J|A|C|S
ARTICLES
Published on Web 03/11/2003

Reversible Dioxygen Binding to Hemerythrin

Maria Wirstam,[†,‡] 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 December 6, 2001; Revised Manuscript Received January 21, 2003; E-mail: rich@chem.columbia.edu, lippard@lippard.mit.edu

Abstract: A combination of conventional quantum chemical methods and a recently developed mixed quantum mechanical/molecular mechanical (QM/MM) method (QSite) is used to determine the different energetic components involved in reversible binding of O_2 to hemerythrin. The use of an accurate quantum chemical description of the active site and the inclusion of effects from the surrounding protein environment are both essential to achieve reversibility and thus to model accurately the binding of O_2 to the carboxylate-bridged diiron center in the protein. The major contributions from the protein environment stabilizing dioxygen binding are (1) the van der Waals interaction between the bound dioxygen and the protein atoms and (2) an increase in the hydrogen bonding energy of an imidazole group, ligated to one of the iron atoms, and a neighboring carboxylate side chain in the second coordination sphere. The protein strain energy for this system is negligible. The calculated total O_2 binding free energy is in good agreement with that derived from the experimental equilibrium constant.

I. Introduction

Hemerythrin (Hr) proteins function as dioxygen carriers for certain marine invertebrates. Found in several species and oligomeric forms, Hr proteins have chemically similar active sites despite sequence differences.[1] The active site structures of the deoxy and oxy forms of Hr, obtained by X-ray crystallography,[2] are depicted in Figure 1. The core contains two iron atoms connected by two carboxylate groups and an oxygen-containing ligand derived from water.[3] The Fe−O bond distances,[2,4] and the measured magnetic exchange coupling constant[5] of the deoxy form, indicate that the bridging oxygen ligand is a hydroxo group. In addition, there are five histidine ligands bound to terminal coordination positions on the two iron atoms and a glutamate amino acid in the second coordination shell that forms a hydrogen bond to one of the histidine ligands. Apart from the components shown in Figure 1, there are only hydrophobic amino acids in the active site region.

In deoxyHr the iron atoms are both in the Fe(II) oxidation state.[6,7] When dioxygen is added, the Fe−O bond distances shorten[2] and the magnetic coupling increases,[8] indicating formation of an oxo bridge. The presence of a μ-oxo ligand in oxyHr is further supported by absorption spectra.[9] There is experimental evidence that the proton on the bridging hydroxide ion in the deoxy form transfers to and stabilizes the bound peroxo ion,[10] with concomitant oxidation of both iron atoms to

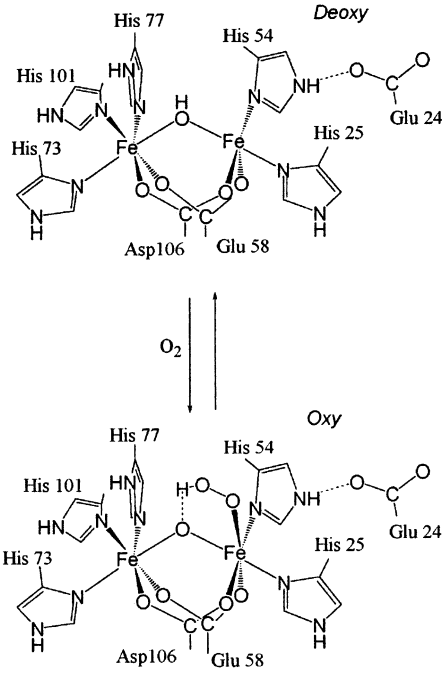


Figure 1. Active sites of the deoxy and oxy forms of hemerythrin.

[†] Columbia University.
[‡] Current address: Biovitrum AB, SE-112 76 Stockholm, Sweden.
[§] Massachusetts Institute of Technology.

(1) Stenkamp, R. E. *Chem. Rev.* **1994**, *94*, 715−726.
(2) Holmes, M. A.; Letrong I.; Turley S.; Sieker L. C.; Stenkamp R. E. *J. Mol. Biol*. **1991**, *218*, 583−593.
(3) Shiemke A. K.; Loehr T. M.; Sanders-Loehr J. *J. Am. Chem. Soc*. **1984**, *106*, 4951−4956.
(4) Zhang K.; Stern E. A.; Ellis F.; Sanders-Loehr J.; Shiemke A. K. *Biochemistry* **1988**, *27*, 7470−7479.
(5) Brunold T. C.; Solomon E. I. *J. Am. Chem. Soc*. **1999**, *121*, 8277−8287.
(6) Garbett. K.; Darnall, D. W.; Klotz, I. M.; Williams, R. J. *Arch. Biochem. Biophys*. **1969**, *135*, 419−434.
(7) York, J. L.; Bearden A. J. *Biochemistry* **1970**, *9*, 4549−4554.
(8) Dawson, J. W.; Gray, H. B.; Hoenig, H. E.; Rossman, G. R.; Schredder, J. M.; Wang, R. H. *Biochemistry* **1972**, *11*, 461.
(9) Garbett, K.; Darnall, D. W.; Klotz, I. M.; Williams, R. J. *Arch. Biochem. Biophys*. **1969**, *135*, 419.
(10) Shiemke, A. K.; Loehr, T. M.; Sanders-Loehr, J. *J. Am. Chem. Soc*. **1986**, *108*, 2437−2443.

10.1021/ja017692r CCC: $25.00

Fe(III).[11] An increased understanding of the mechanism of reversible dioxygen binding and of the electronic structures of the two forms of Hr has been provided by extensive studies of synthetic models that mimic the active site.[12,13] These features have also been investigated by molecular orbital studies of small model structures.[5]

Hemerythrin binds dioxygen reversibly (eq 1 and Figure 1). The task of binding O_2 reversibly requires that the free energies of the oxy and deoxy forms of Hr must be within a few kcal/mol of each other at room temperature.

$$\underset{\text{DeoxyHr}}{[\text{Fe(II)}]_2} + \text{O}_2 \rightleftharpoons \underset{\text{OxyHr}}{[\text{Fe(III)}]_2} \qquad (1)$$

To achieve such a precise energetic range for O_2 binding affinity is nontrivial and there is currently no detailed understanding of the features of the system that are critical to tuning the energetics in this manner. In this paper, we elucidate these features by employing ab initio quantum chemistry based techniques to compute the free energy difference between the oxy and deoxy forms of Hr. Our approach is based on a mixed quantum mechanics/molecular mechanics (QM/MM) methodology, developed by coupling the Jaguar suite of ab initio electronic structure programs[14] with the IMPACT protein modeling program,[15] which is implemented in what we now refer to as the QSite program.[16,17] Studies of quantum chemical small models, in conjunction with full QM/MM calculations that realistically model the protein, are performed to dissect key features controlling the free energy differential between the oxy and deoxy forms of Hr.

The remainder of this paper is organized as follows. In section II, we discuss the Hr models and the computational strategy. In section III, we briefly review the quantum chemical and mixed QM/MM methods; the algorithmic details of these approaches have been described elsewhere.[16,17] In section IV, results are presented for purely quantum chemical calculations on model systems, and in section V, QM/MM calculations on the entire protein are discussed. Section V further contains a computation of the absolute free energy of binding using a thermodynamic cycle, which allows direct comparison with experimental measurement of this quantity. Remarkably good agreement between theory and experiment is obtained. Finally, section VI, the conclusion, summarizes the results presented herein.

II. Overview of Hr Models and Computational Strategy

We have investigated dioxygen binding to Hr using both purely QM models of the active site (containing 63−65 atoms) and by QM/MM calculations incorporating an entire Hr subunit. Interaction between subunits most likely has only a very small effect upon dioxygen binding. These calculations are complementary, and information from both is combined, as discussed below, to generate a final result for the binding free energy and to obtain physical insight into the role of the protein in stabilizing the oxy form. In both cases, the B3LYP hybrid density functional is used as the core quantum chemical technology.

The QM/MM calculations are carried out using a double-ζ plus polarization (DZP) basis set of moderate accuracy (6-31G*) and employing a restricted open shell version of DFT (RODFT). The DZP basis is suitable for geometry optimization, which requires the great majority of the computation time. RODFT methods are used because we have not yet implemented unrestricted DFT (UDFT) calculations in the QM/MM methodology. Care has to be taken to make sure that the MM parts of the oxy and deoxy structures are converged to the same local basin of attraction; our approach to accomplishing this goal is discussed below.

After converged structures are obtained, a QM model of the active site is constructed and studied using UDFT methods and large basis sets. These methods provide accurate single point energies, within the limitations of the B3LYP DFT functional, which can be employed to compute energy differences of the model system. Entropy effects are estimated by calculating vibrational frequencies using a smaller model system and basis set, to save computation time. These data are then combined with the QM/MM data to compute the overall free energy difference between the oxy and deoxy forms of the protein. A thermodynamic cycle, taking into account the experimental solvation free energy of dioxygen in aqueous solution, is required to obtain the final numbers. Spin couplings have also been calculated by using the QM models and are compared with experimental data for these quantities.

Finally, we have systematically investigated the accuracy of the QM/MM interface in two ways: (1) using a model small enough to be treated by QM calculations, systematically increasing the number of MM side chains, and observing the effects of adding MM side chains on dioxygen binding to the model; (2) increasing the size of the QM region in the QM/MM protein calculations and again observing the differences in dioxygen binding energetics. Both types of errors are quite small; details of this work will be discussed elsewhere, in the context of a more general assessment of the accuracy of QSite.

In what follows, we begin in section III by describing our QM and QM/MM computational methods in more detail. Results are presented in section IV, first for model system calculations and then for the full protein calculations.

III. Computational Methods

Quantum chemical calculations were performed with the Jaguar suite of ab initio electronic structure programs, using the density functional (DFT) approach to electron correlation. Jaguar employs pseudospectral (PS) methods that substantially accelerate DFT calculations for large molecules, as documented previously.[18] Jaguar also contains specialized methods for constructing a suitable initial guess for transition metal containing systems,[19] a technology that was essential in carrying out the present studies.

(11) Stenkamp, R. E.; Sieker, L. C.; Jensen, L. H., McCallum, J. D.; Sanders-Loehr, J. *Proc. Natl. Acad. Sci. U.S.A.* **1985**, *82*, 713−716.
(12) He, C.; Barrios A. M.; Lee, D.; Kuzelka, J.; Davydov, R. M.; Lippard, S. J.; *J. Am. Chem. Soc.* **2000**, *122*, 12683−12690.
(13) Mizoguchi, T. J.; Kuzelka, J.; Spingler, B.; DuBois, J. L.; Davydov, R. M.; Hedman, B.; Hodgson, K. O.; Lippard, S. J. *Inorg. Chem.* **2001**, *40*, 4662−4673.
(14) *Jaguar 4.1*; Schrödinger, Inc.: Portland, OR, 2000.
(15) *IMPACT Version 17014*; Schrödinger, Inc.: Portland, OR, 2000.
(16) Philipp, D. M.; Friesner, R. A. *J. Comput. Chem.* **1999**, *20*, 1468−1494.
(17) Murphy, R. B.; Philipp, D. M.; Friesner, R. A. *J. Comput. Chem.* **2000**, *16*, 1442.
(18) Murphy, R. B.; Cao, Y.; Beachy, M. D.; Ringnalda, M. N.; Friesner, R. A. *J. Chem. Phys.* **2000**, *112*, 10131−10141.
(19) Vacek, G.; Perry, J. K.; Langlois, J.-M. *Chem. Phys. Lett.* **1999**, *310*, 189−194.

QM model calculations were performed using spin-unrestricted DFT methods (UDFT), and broken symmetry methods[20] were employed to allow modeling of antiferromagnetic (AF) as well as ferromagnetic (F) spin couplings. We previously used this approach successfully to model spin couplings of the diiron system in methane monooxygenase;[21] here, the use of UDFT as opposed to RODFT is important in getting the relative energies of the oxy and deoxy forms correct.

For the QM/MM calculations, the frozen orbital approach as implemented in the program package QSite was used.[16,17] QSite is based on a tight coupling of Jaguar with the IMPACT protein molecular modeling code.[15] The fundamental theory of the frozen orbital QM/MM method used in QSite is described in full detail in refs 16 and 17 and hence is only outlined here. The hardest task for any QM/MM method is to obtain an adequate description for covalent bonds connecting atoms of the QM region with those of the MM region. In the method of QSite these bonds are modeled by localized frozen orbitals obtained from calculations on small molecule models, in this case blocked dipeptides. Only the basis functions localized at the bond connecting the two regions are included in the orbital, while all other functions are set to zero. The frozen orbitals are included when obtaining the wave function in the QM calculations. The QM atoms interact with the MM atoms by including the MM point charges in the quantum mechanical one-electron Hamiltonian.

For the molecular mechanics part of the QM/MM calculations the OPLS all-atom (AA) force field was employed.[22] This force field has been extensively tested and gives good agreement with ab initio calculations for conformational energies of peptides as well as for thermodynamic and structural properties of proteins. Both for the pure QM calculations and the quantum chemical part of the QM/MM calculations the density-functional theory (DFT) functional B3LYP[23−25] was used for all results reported in this paper. This functional provides excellent results for thermodynamic properties of both organic molecules[26] and metal-containing systems.[27] Although there are new functionals[28] that do not contain an exact exchange contribution, which may be marginally superior for organic compounds and potentially offer computational advantages, the performance of these functionals for inorganic molecules has not yet been extensively explored.

QSite contains a number of features that greatly facilitated the ability to carry out the calculations described below in a reasonable time frame, despite the large size of the QM subsystem and the demanding objective of computing the absolute free energy of binding to a reasonably high precision. The most important of these features is adiabatic minimization of the MM region after each QM geometry step. This algorithmic approach essentially renders a QM/MM computation comparable in CPU time to a purely QM calculation on the QM subsystem. There is some differential due to the typically greater difficulty of converging QM/MM wave functions and geometries.

The accuracy obtained using QSite has been tested for a variety of small model systems, principally small peptides.[17] These tests encompass relative conformational energetics, deprotonation energies, and hydrogen-bonding energies of molecular pairs. They are described in detail in ref 17. Errors reported therein are uniformly below 2 kcal/mol and predominantly below 1 kcal/mol. Because reasonably large data sets, including all 20 amino acids, have been used in fitting and testing, there is confidence that the basic technology is accurate and reliable at the level of a single QM/MM interface.

The present application to hemerythrin requires the use of multiple QM/MM cuts in close physical proximity, as well as incorporation of a transition metal species into the calculation. To determine how well the method performs under these more demanding conditions, we constructed a series of test cases directly examining the effects of the QM/MM interface on dioxygen binding. One set of test cases uses a small model system that can be studied at the pure QM level; here, we systematically increased the number of QM/MM cuts and examined the energy difference between the oxy and deoxy forms. In the second, we systematically enlarged the QM region in QM/MM calculations on the full protein, for example replacing side chain cuts with backbone cuts. In both cases, the deviations in binding energy observed between the different models were uniformly less than 2 kcal/mol; for the most realistic tests of the full QM/MM model, the deviations were less than 1 kcal/mol. The details of this work will be described in another publication.[29] Here, we summarize these studies by concluding that the error induced by the use of the QM/MM model, as opposed to a pure QM calculation, is quite likely to be less than the intrinsic errors in the DFT calculations and physical model assumptions.

For the QM/MM calculations involving the entire protein, the LACVP* basis set was used. The QM region in the QM/MM models contained between 100 and 118 atoms, leading to a maximal basis set size of 1229 basis functions. This basis set is entirely adequate for carrying out geometry optimizations and provides reasonable single point energies. To obtain high accuracy single point energies after geometry optimization, the use of a larger basis set, e.g. the triple-ζ basis described above for the QM calculations, is required and can be handled by the QM/MM code. Given that we are in any case using the pure QM model calculations to provide corrections for the spin state description, however, we decided to make basis set corrections in this manner. A specific formulation of how these corrections are accomplished is provided below (section V).

At present, only RODFT methods are programmed to work within the QM/MM framework. Thus, we were unable to carry out UDFT calculations for the QM/MM models. By using the QM calculations as a correction, however, we are able to incorporate the UDFT results into the final estimation of binding free energy with negligible error. The specific formulation of the correction protocol is described in section V.

IV. Model System Calculations

In this section, results obtained using small quantum chemical models are discussed. We analyze the magnetic coupling of the two iron atoms in the active site, compare the small model with X-ray structures, and calculate accurately (within the limitations of our QM methdology) the O_2 binding energy in the gas phase without the protein environment. In the notation that will be used throughout the paper, Fe1 corresponds to the iron atom binding to 3 histidine groups and Fe2 to that binding to 2 histidine groups.

IV.1. Models and Computational Details. In the purely quantum mechanical model used for the great majority of calculations, the five terminal histidine side chains coordinating to the iron atoms in the active site (Figure 1) were replaced by neutral imidazole groups and the bridging aspartate and glutamate carboxylate side-chains were substituted by negatively charged acetate groups. The model is shown in Figure 2a. The hydrogen bonding Glu 24 shown in Figure 1 was not included in the quantum mechanical model. The resulting model, 63−65 atoms in total, has a net charge of +1 for both oxidized and

(20) Noodleman, L. J. *Chem. Phys.* **1981**, *74*, 5737.
(21) Gherman, B. F.; Dunietz, B. D.; Whittington, D. A.; Lippard, S. J.; Friesner, R. A. *J. Am. Chem. Soc.* **2001**, *123*, 3836−3837.
(22) Jorgensen, W. A.; Maxwell, D. S.; Tirado-Rives, J. *J. Am. Chem. Soc.* **1996**, *118*, 11225.
(23) Becke, A. D. *Phys. Rev. A* **1998**, *38*, 3098.
(24) Vosko, S. H.; Wilk, L.; Nusair, M. *Can. J. Phys.* **1980**, *58*, 1200.
(25) Lee, C.; Yang, W.; Parr, R. G. *Phys. Rev. B* **1988**, *37*, 785.
(26) Curtiss, L. A. Raghavachari, K.; Redfern, P. C.; Pople J. A. *J. Chem. Phys.* **1997**, *106*, 1063.
(27) Siegbahn, P. E. M.; Blomberg, M. R. A. *Annu. Rev. Phys. Chem.* **1999**, *50*, 221−249.
(28) Perdew, J. P.; Burke, K.; Ernzerhof, M. *Phys. Rev. Lett.* **1996**, *77*, 3865−3868.
(29) Wirstam, M.; Gherman, B. F.; Guallar, V.; Murphy, R. B.; Friesner, R. A. Manuscript in preparation.

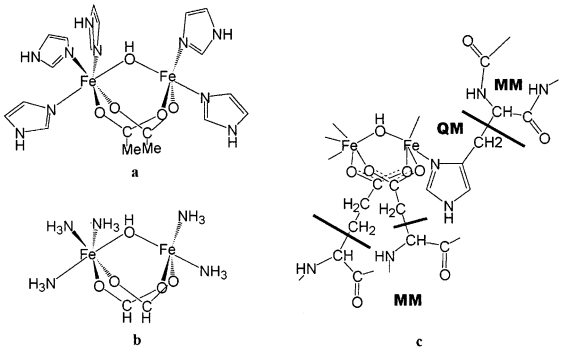


Figure 2. Models used for (a) energy evaluations at the purely QM level, (b) frequency determinations, and (c) QM/MM calculations. For clarity, only one His is shown in the QM/MM model. In the calculations the same type of cut was used for all active site histidine residues.

reduced forms. Full geometry optimizations of these models were performed applying both antiferromagnetic (AF) and ferromagnetic (F) coupling of the spins of the two iron atoms. In the case of the F coupled system, spin multiplicities of 9 and 11 were used for the deoxy and oxy forms, respectively. For the AF coupled system the spin multiplicity of 1 was used for both forms.

For the calculation of the magnetic splitting, geometry optimizations were performed using the LACV3P basis set, and for the single point energy calculations in the converged geometries, the LACV3P** basis set was used.[30,31] For calculations used to compute the O_2 binding energies (section IV.4), the LACV3P** basis set was used for geometry optimization. For the final single-point energy calculations in the converged geometries the very large cc-pVTZ basis set was used for C, N, O, and H. The cc-pVTZ is not defined for Fe, and instead the LACV3P** basis set was used for the two Fe atoms. Frequency calculations were performed to correct for zero-point and entropy effects. These computations are very expensive and were therefore performed by using a smaller model (see section IV.2) and the LACV3P basis set.

We begin by discussing the spin densities and couplings, because these quantities are obtained from a single optimized model structure and are the least complex to analyze. We then discuss the structures obtained from various types of optimization, comparing the small model results with QM/MM and experimental structures. Finally, the free energy of dioxygen binding at the small QM model level is calculated and shown to be in serious disagreement with the experimental data for the protein, indicating the need for QM/MM calculations to recover quantitative results.

IV.2. Spin Densities and Couplings. Our calculations agree with the experimental results that the two iron atoms are AF coupled. The calculated splitting between AF and F coupled states favors the former by 0.4 kcal/mol for the deoxy and 6.1 kcal/mol for the oxy form. These results parallel the experimental findings that the magnetic coupling is larger for the oxy ($J = -77$ cm^{-1})[8] than the deoxy ($J = -14(2)$ cm^{-1})[5] form. An estimate for the *J*-coupling constants can be computed from the theoretical splittings as described in ref 20. Using this approach, *J* can be expressed as indicated in eq 2, where S is the total spin for the F state and E_F and E_{AF} are the calculated energies:

$$J = \frac{2(E_{AF} - E_F)}{S(S+1)} \qquad (2)$$

This procedure has previously given reasonable *J* values for transition metal ions.[32] From eq 2, the computed splitting for the oxy form gives a coupling constant of -142 cm^{-1}, which disagrees with the experimentally measured value. It is more in line with values for synthetic (μ-oxo)diiron(III) model complexes, however.[33] The computed *J*-value for the deoxy form of -14 cm^{-1} is in excellent agreement with experiment. The AF coupling was used for all results on the pure QM model discussed below unless otherwise stated.

The spin populations on the two iron atoms *S*(Fe1) and *S*(Fe2) are +3.83 and −3.78, respectively, in the deoxy form and +4.16 and −4.11 in the oxy form. These values are typical for formal $[Fe(II)]_2$ and $[Fe(III)]_2$ systems considering that there is substantial radical character on the ligands in the latter, with the largest fraction (−0.35) being on the coordinated oxygen atom of the peroxo ligand.

Table 1. Selected Structural Parameters Obtained Using Different Theoretical Models[a]

	QM[level−1 c]		QM[level−2 d]		QM/MM[level−2 e]		expt[f]	
	deoxy	oxy	deoxy	oxy	deoxy	oxy	deoxy	oxy
Fe1−Fe2	3.34	3.24	3.36	3.30	3.32	3.30	3.34	3.28
Fe−N[b]	2.24	2.22	2.26	2.26	2.23	2.23	2.22	2.20
Fe−O[b]	2.11	2.01	2.10	2.02	2.11	2.07	2.12	2.07
O−O		1.42		1.40		1.36		1.49
O_μ−O(OFe)		3.30		2.57		2.47		2.72
O_μ−H	0.96	1.80	0.97	1.66	0.97	1.50		
∠Fe−O_μ−Fe	107.9	124.2	111.3	122.5	111.5	122.6	116.7	125.4

[a] The corresponding parameters from the X-ray structures[2] are reported for comparison. Distances are given in Å. [b] Average value for bond distances within the first coordination shell. [c] QM model, AF coupling and UB3LYP/LACV3P**. [d] QM model, F coupling and ROB3LYP/LACVP*. [e] QM/MM model, F coupling and ROB3LYP/LACVP*. [f] The average value of the four subunits computed directly from the PDB files 1HMO and 1HMD.[2]

IV.3. Structural Optimizations of QM and QM/MM Models and Comparison with Experimental Data for the Protein. The first main column of Table 1 lists a selection of computed geometrical parameters at the highest level of theory (UB3LYP/LACV3P**), which will be referred to as *level-1* in the discussion below. The second main column refers to the same level of theory as used for the QM/MM calculations and will be discussed in the next subsection. The resulting QM geometries are in good agreement with the X-ray crystal structure determinations,[2] listed in the last main column.

The experimentally observed shortening of the Fe−Fe distance upon O_2 binding is reproduced, within error. Our calculations further support conclusions from previous studies suggesting that the bridging oxygen atom is protonated in the deoxy form and that this proton is transferred to the dioxygen group when binding to iron. The modeled deoxy structure without the proton no longer agrees with the experimental structure. The Fe−Fe distance becomes significantly shortened to a value of only 3.05 Å. When the proton is placed on the bridging oxygen atom in the oxy form, it spontaneously

(30) The LACV3P basis set is a triple-ζ contraction of the LACVP basis set developed and tested at Schrödinger, Inc. The LACVP basis set is described in ref 31.
(31) Hay, P. J.; Wadt, W. R. *J. Chem. Phys.* **1985**, *82*, 299.
(32) Desplanches, C.; Ruiz, E.; Rodriguez-Fortea, A.; Alvarez, S. *J. Am. Chem. Soc.* **2002**, *124*, 5197−5205.
(33) Kurtz, D. M., Jr. *Chem. Rev.* **1990**, *90*, 585−606.

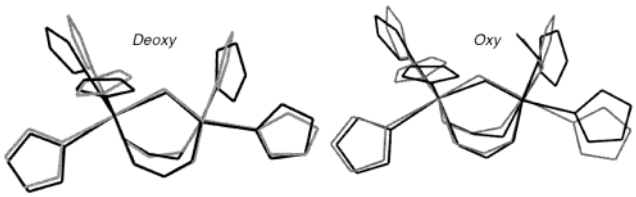


Figure 3. Fully optimized QM structures (gray) and X-ray crystal structures (black).

translocates to the Fe−O−O group. The energy difference between the most stable QM model for the oxy form and for a structure where the distance between the bridging oxo and the proton is frozen at the value of the deoxy form (0.96 Å) is 6.7 kcal/mol. Also, for the oxy form complete removal of the proton from the model leads to a structure that is in poor agreement with the X-ray results. The proton is less well bound in the oxy form as compared to the deoxy form in their respective minimum configurations, by 4 kcal/mol, as expected from the increased charge on iron.

EXAFS studies of hemerythrin reveal an Fe−Fe distance of 3.57 Å for the deoxy and 3.24 Å for the oxy form.[4] Although the measured EXAFS distance for the oxy form is in good agreement with the crystallographic study, the Fe−Fe distance of the deoxy form differs by as much as 0.24 Å. Thus, for both forms the computed Fe−Fe distances are within 0.04 Å of those obtained from crystallography and disagreement is with the distance obtained from EXAFS.

As is shown in Table 1, the distance that is the most challenging to describe with the present methods involves the hydrogen bond between the bridging oxygen and the protonated peroxo moiety (O_μ−H). Large variations in that distance occur among the different methods, and there is also a large discrepancy between the experimental and theoretical values for the O_μ−O(OFe) distance. The X-ray structure shows large variations of the O_μ−O(OFe) distance over four subunits, having minimum and maximum values of 2.62 and 2.98 Å, respectively.[2] These variations in both experimental and theoretical structures may suggest a shallow energy well. Another discrepancy between the experimental and theoretical structures concerns the torsion angles of the imidazole rings of the histidine ligands. This feature can be visualized in Figure 3, where the theoretical and the experimental structures are overlaid. Presumably, the torsion angles are held in specific positions by atoms in the protein structure that are not included in our QM theoretical model. Similar observations were made for ribonucleotide reductase.[34] One question that will be addressed below is to what extent such constrained imidazole rings affect the O_2 binding in hemerythrin.

IV.4. QM Model Calculation of the Free Energy of Dioxygen Binding. When only the electronic DFT energies are included, O_2 is bound by 5.3 kcal/mol (Table 2). This value is obtained by taking the DFT computed energy difference between the oxy form and the deoxy form + free O_2 for the isolated active site model in the gas phase. To compute the free energy of dioxygen binding at room temperature, however, we must include entropic factors associated with movement of the molecule from the gas phase or solution to the protein core. Ordinarily, calculation of this component of the absolute free energy of binding of a small ligand to a protein would be an extremely difficult task. Reorganization of the side chains in the protein and the associated entropy changes must be analyzed, and the sampling of phase space in the protein by the ligand must be properly taken into account, for example by free energy perturbation methods. In the present system, however, the problem is simplified because bound O_2, an exceptionally small ligand, occupies an existing cavity in the deoxy structure and is tightly bound to one of the iron atoms. In this situation there is only one important minimum in the protein−ligand potential energy surface, and its differential entropy, as compared to the deoxy form, can be approximated as harmonic since there are no low-frequency anharmonic modes associated with O_2. We therefore performed analytical second derivative calculations on the model complex using ammonia instead of histidine as ligands and formate instead of the bridging amino acids (Figure 2b). This approach saved computation time, given the fact that the energy gap using this model differs by only 1.5 kcal/mol from the full complement of histidine ligands, and we do not expect the vibrational frequencies associated with O_2 binding to be substantially different. The zero-point effect destabilizes the bound form by 3.1 kcal/mol. Also the entropy change upon binding O_2 was obtained from vibrational frequency calculations in Jaguar and then multiplied with room temperature to give the contribution to the free energy of binding. The entropy was found to destabilize the bound form by an additional 10.7 kcal/mol. Together, the thermal and zero-point effects thus yield a destabilization of the bound form of 13.8 kcal/mol (room temperature). The large thermal effect is exactly what is expected for a binding reaction, during which the translational and rotational degrees of freedom of a molecule are removed. Taken together, this means that O_2 is *unbound* by 8.5 kcal/mol (Table 2) using the gas-phase quantum mechanical model.

Table 2. Energy Contributions (kcal/mol) for Dioxygen Binding to Hemerythrin

DFT[a]	ΔG(theory)[b]	ΔG(gas)[c]	ΔQM/MM[d]	$\Delta G(O_2, g)$[e]	ΔG(soln)[f]	$\Delta G(O_2, s)$[g]
−5.3	+13.8	+8.5	−9.8	−1.3	−3.9	−5.2

[a] O_2 binding at the UB3LYP level using small QM models. [b] Correction for zero-point and entropy effects for dioxygen binding in the gas phase. [c] Total O_2 binding in the gas phase using small QM models. [d] QM/MM correction for the protein environment. [e] Resulting total free energy for dioxygen binding in the gas phase. [f] Correction for the solubility of O_2 in water. [g] Resulting total free energy for dioxygen binding in solution phase.

The large endothermicity obtained for eq 1 suggests that our small QM model is inadequate to explain the experimental results and that there must be compensating interactions in the protein environment that stabilize the bound form. The search for the nature of these interactions has been at the heart of this study and will be addressed in the next subsection. With comparison of the quantum mechanical with the experimental structures (Figure 3), it is clear that one possibility is that the positions of the imidazole rings in the protein facilitate O_2 binding compared to their orientation in the fully relaxed model structure. An indication that these torsion angles are relatively unimportant, however, is given by comparison with full optimizations of a smaller model, where ammonia groups substitute the histidine ligands and formate groups substitute the bridging amino acid ligands (Figure 2b). Switching to this smaller model only had a minor effect of 1.5 kcal/mol on binding.

(34) Torrent, M.; Vreven, T.; Musaev, D. G.; Morokuma, K.; Farkas, O.; Schlegel, H. B. *J. Am. Chem. Soc.* **2002**, *124*, 192−193.

V. O_2 Binding Using Large Quantum Mechanical/Molecular Mechanical Models

As discussed above, small gas-phase QM models of the active site provide a binding free energy for dioxygen that is in poor agreement with experiment. QM/MM calculations were therefore performed to determine the energetic correction to the pure QM results when one entire subunit of hemerythrin is included in the model. The cuts for the seven amino acids coordinating to the dimetallic center were made in the side chains, and the same total charge state (+1) as used for the small QM models was also employed for the QM region of the QM/MM calculation. When not otherwise stated, Glu 24 of Figure 1 was included in the MM region. The resulting model is shown in Figure 2c.

Since the current implementation of QSite does not support unrestricted (U) calculations, restricted open shell (RO) DFT was used instead, which means that all spins are coupled ferromagnetically (F). This procedure introduces errors for energy calculations on a complicated system involving high-spin transition metal atoms. The use of F instead of AF coupling for small active site models stabilizes the O_2 bound form relative to the unbound form of hemerythrin by 5.7 kcal/mol when using UB3LYP in both cases. Going from an F UB3LYP to an F ROB3LYP treatment introduces an even larger error of 6.2 kcal/mol. In the case of hemerythrin, however, this discrepancy does not cause a major problem since the error can be determined and corrected for by performing calculations on small models. The final energy for the O_2 binding (eq 1) can then be expressed as indicated in eq 3, where ΔE is the reaction energy, in this study the O_2 binding energy:

$$\Delta E_{\text{final}} = \Delta E(\text{QM}^{\text{level}-1}) + \Delta E(\text{QM}^{\text{level}-2}/\text{MM}) - \Delta E(\text{QM}^{\text{level}-2}) \quad (3)$$

Level-2 is the QM method used for the QM/MM calculations, and level-1 is the high level QM method used for the applications discussed in the previous subsection. For such a correction to be valid, it is required that the protein environment does not have a severe impact on the energetics coupled to structural differences between the QM region in the QM/MM model relative to the QM region in the absence of the surrounding protein. No such large structural differences are present in hemerythrin, as will be discussed below. The final O_2 binding energy (ΔE_{final}) in hemerythrin was thus computed by using the following procedure. First, two sets of pure QM calculations were performed on small models for the oxy and deoxy forms. One was aimed at computing a highly accurate value for the O_2 binding energy for the small hemerythrin model ($\text{QM}^{\text{level}-1}$) as discussed in the previous section. For the other set of QM calculations, exactly the same quantum mechanical methods that are used for the QM/MM method (LACVP* basis set and ROB3LYP) were employed ($\text{QM}^{\text{level}-2}$). After $\Delta E(\text{QM}^{\text{level}-1})$ and $\Delta E(\text{QM}^{\text{level}-2})$ were obtained, QM/MM calculations were performed at the level-2, to give ΔE_{final} of eq 3.

The accuracy of the QM/MM method in general,[17] and for applications to hemerythrin in particular,[29] has been extensively tested. On the basis of these tests the total error for the O_2 binding energy due to the QM/MM interface is estimated to be $\leq$2 kcal/mol. The resulting model includes 2005−2007 atoms of which 100−102 are in the QM region. Unless forming a salt bridge, charged amino acids on the surface were made neutral to avoid artificial electrostatic interactions, which to a large extent would be screened in the protein due to interactions with the solvent. Crystallographic waters surrounding the subunit are included in the results below. Removal of these waters at an initial stage of the investigation had no notable effect on the relative energies, however, or on the active site structures, indicating that the surface properties are relatively unimportant for O_2 binding. This result is expected since the active site is buried in a hydrophobic pocket, which effectively screens interactions with the surface charges.

The hydrogen atoms were first added automatically by using the Maestro[35] software and were then modified manually according to the procedure that will be described below. In addition to the charged residues facing the surface there is only one charged residue near the active site, a glutamate (Glu 24) that is hydrogen bonded to His 54. This glutamate was included in the MM region and formed a hydrogen bond to His 54 both in its neutral and in anionic forms. The anionic form of Glu 24 stabilized the O_2 bound state by 3.2 kcal/mol relative to the neutral form. The anionic form was adopted both because it is the most common form for glutamate residues in protein environments and because it would neutralize the net positive charge on the diiron core. The result is a net charge of zero for the core region, which is thermodynamically favored in an active site such as hemerythrin that is buried inside the protein, away from the aqueous solvent. The negatively charged form of Glu 24 was thus used for all calculations discussed below. Apart from the surface residues and Glu 24 there are no groups in the protein that would be charged under typical conditions. Furthermore, except for two histidine residues facing the surface, which were made neutral by having ND1 protonated and NE2 unprotonated, there are no histidine residues other than the five coordinating to the active site iron atoms. Except for the crystallographic waters discussed above, no solvent was included in the calculations. Since the results were rather insensitive to the surface properties as discussed above, this choice should not have a major effect for O_2 binding.

More critical for the O_2 binding energy than the treatment of the surface charges are the MM geometries in which the total QM/MM energy is evaluated. To obtain meaningful relative free energies it is essential that the MM configuration be in the same basin of attraction, which does not have to be the global energy minimum, for the two forms. The rmsd value between the oxy and deoxy crystal structures is 0.15 Å for the entire protein and 0.09 Å for the residues coordinating to the iron core, indicating that the protein has essentially the same configuration in both forms. Although crystal structures exist for both forms, only one of them, that of the deoxyHr, was used to avoid convergence to different local minima. The following procedure was employed to obtain a comparison between the two states in which cancellation of errors in the MM region was maximized while at the same time searching in a restricted fashion for the lowest energy of both states. In the first step, a full geometry optimization, having no frozen parameters, of a QM/MM model of the deoxy form was performed, starting out from one of the subunits (A) of the X-ray structure.[2] In the converged structure, the iron core was replaced by that of the oxy form obtained

(35) *Maestro 4.0*; Schrödinger, Inc.: Portland, OR, 2000.

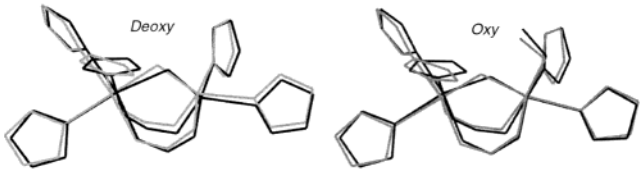


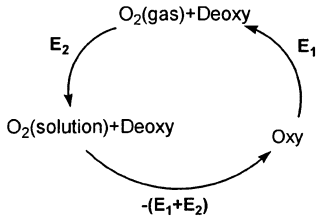
Figure 4. Fully optimized QM/MM structures (gray) and X-ray crystal structures (black) superimposed.

from the small QM models while all other coordinates from the minimization of the deoxy form were maintained. Starting from this structure, a geometry optimization of the oxy form was performed. In the converged geometry the active site was replaced by the corresponding QM/MM optimized coordinates from deoxy form, and a new geometry optimization of the deoxy form was performed since a change in the MM part might have occurred when the oxy form was optimized. Then, that MM geometry was in turn used for a new optimization of the oxy form. The procedure of extracting the MM part from the completed minimization of the other form and using it as an initial structure for a new minimization was repeated until no further changes, in either the converged geometry or final QM/MM energy, of either form occurred, which required four iterations. Thus, most probably the same local minimum was reached for both structures. The active sites of the converged QM/MM structures and the X-ray structures[2] are overlaid in Figure 4. When these results are compared with those of Figure 3, it is obvious that the orientation of the imidazole planes are much better reproduced in the QM/MM structure.

A similar observation was recently made for ribonucleotide reductase.[34] The rmsd value obtained is 0.7 Å. This value can be compared with the rmsd of 0.4 Å between the different subunits of the experimental structure. For the active site an rmsd of only 0.2 Å was obtained, whereas a significantly larger rmsd value of 0.5 Å was obtained for the small QM model. For structural parameters that do not involve the imidazole planes, only small differences between the QM model and the QM/MM model are observed (Table 2). Small differences are also observed for the spin distribution. The spin distributions for the $QM^{level-2}$ deoxy and oxy model structures are $S(Fe1) = 3.77$, $S(Fe2) = 3.75$ and $S(Fe1) = 4.14$, $S(Fe2) = 4.03$, whereas the corresponding values for $QM/MM^{level-2}$ are $S(Fe1) = 3.77$, $S(Fe2) = 3.75$ and $S(Fe1) = 4.17$, $S(Fe2) = 3.90$, indicating only a minor change of the electronic structure of the diiron core due to the protein environment. Since the protein does not have a severe impact on the electronic structure in this case, it is reasonable to assume that the quantum mechanical error made for computing the O_2 binding energy is equal in the QM and QM/MM calculations. The energy difference between these two therefore gives a good estimate of the *effect* of including the surrounding protein in the calculations. This effect can then be added as a correction to the high level (level-1) QM results discussed in the previous section. This correction was found to be 9.8 kcal/mol in favor of the oxy form, giving an exothermicity for the O_2 binding of 1.3 kcal/mol (Table 2).

There is one remaining factor to be considered in computing the free energy difference measured experimentally, namely, that the desired equilibrium is between O_2 bound to the protein and that in solution and not in the gas phase. This equilibrium is easily computed from the gas phase results by employing the thermodynamic cycle depicted in Scheme 1.

Scheme 1



The free energy of solvation of O_2 in water is 3.9 kcal/mol as computed from Henry's law constant at room temperature, yielding a final result for the predicted in vivo free energy of binding of −5.2 kcal/mol. This value can be compared with that computed from the equilibrium constant of $(2.5 \pm 0.5) \times 10^5$ M^{-1} obtained from stopped-flow experiments carried out in solution.[36] Using the fact that the experiment was carried out under the standard state pressure of O_2 (1 atm) and at a temperature of 21.5 °C, we compute the corresponding free energy to be −7.3 kcal/mol. The QM/MM result of −5.2 kcal/mol is in remarkably good agreement with this experimental value.

An analysis of the output from the QM/MM calculations reveals that the largest contributor to the energy difference is the molecular mechanics van der Waals energy computed to arise from interactions of the QM and MM regions. The van der Waals interactions stabilize the oxy form by 6.3 kcal/mol. This contribution is directly attributable to van der Waals contacts of bound dioxygen with the protein atoms. The oxy form is additionally stabilized by 3.2 kcal/mol owing to the difference in electrostatic interaction between the MM and QM regions. The major component of the electrostatic effect is due to the anionic Glu24 residue located near the positively charged iron core (Figure 1). No significant electrostatic effect for O_2 binding was observed when this residue was made neutral. Furthermore, the effect of including Glu24 (modeled by formate) in a purely quantum mechanical model yields a stabilization of the oxy form of 2.7 kcal/mol relative to a quantum mechanical model without the glutamate. The physical explanation of this result is straightforward. Upon dioxygen binding, charge migrates into what is now a peroxo species, leaving other groups in the core such as the histidine, a hydrogen bonding partner to Glu24, more positive. This additional positive charge enhances the strength of the resulting hydrogen bond as compared to the deoxy form, leading to energy stabilization of dioxygen binding. This motif of charge reorganization upon binding of a small molecule ligand, resulting in differential stabilization of hydrogen bonds to charged amino acids, may be employed in many metalloenzymes. A second example that we have recently discovered occurs in cytochrome P450cam and will be reported elsewhere.

Both the fact that the energetic effect due to the protein environment mainly derives from the molecular mechanics and that the relative QM energies are rather insensitive to the choice of first sphere ligands suggest that the contribution of protein strain to the O_2 binding is small. To investigate this matter further, the active site components were extracted from the converged QM/MM structures and QM single point energy calculations were performed for these structures and compared

(36) Lloyd, C. R.; Eyring, E. M.; Ellis, W. R., Jr. *J. Am. Chem. Soc.* **1995**, *117*, 11993−11994.

to the QM optimized energies at the same level. The O_2 binding energy difference between QM and QM/MM optimized structures for the active site was only 0.9 kcal/mol, giving additional support for the previous indications that strain effects are unimportant. This result means that the protein environment of the active site of Hr is very flexible and adjusts readily and with almost no change in energy to the rather substantial change in ligand coordination around Fe2, which converts from octahedral to trigonal bipyramidal local symmetry.

VI. Conclusions

This study of O_2 binding in Hr shows that reversible dioxygen binding in hemerythrin can be reproduced by a theoretical model if the electronic structure, protein environment, and thermal effects all are taken into account. The major contribution from the protein environment arises from the van der Waals interaction between O_2 and protein atoms. One could arguably thus have arrived at a qualitatively reasonable result for the binding energy without the use of QM/MM methods by estimating the noncovalent component of O_2 binding from experimental data on the binding of small nonpolar molecules such as methane to hydrophobic protein cavities. This approach would not have provided the same assurance as the methodology used here. We have determined that strain energy can be neglected despite substantial change in iron−ligand coordination. Furthermore, we provide a quantitative evaluation of the electrostatic stabilization of the oxy form and compute the nonpolar component of the interaction without invoking any data fitting or estimation. The most important aspect of a robust and accurate QM/MM methodology is that it makes it unnecessary to formulate uncontrolled assumptions; the right answer should arise naturally from the calculations. The present results afford a successful application of this kind and highlight how the reversible O_2 binding function in a metalloprotein requires energetic contributions both from metal coordination as well as the protein environment.

Acknowledgment. This work was supported by grants from the National Institute of General Medical Sciences (GM40526 to R.A.F. and GM32134 to S.J.L.). This work was partially supported by the National Computational Science Alliance under Grant MCA95C007N and utilized the NCSA SGI/CRAY origin2000. M.W. thanks The Swedish Foundation for International Cooperation in Research and Higher Education (STINT) for financial support.

JA017692R