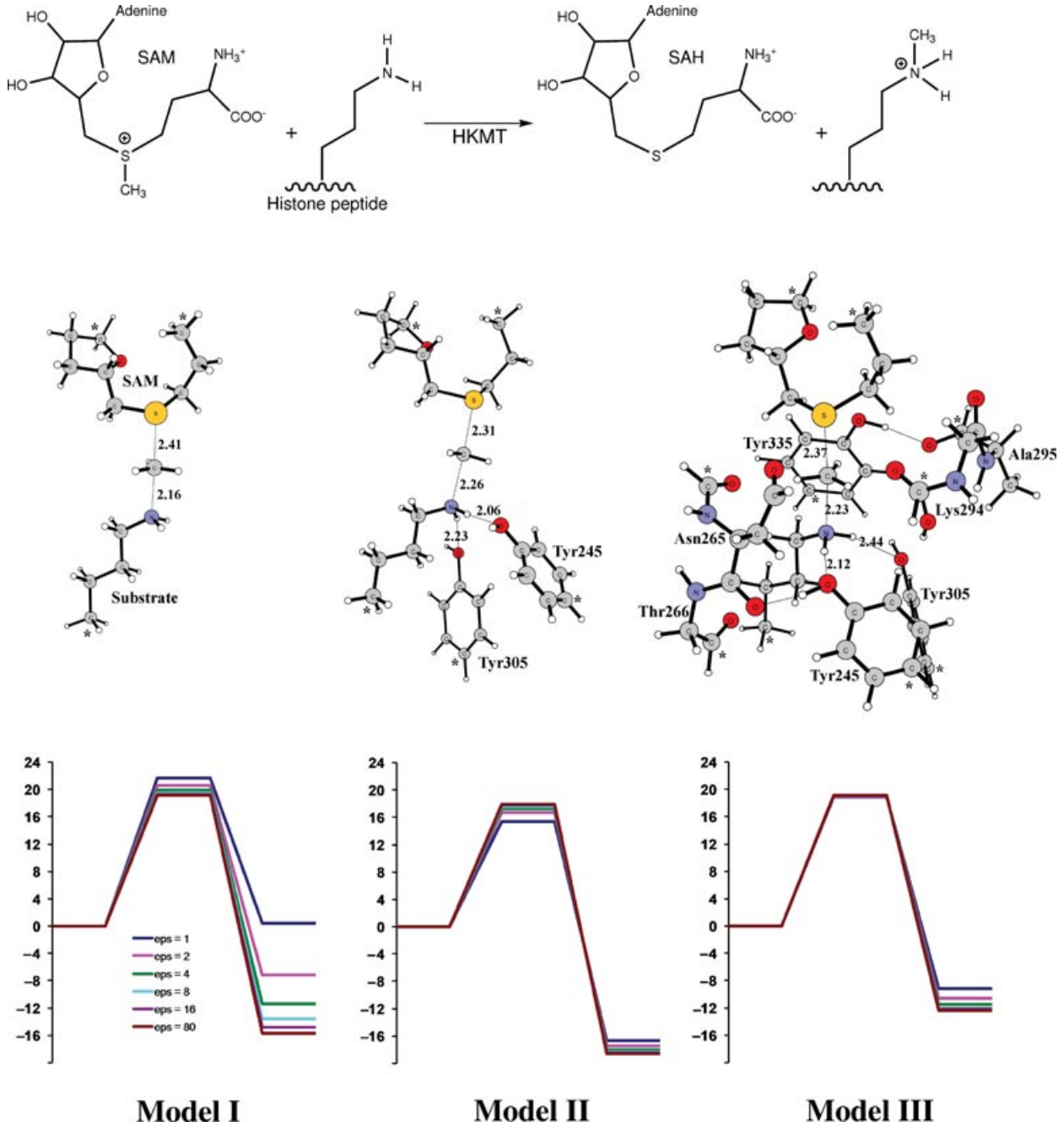

FIGURE 1 | Top: Reaction catalyzed by HKMT SET7/9. Middle: Optimized methyl transfer transition states for the three models of HKMT. Bottom: The resulting potential energy profile using different dielectric constants. Stars indicate fixed centers. HKMT, histone lysine methyltransferase.

Tyr305, that form hydrogen bonds with the amino group of the substrate. These are modeled by phenol rings. Model III contains additional groups that form a ring around the substrate and interact with the transferred methyl group. These are parts of the amino acids Tyr335, Asn265, Thr266, Lys294, and Ala295, and also a crystallographic water molecule (see Figure 1). The three models thus consist of 46, 72, and 132 atoms, respectively, and have a total overall charge of +1.

Reactant, transition state, and product structures were optimized for each model at the B3LYP/6–31G(d,p) level and energies were calculated using the larger basis set 6–311+G(2d,2p). Solvation effects were calculated as single-points at the same level of theory as the geometry optimization using a number

of dielectric constants, $\varepsilon = 2, 4, 8, 16, 80$. The resulting energy profiles for the three models are also shown in Figure 1.

Several important conclusions can be drawn from these results. First, we note that the local geometry of the transition state (in this case the critical S−C and C−N distances) is quite similar between the models. This is a general observation that can be exploited practically. According to this, the tedious work of locating the transition states for the reactions can be performed using a computationally less demanding small model, and the geometric information can then be used to quite quickly find the corresponding transition states for the much larger models.

Another interesting observation is that the calculated energies are not very different in the different models. In particular, the barriers of all models are quite close to the experimental barrier of 20.9 kcal/mol. The fact that the calculated barrier of Model II is slightly lower than the ones calculated for Models I and III, and is hence in less good agreement with the experimental barrier, indicates that the effect of the tyrosines is somewhat exaggerated when they are included alone.

However, the most striking observation seen from Figure 1 is that the solvation effects saturate very quickly, and at a model size of 132 atoms (Model III), they almost completely vanish. Similar observations were also made for two other important and challenging types of reactions, namely ion-pair formation in 4-oxalocrotonate tautomerase[45] and chloride release in haloalcohol dehalogenase (HheC).[46] This seems thus to be a general feature of the cluster approach. A corollary to this result is of course that the particular choice of the dielectric constant becomes less important as the cluster size increases.

Here it is very important to emphasize that geometry optimization is a fundamental element of the cluster approach. Taking coordinates from X-ray structures and using them to evaluate energies without optimizing the geometries will for most applications lead to wrong energy profiles.

A more subtle consequence of the geometry optimization is that the optimized gas phase geometries will adjust to, in part, compensate for the lack of surrounding residues. That is, when only a small part of the enzyme is included explicitly in the model, the internal electrostatic effects of the model will be somewhat exaggerated and thus the local geometry, such as for hydrogen bonds, will be somewhat changed as a result of this. When additional groups are added, this effect will be smaller. Note, for example, how the two hydrogen bonds between the substrate and the tyrosines (Figure 1) become slightly longer when more groups are added in going from Model II to Model III of HKMT. This effect will thus, at least in part, lead to the fast convergence observed when geometries are optimized, in contrast to the slow convergence observed when the geometries are not optimized.[50,51]

It has furthermore been argued that although the solvation effects converge, the relative energies do not, and that charged groups that are as far away as 20 Å from the active site can affect the energies significantly.[50] However, as seen from the example above, and also from the other systematic studies performed thus far,[45,46] the energies obtained when the models become sufficiently large get quite close to each other, and more importantly close to the experimental energies, within a couple of kilocalories per mole, as dictated by the accuracy of the underlying DFT methods.

The combination of coordinate-locking scheme and continuum solvation thus provides a simple but very fruitful way to account for the parts of the enzyme that are not included in the model. The calculated energies have proven to be sufficiently accurate to be able to substantiate or reject various mechanistic proposals. To do this, usually a large number of calculations are needed. The cluster approach is robust and fast enough to allow this kind of studies.

One final comment about the methodology concerns entropy effects. The coordinate-locking procedure described above leads to a number of small imaginary frequencies, typically in the order of $<30i\ \mathrm{cm}^{-1}$. These do not contribute to the energetics in any significant way and can simply be ignored. However, they render the calculation of the harmonic entropy effects inaccurate. Therefore, in most applications using the cluster approach, the reported energies correspond to enthalpies and not free energies. Entropy effects are usually quite small, except in some very specific cases, and do not change the mechanistic conclusions. A quantum mechanics/molecular mechanics (QM/MM) study on the HKMT reaction discussed above found that the free energy barriers differ by only ca 1 kcal/mol from the potential energy barriers.[52,53]

APPLICATIONS

In this section, we will discuss some selected applications of the cluster approach. The examples are chosen to give an idea about the scope of this methodology. First, a quite common scenario is shown in which the methodology helps to distinguish between two mechanistic proposals. The second example shows an

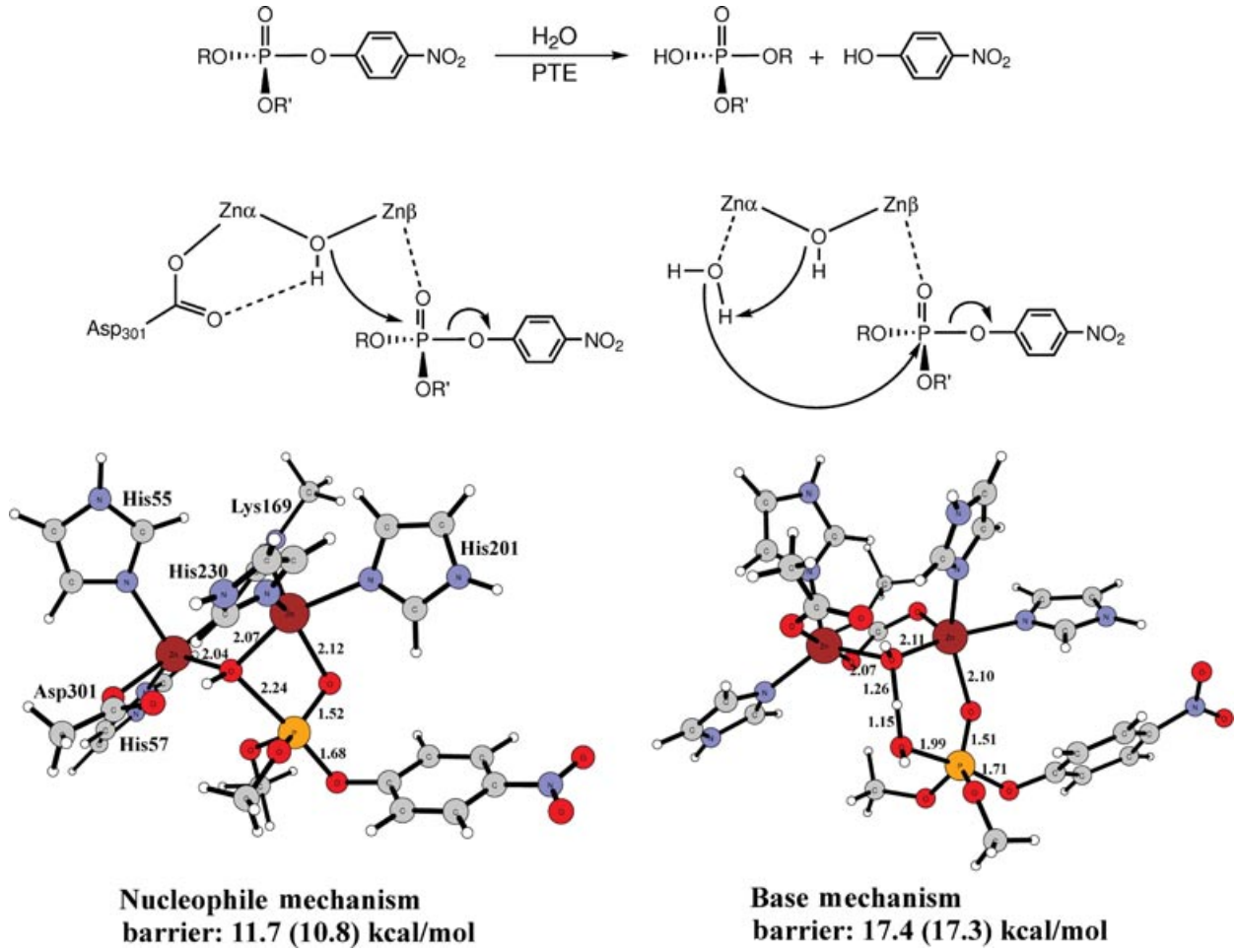

FIGURE 2 | Phosphotriester hydrolysis catalyzed by phosphotriesterase (PTE, top), suggested nucleophile and base mechanisms (middle), and optimized transition state structures for both mechanisms (bottom). In parentheses are the barriers including solvation effects ($\varepsilon = 4$).

application beyond this, namely to determine the factors governing selectivity in the active site. The final example describes one of the most extensive applications of the cluster approach, namely the study of oxygen evolution in photosynthesis. In this Overview, we will focus only on how the cluster approach can be used to elucidate the detailed structure of the manganese cluster of the oxygen-evolving complex (OEC).

Reaction Mechanism of Phosphotriesterase

One of the most common applications of the cluster methodology is to establish enzymatic reaction mechanisms and distinguish between various mechanistic scenarios based on their energetic feasibility. This is indeed one of the most fundamental tasks that the approach must be able to perform. In this respect, this methodology over the past decade has been very successful and helped to determine reaction mechanisms of numerous enzymes, with a wide spectrum of reactivities.[22]

Here, only one brief example will be given to illustrate this issue, namely the study concerning the reaction mechanism of phosphotriesterase (PTE). This bacterial enzyme catalyzes the hydrolysis of a wide variety of phosphotriesters (Figure 2), which are highly toxic compounds used as pesticides and also as chemical warfare agents.[54,55] PTE is a homodimer and each subunit contains an active site that harbors a dinuclear zinc center. The two zinc ions are about 3.4 Å from each other and are liganded by the His55, His57, His201, His230, and Asp301 residues. In addition, a hydroxide and a carboxylated lysine residue, Lys169, bridge the two zinc ions.[56] One important mechanistic question for PTE, and indeed for the whole class of dinuclear zinc enzymes, is whether the bridging hydroxide acts as nucleophile or as a base in the reaction

(see Figure 2). In the first mechanism, the bridging hydroxide attacks the substrate directly to form a pentacoordinated intermediate that subsequently collapses to release the product.[57] In the alternative suggestion, the bridging hydroxide acts as a base, abstracting a proton from a water molecule coordinated to the α-zinc ion, which then performs the nucleophilic attack.[58,59]

Both these possibilities were considered by quantum chemical calculations.[60,61] A model of the active site consisting of 82 atoms was constructed as shown in Figure 2. It contains the two zinc ions along with their first-shell ligands. The histidines were modeled using imidazoles, the aspartate using an acetate, and the carboxylated lysine using a carboxylated methylamine. As a representative substrate, dimethyl 4-nitrophenyl phosphate was used and it was manually placed to bind the more solvent-exposed β-Zn ion. In this model, no atoms needed to be fixed during the geometry optimizations, because all groups are anchored to the bimetal site and no major artificial movements are possible.

On the basis of this model, the stationary points for the two mechanisms can be optimized and their energies can be compared. In Figure 2, the optimized transition states for the initial steps of the two mechanisms are given, along with the calculated barriers. For the nucleophile mechanism, the barrier is calculated to be 11.7 kcal/mol (10.8 kcal/mol when solvation effects are added with $\varepsilon = 4$). The attack is shown to occur directly from bridging position, without the need to first become terminal. The bridging hydroxide is thus sufficiently nucleophilic to perform the attack on the phosphorus center of organophosphate triester substrates.[60]

To investigate the base mechanism, the suggested water nucleophile was added to the model and placed to bind to the α-Zn.[61] However, it was not possible to locate a stable structure in which the water remains bound to the zinc ion. It turns out that a hexacoordinated α-Zn is not likely, as every attempt to optimize structures bound with the water resulted in its dissociation from the metal to form a hydrogen bond to the bridging hydroxide instead. Nevertheless, starting from this structure, the transition state for a base mechanism can be optimized (Figure 2). The barrier is calculated to be 17.4 kcal/mol without solvation and 17.3 kcal/mol including solvation with $\varepsilon = 4$.[61]

The calculations thus give strong support to the nucleophile mechanism as the barrier for this scenario is computed to be considerably lower than that for the base mechanism. These results are furthermore in agreement with the available biochemical, kinetic, and spectroscopic data, and also with recent crystal structures obtained for the product state.[61]

Selectivity in Haloalcohol Dehalogenase

The second example given here concerns the use of the cluster methodology to answer questions beyond reaction mechanisms, namely the origins of selectivity in enzyme active sites. This is a more challenging task that has become possible to tackle only recently. This is mainly because the size of the active site models has grown to become sufficiently large to contain groups that are responsible for binding and positioning of the substrate(s).

To reproduce regio- and enantioselectivity, one typically has to reproduce energy differences in the order of 1–2 kcal/mol. This might seem impossible to achieve with the cluster approach, considering the approximations made and also the accuracy of the underlying quantum chemical methods. However, in these kinds of studies, one usually considers structures that are very similar except for a few parameters. This means that there is typically a large degree of systematic error cancellation, and higher accuracy can thus be achieved. This kind of methodology has already proven to be a very successful tool in the field of asymmetric homogenous catalysis.

The specific example discussed here is concerned with the regioselectivity of the epoxide opening reaction in HheC. This enzyme catalyzes the reversible dehalogenation of vicinal haloalcohols to form epoxides.[62–64] In the reverse direction, it can also use nonhalide nucleophiles, such as N_3^-, CN^-, and NO_2^-, to catalyze the irreversible formation of substituted alcohols (Figure 3).[65–67] The regioselectivity of this reaction is quite high, and in some cases, it is the opposite of the selectivity observed in the solution reaction. For instance, in the case of azide nucleophile and styrene oxide substrate, the attack is to 97% at the benzylic α-position in the nonenzymatic case, whereas in the enzymatic reaction it is to 79% at the terminal β-position.[66]

Both the dehalogenation and the epoxide opening reactions have been investigated using the cluster methodology described above.[46,68] We will here focus only on the selectivity aspect to examine whether this methodology also can be used to reproduce and rationalize the observed selectivity switch.

For comparison, the nonenzymatic reaction was first modeled using the substrate, the azide nucleophile, and a phenol group that functions as an acid in the reaction (see Figure 3). The enzymatic reaction was studied using an active site model consisting of 130 atoms as also shown in Figure 3.[68]

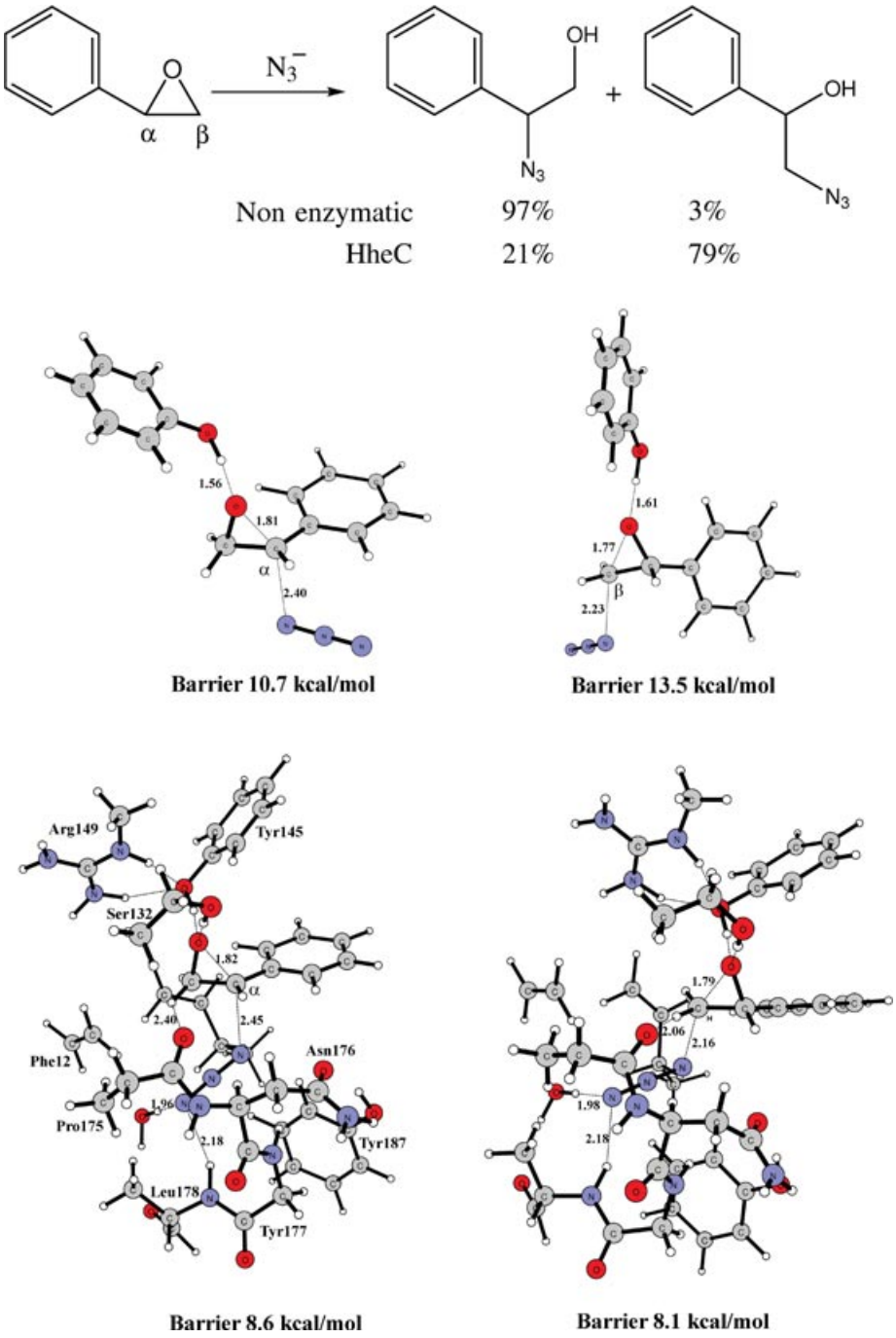

FIGURE 3 | Selectivity of epoxide-opening reaction with and without the haloalcohol dehalogenase (HheC) enzyme (top) and optimized transition state structures azide attack at the α- and β-carbons of styrene oxide in both the nonenzymatic (middle) and enzymatic (bottom) reactions.

The model contains the important groups around the epoxide and the nucleophile, namely the Ser132–Tyr145–Arg149 catalytic triad and the halide-binding site. The catalytic triad helps binding and orienting the epoxide substrate and also provides a proton to the developing alkoxide to form the alcohol product. The halide-binding site consists of several groups (Pro175, Asn176, Tyr177, Leu178, Phe12, Phe186, and Tyr187) and its role is to stabilize the released halide anion in the natural reaction. In the reverse

TABLE 1 | Calculated Barriers (kcal/mol) for Azidolysis of Styrene Oxide. Experimentally-Observed Ratios are Given in Parentheses

Model	Barrier (Cα)	Barrier (Cβ)	Cα–Cβ
Nonenzymatic reaction	10.7	13.5	−2.8 (97:3)
Enzyme Model	8.6	8.1	+0.5 (21:79)
Mutation: Ser132	13.1	13.8	−0.7
Mutation: Leu178-Peptide Bond	10.4	10.5	−0.1
Mutation: Pro175-Peptide Bond	11.2	11.5	−0.3
Mutation: Tyr187	9.2	11.6	−2.4

epoxide-opening reaction, it is harboring the nucleophile. These groups were not included in their entirety but rather using the functional groups or the parts that actually form the site (see Figure 3). The solvation effects were added using the standard $\varepsilon = 4$ for the enzyme model and $\varepsilon = 80$ for the solution model.

The transition states for both the α- and β-attacks were optimized for both cases and are displayed in Figure 3. For the nonenzymatic reaction, the barriers are calculated to be 10.7 and 13.5 kcal/mol for attacks at the Cα and Cβ positions, respectively. The calculated difference of 2.8 kcal/mol in favor of Cα is in excellent agreement with the experimental difference of 2.1 kcal/mol, estimated from the observed 97% α-preference. For the enzyme model, the calculated barriers are 8.6 and 8.1 kcal/mol for attacks at Cα and Cβ positions, respectively. Also, here the calculated difference of 0.5 kcal/mol in favor of Cβ is in excellent agreement with the experimental difference of 0.8 kcal/mol, estimated from the 79% preference for β-attack. The active site model can thus reproduce the experimentally-observed regioselectivity switch with very high accuracy.

A natural question is then what specifically causes this regioselectivity in the enzyme. To address this, one can perform a number of *in silico* mutations in which selected active site features are altered and the barriers are recalculated. Some of these alterations could be very hard or even impossible to perform experimentally. In that sense, calculations can be advantageous and provide information not easily accessible otherwise.

Examination of the transition state geometries gives hints as to which groups could be involved in governing the selectivity. As seen from Figure 3, the active site binds the nucleophile and the substrate in such a way that the nucleophile is closer to the β-carbon. This is achieved by several hydrogen bonds, and mutation of these groups can give quantitative assessment of their importance for the selectivity. For example, the hydroxyl group of Ser132 can be replaced by a hydrogen atom, corresponding to its mutation to an alanine residue. This mutation leads to a significant raise of both barriers (Table 1). The β-selectivity is reduced by 1.2 kcal/mol. Similarly, removal of the hydrogen bond formed from the Leu178 backbone peptide bond to the azide nucleophile (by changing the peptide bond into an ester bond) leads to reduced β-selectivity by 0.6 kcal/mol. A group that was observed to electrostatically stabilize the β transition state but not the α transition state is the carbonyl group peptide bond of Pro175. Replacing this carbonyl by a methylene resulted in a decreased β-selectivity by 0.8 kcal/mol. Finally, a modification that turned out to have major impact on the selectivity was the removal of the side chain of Tyr187 (effectively mutating it into a glycine residue). This mutation results in α-selectivity that is very similar to the nonenzymatic case. By removing this group, the nucleophile is sterically less hindered from reaching the α-carbon and the selectivity is thus greatly affected. Very similar results were obtained for all the mutations when CN^- was used as a nucleophile in the calculations.[68]

Thus, the cluster model of the HheC active site not only helped to establish the mechanism and reproduced the regioselectivity of the reaction quite accurately but it also proved very useful in dissecting the various contributions. This kind of detailed analysis of active site models will no doubt continue and become increasingly useful as the models get larger.

Structure of Oxygen-Evolving Complex in Photosystem II

One of the largest projects using the cluster model so far has been the attempt to understand water oxidation in photosynthesis in plants. It is not possible to describe this project in total here, and only one aspect will therefore be discussed, namely the structure of the OEC. The OEC contains four manganese and one calcium atom. X-ray diffraction studies during the

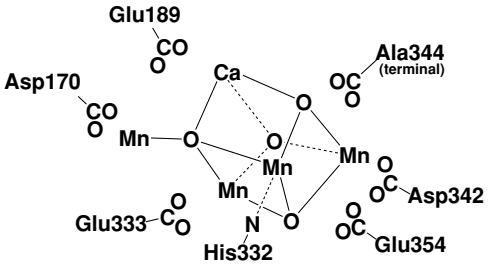

FIGURE 4 | Simplified schematic picture of the structure of the oxygen evolving complex, suggested by X-ray crystallography.

past few years have considerably clarified the detailed structure of the OEC.[69–71] These studies showed that three of the manganese and the calcium atoms form a cuboidal structure, with the fourth manganese situated outside the cube, see Figure 4. The amino acids most likely to be ligated to the complex were also assigned, and are also shown in the figure. The resolution has been rather low, 2.9–3.5 Å, and many key questions still remain, such as the exact ligation pattern and the positions of the metal atoms. The task is to try to improve this structure by using the quantum chemical cluster approach.

The earliest cluster studies of water oxidation focused on general principles for O-O bond formation, using rather small and approximate models. These studies all led to the conclusion that an oxygen radical is needed for a low barrier formation of O_2.[72] An important step forward was taken when essentially all possibilities were tested for Model A in Figure 5. It was surprisingly found that the lowest barrier occurs when the oxygen radical reacts with a bridging μ-oxo ligand.[73] This type of transition state (TS) was therefore the starting point for obtaining a better structure. The next step is to try to fit this TS into the X-ray structures in some way.

In order to proceed, two important assumptions were made. It was concluded that the most accurate positions determined by the X-ray analysis are the positions of the backbone atoms. Their positions are, for example, rather close in the two X-ray structures even though the OEC metal complexes are quite different.[69,70] The first important assumption made was therefore that these positions are sufficiently well described by the low-resolution X-ray structures, and that a further optimization of the structure is meaningful. To ensure that the optimized models should stay as closely as possible to the X-ray structures, the backbone atoms were therefore fixed at the positions given by the X-ray structure.[69] This meant that 19 atoms, of a total of 128 in Model B in Figure 5, were fixed, whereas the rest were optimized. With this approach, a search for a better structure of the OEC was initiated. Still, there are a very large number of possibilities, so only the chemically most reasonable ones were fully investigated. The second important assumption made was based on experience during the past decade. It was assumed that the lowest energy structure obtained in this way, with the backbone constraints, is the one that is adopted by nature. Against this assumption, it can be argued that in principle it is possible that a higher energy structure is used in the enzyme if the barriers for its decay are high enough, but this is considered as very unlikely, see further Ref 42.

It turned out that it was relatively easy to fit the TS structure, obtained in the earlier studies, into the backbone structure given by the X-ray analysis.[74] In fact, this protein-fitted structure gives an even lower barrier than before for O-O bond formation. However, to obtain a more direct comparison to the measured electron densities and suggested X-ray structures, a structure for the resting state is needed. Electrons and protons were therefore added to the TS structure and the geometries were reoptimized until the resting state was reached. This structure, for the largest model used so far with about 190 atoms, is shown as Model C in Figure 5, illustrating the size of the model and the amino acids included in the model.

The optimized structure for the cluster model of the resting state has been placed into the electron density from the X-ray density measurement,[74] and it was shown that the structure fits very well into this density. It should again be emphasized that the structure was not fitted to the density but was optimized only with the constraints of fixed backbone atoms. There are quite significant differences between the optimized metal positions and those obtained from the X-ray analysis, with deviations of up to 1.8 Å. This illustrates the problem with the low resolution, because all three complexes fit well into the density. The optimal cluster structure has recently been compared with structures obtained in different ways, for example from QM/MM, and was shown to be significantly better.[75]

The DFT structure agrees in many aspects with structures suggested by extended X-ray absorption fine-structure (EXAFS) studies,[76–78] in which it had been emphasized that substantial discrepancies exist between EXAFS and X-ray structures. One characteristic feature of the suggested EXAFS structures is that there should be a short Mn–Mn distance between the outer manganese and the nearest manganese in the

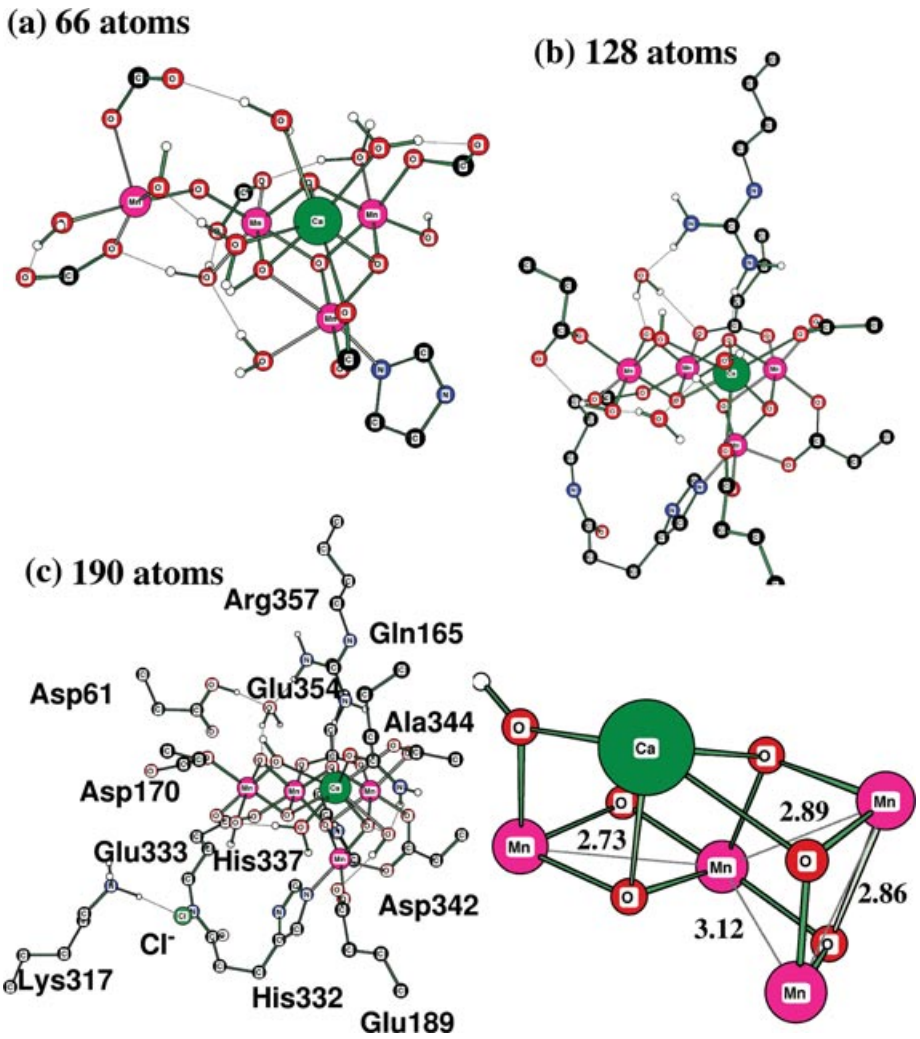

FIGURE 5 | Three different models used to study water oxidation. The structures are for the resting state and have been fully optimized, with backbone constraints for the larger models. For clarity, most amino acid protons have been left out from the figure. A close-up of the metal cluster in Model C is shown at the bottom right.

Mn_3Ca cube. These manganese atoms should therefore be connected by two μ-oxo bonds. This agrees with what is found in the optimized structure, see, the close-up in Figure 5. The number of short Mn–Mn distances in the resting state have been a controversial issue with different suggestions. The best optimized structure at present has Mn–Mn distances of 2.73, 2.86, 2.89, and 3.11 Å. This is in as good agreement with EXAFS suggestions as one can hope for, using DFT. In one EXAFS interpretation,[76] three distances of 2.7–2.8 Å and one of 3.3 Å are suggested, whereas in another one only two short 2.7 Å distances are suggested,[78] illustrating the remaining EXAFS uncertainties. Recently, a comparison has also been made between the calculated solution EXAFS spectrum for the cluster model with the corresponding experimental spectrum showing good agreement.[79]

CONCLUSION

As demonstrated by this Overview, the cluster approach has already proven to be a quite powerful and useful tool in the study of enzymatic reactions. A number of important problems have already been solved and it is not difficult to predict that many more questions will be answered in the future.

All aspects of the cluster model are likely to undergo gradual developments in the future. The electronic structure methods used will be more accurate. For example, van der Waals effects are already starting to be routinely included. At the present stage, errors due to van der Waals effects have often been confused with errors due to near-degeneracy (multideterminantal effects). With the inclusion of van der Waals effects, the understanding of the near

degeneracy effects will therefore automatically increase with more experience. The main tool to calibrate the near degeneracy effects is to vary the amount of exact exchange, and this type of calibration will be more certain.

The size of the models will gradually increase as the limitations of previous models are better understood. Systematic size studies in different cases will be further investigated and the conditions and physical grounds for the cluster convergence will be established.

As the model size increases, more advanced applications will be possible. For example, various selectivities will be able to be rationalized and also predicted. Techniques like the *in silico* mutations shown above will be more common, which in turn is likely to lead to applications in rational computational protein design.

REFERENCES

1. Blomberg MRA, Siegbahn PEM. A quantum chemical approach to the study of reaction mechanisms of redox-active metalloenzymes. *J Phys Chem B* 2001, 105:9375–9386.
2. Himo F, Siegbahn PEM. Quantum chemical studies of radical-containing enzymes. *Chem Rev* 2003, 103:2421–2456.
3. Noodleman L, Lovell T, Han WG, Li J, Himo F. Quantum chemical studies of intermediates and reaction pathways in selected enzymes and catalytic synthetic systems. *Chem Rev* 2004, 104:459–508.
4. Himo F. C-C bond formation and cleavage in radical enzymes, a theoretical perspective. *Biophys Biochim Acta* 2005, 1707:24–33.
5. Himo F. Quantum chemical modeling of enzyme active sites and reaction mechanisms. *Theor Chem Acc* 2006, 116:232–240.
6. Siegbahn PEM, Borowski T. Modeling enzymatic reactions involving transition metals. *Acc Chem Res* 2006, 39:729–738.
7. Ramos MJ, Fernandes PA. Computational enzymatic catalysis. *Acc Chem Res* 2008, 41:689–698.
8. Siegbahn PEM, Himo F. Recent developments of the quantum chemical cluster approach for modeling enzyme reactions. *J Biol Inorg Chem* 2009, 14:643–651.
9. Krauss M, Garmer DR. Active site ionicity and the mechanism of carbonic anhydrase. *J Am Chem Soc* 1991, 113:6426–6435.
10. Daggett V, Schröder S, Kollman P. The catalytic pathway in trypsin: classical and quantum mechanical calculations *J Am Chem Soc* 1991, 113:8926–8935.
11. Turi L, Náray-Szabó G. Computational studies on aspartic proteases. I. Active-site protonation and hydration in the substrate-free crystalline state. *Int J Quant Chem* 1992, 42:1537–1551.
12. Zheng YJ, Merz KM. Mechanism of the human carbonic anhydrase II catalysed hydration of carbon dioxide. *J Am Chem Soc* 1992, 114:10498–10507.
13. Alex A, Clark T. MO-studies of enzyme reaction mechanisms 1. model molecular orbital study of the cleavage of peptides by carboxypeptidase A. *J Comp Chem* 1992, 13:704–717.
14. Wilkie J, Williams IH. Transition-state structural variation in a model for carbonyl reduction by lactate dehydrogenase: computational validation of empirical predictions based upon Albery–More O'Ferrall–Jencks diagrams. *J Am Chem Soc* 1992, 114:5423–5425.
15. Mulholland AJ, Grant GH, Richards WG. Computer modelling of enzyme catalysed reaction mechanisms. *Protein Eng* 1993, 6:133–147.
16. Siegbahn PEM. Electronic structure calculations for molecules containing transition metals. In: Prigogine I, Rice SA, Wiley J, eds. *Advances in Chemical Physics: New Methods in Computational Quantum Mechanics*. Vol. 63. New York: John Wiley & Sons; 1996, 333.
17. Siegbahn PEM. The current status of the MC-CI method as applied to molecules containing transition metal atoms. *Faraday Symp Chem Soc* 1984, 19:97–107.
18. Blomberg MRA, Siegbahn PEM, Svensson M. Comparisons of results from parametrized configuration interaction (PCI-80) and from hybrid density functional theory with experiments for first row transition metal compounds. *J Chem Phys* 1996, 104:9546–9554.
19. Siegbahn PEM, Crabtree RH. The mechanism of C-H activation by di-iron methane monooxygenases: quantum chemical studies. *J Am Chem Soc* 1997, 119:3103–3113.
20. Siegbahn PEM. Modeling aspects of mechanisms for reactions catalyzed by metalloenzymes. *J Comp Chem* 2001, 22:1634–1645.
21. Pelmenschikov V, Blomberg MRA, Siegbahn PEM. A theoretical study of the mechanism for peptide hydrolysis by thermolysin. *J Biol Inorg Chem* 2002, 7:284–298.
22. Hopmann KH, Himo F. Theoretical study of the full reaction mechanism of human soluble epoxide hydrolase. *Chem Eur J* 2006, 12:6898–6909.
23. Kumar D, Hirao H, Shaik S, Kozlowski PM. Proton-shuffle mechanism of O−O activation for formation

of a high-valent oxo-iron species of bleomycin. *J Am Chem Soc* 2006, 128:16148–16158.

24. Hopmann KH, Guo JD, Himo F. Theoretical investigation of the first-shell mechanism of nitrile hydratase. *Inorg Chem* 2007, 46:4850–4856.
25. Aluri S, de Visser SP. The mechanism of cysteine oxygenation by cysteine dioxygenase enzymes. *J Am Chem Soc* 2007, 129:14846–14847.
26. Guell M, Siegbahn PEM. Theoretical study of the catalytic mechanism of catechol oxidase. *J Biol Inorg Chem* 2007, 12:1251–1264.
27. Fee JA, Case DA, Noodleman L. Toward a chemical mechanism of proton pumping by the B-type cytochrome c oxidases: application of density functional theory to cytochrome ba3 of *Thermus thermophilus*. *J Am Chem Soc* 2008, 130:15002–15021.
28. Liao RZ, Yu JG, Raushel FM, Himo F. Theoretical investigation of the reaction mechanism of the dinuclear zinc enzyme dihydroorotase. *Chem Eur J* 2008, 14:4287–4292.
29. Carvalho ATP, Swart M, van Stralen JNP, Fernandes PA, Ramos MJ, Bickelhaupt FM. Mechanism of thioredoxin-catalyzed disulfide reduction. activation of the buried thiol and role of the variable active-site residues. *J Phys Chem B* 2008, 112:2511–2523.
30. Amata O, Marino T, Russo N, Toscano M. Human insulin-degrading enzyme working mechanism. *J Am Chem Soc* 2009, 131:14804–14811.
31. Sevastik R, Himo F. Reaction mechanism of cis-3-chloroacrylic acid dehalogenase: a theoretical study. *Biochemistry* 2009, 48:9641–9649.
32. Yang L, Liao RZ, Yu JG, Liu RZ. DFT study on the mechanism of *Escherichia coli* inorganic pyrophosphatase. *J Phys Chem B* 2009, 113:6505–6510.
33. Roos K, Siegbahn PEM. Density functional theory study of the manganese-containing ribonucleotide reductase from *Chlamydia trachomatis*: why manganese is needed in the active complex. *Biochemistry* 2009, 48:1878–1887.
34. Lill SON, Siegbahn PEM. An autocatalytic mechanism for NiFe-hydrogenase: reduction to Ni(I) followed by oxidative addition. *Biochemistry* 2009, 48:1056–1066.
35. Becke AD. Density-functional thermochemistry. 3. The role of exact exchange. *J Chem Phys* 1993, 98:5648–5652.
36. Siegbahn PEM, Blomberg MRA. Bond-dissociation using hybrid DFT. *Intern J Quantum Chem* 2010, 110:317–322.
37. Reiher M, Salomon O, Hess BA. Reparameterization of hybrid functionals based on energy differences of states of different multiplicity. *Theor Chem Acc* 2001, 107:48–55.
38. Siegbahn PEM. Hybrid density functional study of the oxidized states of NiFe-hydrogenase. *Comptes Rendus Chimie* 2007, 10:766–774.
39. Grimme S. Semiempirical hybrid density functional with perturbative second-order correlation. *J Chem Phys* 2006, 124:034108.
40. Schwabe T, Grimme S. Towards chemical accuracy for the thermodynamics of large molecules: new hybrid density functionals including non-local correlation effects. *Phys Chem Chem Phys* 2007, 9:3397–3406.
41. Lewin JL, Heppner DE, Cramer CJ. Validation of density functional modeling protocols on experimental bis(μ-oxo)/μ-η^2:η^2-peroxo dicopper equilibria. *J Biol Inorg Chem* 2007, 12:1221–1234.
42. Gherman BF, Cramer CJ. Quantum chemical studies of molecules incorporating a $Cu_2O_2^{2+}$ core. *Coord Chem Rev* 2009, 253:723–753.
43. Radon M, Pierloot K. Binding of CO, NO, and O2 to heme by density functional and multireference *ab initio* calculations. *J Phys Chem A* 2008, 112:11824–11832.
44. Jensen KP, Ryde U. Theoretical prediction of the Co-C bond strength in cobalamins. *J Phys Chem A* 2003, 107:7539–7545.
45. Sevastik R, Himo F. Quantum chemical modeling of enzymatic reactions: the case of 4-oxalocrotonate tautomerase. *Bioorg Chem* 2007, 35:444–457.
46. Hopmann KH, Himo F. Quantum chemical modeling of the dehalogenation reaction of haloalcohol dehalogenase. *J Chem Theor Comp* 2008, 4:1129–1137.
47. Georgieva P, Himo F. Quantum chemical modeling of enzymatic reactions: the case of histone lysine methyltransferase. *J Comp Chem* 2010, 31:1707–1714.
48. Wilson JR, Jing C, Walker PA, Martin SR, Howell SA, Blackburn GM, Gamblin SJ, Xiao B. Crystal structure and functional analysis of the histone methyltransferase SET7/9. *Cell* 2002, 111:105–115.
49. Xiao B, Jing C, Wilson JR, Walker PA, Vasisht N, Kelly G, Howell SA, Taylor IA, Blackburn GM, Gamblin SJ. Structure and catalytic mechanism of the human histone methyltransferase SET7/9. *Nature* 2003, 421:652–656.
50. Hu LH, Eliasson J, Heimdal J, Ryde U. Do quantum mechanical energies calculated for small models of protein-active sites converge? *J Phys Chem A* 2009, 113:11793–11800.
51. Sumowski CV, Ochsenfeld C. A convergence study of QM/MM isomerization energies with the selected size of the QM region for peptidic systems. *J Phys Chem A* 2009, 113:11734–11741.
52. Hu P, Zhang Y. Catalytic mechanism and product specificity of the histone lysine methyltransferase SET7/9: an *ab initio* QM/MM-FE study with multiple initial structures. *J Am Chem Soc* 2006, 128:1272–1278.
53. Wang S, Hu P, Zhang Y. *Ab initio* quantum mechanical/molecular mechanical molecular dynamics simulation of enzyme catalysis: the case of histone lysine methyltransferase SET7/9. *J Phys Chem B* 2007, 111:3758–5764.

54. Donarski WJ, Dumas DP, Heitmeyer DP, Lewis VE, Raushel FM. Structure–activity relationships in the hydrolysis of substrates by the phosphotriesterase from *Pseudomonas diminuta*. *Biochemistry* 1989, 28:4650–4655.
55. Caldwell SR, Newcomb JR, Schlecht KA, Raushel FM. Limits of diffusion in the hydrolysis of substrates by the phosphotriesterase from *Pseudomonas diminuta*. *Biochemistry* 1991, 30:7438–7444.
56. Benning MM, Shim H, Raushel FM, Holden HM. High resolution X-ray structures of different metal-substituted forms of phosphotriesterase from *Pseudomonas diminuta*. *Biochemistry* 2001, 40:2712–2722.
57. Aubert SD, Li Y, Raushel FM. Mechanism for the hydrolysis of organophosphates by the bacterial phosphotriesterase. *Biochemistry* 2004, 43:5707–5715.
58. Wong KY, Gao J. The reaction mechanism of paraoxon hydrolysis by phosphotriesterase from combined QM/MM simulations. *Biochemistry* 2007, 46:13352–13369.
59. Jackson CJ, Foo JL, Kim HK, Carr PD, Liu JW, Salem G, Ollis DL. In crystallo capture of a Michaelis complex and product-binding modes of a bacterial phosphotriesterase. *J Mol Biol* 2008, 375:1189–1196.
60. Chen SL, Fang WH, Himo F. Theoretical study of the phosphotriesterase reaction mechanism. *J Phys Chem B* 2007, 111:1253–1255.
61. Kim J, Tsai PC, Chen SL, Himo F, Almo SC, Raushel FM. Structure of diethyl phosphate bound to the binuclear metal center of phosphotriesterase. *Biochemistry* 2008, 47:9497–9504.
62. van den Wijngaard AJ, Reuvekamp PTW, Janssen DB. Purification and characterization of haloalcohol dehalogenase from *Arthrobacter sp*. strain AD2. *J Bacteriol* 1991, 173:124–129.
63. Nakamura T, Nagasawa T, Yu F, Watanabe I, Yamada H. Characterization of a novel enantioselective halohydrin hydrogen-halide-lyase. *App Environ Microbiol* 1994, 60:1297–1301.
64. de Jong RM, Tiesinga JJ, Rozeboom HJ, Kalk KH, Tang L, Janssen DB, Dijkstra BW. Structure and mechanism of a bacterial haloalcohol dehalogenase: a new variation of the short-chain dehydrogenase/reductase fold without an NAD(P)H binding site. *EMBO J* 2003, 22:4933–4944.
65. Hasnaoui G, Lutje Spelberg JH, de Vries E, Tang L, Hauer B, Janssen DB. Nitrite-mediated hydrolysis of epoxides catalyzed by halohydrin dehalogenase from *Agrobacterium radiobacter* AD1: a new tool for the kinetic resolution of epoxides. *Tetrahedron Asym* 2005, 16:1685–1692.
66. Lutje Spelberg JH, van Hylckama Vlieg JET, Tang L, Janssen DB, Kellogg RM. Highly enantioselective and regioselective biocatalytic azidolysis of aromatic epoxides. *Org Lett* 2001, 3:41–43.
67. Majerić-Elenkov M, Hauer B, Janssen DB. Enantioselective ring opening of epoxides with cyanide catalysed by halohydrin dehalogenases: a new approach to non-racemic β-hydroxy nitriles. *Adv Synth Catal* 2006, 348:579–585.
68. Hopmann KH, Himo F. Cyanolysis and azidolysis of epoxides by haloalcohol dehalogenase: theoretical study of the reaction mechanism and origins of regioselectivity. *Biochemistry* 2008, 47:4973–4982.
69. Ferreira KN, Iverson TM, Maghlaoui K, Barber J, Iwata S. Architecture of the photosynthetic oxygen-evolving center. *Science* 2004, 303:1831–1838.
70. Loll B, Kern J, Saenger W, Zouni A, Biesiadka J. Towards complete cofactor arrangement in the 3.0 angstrom resolution structure of photosystem II. *Nature* 2005, 438:1040–1044.
71. Guskov A, Kern J, Gabdulkhakov A, Broser M, Zouni A, Saenger W. Cyanobacterial photosystem II at 2.9-angstrom resolution and the role of quinones, lipids, channels and chloride. *Nat Struct Mol Biol* 2009, 16:334–342.
72. Siegbahn PEM. Crabtree RH. Manganese oxyl radical intermediates and O-O bond formation in photosynthetic oxygen evolution and a proposed role for the calcium cofactor in photosystem II. *J Am Chem Soc* 1999, 121:117–127.
73. Siegbahn PEM. O-O bond formation in the S-4 state of the oxygen-evolving complex in photosystem II. *Chem Eur J* 2006, 12:9217–9227.
74. Siegbahn PEM. A structure-consistent mechanism for dioxygen formation in photosystem II. *Chem Eur J* 2008, 27:8290–8302.
75. Siegbahn PEM. An energetic comparison of different models for the oxygen evolving complex of photosystem II. *J Am Chem Soc* 2009, 131:18238–18239.
76. Yano J, Kern J, Sauer K, Latimer MJ, Pushkar Y, Biesiadka J, Loll B, Saenger W, Messinger J, Zouni A, Yachandra VK. Where water is oxidized to dioxygen: structure of the photosynthetic Mn4Ca cluster. *Science* 2006, 314:821–825.
77. Yano J, Kern J, Pushkar Y, Sauer K, Glatzel P, Bergmann U, Messinger J, Zouni A, Yachandra VK. High-resolution structure of the photosynthetic Mn4Ca catalyst from X-ray spectroscopy. *Philos Trans R Soc Lond B Biol Sci* 2008, 363:1139–1147.
78. Dau H, Grundmeier A, Loja P, Haumann M. On the structure of the manganese complex of photosystem II: extended-range EXAFS data and specific atomic-resolution models for four S-states. *Philos Trans R Soc Lond B Biol Sci* 2008, 363:1237–1244.
79. Ryde U, Siegbahn PEM. EXAFS refinements of a theoretical model for the oxygen evolving complex in photosystem II 2010. In preparation.

