J|A|C|S
ARTICLES
Published on Web 12/29/2004

Substrate Hydroxylation in Methane Monooxygenase: Quantitative Modeling via Mixed Quantum Mechanics/ Molecular Mechanics Techniques

Benjamin F. Gherman,[†] Stephen J. Lippard,*[,‡] and Richard A. Friesner*[,†]

Contribution from the Department of Chemistry and Center for Biomolecular Simulation, Columbia University, New York, New York 10027, and Department of Chemistry, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139

Received January 9, 2004; E-mail: rich@chem.columbia.edu; lippard@mit.edu

Abstract: Using broken-symmetry unrestricted density functional theory quantum mechanical (QM) methods in concert with mixed quantum mechanics/molecular mechanics (QM/MM) methods, the hydroxylation of methane and substituted methanes by intermediate Q in methane monooxygenase hydroxylase (MMOH) has been quantitatively modeled. This protocol allows the protein environment to be included throughout the calculations and its effects (electrostatic, van der Waals, strain) upon the reaction to be accurately evaluated. With the current results, recent kinetic data for CH_3X (X = H, CH_3, OH, CN, NO_2) substrate hydroxylation in MMOH (Ambundo, E. A.; Friesner, R. A.; Lippard, S. J. *J. Am. Chem. Soc.* **2002**, *124*, 8770−8771) can be rationalized. Results for methane, which provide a quantitative test of the protocol, including a substantial kinetic isotope effect (KIE), are in reasonable agreement with experiment. Specific features of the interaction of each of the substrates with MMO are illuminated by the QM/MM modeling, and the resulting effects upon substrate binding are quantitatively incorporated into the calculations. The results as a whole point to the success of the QM/MM methodology and enhance our understanding of MMOH catalytic chemistry. We also identify systematic errors in the evaluation of the free energy of binding of the Michaelis complexes of the substrates, which most likely arise from inadequate sampling and/or the use of harmonic approximations to evaluate the entropy of the complex. More sophisticated sampling methods will be required to achieve greater accuracy in this aspect of the calculation.

I. Introduction

The hydroxylase component (MMOH) of the soluble methane monooxygenase (sMMO) enzyme is a non-heme diiron protein that catalyzes the hydroxylation of C−H bonds by dioxygen in a wide range of substrates, including methane.[1−3] Over the past decade, extensive experimental and theoretical studies of this enzyme have been carried out, thereby contributing substantially toward the elucidation of various steps in the catalytic cycle at an atomic level of detail.[4−7] In particular, the use of ab initio quantum chemical methods based on density functional theory (DFT), in conjunction with the use of large basis sets and large (∼100-atom) models for the active site, has led to remarkably good qualitative agreement of theory and experiment. When detailed reaction rates are available, our laboratory has demonstrated reasonable agreement between theory and experiment for a number of MMOH reactions, for example, those involving dioxygen activation.[8]

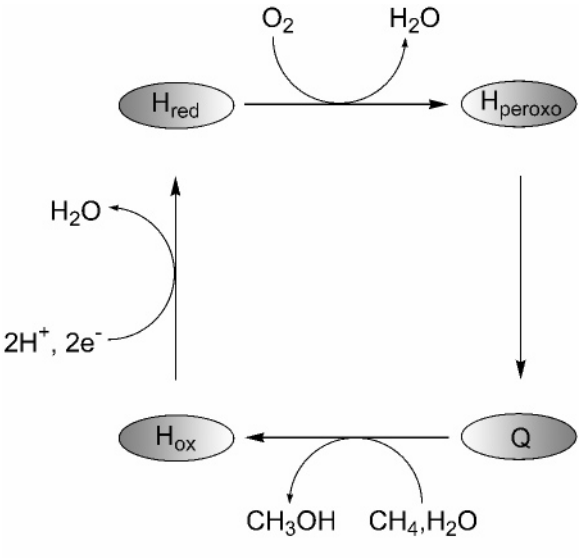
Although these results are highly encouraging, the calculations to date contain a significant number of approximations. Upgrading the level of theory and demanding agreement with experiment for a wider range of data will strengthen the bridge that we are attempting to construct between theory and experiment with increased rigor and robustness. A major improvement to our previous protocol is to incorporate the entire protein in the calculations, using mixed quantum mechanics/molecular mechanics (QM/MM) techniques. The MMOH active site is relatively compact, buried in the interior of the protein and shielded from bulk solvent.[9−11] Nevertheless, peripheral protein groups may have significant effects, as they do in other systems we have investigated, such as cytochrome P450[12,13] and hemerythrin.[14]

In the present article, we apply our QM/MM methodology,[15−17] implemented in the QSite program,[18] to study the hydroxylation

[†] Columbia University.
[‡] Massachusetts Institute of Technology.

(1) Hanson, R. S.; Hanson, T. E. *Microbiol. Rev.* **1996**, *60*, 439−471.
(2) Lipscomb, J. D. *Annu. Rev. Microbiol.* **1994**, *48*, 371−399.
(3) Liu, K. E.; Lippard, S. J. *Adv. Inorg. Chem.* **1995**, *42*, 263−289.
(4) Feig, A. L.; Lippard, S. J. *Chem. Rev.* **1994**, *94*, 759−805.
(5) Wallar, B. J.; Lipscomb, J. D. *Chem. Rev.* **1996**, *96*, 2625−2657.
(6) Merkx, M.; Kopp, D. A.; Sazinsky, M. H.; Blazyk, J. L.; Müller, J.; Lippard, S. J. *Angew. Chem., Int. Ed.* **2001**, *40*, 2782−2807.
(7) Baik, M.-H.; Newcomb, M.; Friesner, R. A.; Lippard, S. J. *Chem. Rev.* **2003**, *103*, 2385−2420.
(8) Gherman, B. F.; Baik, M.-H.; Lippard, S. J.; Friesner, R. A. *J. Am. Chem. Soc.* **2004**, *126*, 2978−2990.
(9) Rosenzweig, A. C.; Frederick, C. A.; Lippard, S. J.; Nordlund, P. *Nature* **1993**, *366*, 537−543.
(10) Rosenzweig, A. C.; Nordlund, P.; Takahara, P. M.; Frederick, C. A.; Lippard, S. J. *Chem. Biol.* **1995**, *2*, 409−418.
(11) Whittington, D. A.; Lippard, S. J. *J. Am. Chem. Soc.* **2001**, *123*, 827−838.
(12) Guallar, V.; Gherman, B. F.; Lippard, S. J.; Friesner, R. A. *Curr. Opin. Chem. Biol.* **2002**, *6*, 236−242.
(13) Guallar, V.; Baik, M.-H.; Lippard, S. J.; Friesner, R. A. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 6998−7002.
(14) Wirstam, M.; Lippard, S. J.; Friesner, R. A. *J. Am. Chem. Soc.* **2003**, *125*, 3980−3987.
(15) Philipp, D. M.; Friesner, R. A. *J. Comput. Chem.* **1999**, *20*, 1468−1494.

Scheme 1



reaction for a number of substituted methanes, CH_3X (X = H, CH_3, OH, CN, NO_2, or F), that react with MMOH. Recently, we reported a kinetic analysis of the hydroxylation of most of these substrates by the di(μ-oxo)diiron(IV) intermediate Q in the catalytic cycle. This investigation contained some rather surprising results with regard to variations in rate constants and observed kinetic isotope effects.[19] Our objective here is to utilize QM/MM methods to explain these effects in a satisfactory fashion. This goal provides a highly demanding test of the methodology, since the distinctions among the substrates are subtle, all being roughly the same size, yet substantial differences in behavior are observed experimentally.[19] Success in explaining these data would significantly enhance our confidence that the computational machinery yields reliable results even for the fine details of metalloprotein active site chemistry.

The remainder of this article is organized as follows. Section II briefly reviews experimental and theoretical work on MMOH that is relevant to the present investigation. Section III reviews our computational methodology, including the extensive validation studies that we have performed to demonstrate that our QM/MM technology is accurate and robust when compared to fully QM calculations on the same system. Section IV describes the computational model of the MMOH active site, as well as the substates to be investigated. Section V presents the results and discussion, and section VI, the conclusion, summarizes the results and overall efficacy of the QM/MM methodology.

II. Review of MMOH Catalytic Chemistry

The MMOH catalytic cycle consists of four principal steps, depicted in Scheme 1. Catalysis commences with the activation of dioxygen by H_{red}, a diiron(II) form of the protein, to generate H_{peroxo}, a diiron(III) species,[20−23] which subsequently decays to give the diiron(IV) intermediate Q.[23−25] Q is competent to hydroxylate methane, various substituted methanes, and other hydrocarbons.[19,21−23,26,27] The present article focuses on the initial portion of the reaction of Q with methane and substituted methanes up to and including the transition state for hydrogen abstraction.[28]

Experiments performed on enzymes isolated from *Methylococcus capsulatus* Bath (MMO Bath) and *Methylosinus trichosporium* OB3b (MMO OB3b) have provided a significant body of knowledge concerning substrate hydroxylation. Rate constants for the hydroxylation of methane and ethane have previously been determined,[22,23,26,27,29] and a systematic study of the reaction of MMOH with substituted methanes has been carried out.[19,30] A key source of information has come from the deuterium kinetic isotope effect (KIE) measurements for these reactions. The KIE for the methane reaction was determined to be 23−28 in MMO Bath[19,27] and 42 in MMO OB3b.[29] Given that the largest classical KIE for such a reaction is ∼7, corresponding to a symmetrical transition state in which the C−H bond-stretching mode is completely lost, it has been hypothesized that hydrogen atom tunneling occurs. KIEs for the reaction of other substrates show a surprising range: ethane, ∼1;[19,22,29] methanol, ∼1;[19] acetonitrile, 46.4;[19] nitromethane, 8.1.[19] The lack of KIEs for ethane and methanol clearly indicates that the C−H bond cleavage step is not rate-determining in these cases, and it has been suggested that it is instead substrate diffusion/binding. Substrate access to the active site has been hypothesized to be controlled in part by the regulatory component MMOB of MMO.[29,31,32] Mutagenesis of residues in MMOB important to the formation of the MMOB−MMOH complex has suggested the existence of a channel or region in the complex tuned to methane as a substrate; upon diminution in the size of certain side chains in MMOB there is an acceleration in the admittance of larger substrates. Experiment, however, has yet to yield a more detailed molecular basis for the KIE differences that reflects the specific properties of the substrates and depends on specific substrate/protein interactions.

Theoretical studies of the hydroxylation of methane and its derivatives to provide an explanation for these data are currently unavailable. If accuracy is to be achieved, the required calculations are challenging from a technical point of view, necessitating the use of, for example, relatively large models of the system and appropriate treatment of the antiferromagnetic spin coupling, features that have not been incorporated in many prior computational efforts.[33−42] A study by us addressing the

(16) Murphy, R. B.; Philipp, D. M.; Friesner, R. A. *J. Comput. Chem.* **2000**, *21*, 1442−1457.
(17) Murphy, R. B.; Philipp, D. M.; Friesner, R. A. *Chem. Phys. Lett.* **2000**, *321*, 113−120.
(18) *Qsite*; Schrödinger, Inc.: Portland, OR, 2000.
(19) Ambundo, E. A.; Friesner, R. A.; Lippard, S. J. *J. Am. Chem. Soc.* **2002**, *124*, 8770−8771.
(20) Liu, K. E.; Wang, D. L.; Huynh, B. H.; Edmondson, D. E.; Salifoglou, A.; Lippard, S. J. *J. Am. Chem. Soc.* **1994**, *116*, 7465−7466.
(21) Lee, S. K.; Nesheim, J. C.; Lipscomb, J. D. *J. Biol. Chem.* **1993**, *268*, 21569−21577.
(22) Brazeau, B. J.; Lipscomb, J. D. *Biochemistry* **2000**, *39*, 13503−13515.
(23) Liu, K. E.; Valentine, A. M.; Wang, D. L.; Huynh, B. H.; Edmondson, D. E.; Salifoglou, A.; Lippard, S. J. *J. Am. Chem. Soc.* **1995**, *117*, 10174−10185.
(24) Shu, L. J.; Nesheim, J. C.; Kauffmann, K.; Münck, E.; Lipscomb, J. D.; Que, L. *Science* **1997**, *275*, 515−518.
(25) Lee, S. K.; Fox, B. G.; Froland, W. A.; Lipscomb, J. D.; Münck, E. *J. Am. Chem. Soc.* **1993**, *115*, 6450−6451.
(26) Nesheim, J. C.; Lipscomb, J. D. *Biochemistry* **1996**, *35*, 10240−10247.
(27) Valentine, A. M.; Stahl, S. S.; Lippard, S. J. *J. Am. Chem. Soc.* **1999**, *121*, 3876−3887.
(28) Gherman, B. F.; Dunietz, B. D.; Whittington, D. A.; Lippard, S. J.; Friesner, R. A. *J. Am. Chem. Soc.* **2001**, *123*, 3836−3837.
(29) Brazeau, B. J.; Wallar, B. J.; Lipscomb, J. D. *J. Am. Chem. Soc.* **2001**, *123*, 10421−10422.
(30) Muthusamy, M.; Ambundo, E. A.; George, S. J.; Lippard, S. J.; Thorneley, R. N. F. *J. Am. Chem. Soc.* **2003**, *125*, 11150−11151.
(31) Wallar, B. J.; Lipscomb, J. D. *Biochemistry* **2001**, *40*, 2220−2233.
(32) Brazeau, B. J.; Lipscomb, J. D. *Biochemistry* **2003**, *42*, 5618−5631.
(33) Siegbahn, P. E. M. *Inorg. Chem.* **1999**, *38*, 2880−2889.
(34) Siegbahn, P. E. M. *J. Biol. Inorg. Chem.* **2001**, *6*, 27−45.
(35) Basch, H.; Mogi, K.; Musaev, D. G.; Morokuma, K. *J. Am. Chem. Soc.* **1999**, *121*, 7249−7256.

mechanism of methane hydroxylation in MMO avoided these problems,[28] but nonetheless provides little help in understanding the role of protein environment required for the present purpose. To date, the only mixed quantum mechanics/molecular mechanics calculation has been a preliminary study of H_{red} and H_{ox} using the ONIOM method.[43] The present work aims at providing not only the sought after theoretical results, but also at generating them in the most rigorous manner possible, given currently available computational tools. Using such means, we wish to provide insight into how substrate properties and protein/substrate interactions govern the hydroxylation of substituted methanes by MMOH.

III. Computational Methods

A. QM Methods. Our core computational method for stand-alone QM calculations, and for the QM region in QM/MM calculations, is hybrid density functional theory (DFT), specifically the B3LYP functional.[44−46] These methods yield excellent results for atomization energies for a wide range of chemical systems.[47−49] The testing of metal-containing systems has been less systematic than for organic molecules, but there is nevertheless a substantial body of information demonstrating that errors are unlikely to exceed a range of 3−5 kcal/mol in typical cases.[50−54] Our own work over the past several years on both organometallic complexes[55−60] and protein active sites[12,14,28,61,62] has been consistent with this estimate.

To achieve results of this quality, it is necessary to employ large basis sets. We have extensively tested a protocol in which geometry optimizations are carried out with a smaller, but still quite respectable, mixed basis set (LACVP** [63−65] for the iron atoms, 6-31G** near the model's core, 6-31G otherwise), followed by single-point calculations using LACV3P** [63−65] for the iron atoms, the cc-pVTZ(-f) correlation consistent basis set of Dunning[66] near the core, and 6-31G** otherwise. Comparisons with the use of the larger mixed basis set for the entire set of calculations suggest that the errors induced by the present protocol, as compared to using the largest basis set discussed earlier for all atoms of the system and in geometry optimization as well as for single-point energy evaluation, are quite small, typically less than 1 kcal/mol.

Within the stand-alone QM calculations, an unrestricted DFT (UDFT) methodology was employed to model effectively the open shell orbitals on the two iron atoms. In addition, antiferromagnetic (AF) coupling was included through the use of broken symmetry (BS) UDFT wave functions.[67]

The Jaguar suite of ab initio quantum chemical programs[68] is used to carry out all QM calculations. Jaguar is particularly efficient for treating large molecules with large basis sets,[69] and it is able to handle geometry optimizations and transition-state searches for the 75−150 atom QM regions that are investigated in the present article. Vibrational frequency calculations, carried out by using analytical second derivatives, are performed on an approximately half-sized version of the full QM model[70] with the small mixed basis set. The errors induced as a result of the use of a small model and basis are estimated to be negligible.[71] Enthalpy and entropy terms and zero-point energy (ZPE) corrections are then evaluated via standard gas phase formulas for the partition function, leading to a total free energy given by eq 1 for the stand-alone QM model.

$$G_{QM} = E_{quantum} + H - T\cdot S + ZPE \qquad (1)$$

This approximation would be problematic for larger, more complex substrates where anharmonic effects could be substantial, but is reasonable for the small substrates and relatively rigid protein cavity that we are dealing with in the present case.

B. QM/MM Methods. Over the past several years, we have developed methods for QM/MM calculations based on the use of frozen orbitals as the interface between the QM and MM regions.[15−17] The methodology has been specifically optimized for modeling protein active sites, with parametrization enabling the QM region to be defined via a series of cuts in the backbone and at side chains between C_α and C_β. Details of the energy function and analytical gradients, as well as the numerical implementation, are presented in refs 15−17.

The methods are implemented in the QSite program,[18] which has been constructed via a tight coupling of the Jaguar suite of electronic structure programs[68] and the IMPACT protein modeling program of

(36) Basch, H.; Musaev, D. G.; Mogi, K.; Morokuma, K. *J. Phys. Chem. A* **2001**, *105*, 3615−3622.
(37) Yoshizawa, K.; Ohta, T.; Yamabe, T.; Hoffmann, R. *J. Am. Chem. Soc.* **1997**, *119*, 12311−12321.
(38) Yoshizawa, K.; Yamabe, T.; Hoffmann, R. *New J. Chem.* **1997**, *21*, 151−161.
(39) Yoshizawa, K.; Ohta, T.; Yamabe, T. *Bull. Chem. Soc. Jpn.* **1998**, *71*, 1899−1909.
(40) Yoshizawa, K.; Suzuki, A.; Shiota, Y.; Yamabe, T. *Bull. Chem. Soc. Jpn.* **2000**, *73*, 815−827.
(41) Yoshizawa, K. *J. Inorg. Biochem.* **2000**, *78*, 23−34.
(42) Yoshizawa, K. *Coord. Chem. Rev.* **2002**, *226*, 251−259.
(43) Torrent, M.; Vreven, T.; Musaev, D. G.; Morokuma, K.; Farkas, O.; Schlegel, H. B. *J. Am. Chem. Soc.* **2002**, *124*, 192−193.
(44) Johnson, B. G.; Gill, P. M. W.; Pople, J. A. *J. Chem. Phys.* **1993**, *98*, 5612−5626.
(45) Lee, C. T.; Yang, W. T.; Parr, R. G. *Phys. Rev. B* **1988**, *37*, 785−789.
(46) Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 1372−1377.
(47) Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 5648−5652.
(48) Curtiss, L. A.; Raghavachari, K.; Redfern, P. C.; Pople, J. A. *J. Chem. Phys.* **1997**, *106*, 1063−1079.
(49) Bauschlicher, C. W. *Chem. Phys. Lett.* **1995**, *246*, 40−44.
(50) Ricca, A.; Bauschlicher, C. W. *Theor. Chim. Acta* **1995**, *92*, 123−131.
(51) Ricca, A.; Bauschlicher, C. W. *J. Phys. Chem.* **1995**, *99*, 5922−5926.
(52) Ricca, A.; Bauschlicher, C. W. *J. Phys. Chem. A* **1997**, *101*, 8949−8955.
(53) Glukhovtsev, M. N.; Bach, R. D.; Nagel, C. J. *J. Phys. Chem. A* **1997**, *101*, 316−323.
(54) Blomberg, M. R. A.; Siegbahn, P. E. M.; Svensson, M. *J. Chem. Phys.* **1996**, *104*, 9546−9554.
(55) Hascall, T.; Baik, M. H.; Bridgewater, B. A.; Shin, J. H.; Churchill, D. G.; Friesner, R. A.; Parkin, G. *Chem. Commun.* **2002**, 2644−2645.
(56) Shin, J. H.; Bridgewater, B. M.; Churchill, D. G.; Baik, M. H.; Friesner, R. A.; Parkin, G. *J. Am. Chem. Soc.* **2001**, *123*, 10111−10112.
(57) Bergquist, C.; Storrie, H.; Koutcher, L.; Bridgewater, B. M.; Friesner, R. A.; Parkin, G. *J. Am. Chem. Soc.* **2000**, *122*, 12651−12658.
(58) Bridgewater, B. M.; Fillebeen, T.; Friesner, R. A.; Parkin, G. *J. Chem. Soc., Dalton Trans.* **2000**, 4494−4496.
(59) Bergquist, C.; Bridgewater, B. M.; Harlan, C. J.; Norton, J. R.; Friesner, R. A.; Parkin, G. *J. Am. Chem. Soc.* **2000**, *122*, 10581−10590.
(60) Hascall, T.; Rabinovich, D.; Murphy, V. J.; Beachy, M. D.; Friesner, R. A.; Parkin, G. *J. Am. Chem. Soc.* **1999**, *121*, 11402−11417.
(61) Baik, M. H.; Lee, D.; Friesner, R. A.; Lippard, S. J. *Isr. J. Chem.* **2001**, *41*, 173−186.
(62) Dunietz, B. D.; Beachy, M. D.; Cao, Y. X.; Whittington, D. A.; Lippard, S. J.; Friesner, R. A. *J. Am. Chem. Soc.* **2000**, *122*, 2828−2839.
(63) Hay, P. J.; Wadt, W. R. *J. Chem. Phys.* **1985**, *82*, 270−283.
(64) Wadt, W. R.; Hay, P. J. *J. Chem. Phys.* **1985**, *82*, 284−298.
(65) Hay, P. J.; Wadt, W. R. *J. Chem. Phys.* **1985**, *82*, 299−310.
(66) Dunning, T. H. *J. Chem. Phys.* **1989**, *90*, 1007−1023.
(67) Noodleman, L. *J. Chem. Phys.* **1981**, *74*, 5737−5743.
(68) *Jaguar*, version 4.1; Schrödinger, Inc.: Portland, OR, 2000.
(69) Friesner, R. A.; Murphy, R. B.; Beachy, M. D.; Ringnalda, M. N.; Pollard, W. T.; Dunietz, B. D.; Cao, Y. X. *J. Phys. Chem. A* **1999**, *103*, 1913−1928.

(70) Vibrational frequencies were computed for truncated models only owing to the prohibitive cost of computing frequencies for the entire ∼100 atom model. Glutamic acid residues were cut between C_α and C_β and capped with hydrogen atoms. NH_3 ligands were used to replace histidine residues. Before frequencies were computed, the positions of the capping hydrogen atoms from the glutamic acid residues and ammonia hydrogen atoms were optimized, while freezing the rest of the structure.
(71) Use of a small model can lead to the introduction of artificial imaginary frequencies. Some small imaginary frequencies are due to bending and stretching modes of the capping or ammonia hydrogens. Other imaginary frequencies are merely numerical noise from the computation. Some imaginary frequencies are due to instabilities created in the small model upon removal from the protein environment. Given the inability at present to carry out frequency calculations within the QM/MM framework, this last problem is unavoidable when it surfaces. Careful examination of the vibrational modes allows these imaginary frequencies to be sifted through and structures to be identified as valid stationary points. The impact of the erroneous negative frequencies on the zero-point energy is minimal since low frequency vibrations contribute least to this quantity. Thermodynamic quantities are also little affected since excited vibrational modes are largely unpopulated at room temperature. Energetic errors thus caused are expected to be much smaller than those in the electronic energy from the DFT calculation.

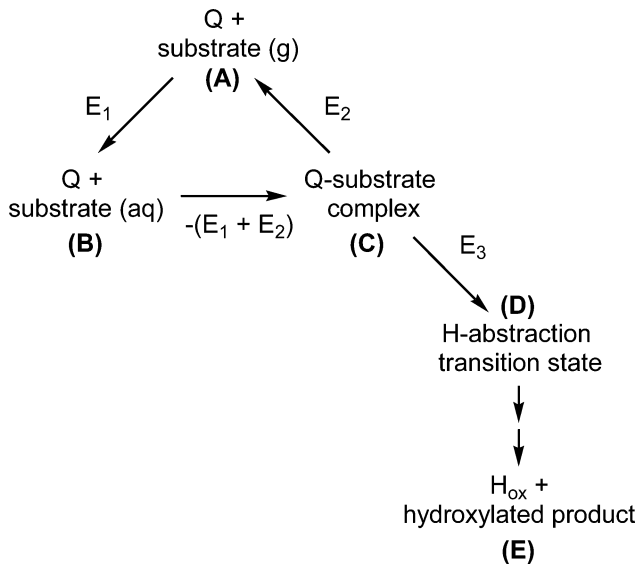
Levy and co-workers.[72] The OPLS-AA protein molecular mechanics force field of Jorgensen and co-workers is used to represent the MM region.[73,74] Capabilities for efficient minimization and transition-state optimization have been implemented and extensively tested. A key feature of the methodology is the use of an adiabatic approach to optimization in which the MM region is fully optimized after each QM step, resulting in very large reductions in computational effort as compared to carrying out a QM gradient evaluation at each geometry step. The CPU times are a small multiple of those required for an equivalent purely QM calculation.

We have extensively validated the QSite methodology over a wide range of test cases, including relative conformational energies of dipeptides,[15−17] deprotonation energies of amino acid side chains,[16,17] addition of dioxygen to a large model of the hemerythrin active site,[75] and computations of binding energies and energies of transition states and intermediates for a number of enzymes, including hemerythrin,[14,76] cytochrome P450,[12,13,16,76] triosephosphate isomerase (TIM),[77] and a class C β-lactamase and a penicillin binding protein (PBP).[78] The first suite of tests demonstrates that the difference between the QM/MM and fully QM calculations for model systems are in the great majority of cases less than 1 kcal/mol and uniformly less than 2 kcal/mol. Comparison with experimental free energies of stable intermediates and activation free-energy barriers, estimated from rate constants via simple transition-state theory approximations, has yielded agreement within the range of 2−3 kcal/mol for all of the systems we have examined to date.[8,14,77,78]

Although this experience is highly encouraging, the number of explicit comparisons that we have been able to make with experiment is still anecdotal. It is critical to continue to confront theory and experiment in as quantitative a fashion as possible so that any problems with the methodology can be uncovered. This objective is a significant motivation for the present study, particularly since the experimental data for MMOH are among the most extensive and reliable for metal-containing proteins. The ability to treat MMOH using QM/MM techniques will enable us to continue the process of rigorously calibrating the performance of both our QM/MM techniques and the underlying DFT formalism.

C. Protocol for Obtaining and Analyzing Free Energies of Intermediates and Transition States Using QM and QM/MM Methods. To analyze quantitatively in detail the interaction of small molecule substrates with MMO, we construct the thermodynamic cycle depicted in Scheme 2. We begin with reactants at infinite separation in the gas phase, which is assigned a free energy of zero. The substrate is then transferred from the gas phase to aqueous solution, and the free energy associated with this step is available from experimental solvation free-energy data.[79,80] The next step is to transfer the substrate from solution into the binding cavity of the protein, thus forming a protein−substrate complex prior to hydrogen atom abstraction. We have constructed complexes of this type for each of the substrates considered herein (vide infra) and evaluated their total free energy by a combination of QM and QM/MM calculations, discussed in more detail later. Finally, the total energy of the protein−substrate hydrogen abstraction transition state is evaluated, using the same sequence of calculations employed to calculate the initial protein−ligand complex. For all QM/MM calculations, we make the approximation that bulk solvation effects

Scheme 2



are unimportant, other than screening of surface ionizable groups of the protein, because the active site is buried in the protein interior.[9]

As discussed previously,[14] our QM/MM methodology is at present implemented exclusively using a restricted open-shell DFT (RODFT), rather than UDFT, treatment of spin. Accurate modeling of the diiron core in MMOH requires the use of UDFT methods, however. To overcome this difficulty, we employ a strategy first described, and used successfully, in our study of the free energy of dioxygen binding in hemerythrin.[14] QM/MM geometry optimizations are performed with the small mixed basis set, and RODFT methods are used to obtain reactant and transition-state structures in the protein environment. The QM regions of these structures are then used to build ∼75-atom models of the diiron core region, including substrate, by capping the relevant bonds of the model with hydrogen atoms. These hydrogen atom positions are subsequently optimized by using the small mixed basis set and RODFT methods. Finally, single-point large mixed basis set RODFT calculations are carried out on each QM/MM structure and the corresponding QM models. The energy difference (eq 2) constitutes the QM/MM energy correction to the purely QM results obtained by using the methods in section A.

$$E_{\mathrm{QM/MM}} = \mathrm{H}_{\mathrm{QM/MM}} - \mathrm{H}_{\mathrm{QM(RODFT)}} \quad (2)$$

The total free energy of the system obtained from this protocol is then simply given by eq 3.

$$G_{\mathrm{TOT}} = G_{\mathrm{QM}} + E_{\mathrm{QM/MM}} \quad (3)$$

The errors induced by this protocol are small compared to intrinsic errors in the DFT methods. Confirmation of this assertion is provided by success in comparisons with experimental data, both in the case of hemerythrin, where the experimental absolute free energy of dioxygen binding was reproduced to within ∼2 kcal/mol,[14] and in the present article, as discussed later.

In addition to enabling proper treatment of spin and the use of a sufficiently large, flexible basis set to converge DFT energy differences, the employment of both QM and QM/MM models enables an evaluation of the strain energy induced by the protein, obtained by relaxing the QM/MM-derived QM model system. Analysis of the QM/MM Hamiltonian further enables decomposition of the QM/MM interaction energy into van der Waals and electrostatics components, evaluation of the MM reorganization energy upon hydrogen atom abstraction, and other contributions.

For several of the substrates, comparison with experiment in essence requires evaluation of the free energy of binding of the ligand to the

(72) *Impact*, version 1.7; Schrödinger, Inc.: Portland, OR, 2000.
(73) Jorgensen, W. L.; Maxwell, D. S.; Tirado-Rives, J. *J. Am. Chem. Soc.* **1996**, *118*, 11225−11236.
(74) Kaminski, G. A.; Friesner, R. A.; Tirado-Rives, J.; Jorgensen, W. L. *J. Phys. Chem. B* **2001**, *105*, 6474−6487.
(75) Wirstam, M.; Gherman, B. F.; Guallar, V.; Murphy, R. B.; Friesner, R. A. Manuscript in preparation.
(76) Friesner, R. A.; Baik, M.-H.; Guallar, V.; Gherman, B. F.; Wirstam, M.; Murphy, R. B.; Lippard, S. J. *Coord. Chem. Rev.* **2003**, *238*−*239*, 267−290.
(77) Guallar, V.; Jacobson, M.; McDermott, A.; Friesner, R. A. *J. Mol. Biol.* **2004**, *337*, 227−239.
(78) Gherman, B. F.; Goldberg, S. D.; Cornish, V. W.; Friesner, R. A. *J. Am. Chem. Soc.* **2004**, *126*, 7652−7664.
(79) Cabani, S.; Gianni, P.; Mollica, V.; Lepori, L. *J. Solution Chem.* **1981**, *10*, 563−595.
(80) Pye, C. C.; Ziegler, T. *Theor. Chem. Acc.* **1999**, *101*, 396−408.

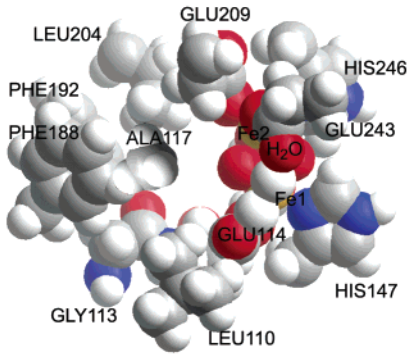


Figure 1. Space-filling representation of the MMOH active site, revealing the substrate binding pocket.

protein, as compared with the ligand free in solution. Achievement of high accuracy in the computation of this quantity is extremely difficult, and we are aware that the methods described earlier are formally deficient in several respects. The results presented for these cases can be viewed as an attempt to calibrate the degree to which the approximations that we have made, which can be viewed as a "model chemistry," are quantitatively deficient. Future work will involve a more rigorous evaluation of the protein−ligand binding affinity, with a particular focus on sampling and entropic effects. The likely results of such improvements are briefly discussed later.

IV. Physical Model

Construction of the QM/MM computational models of the MMOH active site began with crystallographic data, which are available for H_{red} and H_{ox}.[9−11,81,82] Given the proposed similarity in core structures between Q and H_{ox},[62] the H_{ox} crystal structure from MMO Bath was taken as the base QM/MM model.[9] MMOH is a dimeric protein in which each monomer comprises three polypeptides with $\alpha\beta\gamma$ stoichiometry. Including hydrogen atoms, MMOH contains ∼38 600 atoms. To make the QM/MM calculations manageable, we take advantage of the fact that the dinuclear iron centers are located entirely within the α-subunits and restrict our attention to only one such α-subunit (∼9400 atoms) in the calculations. Since the current QM/MM code is capable of handling at most 8000 MM atoms, the QM/MM model was further restricted to include only residues and crystallographic waters within 35 Å of the active site, leading to a final model with ∼7200 atoms. Charged surface amino acids not engaged in salt bridges were made neutral to mimic the screening effect of solvent. The QM region was set to include the two iron atoms, the water ligated to Fe1, the two bridging oxygen atoms, and the side chains of Glu114, Glu144, Glu209, Glu243, His147, and His246, for a total of ∼100 QM atoms with a net charge of zero. Last, the initial geometry of the core region was adjusted to resemble best the core geometry of intermediate Q from the purely QM calculations. A thorough QM/MM optimization of this final QM/MM model assured that an optimal protein structure for Q was established and that the MM configuration had settled into the bottom of one particular basin of attraction.

Examination of the protein structure reveals the substrate binding pocket to comprise principally of two halves (Figure 1). The first is formed by the iron atoms and coordinating ligands. The second side is formed by Leu110, Gly113, Ala117,

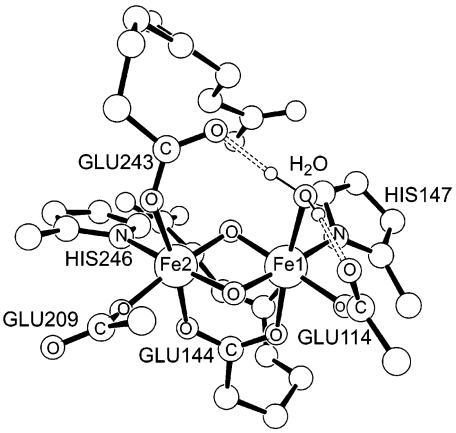


Figure 2. Model of Q used in purely QM calculations. Only residues coordinated to the iron atoms are labeled. Hydrogen atoms on the protein residues are omitted for clarity.

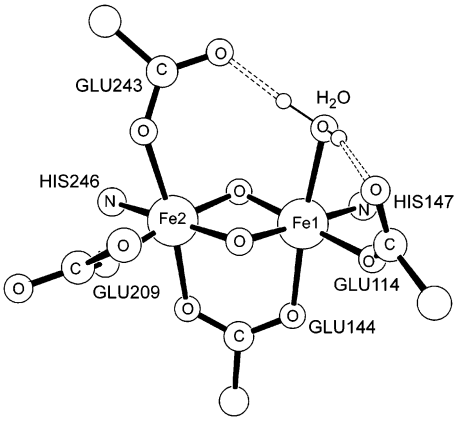


Figure 3. 50-Atom model of Q used for frequency calculations. Hydrogen atoms on the protein residues are omitted for clarity.

Phe188, Phe192, and Leu204, all which have their hydrophobic side chains directed toward the pocket, except for Gly113, the backbone carbonyl of which is directed inward.

All purely QM calculations for Q and the H-abstraction transition states were based upon the ∼100-atom model previously utilized by us in our study of methane hydroxylation in MMOH (Figure 2).[28] QM/MM H-abstraction transition states were obtained by inserting substrate into the hydrophobic cavity of the Q protein structure so that the geometry of substrate relative to the diiron core was the same as that in the purely QM calculations. QM/MM van der Waals complexes were then obtained by moving the abstracted H atom back to the equilibrium C−H distance and optimizing. Small structural adjustments were made during the geometry optimization protocol to achieve realistic energies along the reaction profile. By taking the QM regions from these two sets of calculations and capping the relevant bonds with H atoms, ∼75-atom QM/MM-derived quantum models were obtained for use in the RODFT QM calculations. Purely QM van der Waals complex geometries were the same as the QM/MM-derived quantum models of the van der Waals complexes. Frequency calculations were carried out with ∼50-atom models extracted from the purely QM geometries for Q, the van der Waals complexes, and the H-abstraction transition states (Figure 3).

V. Results

A. Overview. Free energies of the protein−substrate complexes and transition states relative to infinitely separated Q

(81) Rosenzweig, A. C.; Brandstetter, H.; Whittington, D. A.; Nordlund, P.; Lippard, S. J.; Frederick, C. A. *Proteins: Struct., Funct., Genet.* **1997**, *29*, 141−152.
(82) Elango, N.; Radhakrishnan, R.; Froland, W. A.; Wallar, B. J.; Earhart, C. A.; Lipscomb, J. D.; Ohlendorf, D. H. *Protein Sci.* **1997**, *6*, 556−568.

Table 1. Detailed Breakdown of the QM/MM Relative Free Energies (at 25 °C) for All Substrates Investigated[a]

	CH_4	CH_3CH_3	CH_3OH	CH_3CN	CH_3NO_2	CH_3F
Q-Substrate Complex						
ΔG_{TOT}	−0.44	−2.38	3.48	1.08	5.03	11.78
binding affinity (M^{-1})	2.10	55.5	2.81×10^{-3}	0.16	2.06×10^{-4}	2.32×10^{-9}
Purely QM Components						
$\Delta E_{quantum}$	−1.26	−2.26	−3.45	−5.58	−0.38	−3.33
ΔZPE	0.70	−1.45	−0.59	−1.38	0.14	−0.30
ΔH	0.83	0.66	0.33	0.64	0.53	0.92
ΔS	−28.96	−37.18	−39.83	−38.08	−48.12	−35.21
ΔG_{QM}	8.90	8.04	8.17	5.03	14.64	7.79
QM/MM Components						
$\Delta E_{QM/MM}$	−7.34	−8.59	−9.79	−7.84	−13.31	3.77
strain energy	2.95	−2.64	1.38	2.95	−2.32	−0.94
QM/MM electrostatic energy	−8.99	4.15	−5.47	−3.39	0.61	10.91
ΔE_{MM} total	−1.3	−10.1	−5.7	−7.4	−11.6	−6.2
ΔE_{MM} electrostatic	−3.7	−1.8	−7.6	−4.2	−8.1	−7.5
ΔE_{MM} van der Waals	0.81	−7.47	0.62	−4.21	−5.58	−0.63
Other Terms						
substrate desolvation energy	−2.00[79]	−1.83[79]	5.10[79]	3.89[79]	3.70[80]	0.22[79]
H-Abstraction Transition State						
ΔG_{TOT}[b]	15.32	12.15	16.67	14.59	23.20	23.62
Purely QM Components						
$\Delta E_{quantum}$	17.88	15.81	16.63	15.62	19.45	16.57
ΔZPE	−3.19	−3.48	−4.01	−3.48	−3.07	−3.08
ΔH	0.65	0.85	1.00	1.20	0.44	0.60
ΔS	−28.96	−37.18	−39.83	−38.08	−48.12	−35.21
ΔG_{QM}	23.97	24.27	25.50	24.69	31.17	24.59
QM/MM Components						
$\Delta E_{QM/MM}$	−6.65	−10.29	−13.93	−13.99	−11.67	−1.19
strain energy	−1.51	0.12	−0.31	−0.82	−2.13	−7.15
QM/MM electrostatic energy	1.36	−0.81	−3.32	−2.47	5.16	9.06
ΔE_{MM} total	−6.5	−9.6	−10.3	−10.7	−14.7	−3.1
ΔE_{MM} electrostatic	−5.7	−3.1	−9.4	−7.0	−9.5	−8.9
ΔE_{MM} van der Waals	−1.15	−6.71	−1.69	−4.69	−7.22	1.35
Other Terms						
substrate desolvation energy	−2.00[79]	−1.83[79]	5.10[79]	3.89[79]	3.70[80]	0.22[79]

[a] All energy values are expressed relative to Q + substrate (aq). Units are kcal/mol, except for entropy, which has units of cal/mol·K. [b] Hydrogen atom tunneling corrections are not included.

Table 2. Detailed Breakdown of the QM/MM Transition-State Energies (at 25 °C) Relative to the Q-Substrate Complexes[a]

	CH_4	CH_3CH_3	CH_3OH	CH_3CN	CH_3NO_2	CH_3F
ΔG_{TOT}[b]	15.76	14.53	13.19	13.51	18.17	11.84
Purely QM Components						
$\Delta E_{quantum}$	19.14	18.07	20.08	21.20	19.83	19.90
ΔZPE	−3.89	−2.03	−3.42	−2.10	−3.21	−2.78
ΔH	−0.18	0.19	0.67	0.56	−0.09	−0.32
ΔS	0.00	0.00	0.00	0.00	0.00	0.00
ΔG_{QM}	15.07	16.23	17.33	19.66	16.53	16.80
QM/MM Components						
$\Delta E_{QM/MM}$	0.69	−1.70	−4.14	−6.15	1.64	−4.96
strain energy	−4.46	2.76	−1.69	−3.77	0.19	−6.21
QM/MM electrostatic energy	10.35	−4.96	2.15	0.92	4.55	−1.85
ΔE_{MM} total	−5.2	0.5	−4.6	−3.3	−3.1	3.1
ΔE_{MM} electrostatic	−2.0	−1.3	−1.8	−2.8	−1.4	−1.4
ΔE_{MM} van der Waals	−1.96	0.76	−2.31	−0.48	−1.64	1.98
Other Terms						
substrate desolvation energy	0.00	0.00	0.00	0.00	0.00	0.00

[a] Units are kcal/mol, except for entropy, which has units of cal/mol·K. [b] Hydrogen atom tunneling corrections are not included.

and aqueous substrate, for all substrates investigated, are presented in Table 1, while energies of the transition states relative to the protein−substrate complexes can be found in Table 2. A detailed breakdown of the QM/MM free energies, including purely QM ZPE corrections and enthalpy and entropy terms, as well as the strain energy and components yielded from analysis of the QM/MM Hamiltonian, are also provided in the tables. Substrate binding affinities are given in Table 1. Figure 4 displays a reaction profile showing the relative free energies for each of these states as well as estimates (vide infra) of the

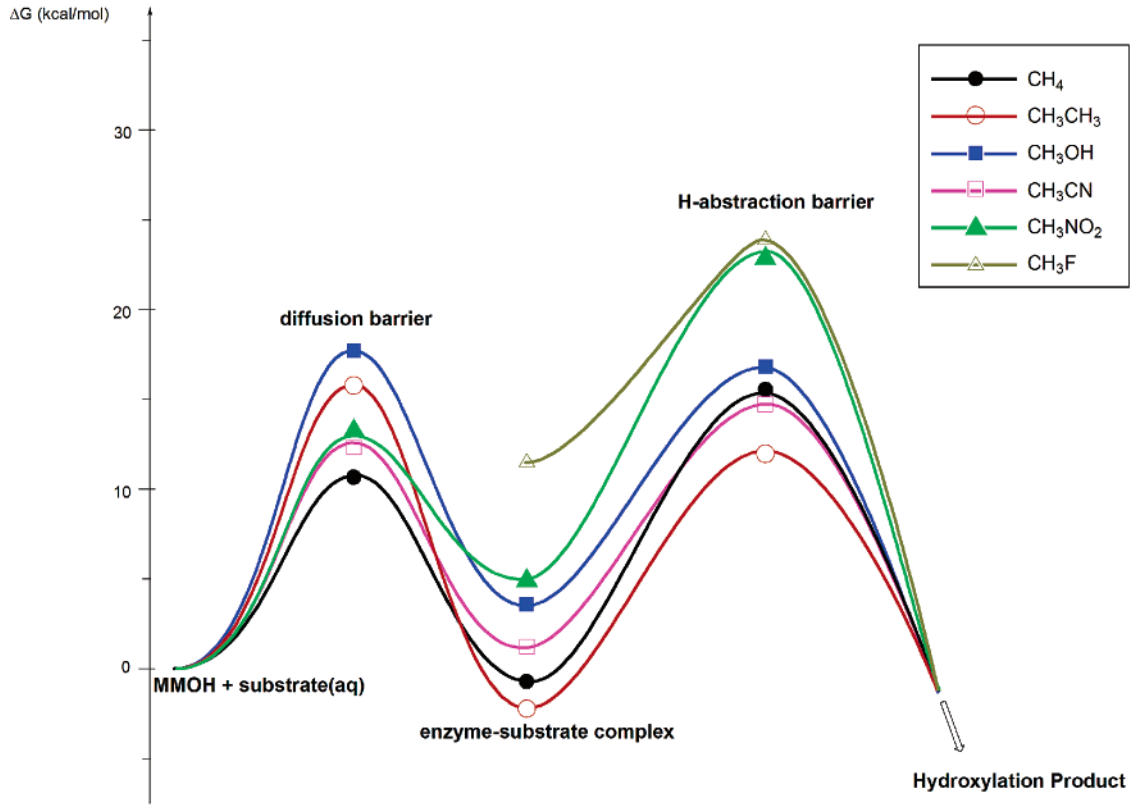


Figure 4. Profile for the reaction of Q with aqueous substrate. The diffusion barrier for CH_3F cannot be estimated because of a lack of experimental data for this substrate.

barrier for diffusion of the substrates into the protein. Geometric information (Table S1, Figure S1) and Mulliken spin populations (Table S2) for the purely QM hydrogen abstraction transition-state structures can be found in Supporting Information.

Having computed free energies for the various points on the thermodynamic cycle, we can now compare these results with the measured experimental rate constants.[19] Using the kinetic model in eq 4,

$$\mathrm{Q} + \mathrm{S} \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} \mathrm{QS} \xrightarrow{k_2} \mathrm{MMOH_{ox}} + \text{product} \qquad (4)$$

the experimentally observed rate for Q decay is in general given by eq 5.

$$k_{\mathrm{obs}} = \frac{k_2[\mathrm{S}]}{k_{-1}/k_1 + [\mathrm{S}]} \qquad (5)$$

The data reported in ref 19 fall into two categories, however. For methane, ethane, and methanol, the overall rate for the transformation of substrate in solution into product was measured (points B and E in Scheme 2). In these cases, with the substrate concentrations used in the experiments, $k_{-1}/k_1 \gg [\mathrm{S}]$ and k_{obs} simplifies to the expression in eq 6.

$$k_{\mathrm{obs}} = \frac{k_2[\mathrm{S}]}{k_{-1}/k_1} \qquad (6)$$

This observed rate corresponds to steps governed by the free-energy difference between the H-abstraction transition state and Q plus aqueous substrate. For acetonitrile and nitromethane, on the other hand, saturating concentrations of substrate were used, leading to a situation in which nearly all Q existed as the enzyme−substrate complex. In these cases, $[\mathrm{S}] \gg k_{-1}/k_1$, and the measured rate is equivalent to the rate constant k_2, corresponding to the steps characterized by the activation free energy starting from the bound enzyme−substrate complex and culminating in production of hydroxylated product (points C and E in Scheme 2). Finally, no experiments have yet been performed for methyl fluoride. The computational results we have obtained for this substrate therefore constitute a prediction of the expected reaction rate under various experimental conditions.

In comparing theory and experiment (Table 3), we convert the experimental rates to pseudo-free-energy barriers using transition-state theory according to eq 7,

$$\Delta G^{\ddagger} = -RT \ln\left(\frac{k_{\mathrm{obs}} h}{k_{\mathrm{B}} T \kappa}\right) \qquad (7)$$

where k_{obs} is the observed rate, $\Delta G^{\ddagger}$ is the activation barrier, k_{B} is Boltzmann's constant, T is temperature, κ is the transmission coefficient (taken to be unity), h is Planck's constant, and R is the gas constant. Theoretical values for k_{obs} and the corresponding pseudo-activation barriers are calculated for each substrate from the data in Tables 1 and 2. The ratio k_{-1}/k_1 is equal to the inverse of the equilibrium constant (K_{eq}) for protein−substrate complexation, which can be calculated from the free-energy change ΔG for this process via eq 8.

$$K_{\mathrm{eq}} = \exp\left(\frac{-\Delta G}{RT}\right) \qquad (8)$$

The k_2 term is obtained from the free-energy difference between

Table 3. Corresponding Theoretical and Experimental Free Energies of Activation and k_{obs} Values for Substrate Hydroxylation[90]

	theoretical		experimental[a]		
	$\Delta G^{\ddagger}$ at 25 °C (kcal/mol)	k_{obs} at 25 °C (s^{-1})	$\Delta G^{\ddagger}$ at 20 °C (kcal/mol)	k_{obs} at 20 °C (s^{-1})	KIE
CH_4 (0.7 mM)[b]	18.57	0.152	15.4	19.4	23.1−42[27,29]
CH_3CH_3 (0.95 mM)	16.23	7.85	15.3	24.5	1.00[22,29]
CH_3OH (15.1 mM)	19.13	0.059	15.7	11.5	1.01
CH_3CN	13.51	777	13.9	282	46.4
CH_3NO_2	18.17	0.30	16.2	5.34	8.1

[a] All k_{obs} values and KIEs are from ref 19 as well as additional references as noted. [b] For methane, hydrogen atom tunneling corrections have been included.

the protein−substrate complex and H-abstraction transition state according to eq 9.

$$k = \left(\frac{k_B T \kappa}{h}\right) \exp\left(\frac{-\Delta G^{\ddagger}}{RT}\right) \tag{9}$$

In computing the free energy of the protein−substrate complexes relative to the reactants (Q + CH_3X (aq)) and the transition states, we have had to make one simplifying approximation. Calculation of the entropy of the QS complex is complicated by the fact that, without the protein, the complex has several very low frequency vibrational modes, owing to the very weak interaction between the substrate and diiron core. These modes are not accurately represented by this calculation, however, because the protein provides a tight constraint on substrate motion, which is entirely missing in the simplified model. A full second derivative calculation including the protein would solve this problem, although even here the use of a harmonic approximation to compute entropies would be problematic. Since this capability is presently not available in QSite, we have approximated the entropy of the complex as being negligibly different from that of the transition state. The justification of such an approximation is straightforward. The principal contribution to the entropy difference between the infinitely separated enzyme and substrate and the transition state is the translational and rotational entropy restriction on the substrate in the latter. This restriction will be more or less equivalent in the QS complex. In addition, computation of entropies based upon a harmonic approximation, coupled with anharmonicities that are most likely present for the substrate within the active site and relatively limited sampling of the phase space for ligand binding within the active site pocket, leads to a systematic underestimation of the entropy and overestimation of the free energies (on the order of 2−3 kcal/mol) for the protein−ligand complexes and transition states.

For the three cases where overall rates were measured and k_{obs} is given by eq 6, comparison with the experimental data[19] must also address the observation that methane exhibits a very large KIE, whereas ethane and methanol exhibit essentially no KIE (Table 3). We have interpreted this result as indicating that hydrogen atom abstraction is the rate-determining step in substrate hydroxylation only for the case where a large KIE is observed. This assignment requires that a second process must be rate-determining in the remaining cases. A reasonable candidate for this second process is diffusion of the substrate into the active site of MMOH. The physical picture is that the native protein must undergo conformational fluctuations to admit the substrate and, furthermore, that the substrate may have to pass through a region of the protein where the interaction energy is unfavorable. Modeling of the diffusion process could in principle be carried out by molecular dynamics simulations; however, we do not pursue such calculations in the present article. Rather, we treat the alternative process in a phenomenological fashion, inferring the free-energy barriers for substrate diffusion from the experimental data and the QM/MM calculations. As will be shown later, there are some simple consistency checks that can be made in carrying out this analysis, which suggest that the approach is reasonable.

B. Methane. We begin with a discussion of the methane reaction (Figure S2), where the KIE indicates that hydrogen atom abstraction is rate-determining.[19] As can be seen in Table 3, the pseudo-free-energy barriers calculated from the experimental and theoretical k_{obs} agree to within 3.2 kcal/mol. This result includes a simple tunneling correction, discussed later. This discrepancy is somewhat larger than what we have observed in other QM-[8] and QM/MM-based[13,14] activation free-energy calculations. Furthermore, as will be seen later in the discussions of ethane and methanol, the barriers computed for all three cases where the protein−ligand binding affinity is a component of the overall free energy of activation are consistently too large as compared to experiment. Thus, we believe that there is a small, but definite, systematic error in the protocol that we have employed, leading to overestimation of the free energy of the hydrogen atom abstraction transition state. Possible sources of this systematic error can be enumerated as follows:

(1) Intrinsic errors in the DFT and/or QM/MM methodology, with the contributions from the latter probably being small.

(2) Failure to sample rigorously the protein−ligand complex phase space. Possible problems include both location of the lowest energy structure for the transition state, a quite difficult task requiring investigation of multiple initial guesses to achieve the current results, and accurate computation of the entropy of protein−ligand binding, which is probably underestimated by the simple harmonic approximation used here.

(3) Failure to model rigorously displacement of a bound water upon substrate binding. It is not clear whether a suitable water, i.e., one in the appropriate location with a high residence time for displacement, exists, but this possibility should be carefully investigated.

(4) Effects of the partner protein MMOB on the rate constants. Experimental data demonstrate that, for a key dioxygen activation step, the presence of MMOB reduces the barrier height by 2−3 kcal/mol compared to the reaction in the absence of MMOB.[83] Because our computational models do not include MMOB, there being at present no crystallographic data that

(83) Liu, Y.; Nesheim, J. C.; Lee, S. K.; Lipscomb, J. D. *J. Biol. Chem.* **1995**, *270*, 24662−24665.

would enable its effect to be accounted for in the calculations, its absence may contribute to the systematic bias toward calculating a slower rate constant than experimentally observed.

In view of the above issues, we regard the present agreement between theory and experiment as encouraging and consistent with our previous results.[12−14,77,78] A more extensive discussion of the comparison in the context of the overall methodology is presented in the Discussion section.

In the case of methane, where both hydrogen atom abstraction is rate-determining and k_{obs} and the KIE were determined for the reaction of substrate originating in solution, theoretical and experimental KIE values can be compared. The theoretical KIE is obtained by substituting CD_4 for CH_4 in the purely QM model, leading to significant changes in the ZPE term and smaller changes in H and S ($E_{quantum}$ remains unchanged). We assume that differences in the solvation energy and QM/MM interaction energy $E_{QM/MM}$ are negligible between CD_4 and CH_4. Based upon the difference in G_{QM}, then, the KIE can be directly determined as 10.44, assuming the transmission coefficient κ is equal to unity for both CD_4 and CH_4. The effects of quantum tunneling on κ can be estimated by using the Skodje and Truhlar approximation[84] (eqs 10−12)

$$\alpha = \frac{2\pi}{h\mathrm{Im}(\nu^{\ddagger})} \tag{10}$$

$$\beta = 1/k_B T \tag{11}$$

$$\kappa(T) = \frac{\beta\pi/\alpha}{\sin(\beta\pi/\alpha)} - \frac{\beta}{\alpha - \beta}\exp[(\beta - \alpha)(\Delta G^{\ddagger}_{TOT} - \Delta G_{TOT})] \tag{12}$$

where $\nu^{\ddagger}$ is the imaginary frequency associated with the reaction coordinate, $\Delta G^{\ddagger}_{TOT}$ is the energy difference between the transition state and protein−substrate complex, and ΔG_{TOT} is the endothermicity for radical intermediate formation estimated from ref 28.[85] The resulting ratio κ_H/κ_D is 2.57 at 25 °C for methane, yielding a final KIE of 26.79 when tunneling is included. By comparison, experimental KIE values range from 23 to 28 in MMO Bath[19,27] to 42 in the MMO OB3b system.[29] If the effect of hydrogen atom tunneling is then included in the theoretical value for k_{obs}, the final calculated rate of reaction for methane is $k_{obs} = 0.152\ s^{-1}$, which leads to a tunneling-corrected pseudo-activation energy of 18.57 kcal/mol for CH_4 (Table 3).

C. Ethane. In the case of ethane as substrate (Figure S3), the KIE measurement indicates that hydrogen atom abstraction is not the rate-determining step.[19,22,29] Experimental studies examining the effect of MMOB on substrate access to the active site,[29,31,32] furthermore, support the notion that substrate diffusion may be rate-determining in the case of ethane. The only comparison we can make between our calculation of the free energy of activation for hydrogen atom abstraction for ethane and the experimental data is to evaluate whether our results are consistent with the experimentally observed reaction rate for this substrate. The first step is to convert the measured rate at a specified concentration (chosen to be [S] = 0.95 mM) into an effective free energy of activation barrier, using a simple transition-state theory approximation in which a transmission coefficient of unity is assumed (eq 7). The value obtained, cf. Table 3, is 15.3 kcal/mol. Next, we use our QM/MM protocol (identical to that used for methane) to determine the pseudo-activation energy of hydrogen atom abstraction at the same concentration, and this yields a value of 16.24 kcal/mol at [S] = 0.95 mM. This result does not include tunneling corrections, because the simple approximation employed for methane cannot be used for ethane because of parameters lying outside its domain of applicability. We have therefore chosen at present not to pursue the use of more sophisticated computations to evaluate the tunneling correction, in part because of the inability to make direct contact with experiment for this substrate in any case. If the magnitude of the tunneling correction were equal to that obtained for methane, the resulting barrier of 15.19 kcal/mol would be very close to the experimentally determined barrier. One would further expect this value to be reduced by many of the same factors affecting the methane calculation, discussed earlier, thus leading to a result clearly consistent with the experimental observation that hydrogen atom abstraction is not rate-limiting for ethane.

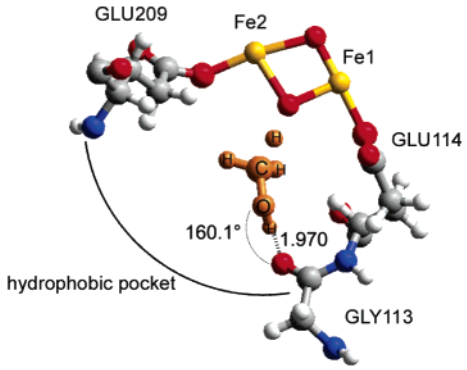


Figure 5. QM/MM structure for Q−CH_3OH hydrogen abstraction transition state. Single numbers indicate distances (angstroms). Color coding: C-gray, N-blue, O-red, H-white, substrate-brown.

The relative free energies of the ethane and methane Michaelis complexes and hydrogen abstraction barriers obtained from the calculations are likely to be more accurate than the absolute barriers, because of cancellation of the systematic sources of error outlined earlier. The principal reasons for the calculated differentials are that, versus methane, ethane has (1) increased van der Waals interactions with the hydrophobic side chains of the residues (e.g., Leu110, Ala117, Phe188, Phe192, Leu204) comprising the substrate binding pocket, (2) a more positive free energy of solvation because of the hydrophobic effect, and (3) a lower barrier to bond breaking in the electronic structure calculations.

D. Methanol. We next consider the case of methanol (Figures 5, S4), for which it is first necessary to examine the transition-state structure that is formed in our QM/MM computations. In ref 19, we hypothesized that the apparently lower free-energy barrier to hydrogen atom abstraction displayed by methanol, to the point where diffusion becomes rate-limiting, was due to the ability of the −OH group to serve as a hydrogen bond donor with the protein, thus lowering the transition-state free energy. This hypothesis is confirmed by the QM/MM calculations. Figure 5 depicts the transition state that is formed when methanol is the substrate. It can be seen that methanol forms a hydrogen bond with a perfectly positioned backbone carbonyl

(84) Skodje, R. T.; Truhlar, D. G. *J. Phys. Chem.* **1981**, *85*, 624−628.
(85) Small changes to the endothermicity in the calculation of the transmission coefficient lead to trivial changes in the KIE for methane, making such an approximation very reasonable.

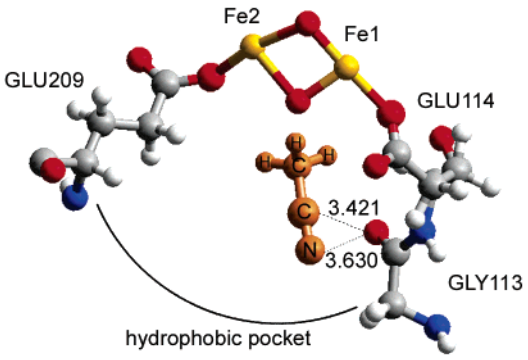


Figure 6. QM/MM structure for Q−CH_3CN complex. Single numbers indicate distances (angstroms). Color coding: C-gray, N-blue, O-red, H-white, substrate-brown.

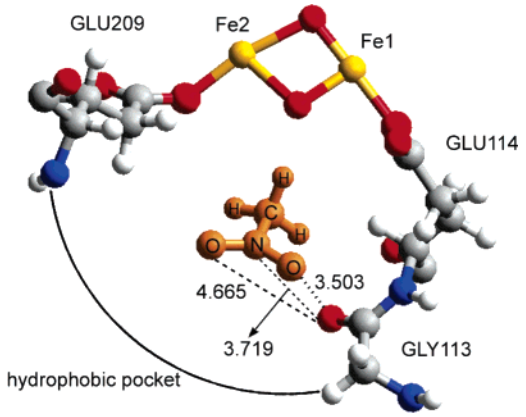


Figure 7. QM/MM structure for Q-CH_3NO_2 complex. Single numbers indicate distances (angstroms). Color coding: C-gray, N-blue, O-red, H-white, substrate-brown.

associated with Gly113. This interaction is reflected in the more favorable QM/MM electrostatic interaction energy at the hydrogen abstraction transition state for methanol as compared to the hydrocarbon substrates (Table 1).

The favorable interaction of methanol with the protein does not result in a lower calculated activation barrier for the overall hydroxylation reaction of this substrate compared to methane, however, owing to the presence of a large number of other differences in relative energetics. The most important of these is the solvation free-energy term. Methanol is much better solvated by water than methane, and this effect more or less cancels the better electrostatic interactions in the protein. The total barrier height depends on a complex combination of factors, which are detailed in Table 1.

Comparison of the methanol results with experiment can be carried out in a fashion analogous to the analysis for ethane. The observed experimental rate for a given [S] (here, 15.1 mM) is converted to an effective free-energy barrier, in this case 15.7 kcal/mol (Table 3). The calculated effective activation barrier at this substrate concentration is computed to be 19.1 kcal/mol, which is too large as compared to the experimental value. However, the theoretical number again does not include a tunneling correction, and is likely to be subject to the same ∼2−3 kcal/mol systematic errors as in the methane calculation. When these factors are taken into account, the methanol results are qualitatively consistent with the experimental data. The larger error may be related in part to the fact that obtaining an optimal geometry for methanol within the substrate binding pocket proved to be a particularly challenging conformational search problem, complicated by the need to accommodate both the methanol-Gly113 hydrogen bond and van der Waals interactions of the $-CH_3$ group with the protein (specifically hydrophobic residues lining the binding pocket). Although the structures presented here represent the most energetically favorable methanol/protein configuration emergent from the current calculations, the existence of a binding mode for methanol that would be significantly lower in energy is a definite possibility. Finding such a geometry would, however, require an extended conformational search, including freedom of movement for the hydrophobic side chains comprising the substrate pocket, which was not carried out for the present work.

E. Acetonitrile and Nitromethane. The cases of acetonitrile (Figures 6, S5) and nitromethane (Figures 7, S6) as substrates provided the challenge of locating the correct position for their polar groups within the active site pocket, which required a nontrivial conformational search. These groups may confer highly unfavorable interactions with the backbone carbonyl oxygen of Gly113, to which methanol, with its polar hydrogen bond donor, forms a strong hydrogen bond. Careful investigation of the possible positioning of these molecules yielded the protein−substrate complexes depicted in Figures 6 and 7, the motifs of which are rather similar from an electrostatic point of view. In the case of acetonitrile, the nitrile group is positioned so that it packs in a side-on fashion to the Gly113 carbonyl oxygen, with the positive C and negative N roughly equidistant from the oxygen atom, the carbon being the slightly closer of the two. If instead the group is oriented so that the linear nitrile group is aimed at the carbonyl oxygen, with the N atom consequently much closer to the O than the C, the interaction is strongly unfavorable. In nitromethane, the positive N is counterbalanced by a negative charge spread equally over the two oxygen atoms of the nitro group. Nitromethane takes advantage of these electronic features by orienting in the binding pocket such that the N and one O atom are nearly equidistant to the carbonyl oxygen, while the other O atom is ∼1 Å more distant. The result is a net partial positive charge relatively closer to, and a partial negative charge relatively farther from, the carbonyl oxygen atom.

It is not possible quantitatively to assess how accurately the QM/MM model evaluates the unusual interactions between the substrate and the enzyme active site described earlier. Because the experimental rate constants for these two substrates assume the enzyme−substrate complex as a starting point, the interaction of the $-CN$ or $-NO_2$ groups with the protein is similar in both the reactants and the transition state, and hence the energy associated with these interactions will cancel when computing the free energy of activation. Because the interaction is between QM (substrate) and MM (protein) regions, we cannot be as confident of our present results as we would be in a fully QM calculation of the strength of the interaction. This problem could be overcome in principle by treating the interacting atoms in the protein at the QM level. To do so would have added significantly to the computational cost, however, and given the unavailability of direct experimental data for comparison, this was not attempted in the present study. Nonetheless, we are confident that the structures obtained are reasonable and that the accommodation of the ligand to the protein structure occurs more or less as presented in Figures 6 and 7.

We now turn to the calculation of the experimentally measured rate constants for acetonitrile and nitromethane, which involves abstraction of a hydrogen atom from the substrate in the QS complex. Agreement with experiment for these two cases is quite reasonable (cf. Table 3), comparable to what we have seen in a wide range of other calculations of activation free energies that do not require computation of protein−ligand binding affinities (maximum error ∼2 kcal/mol), and suggests that the approximation made with regard to the treatment of entropy discussed earlier is valid.

Although the magnitude of the imaginary frequency for the transition states for CH_3CN and CH_3NO_2 does not permit their KIEs to be calculated by using eqs 10−12,[84] their relative experimental KIEs (Table 3) can be rationalized based upon the QM/MM calculations. Small molecule calculations show formation of the $^{\bullet}CH_2CN$ radical from acetonitrile to be 5 kcal/mol less endothermic than formation of the $^{\bullet}CH_2NO_2$ radical from nitromethane. Taking into account that the barrier height versus the QS complex is 4.66 kcal/mol higher for nitromethane (Table 2), the radical intermediate for the acetonitrile case should be ∼5−10 kcal/mol lower in energy than for the nitromethane case. The more symmetric barrier then for acetonitrile would favor hydrogen atom tunneling more than for nitromethane, in accord with the ∼6-fold greater KIE seen experimentally for CH_3CN versus CH_3NO_2.

As noted earlier in our discussion of the energies of interaction of acetonitrile and nitromethane with the protein cavity, we cannot directly compare the overall free energies of activation calculated for these substrates (Table 2) with experimental data. We consider it likely that the errors in these calculations are somewhat larger than those presented earlier for methane and ethane, because of the greater difficulties in positioning acetonitrile and nitromethane in the cavity, the aforementioned problems in computing accurate protein−substrate interaction energies, and the greater chemical complexity of these substrates as compared to the simple hydrocarbons, which may lead to larger errors in estimating entropies due to inadequate sampling and treatment of anharmonicity. These factors have a much smaller effect on the rate constants obtained starting from the Michaelis complex, because there will be significant cancellation of error between the reactant complex and the transition state. In the future, it will clearly be necessary to confront theory and experiment directly for total free-energy calculations for a wider range of systems, quantify the errors in our current protocol, and design improved methods that can overcome the limitations of the approach taken here.

F. Methyl Fluoride. Finally, we have carried out calculations for methyl fluoride, for which there are presently no experimental data. In this case, unlike for acetonitrile and nitromethane, there is no way geometrically for the fluorine atom to escape unfavorable electrostatic interactions with the carbonyl oxygen of Gly113 (Figure 8, S7). Therefore, we predict that the overall activation free energy for methyl fluoride should be substantially higher than for all of the substrates other than nitromethane, which has an unusually high barrier, close to that of methyl fluoride, principally due to a larger entropy loss upon binding than is the case for the other substrates. We also predict a KIE close to the classical KIE with little contribution from hydrogen atom tunneling based upon the high reaction barrier and relative endothermicity for formation of the $^{\bullet}CH_2F$ radical from CH_3F, which is nearly equal to that of $^{\bullet}CH_2NO_2$ from CH_3NO_2.

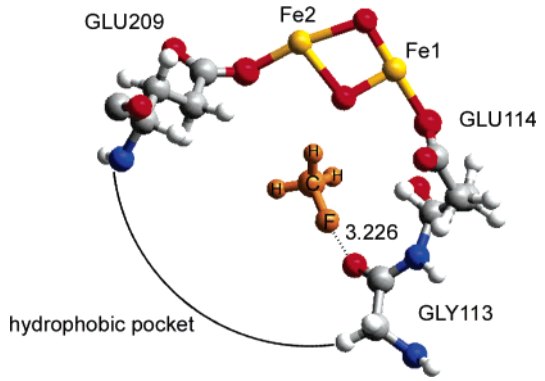


Figure 8. QM/MM structure for Q−CH_3F complex. Single numbers indicate distances (angstroms). Color coding: C-gray, N-blue, O-red, H-white, substrate-brown.

VI. Discussion

The present results provide strong support for the basic MMOH hydrogen atom abstraction mechanism depicted in Figure 5. Although a number of computational groups working on MMOH have converged on a model of this type,[33−35] there are also alternative proposals in the literature, such as those of Yoshizawa,[37−42] which invoke direct interaction of the metal with the substrate as the key step in the catalytic process. However, to our knowledge, those proposing structures along these lines have been unable to obtain from a detailed model results for the activation barrier that are low enough to be qualitatively compatible with the experimental data. By demonstrating that for three different substrates (methane, acetonitrile, nitromethane) reasonable quantitative agreement with experiment is obtained when a realistic computational model is employed, and that the various substrates in the protein cavity are appropriately positioned for reaction employing the proposed mechanism, confidence that the mechanism is in fact correct is substantially enhanced.

We turn next to the question of the overall quantitative accuracy of our hybrid DFT-based QM/MM computational methodology, in the context of the results presented in this article, considering as well the other work cited in section III. There are two major aspects of reproducing experimental kinetic data for enzymatic reactions: (1) calculation of the binding affinity of the substrate to the enzyme to form the enzyme−substrate complex and (2) determination of the activation barrier between the enzyme−substrate complex and the transition state. In the present article, the only quantitative comparison that can be made involving both (1) and (2) is that for methane; the experimental rate constants for acetonitrile and nitromethane involve (2) only. With regard to previous work,[14] the binding of dioxygen to hemerythrin involved only step (1) (only equilibrium free energies were computed, as opposed to a barrier height), whereas the remaining systems (β-lactamase, triose phosphate isomerase, and other MMO reactions) involved exclusively evaluation of step (2), because the starting point was always a specific enzyme−substrate complex of some sort. Thus, most of the results we have accumulated to date have been targeted at step (2), whereas only two examples involve explicit computation of protein−ligand binding affinities.

We begin with a discussion of the calculation of activation barriers starting from a protein−ligand complex (step 2 above). An initial question of fundamental importance is what sort of accuracy one can expect from hybrid DFT methods (specifically B3LYP) for transition state activation barriers in the systems under study. Certainly B3LYP can yield poor results for activation barriers; for example, a recent study by Truhlar and collaborators[86] of 22 small molecule radical reactions (and 40 barrier heights in all, since forward and reverse reactions can be considered independently) yields a mean unsigned error (MUE) of 4.82 kcal/mol and a systematic underestimation of the barrier heights as evidenced by a mean signed error of −4.75 kcal/mol, distributed across virtually all of the reactions under study. On the other hand, Houk and co-workers[87] reported comparisons of B3LYP-based barriers with experiment for nine pericyclic reactions, obtaining an MUE of 1.7−2.9 kcal/mol (depending upon basis set) and a mean signed error of 1.5 kcal/mol larger than experiment. Furthermore, for eight of the nine reactions, errors were uniformly small; for example, for the 6-31G* basis set, there is a maximum error of 3 kcal/mol and an MUE of 1.2 kcal/mol for the eight remaining reactions. Performance for metal-containing systems is also an issue; however, there are little reliable experimental data on reaction barriers for small molecules, and thus a rigorous assessment of this issue is difficult. The enzymatic reactions we have investigated are arguably quite different from the small molecule radical reactions in Truhlar's data set; thus far, the results are much closer to those obtained in ref 87 than those in ref 86 (particularly if the outlier in ref 87 is eliminated), which is reasonable based on the key physical characteristics of each set of test cases. However, there are almost certainly some enzymatic reactions that will fall into the "difficult" regime and display large errors, and therefore each new system must be approached with caution. On the other hand, both our results and those of Houk and co-workers suggest that there is also a "normal" regime where current hybrid DFT does very well and can be used in those cases to discriminate effectively alternative mechanisms and understand details of the reaction and that a substantial number of systems are likely to be in this category.

As we have suggested previously, there are a significant number of other possible sources of error in generating a final number for step 2: use of simplistic harmonic approximations for zero-point energies, neglect of tunneling and dynamical effects of the protein, possible sampling errors, and problems with the QM/MM interface and use of UDFT corrections via a cluster model to the RODFT-based QM/MM calculations. We have extensively tested the QM/MM interface errors and believe that they are small compared to the intrinsic DFT errors. The UDFT/RODFT protocol is potentially problematic, primarily when the structures of the local transition-state region are strongly affected by the treatment of spin; that is not the case for the present calculations or those that we have reported previously. In fact, superposition of the UDFT and RODFT models for Q and the transition states shows negligible differences in the geometries of the Fe_2O_2 core and the atoms coordinated to the metal centers, indicating that the geometry around the iron atoms does not depend on the methodologies employed here. Differences in the peripheral regions of the two models can be directly attributed to the protein environment, since the UDFT models were obtained through purely QM optimizations, whereas the RODFT models were extracted from the QM regions after QM/MM optimizations. The zero-point approximation is often used for small molecule reactions, and there is no particular reason to think it is less accurate in the present case. Inclusion of tunneling would, as shown by the methane example, probably bring theory and experiment into better agreement, assuming the magnitude of the tunneling corrections is similar for other substrates. The effects of protein dynamics, and possible sampling errors on the protein surface, are harder to judge. We have made a serious attempt to investigate possible ligand placements in the active site, but cannot at present guarantee that we have found the lowest energy position in all cases. However, positioning is most likely to affect step (1) rather than step (2), because some of the error associated with an inaccurately positioned ligand would cancel between the reactant complex and the transition state. Finally, the effect of protein dynamics on enzymatic reactions is at present highly controversial; there are quite reasonable arguments that the effects are small (see for example ref 88, which adopts precisely this position), and our results to date can be taken to support this point of view. When the reaction is highly localized, as in the case of a hydrogen atom transfer, it is reasonable to hypothesize that long range protein entropy of the reactant and transition state is similar (and hence cancel in the computation of the free-energy barrier) and that shorter range effects are plausibly described by the harmonic approximation using a cluster model extracted from the QM/MM structure (the protocol employed in the present article). A related, and probably more relevant, question is whether the reaction can proceed through multiple configurations, as opposed to the particular stationary points we have located, and what the effect would be of averaging over these configurations. There are little quantitative data in the literature regarding the effects of protein conformational averaging on transition-state barriers. One article that considers a related question is a study by Warshel and co-workers of redox potentials of plastocyanin and rustocyanin.[89] These authors report that conformations generated in the course of MD simulations vary over a range of 0.09 eV (2.5 kcal/mol) for both reactions. Redox potentials, which involve a difference of a full electronic charge on a metal center in the active site, can be expected to have a much larger sensitivity to protein conformational changes (via Coulombic interactions) than will the localized hydrogen atom transfer that we are considering in MMO. Furthermore, the expected "error" involved in using stationary points cannot be simply extrapolated from the variation in redox potential of conformations generated in an MD simulation. The Boltzmann factor guarantees that the lowest energy states (which we have endeavored to locate) will make substantially larger contributions to the free energy than will initial states with higher energies. Given these two factors, our belief is that the error arising from neglect of conformational sampling is, in the present case, significantly less than the 2.5 kcal/mol value cited earlier.

(86) Lynch, B. J.; Fast, P. L.; Harris, M.; Truhlar, D. G. *J. Phys. Chem. A* **2000**, *104*, 4811−4815.
(87) Guner, V.; Khuong, K. S.; Leach, A. G.; Lee, P. S.; Bartberger, M. D.; Houk, K. N. *J. Phys. Chem. A* **2003**, *107*, 11445−11459.
(88) Warshel, A. *Acc. Chem. Res.* **2002**, *35*, 385−395.
(89) Olsson, M. H. M.; Hong, G. Y.; Warshel, A. *J. Am. Chem. Soc.* **2003**, *125*, 5025−5039.
(90) Theoretical free energies and rate constants are calculated at 25 °C rather than 20 °C, since the experimental values (see refs 79 and 80) for the free energies of solvation for the substrates were determined at 25 °C.

With regard to step 1, the harmonic approximation for the entropy of binding is almost certainly problematic for larger and more complex substrates. The agreement with experiment obtained for methane may well be reflective of its small size and simplicity, as well as the small size of the MMO cavity. A similar argument could be made for our success in reproducing the binding affinity of dioxygen to hemerythrin (ref 14). Given that there are only two points of comparison with experimental data for step 1, no conclusions of statistical significance can be drawn. We defer an extensive discussion of binding affinity computation to other publications; suffice it to say here that in the future considerably more sophisticated methods will almost certainly be required to succeed in the general case, when problems such as displacements of substantial numbers of waters from the active site will have to be addressed.

To summarize, the QM/MM methodology is primarily directed at computing activation barriers starting from an enzyme−substrate complex (step 2), and, for this problem, encouraging results have been obtained. Success in treating step 1 for several very small substrates is somewhat surprising, and future investigations will likely reveal the necessity of a more rigorous approach.

VII. Conclusion

We have carried out QM/MM calculations on the active site of MMOH and investigated the activation free energy of hydrogen atom abstraction for CH_3X (X = H, CH_3, OH, CN, NO_2, F) substrates for this enzyme. For two of these substrates (acetonitrile and nitromethane), the corresponding experiments are performed under saturating conditions, and the appropriate comparison with experiment is the free energy of activation of the reaction starting from the bound enzyme−substrate complex. For both substrates, we achieve good agreement with experiment (1−2 kcal/mol), which is comparable to results we have obtained when computing free-energy barriers for other MMOH reactions[8] and also in a number of other systems.[14,77,78]

For methane, the experimental rate is proportional to the substrate concentration, indicating that the latter is not saturating with respect to protein−ligand binding at the highest achievable concentrations in aqueous buffer. It is therefore necessary to include the free energy of binding to compare theory with experiment. In this case, the pseudo-free energy of activation determined from the calculated value for the reaction rate k_{obs} is in reasonable agreement with the experimental data (∼3 kcal/mol). This difference is somewhat larger than what is seen for the calculations on substrates that achieve saturating concentrations, however. Moreover, the calculated barriers for ethane and methanol, which also exhibit pseudo-first-order experimental kinetics, similarly appear to be systematically higher than the measured values. The increased discrepancy is most likely due to a systematic error in computation of the protein−ligand binding free energies; however, establishing the source of error in a rigorous fashion, and devising computational methods to reduce it, will require substantial further work. When the hypothesized systematic bias, as well as tunneling effects, are taken into account, the results for methanol and ethane are in qualitative agreement with the experimental observation that hydrogen atom abstraction is not rate-limiting for these reactions. Again, substantial further effort is required to model this process.

Inclusion of the protein environment was essential in calculating the absolute free energies of hydrogen atom abstraction and in understanding the differences in kinetic behavior observed for the various substrates. The van der Waals interaction between the protein and substrate plays a key role in stabilizing the transition state. Inclusion of electrostatic interactions is required to understand the experimental results for polar substrates. The success of the QM/MM methodology from both a qualitative and quantitative point of view further confirms that our technology can provide a useful detailed atomic level picture of metalloenzyme catalysis, particularly when the reaction involves a transformation from one protein−substrate complex to another. A more accurate treatment of the protein−substrate binding component of the activation free energy, as indicated earlier, requires new methods development, with the most likely area for improvement being the incorporation of a more extensive sampling algorithm into the computational protocol.

Acknowledgment. This work was supported by Grant GM40526 (to R.A.F.) and GM32134 (to S.J.L.) from the NIH and by an NSF/DOE funded EMSI Grant (CHE-98-10367) to R.A.F. Computational resources were provided by NPACI (to R.A.F.). We acknowledge support from the NIH through a Biophysics Training Grant to B.F.G. We thank Dr. Victor Guallar for fruitful discussions.

Supporting Information Available: Tables S1−S2, Figures S1−S7, and tables of atomic coordinates and vibrational frequencies for each purely QM structure (PDF). Computed protein structures are available upon request from rich@chem.columbia.edu. This material is available free of charge via the Internet at http://pubs.acs.org.

JA049847B