Acc. Chem. Res. **2001,** *34,* 351−358

Large-Scale ab Initio Quantum Chemical Calculations on Biological Systems

RICHARD A. FRIESNER* AND
BARRY D. DUNIETZ
Department of Chemistry and Center for Biomolecular Simulation, Columbia University, 3000 Broadway, Mail Code 3110, New York, New York 10027

Received December 2, 1999

ABSTRACT
In this Account we describe recent advances in two ab initio electronic structure methods, localized perturbation approaches and density functional theory, that allow accurate calculations including electron correlation to be carried out for systems with hundreds of atoms. Application of these methods to large-scale modeling of biological systems is discussed. Localized perturbation methods are best suited to computation of conformational energetics and nonbonded interactions: determination of the relative energetics of various conformations of the alanine tetrapetide is presented. Density functional theory is the method of choice for studying reactive chemistry; investigations of the catalytic cycle of the enzyme methane monooxygenase are reviewed.

I. Introduction

The past 30 years have seen a remarkable enhancement in the capabilities of quantum chemical methods to study complex chemical systems. Advances in computer hardware, the development of sophisticated computer codes with increasingly efficient algorithms, the systematic application of quantum chemical methodology to a wide range of small molecules (where benchmarking against experimental data is possible), and fundamental improvements in the treatment of electron correlation (exemplified by new density functional approaches) have combined to yield powerful new methodologies which can routinely handle systems containing hundreds of atoms and provide near-chemical accuracy in many cases for structures and energies. The 1998 Nobel Prize in Chemistry was awarded to John Pople and Walter Kohn for pioneering several of the key developments outlined above.

Quantum chemical calculations have been applied to biological systems almost from the beginning of these endeavors. Initially one could study only small models of the actual biochemically relevant molecules; however, even this was often capable of yielding significant insights into reaction pathways, conformational preferences, hydrogen-bonding interactions, and a wide range of other important issues. Until recently, however, realistic quantum chemical modeling of biological processes and interactions—that is, treatment of a sufficient number of atoms at a sufficiently reliable level of electron correlation to accurately reproduce structures and energies—was not feasible. Over the past 5 years, this situation has changed dramatically, and we are now entering an era in which it will be possible even for nonexperts to carry out such calculations routinely. The goal of this Account is to describe the approaches we have taken in our laboratory to achieve these objectives, to give some examples of what is possible with current methods, and to discuss likely developments over the next 5 years.

This Account is organized as follows. In section II, we describe several different quantum chemical methods capable of treating large systems, accuracies of the methods as evaluated by comparison with small molecule experimental data, and the problems for which each method is well suited. This section includes a discussion of computational performance for our Jaguar suite of ab initio electronic structure codes. In section III, two large-scale biological applications are presented, each of which uses a different quantum chemical methodology. In section IV, the conclusions, we discuss future directions.

II. Quantum Chemical Methods

Biological systems are characterized by specificity, both in structure and in chemical reactivity. To understand the origin of this specificity, small models of conformational, nonbonded, and reactive energetics can only take us so far. The key size range at which realistic structural models of a biologically specific core region can be constructed is between 20 and 200 atoms. Systems of this size can be treated at the Hartree−Fock (HF) level with current quantum chemical codes; however, for many if not most problems, inclusion of electron correlation is critical if results of sufficient accuracy for biological relevance are to be achieved. Because of the steep scaling of computational effort with system size N, traditional high-accuracy quantum chemical methods such as CCSD (T) (scaling as N^7) or even second-order Moller−Plesset perturbation theory (MP2), scaling as N^5, cannot readily be applied to systems much larger than 20 atoms in a cost-effective fashion. Thus, before biological systems could be ac-

Richard A. Friesner was born in New York, NY, in 1952. He received his B.S. degree in chemistry from the University of Chicago in 1973. He obtained his Ph.D. in 1979 at the University of California, Berkeley, working in the laboratory of Kenneth Sauer. He then spent three years as a postdoctoral fellow working with Robert Silbey at the Massachusetts Institute of Technology. He joined the Chemistry Department at the University of Texas at Austin in 1982 as an Assistant Professor. In 1990, he became Professor of Chemistry at Columbia University. Currently, he is also Director of the Center for Biomolecular Simulation at Columbia, supported by the National Institutes of Health Division of Research Resources. In addition to the application of quantum chemical methods to biological systems, his interests include development of molecular mechanics force fields and models for continuum solvation, computational methods for protein folding, and calculation of electron-transfer rates in complex molecules and materials.

Barry D. Dunietz is a Ph.D. student in chemical physics under the supervision of Professor Richard Friesner at Columbia University. He is currently at his final stages of his studies. He was born in the United States and was raised in Germany and Israel. He served in the IDF army in the years 1987−1990. In 1994, he received his Bs.C. in chemistry and computer sciences from Tel-Aviv University, Israel. His research involves the application and development of quantum chemistry methods.

* To whom correspondence should be addressed. Phone: (212) 854-7606. Fax: (212) 854-7454. E-mail: rich@chem.columbia.edu.

10.1021/ar980111r CCC: $20.00
Published on Web 02/15/2001

curately addressed by quantum chemical methods, new approaches to electron correlation were required. Two such approaches, density functional theory (DFT) and localized perturbation methods, have proven to be exceptionally effective; we briefly describe each below.

A. Localized Perturbation Methods. In quantum chemical methods for electron correlation, two sets of one-electron wave functions are defined. These are the occupied orbitals and the unoccupied orbitals. The unoccupied orbitals span the correlation space orthogonal to the occupied space and are commonly referred to as virtual orbitals. The orbitals are, in general, delocalized over the entire molecule. The delocalization of both occupied and virtual orbitals means that from each occupied orbital, one must include excitations to all virtual orbitals in the correlation calculation.

Physically, however, electron correlation is a localized phenomenon. In the mid-1980s, Pulay and co-workers developed an approach that could profitably exploit this fact.[1] First, the HF orbitals are localized by taking linear combinations of the molecular orbitals, for example via a procedure known as Boys localization. In an ordinary molecular ground-state wave function, each localized orbital corresponds to a bond (localized predominantly on two bonded atoms) or a lone pair (localized predominantly on one atom). Once the occupied orbitals are localized, the virtual (unoccupied) space associated with each occupied orbital can also be localized. For a large molecule, localization of the virtual space leads to a compact description in terms of a restricted set of orbitals which does not grow with molecular size. With the proper implementation, this leads to very large reductions in computational effort as is discussed below; at the same time, the accuracy of localized perturbation calculations is at the very least comparable to that of the corresponding delocalized calculation with the same basis set.

B. Density Functional Theory. Density functional theory (DFT) is based on the Hohenberg−Kohn theorem, which states that the energy of the ground state of a many-electron system is a functional of the electron density. The electron density is a function of three spatial dimensions, as opposed to the N-particle wave function, which must be represented in $3N$ dimensions. Therefore, this approach promises a qualitative reduction in computational effort as compared to traditional ab initio methods for electron correlation. However, the correct functional cannot be rigorously derived from the Schrodinger equation; a combination of heuristic reasoning often based on known physical limits, inspired guesswork, and empirical experimentation is required to produce a functional of improved accuracy. In most DFT methodologies, the unknown parts of the functional are collected in the "exchange-correlation" part. The first widely used exchange-correlation functionals were based on the uniform electron gas and were implemented via the local density approximation (LDA). While the LDA performs remarkably well considering the crudeness of the underlying approximation (on balance about as well as Hartree−Fock theory), it is not capable of quantitatively describing reactive chemistry. For this reason, until the past 10 years, DFT was not extensively used in the chemical community.

Over the past decade, tremendous improvements in the quality of the functionals available have made DFT the method of choice for many quantum chemical applications. The work of Becke,[2−4] Perdew,[5] Parr,[6] Pople,[7] and others[8,9] has produced improved performance of DFT methodology. Two new classes of functionals are based on gradient corrections to the local density approximation: generalized gradient approximation (GGA)[2,5,6] and hybrid functionals[3,4,7] which contain an admixture of exact exchange (Hartree−Fock) and approximated exchange (LDA). Both of these classes provide dramatic improvement as compared to the LDA in all molecular properties, including geometries, atomization energies, descriptions of hydrogen bonding, etc. At present, hybrid methods appear to supply greater accuracy at a moderately larger ($\sim 2\times$) cost, although new functionals are constantly being developed, and this situation may change in the future.

The computational effort involved in carrying out DFT is roughly comparable to that for a Hartree−Fock calculation. Indeed, both methods are most easily formulated as self-consistent field equations for a set of occupied orbitals ("density orbitals" in the case of DFT), with the principal difference being the replacement of the Hartree−Fock exchange operator with an exchange-correlation operator in DFT, the form of which is mandated by the specific functional being employed. With efficient numerical methods, this means that DFT can readily be applied to systems in the 50−200 atom range. Furthermore, DFT methods can easily be deployed across the entire periodic table, for example in transition metal calculations (although accuracy is not uniform for the various transition metal rows and columns). In contrast, traditional quantum chemical methods can lose significant accuracy for heavier elements.

C. Numerical Implementation and Computational Performance. A great deal of effort has been expended in developing efficient numerical methods for quantum chemical calculations on large systems. In ref 10, we have reviewed many of these methods. Here, we shall focus on the methods we have implemented in the Jaguar suite of ab initio programs; furthermore, we shall summarize results rather than present details of the computational algorithms. These methods are based on pseudospectral (PS) numerical techniques, originally developed for numerical simulations of hydrodynamic turbulence. The synthesis of PS techniques with conventional quantum chemical basis sets yields a hybrid approach with substantial computational advantages over standard methods based on analytical two-electron repulsion integrals, such as those found in the Gaussian series of quantum chemistry programs.

For DFT calculations, the timing advantages are quantitative; the scaling of PS methods with system size N is still N^2 for molecules of moderate size (for very large systems, scaling of $N \ln N$ or even N can be reached using fast multipole methods;[11] the reader is referred to ref 12 for a recent review on the O(N) class of methods), but

the prefactor scaling is smaller than that in the conventional DFT methods, and the asymptotic scaling regime is reached more quickly. Detailed comparisons of the relative efficiencies of Jaguar and Gaussian 92 for DFT calculations are presented in ref 10. For the largest systems, a factor of more than 10× performance enhancement is obtained for hybrid functional calculations and more than 20× for GGA calculations. The additional factor of 2 in performance advantage for GGA calculations arises because we have implemented an atom-centered version of the fast multipole algorithm which is quite effective even for a relatively small number of atoms (whereas standard multipole approaches[11] reduce computational effort only for considerably larger systems).

For the PS implementation of localized perturbation methods,[13] the advantages observed are qualitative. For LMP2 they allow one to avoid carrying out the traditional four-index transform needed in conventional methods to evaluate two-electron integrals over molecular orbitals. We observe computational scaling with a system size of $N^{2.5}$ for systems in the size range of interest here; this is in contrast to the N^4-N^5 obtained for canonical MP2 approaches. This leads to large reductions in computational effort as compared to the implementations of MP2 in conventional electronic structure codes. For example, for piperidine, a small organic molecule, a factor of 20× speed-up is obtained as compared to Gaussian 92.[13] For larger molecules, such as those that will be considered below, the differential is extrapolated to be several orders of magnitude. In essence, localized PS methods extend the tractable range of MP2 calculations into a new size regime.

D. Accuracy of Correlated Quantum Chemical Methods for Molecular Properties. DFT methods have become the methods of choice for treatment of reactive systems containing a large number of atoms, particularly those including transition metals. While they do not yet uniformly achieve chemical accuracy (1 kcal/mol maximum error for heats of atomization), the performance of the best functionals when applied to the extended G2 database of Curtiss and co-workers[14] is impressive, with hybrid B3LYP methods achieving an average error of 3.1 kcal/mol for several hundred heats of atomization, ionization potentials, electron affinities, and proton affinities of a wide range of small molecules. This is to be compared with an average error of 1.58 kcal/mol for the G2 methodology and 0.94 kcal/mol for the improved G3 methodology,[15] based on traditional ab initio quantum chemical approaches to electron correlation. Both the G2 and G3 theories require coupled cluster calculations, scaling as N^7, that are intractable for large systems.

LMP2 methods, even when augmented with empirical parameters compensating for basis set incompleteness, are no better than DFT calculations in computing reactive energies such as heats of atomization, and they are more expensive and difficult to apply to transition-state calculations which are essential in studying reaction mechanisms. Therefore, at present, DFT methods are preferable for such studies. However, when considering energy differences that do not involve making or breaking bonds, the situation is quite different. LMP2 methods have been shown to provide substantially better accuracy than DFT for conformational energy differences of small organic molecules when compared with experimental data;[10,13] with a large basis set, the deviation between the LMP2 results and experiment for a test suite of 36 molecules is ~0.25 kcal/mol. For dispersion energies, DFT methods are qualitatively incorrect, whereas LMP2 provides good (although not perfect) estimates of van der Waals interaction.

III. Biological Applications

We consider below two large-scale biological applications, each of which is optimally addressed by different quantum chemical approaches, based on the analysis presented in the previous section. The first is evaluation of relative conformational energies of peptides; here, LMP2 calculations provide high-quality results for a reasonable computational cost. The second is determination of the catalytic intermediates in the active site of the enzyme methane monooxygenase (MMO), for which DFT methods represent the best approach at present.

A. Conformational Energetics of Peptides: Assessment and Improvement of Molecular Modeling Force Fields. Many biologically important chemical processes involve noncovalent interactions. For example, protein and peptide conformations are critical to their physiological functions, and most protein−ligand binding events are driven by nonbonded interactions. Furthermore, a few kilocalories per mole or less of energy is often critical in controlling molecular recognition phenomena. Thus, if one is to build predictive models of biological systems, high accuracy in the treatment of conformational energetics and nonbonded interactions is required. Furthermore, models used to investigate conformational and molecular recognition processes must be able to be rapidly evaluated so that extensive phase space sampling, which is almost always needed, can be carried out.

The requirement for computational efficiency mandates that the model be cast as a classical molecular mechanics force field. The role of quantum chemistry here is in developing parameters for the force field and in testing transferability of the force field from the small molecules (against which it is parametrized) to larger systems. For both of these tasks, computational efficiency without sacrificing accuracy is demanded.

We focus here on conformational energetics (which of course implicitly includes nonbonded interactions as a key determinant of relative energies). As was mentioned above, DFT methods display significant discrepancies with experimental data for small-molecule conformational energetics, whereas LMP2 with a high-quality extended basis set does much better (note that a large basis set is necessary to obtain significant improvement from LMP2 as compared to HF or DFT calculations). Furthermore, for larger systems, dispersion energy will be important, and

Table 1. Comparison of Alanine Tetrapeptide Structures and Relative Conformational Energetics of Various Force Fields with Quantum Chemical Calculations Using HF/6-31G Geometry Optimization and LMP2/cc-pVTZ Single-Point Energies[a]**

	energy RMSDs					
	av geometry RMS	geometry RMS > 0.60	unrestrained minima	restrained minima	pairwise errors > 3	max pairwise error
LMP2/cc-pVTZ(-f)			0.00	0.00	0	0.00
HF/6-31G**	0.00	0	1.10	1.10	1	3.08
MMFF	0.32	0	1.24	1.21	5	3.35
OPLS-AA(2,2)	0.16	0	1.31		6	4.72
GROMOS	0.39	1	1.60	2.42	10	5.67
CFF95	0.41	3	1.86	1.91	14	6.48
CHARMM 22	0.86	5	2.56	3.78	13	10.20
AMBER94	0.58	4	3.42	1.42	20	12.53

[a] All energy differences are given in kilocalories per mole; geometry RMSD values are in angstroms.

LMP2 again will represent this much more accurately than either HF or DFT. Therefore, in developing and testing molecular modeling force fields, we have chosen to standardize on the following protocol: HF geometry optimization using the 6-31G** basis set, followed by a single-point LMP2 calculation using the cc-pVTZ(-f) correlation-consistent basis set.

Our first application of quantum chemical methods to test molecular modeling force fields examined the alanine tetrapeptide.[16] Many protein force fields have been parametrized to fit quantum chemical results (although typically not at a converged level of accuracy) for the alanine dipeptide. By examining the ability of force fields to predict the relative energies of 10 conformations of the alanine tetrapeptide, one can test the transferability of the force field to a larger peptide structure. The conformations were taken from a molecular mechanics conformational search (at a variety of energies, so as to have a sampling over a wide range of phase space), subjected to our standard quantum chemical protocol discussed above, and then minimized using each of the force fields to be examined.

The performance of different force fields in calculating the relative energies of the 10 tetrapeptides conformation is detailed in ref 16; a subset of these results is presented in Table 1. The results of the force fields were compared to the HF/6-31G** optimized structures and LMP2/cc-pVTZ(-f) single-point energies evaluated using the HF/6-31G** optimized structures; we also compared the HF energies at the minimum energy geometry with the LMP2 energies to display the effects of the inclusion of electron correlation on conformational energetics. We have been able to improve the accuracy of one of the best performers, the OPLS-AA force field of Jorgensen and co-workers.[17] The improvement has been achieved by refitting the backbone torsional parameters of the dipeptide to LMP2 data. The improvement results in reducing the root-mean-square deviation (RMSD) of the energy to ~0.5 kcal/mol; this is a qualitative improvement over previous efforts. Development of a complete version of this force field, covering all 20 amino acid side chains, is currently underway in our laboratory.

B. Modeling of the Catalytic Mechanism of Methane Monooxygenase. We consider here, as a prototypical problem in bioinorganic catalysis, the functioning of the enzyme methane monooxgenase (MMO), which we have been studying in collaboration with Prof. Steve Lippard and his research group at the Massachusetts Institute of Technology. This enzyme catalyzes the reaction of dioxygen and methane to form methanol, via a diiron site which undergoes a series of transitions through several different intermediate states. The system has been extensively studied experimentally (reviewed in refs 18 and 19), and we briefly summarize the results below. For each intermediate in the catalytic cycle, the spin and charge states of the iron atoms are altered; there are other significant changes, in structure and protonation states, as well. The resting state of the enzyme, which we refer to here as the oxidized state, has two high-spin Fe(III) atoms which are antiferromagnetically (AF) coupled. Via long-range electron transfers from an auxiliary protein, this state is converted to the reduced state of the enzyme, the highest energy state in the cycle, which has two Fe(II) irons coupled ferromagnetically. Next, dioxygen adds to the complex, producing the intermediate state H_{peroxo} (which will be denoted as intermediate P), which contains two Fe(III) atoms with AF coupling. P then converts into a final intermediate, Q, which is believed to carry out the catalytic reaction with methane to produce methanol. Q contains two Fe(IV) atoms, again high spin and AF coupled. From recent EXAFS measurements,[20] Q is also thought to have a very short Fe−Fe distance, on the order of 2.5 Å, a distance which is not easy to achieve in model systems. After reacting with methane, Q is converted back to the oxidized state of the enzyme and the final products of the reaction, water and methanol. The entire cycle is presented in schematic form in Figure 1. The charges and spin coupling of the irons defining the different intermediates are provided in ref 25.

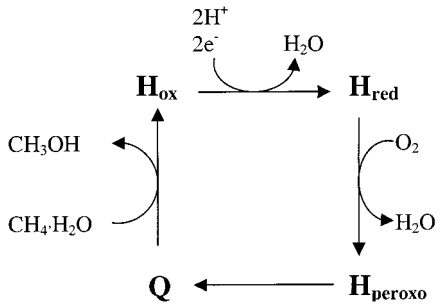

FIGURE 1. Thermodynamic reaction cycle for MMOH.

Lippard and co-workers have obtained high-resolution structures of both the oxidized and reduced forms of the enzyme via X-ray crystallography.[21,22] These structures provide an initial test for ab initio quantum chemical calculations; can the model and methodology reproduce the X-ray structure around the active site? Previous ab initio work on MMO,[23,24] while employing high-quality DFT methods and yielding interesting insights into the electronic structure of the diiron core, were restricted to a relatively small model of the enzyme active site, the largest such model[24] being one which retained only side-chain groups directly ligated to the metals, and hence did not include any atoms in the second coordination sphere. Our objective was to investigate larger structural models of the active site while retaining the high-quality electronic structure methods used in refs 23 and 24. Below we summarize our results, which are reported in detail in ref 25.

For the geometry optimizations the 6-31G** basis set for organic atoms and the LACVP** basis set for the irons were used. In cases where crystal structures were available, these were used as the starting guess for the geometry; for intermediates where there is no crystal structure information, a large number of initial guess structures were formulated in collaboration with the Lippard group and ultimately ranked according to energetic criteria. Single-point energy calculations were then performed on the optimized structures using the cc-PVTZ(-f) basis set for organic atoms and the LACV3P** basis set for the iron atoms; these larger basis sets were necessary to obtain reliable energetic comparisons for the different intermediates.

Initially, we carried out geometry optimizations for the oxidized and reduced form of the enzyme with a model containing the diiron center and the immediate ligands of the irons (referred to below as the "small" model; this model is comparable to that used in ref 24). The oxidized enzyme yielded qualitative agreement with the X-ray structure, although the details of ligand geometry were quantitatively inaccurate. For the reduced enzyme, however, the optimized structure bore little resemblance to the X-ray structure. The Fe−Fe distance expanded by ∼1 Å, and the structure of one of the bridging moieties (an unusually positioned carboxylate with one oxygen atom serving as a μ-oxo bridge between the two irons) was transformed to an entirely different motif. Figure 2 presents the starting X-ray geometry for the reduced enzyme, while Figure 3 displays the results of optimizing the small model. The gross deviations of the small-model optimization from the experimental structure, discussed above, are apparent.

These results indicate that a structural model of the size of the small model is inadequate for the study of MMO if structural fidelity with the crystallographic data is to be maintained. It also suggests that the reduced enzyme is prepared in a strained configuration which is kept in place only by the surrounding protein residues; this fits in well with the architecture of the catalytic cycle, as the reduced enzyme is the highest energy point on the cycle, and the strained configuration can be interpreted as a device for energy storage which is used in later steps to carry out the desired catalytic reaction.

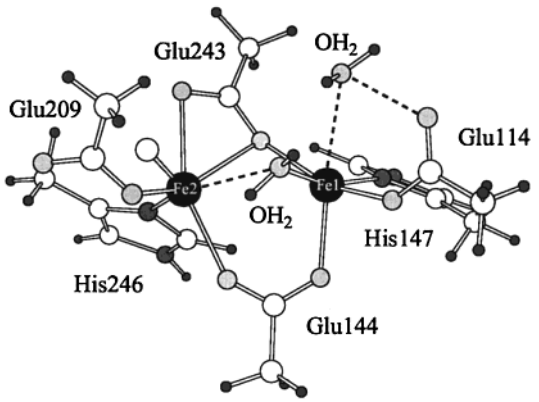

FIGURE 2. Reduced small-model starting structure (taken from the crystal structure). In the figures white circles indicate carbon atoms, the light gray circles are oxygen atoms, and the darker gray circles are nitrogen atoms. The small solid black circles are hydrogen atoms.

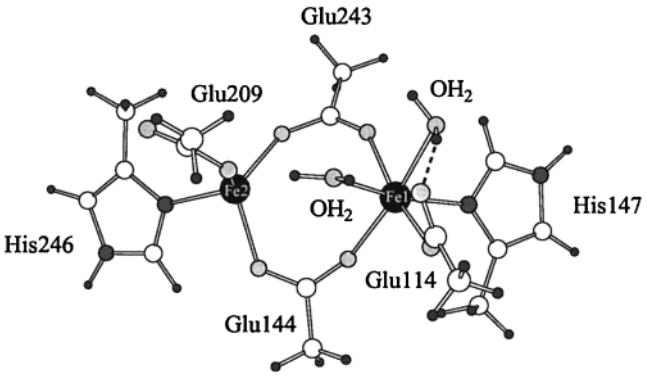

FIGURE 3. Optimized reduced small model. Significant structural differences are noticed when comparing the optimized structure to the initial structure taken from the X-ray crystallography data.

We next examined the X-ray structure at larger distances from the core of the active site and constructed a second model (the "large" model) containing a hydrogen-bonded network which we hypothesized was the mechanism for holding the reduced configuration in place. This minimal model containing the key hydrogen bonds is, indeed, stable for both the oxidized and reduced enzyme and yields satisfactory RMSDs from the experimental X-ray structural data. Figure 4 presents the results of geometry optimization of the large model for the reduced enzyme; it can be seen that, in addition to the RMSD from the X-ray structure being acceptable, the key structural motifs (Fe−Fe distance, positioning of the μ-oxo bridging carboxylate) are properly preserved. Table 2 summarizes the RMSD values observed for the large- and small-model optimizations as compared to those for the X-ray crystal structures. We conclude that the large model is suitable for realistic modeling of the MMO catalytic cycle.

Generation of a structure for the P intermediate from the reduced enzyme structure can be accomplished via addition of dioxygen and removal of the one of the structural waters which is weakly bound to the complex. We obtain both superoxo and peroxo structures, for which

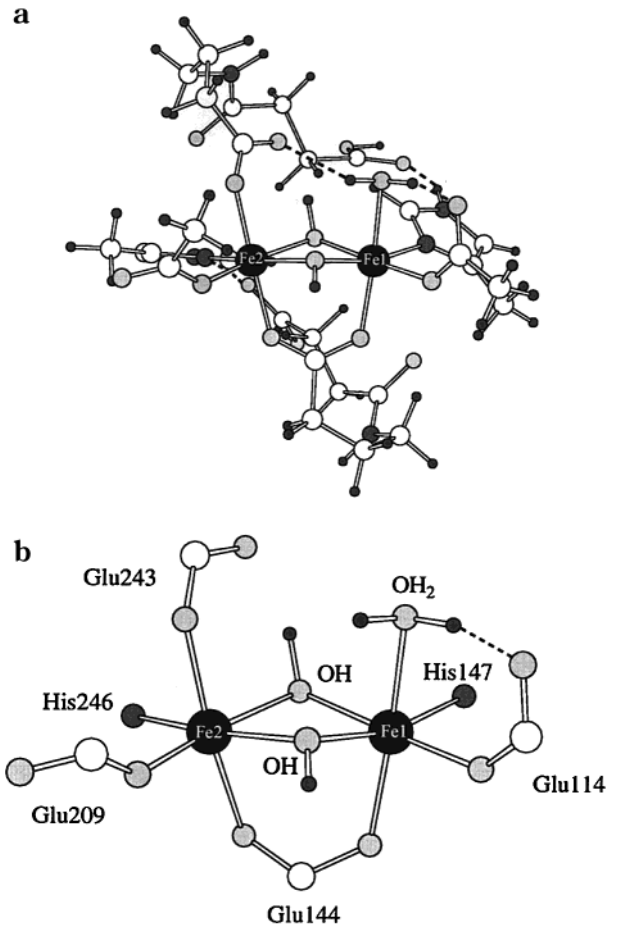

FIGURE 4. (a) Optimized reduced large model. The larger optimized structure demonstrates greater stability in reproducing the X-ray data, as opposed to the smaller model. (b) Core of the optimized reduced large model.

Table 2. RMSD of the Oxidized and Reduced Models from the MMOH Crystal Structure

model	species	all atoms	no hydrogens
small	reduced	1.653	1.455
large	reduced	0.411	0.320
small	oxidized	1.243	0.894
large	oxidized	0.634	0.501

the products are lower in energy than the reactants by a few kilocalories per mole. Studies of barrier heights, currently ongoing, will be needed to determine how these structures participate in the catalytic cycle; their relative energies with respect to each other are smaller than the accuracy of the DFT methodology.

Finally, we have examined a number of possibilities for the Q intermediate, which has been the subject of intense speculation by many workers. Many of the current models for Q assume that there are two dibridging carboxylates, i.e., that the "unusual" labile carboxylate is displaced from the μ-oxo position by dioxygen, and that it, in turn, displaces the structural water to form a dibridging structure. A representative picture of the core of a model of this type is given in Figure 5. However, such a model has serious difficulties, at least in the context of our investigations, for the following simple reason: the remaining structural water forms two strong hydrogen bonds to

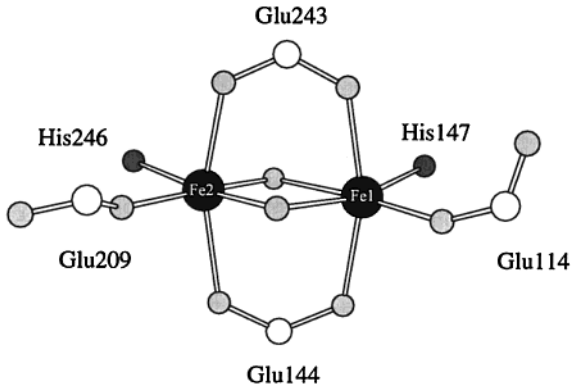

FIGURE 5. Core of the optimized alternative Q structure. This structure has the typical diamond core structure, which does not allow the additional hydrogen-bonding network observed in the currently suggested Q model.

charged moieties as well as being ligated to one of the irons, and hence is "solvated" by on the order of 30−40 kcal/mol. Displacement of this water therefore results in Q models with energies much higher than that of P plus a water returning to the bulk. Our calculations using a large model of this structure are in quantitative agreement with this analysis. There is, however, a simple alternative which possesses the appropriate energetics and at the same time is consistent with the experimental data, including the ultrashort Fe−Fe distance measured in the EXAFS experiments. In this model, shown in Figure 6, the labile carboxylate is still displaced from the core (and forms an opportunistic hydrogen bond with a backbone amide group), but the structural water remains and forms a hydrogen bond with the ligated oxygen on the other iron. This hydrogen bond draws the two irons together and is responsible for the short Fe−Fe distance. The value of this distance resulting from the calculations is somewhat larger than the experimental value (2.67 vs 2.5 Å), but there are several quite plausible reasons for this, including the flatness of the Fe−Fe potential surface (which we have verified explicitly in ref 25), minor errors in the DFT methodology and use of a finite basis set, effects of the remainder of the protein which we have not included in the model, and inaccuracies of the experimental data.

Table 3 provides energy comparisons of the different steps in the cycle, demonstrating that each step going from reactants (reduced enzyme) to products (oxidized enzyme plus methanol and water) is exothermic. The solvation free energy in water has been approximated to be 10 kcal/mol. These energy comparisons can be viewed as approximate enthalpy differences, because the differences in the zero-point energies between the product and reactant structures are expected to be negligible, as has already been found for the smaller models studied in ref 23. Our results are fully consistent with all of the available structural and thermodynamic data, in the context of a highly realistic model of the active site, although it should be noted that the precision of the DFT is not sufficient to provide complete confidence in the exothermicity of some of the calculated transformations, given that there are energy differences of less than 5 kcal/mol. In future work,

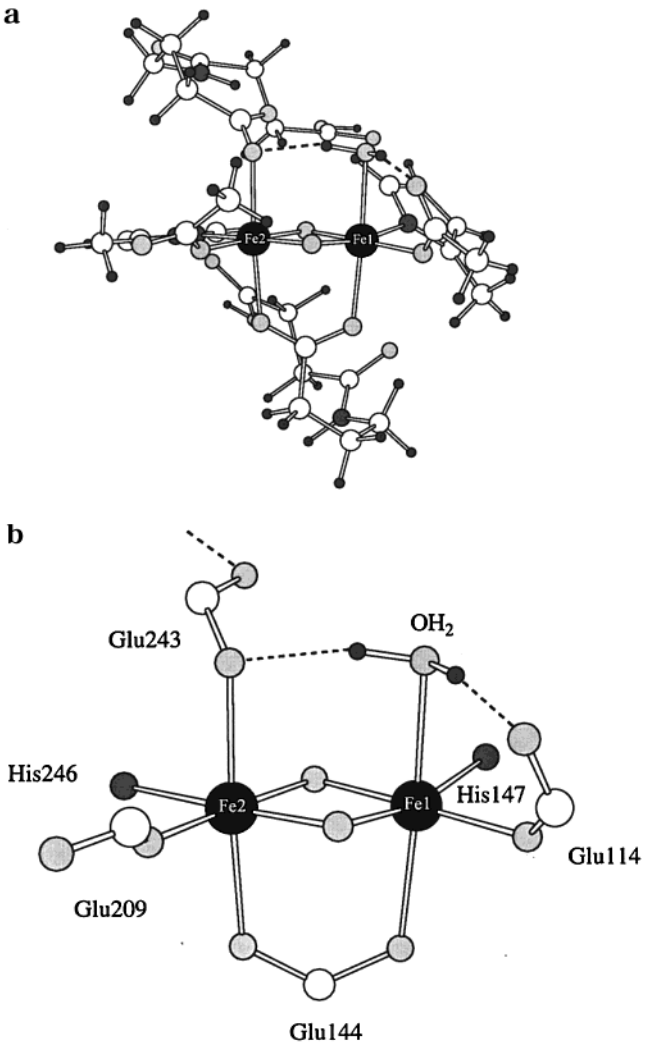

FIGURE 6. (a) Optimized Q structure. A water molecule is added in a position to gain the additional energetic stabilization by allowing the formation of the hydrogen-bonding network. (b) Core of the optimized Q structure. The hydrogen-bonding network introduced by the additional water molecule is clearly demonstrated.

Table 3. Summary of Energy Comparisons

reactants	products	enthalpy difference (kcal/mol)
reduced + O_2	H_{peroxo} + H_2O(aq)	−2.2
Q-alternate + H_2O	Q-best	−54
H_{peroxo}	Q-best	−4.1
Q-best + H_2O + CH_4	oxidized + CH_3OH	−71

we intend to employ the same model and methods to study the transition states of the system, including the methane conversion reaction.

IV. Conclusion

The two applications presented above represent the current state of the art in applying large-scale quantum chemical methods to biological systems. In both cases, insights were gained that would have otherwise been inaccessible, resulting in significant progress in understanding the relevant biological process. In the force field studies, we elucidated the limitations on current protein molecular mechanics force fields and have provided benchmarks which will be available to calibrate new force field models as they are developed. In the MMO investigations, the most important results of general relevance (i.e., beyond the specifics of the MMO problem itself) can be summarized as follows:

(1) We have provided the first (to our knowledge) example where the highly strained nature of a catalytic intermediate in an enzymatic reaction has been rigorously deduced by computational methods.

(2) We have shown that incorporation of the second coordination shell is crucial in studying transition-metal-containing active sites if the role of the protein in controlling the catalytic intermediates is to be understood.

(3) We have presented an atomic-level description of the structural and energetic parameters involved in making a carboxylate shift, a common motif in enzymatic catalysis.

(4) The role of the hydrogen-bonded network in controlling the active site structure, and the shifts in this network as one cycles through the various catalytic intermediates, have been described in structural and energetic detail.

The near future will see not only quantitative enhancements in computational methods based on faster computers and algorithmic improvements, but also qualitative augmentation of methodology that increases the level of realism while at the same time reducing computational effort. An example of an approach along these lines is mixed quantum mechanics/molecular mechanics (QM/MM) methods,[26−29] in which a small part of the system is treated via quantum chemistry and the remainder via a molecular modeling force field. Such methods are particularly well suited to studying problems such as metalloprotein active sites, as the reactive chemistry can be treated via quantum chemistry and the remainder of the protein via molecular mechanics.

This work was supported in part by grants from the NIH (Institute of General Medical Science (GM 40526) and Division of Research Resources (P41-RR06892)) and by the Environmental Molecular Sciences Institute at Columbia University, funded by the NSF (CHE-98-10367).

References

(1) Saebo, S.; Pulay, P. Local Treatment of Electron Correlation. *Annu. Rev. Phys. Chem.* **1993**, *44*, 213−236.
(2) Becke, A. D. Density-Functional Exchange-Energy Approximation With Correct Asymptotic-Behavior. *Phys. Rev. A* **1988**, *38*, 3098−3100.
(3) Becke, A. D. A New Mixing of Hartree−Fock and Local Density-Functional Theories. *J. Chem. Phys.* **1993**, *98*, 1372−1377.
(4) Becke, A. D. Density-Functional Thermochemistry. 3. The Role of Exact Exchange. *J. Chem. Phys.* **1993**, *98*, 5648−5652.
(5) Perdew, J. P.; Chvary, J. A.; Vosko, S. H.; et al. Atoms, Molecules, Solids, And Surfaces−Applications Of The Generalized Gradient Appromimation For Exchange And Correlation. *Phys. Rev. B* **1992**, *46*, 6671−6687.
(6) Lee, C. T.; Yang, W. T.; Parr, R. G. Development Of The Colle-Salvetti Correlaiton-Energy Formula Into A Functional of the Electron-Density. *Phys. Rev. B* **1988**, *37*, 785−789.
(7) Johnson, B. G.; Gill, P. M. W.; Pople, J. A. The Performance Of A Family Of Density Functional Methods. *J. Chem. Phys.* **1993**, *98*, 5612−5626.
(8) Rabuck A. D.; Scuseria G. E. Assessment of recently developed density functionals for the calculation of enthalpies of formation in challenging cases. *Chem. Phys. Lett.* **1999**, *309*, 450.

(9) Bose, A. D.; Doltsinis, N. L; Handy, N. C.; Sprik, M. J. New generalized gradient approximation functional. *J. Chem. Phys.* **2000**, *112*, 1670.
(10) Friesner, R. A.; Murphy, R. B.; Beachy, M. D.; Ringnalda, M. N.; Pollard, W. T.; Dunietz, B. D.; Cao, Y. X. Correlated ab Initio Electronic Structure Calculations for Large Molecules. *J. Phys. Chem.* A **1999**, *103*, , 1913−1928.
(11) White, C.; Head-Gordon, M. Derivation and Efficient Implementation of the Fast Miltipole Method. *J. Chem. Phys.* **1994**, *101*, 6593.
(12) Goedecker, S. Linear scaling electronic structure methods. *Rev. Mod. Phys.* **1999**, *71*, 1085.
(13) Murphy, R. B.; Beachy, M. D.; Friesner, R. A.; Ringnalda, M. N. Pseudospectral Localized Moller-Plesset Methods−Theory And Calculation Of Conformational Energies. *J. Chem. Phys.* **1995**, *103*, 1481−1490.
(14) Curtiss, L. A.; Raghavachari, K.; Redfern, P. C.; et al. Assessment of Gaussian-2 and Density Functional Theories for the Computaiton of Enthalpies of Formation. *J. Chem. Phys.* **1997**, *106*, 1063−1079.
(15) Curtiss, L. A.; Raghavachari, K.; Redfern, P. C.; Rassolov, V.; Pople, J. A. Gaussian-3 (G3) Theory For Molecules Containing First And Second-Row Atoms. *J. Chem. Phys.* **1998**, *109*, 7764−7776.
(16) Beachy, M. D.; Chasman, D.; Murphy, R. B.; Halgren, T. A.; Friesner, R. A. Accurate ab Initio Quantum Chemical Determination of the Relative Energetics of Peptide Conformations and Assessment of Empirical Force Fields. *J. Am. Chem. Soc.* **1997**, *119*, 5908−5920.
(17) Jorgensen, W. L.; Maxwell, D. S.; Tirado-Rives, J. Development and Testing of the OPLS All-Atom Force Field on Conformational Energetics and Properties of Organic Liquids. *J. Am. Chem. Soc.* **1996**, *118*, 11225−11236.
(18) Liu, K. E.; Lippard, S. J. Studies of the Soluble Methane Monooxygenase Protein System: Structure, Component Interactions, and Hydroxylation Mechanism. *Adv. Inorg. Chem.* **1995**, *42*, 263−289.
(19) Waller, B. J.; Lipscomb, J. D. Dioxygen Activation by Enzymes Containing Binuclear Non-Heme Iron Clusters. *Chem. Rev.* **1996**, *96*, 2625−2657.
(20) Shu, L.; Nesheim, J. C.; Kauffman, K.; Munck, E.; Lipscomb, J. D.; Que, L. An (Fe_2O_2)-O-IV Diamond Core Structure for the Key Intermediate Q of Methane Monooxygenase. *Science* **1997**, *275*, 515−517.
(21) Rosenzweig, A. C.; Nordlunc, P.; Takahara, P. M.; Frederick, C.; Lippard, S. J. Geometry of the soluble methane monooxygenase catalytic diiron center in two oxidation states. *Chem. Biol.* **1995**, *2*, 409−418.
(22) Rosenzweig, A. C.; Frederick, C. A.; Lippard, S. J.; Nordlund, P. Crystal structure of a bacterial non-haem iron hydroxylase that catalyses the biological oxidation of methane. *Nature* **1993**, *366*, 537−543.
(23) Siegbahn, P. E. M.; Crabtree, R. H. Mechanism of C−H Activation by Diiron Methane Monooxygenases. *J. Am. Chem. Soc.* **1997**, *119*, 3103−3113.
(24) Siegbahn, P. E. M. Theoretical Model Studies of the Iron Dimer Complex of MMO and RNR. *Inorg. Chem.* **1999**, *38*, 2880−2889.
(25) Dunietz, B. D.; Friesner, R. A.; Beachy, M. D.; Cao, Y.; Whittington, D.; Lippard, S. Large Scale Ab Initio Quantum Chemical Calculation of the Intermediates in the Methane Monooxygenase Catalytic Cycle. *J. Am. Chem. Soc.* **1999**, *122*, 2828−2839.
(26) Gao, J. L.; Amara, P.; Alhambra, C.; et al. A Generalized Hybrid Orbital (Gho) Moethod for the Treatment of Boundary Atoms in Combined QM/MM Calculations. *J. Phys. Chem. A.* **1998**, *102*, 4714−4721.
(27) Eurentius, K. P.; Chatfield, D. C.; Brooks, B. R.; Hodoscek, M. Enzyme mechanisms with hybrid quantum and molecular mechanical potentials. I. Theoretical considerations. *Int. J. Quanum Chem.* **1996**, *60*, 1189−1200.
(28) Philipp, D. M.; Friesner, R. A. Mixed Ab Initio QM/MM Modeling Using Frozen Orbitals. *J. Comput. Chem.* **1999**, *20*, 1468−1494.
(29) Vreven, T.; Morokuma, K. J. The accurate calculation and prediction of the bond dissociation energies in a series of hydrocarbon using IMOMO methods. *J. Chem. Phys.* **1999**, *111*, 8799−8803.

AR980111R