

Theoretical Studies of [FeFe]-Hydrogenase: Structure and Infrared Spectra of Synthetic Models

Silviu Zilberman,[†] Edward I. Stiefel,^{†,‡} Morrel H. Cohen,^{†,§} and Roberto Car^{*,†,⊥}

Department of Chemistry, Princeton University, Princeton, New Jersey 08544, Princeton Environmental Institute, Princeton, New Jersey 08544, Department of Physics and Astronomy, Rutgers University, Piscataway, New Jersey 08854, and Princeton Institute for the Science and Technology of Materials, Princeton University, Princeton, New Jersey 08544

Received: November 12, 2005; In Final Form: February 1, 2006

First-principles density functional theory calculations of synthetic models of [FeFe]-hydrogenase are used to show that the theoretical methods reproduce observed structures and infrared spectra to high accuracy. The accuracy is demonstrated for synthetic Fe(I)Fe(I) models ($[(\mu\text{-PDT})\text{Fe}_2(\text{CO})_6]$ and $[(\text{CN})(\text{CO})_2(\mu\text{-PDT})\text{Fe}_2(\text{CO})_2(\text{CN})]^{2-}$), for which we show that their infrared spectra are sensitive to the geometric arrangement of their CO/CN ligands and can be used in conjunction with quantum-mechanical total energies to predict the correct ligand geometry. We then analyze and predict the structure of mixed-valence Fe(II)Fe(I) models ($[(\mu\text{-MeSCH}_2\text{C}(\text{Me})(\text{CH}_2\text{S})_2)\text{Fe}_2(\text{CO})_4(\text{CN})_2]^{x-}$). These capabilities promise to distinguish among the various structural isomers of the enzyme's active site which are consistent with the limited accuracy of the X-ray observations.

1. Introduction

A fundamentally important issue in understanding the functioning of metalloenzymes is the relationship between the structure of the active site and the mechanism of enzyme action. X-ray determination of the structure of the active site region, though a crucial step toward that knowledge, may not have sufficient resolution to determine fine structural features that may be critical for understanding the mechanism. An example of the structural uncertainty is provided by the [FeFe]-hydrogenase, for which the bridging ligand in the active cluster is unknown and has been suggested to be either propane dithiol¹ (PDT) or dithiomethylamine² (DTN).

The study of biochemical reactions is inherently challenging. The reacting systems are typically not well characterized on the microscopic level, and they may undergo changes in conformation, oxidation state, charge transfer, and spin redistribution during the reaction. A wealth of experimental methods is used to decipher these processes: X-ray crystallography, infrared (IR) and Raman spectroscopy, optical spectroscopy, Mössbauer spectroscopy, nuclear magnetic resonance (NMR), electron paramagnetic resonance (EPR), and others. Each has some advantages and some shortcomings, but the combination of information from different sources allows the buildup of a coherent picture. Theoretical analysis can be an invaluable tool in constructing that coherence.

However, the theoretical study of biochemical reactions is also quite challenging.³ Treating chemical reactions in detail theoretically requires a quantum mechanical approach. The combination of modern density-functional theory⁴ (DFT) and the continuing increase in computational and algorithmic power

allows us to study enzyme structure and reactivity quantum mechanically. Unfortunately, we are still a long way from a full quantum-mechanical description of entire proteins, and the theoretical treatment currently must focus on a minimal portion of the protein that includes the presumed active site and as little as possible of its neighbors. In addition, the original first-shell ligands within the active site are often replaced by small functional groups of similar chemical/physical effect. Both of these two steps in forming computational models of the active site, that is, selecting where to truncate and using surrogate ligands, represent a major challenge that requires experience and sometimes trial-and-error. We refer to errors originating from this process as “model errors”.

The difficulty is further increased by the presence of “method errors” associated with the accuracy of the DFT exchange-correlation functional and with errors in the actual numerical implementation, the latter in particular due to the choice of the basis set. These two sources of error are convolved in the computational results, making it difficult to understand why in some cases computations fail to match the experiments. A good solution to this problem would be to test, when possible, the calculations for model systems as similar as possible to the true active site which have been synthesized and characterized experimentally, thus establishing the expected method errors and highlighting model errors in some cases.

In the present paper we are interested in particular in the [FeFe]-hydrogenase,^{5,6} an enzyme that catalyzes the reversible formation of molecular hydrogen from protons and electrons. Understanding the structures and mechanisms of hydrogenases has important biological implications, as well as implications in the quest for biological and bio-inspired catalysts for hydrogen production.^{7,8} We establish that first-principles density-functional computations form a reliable tool for establishing subtle, important structural features of the active cluster in the enzyme. We do so by computations of the total energies, structures, and IR spectra of synthetic model compounds designed to mimic

* Address correspondence to this author. E-mail: rcar@princeton.edu.

[†] Department of Chemistry, Princeton University.

[‡] Princeton Environmental Institute.

[§] Department of Physics and Astronomy, Rutgers University.

[⊥] Princeton Institute for the Science and Technology of Materials, Princeton University.

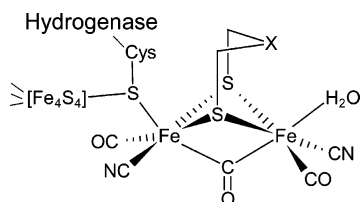


Figure 1. The structure of the [FeFe]-hydrogenase active site.^{1,9}

the active site of the enzyme. We show that our calculations for such model systems not only agree well with measurements of structure and infrared (IR) spectra but also can resolve structural features which are unresolvable by X-ray study of the enzymes.

The structure^{1,10} of the [FeFe]-hydrogenase active site is illustrated in Figure 1. Most (but not all) structural features of the active site are now well established. The active site is composed of a diiron subcluster connected to an Fe₄S₄ cubane. The diiron part contains biologically unusual CO and CN⁻ ligands that maintain a low-spin, low-oxidation-state configuration¹¹ of the iron ions.

Soon after the publication of the crystallographic structures of the *Clostridium pasteurianum* I¹⁰ (CpI) and *Desulfovibrio desulfuricans*¹ (DdHase) hydrogenases, several groups independently reported^{12–14} synthetic models of the diiron cluster: [(μ-PDT)Fe₂(CO)₆] (model A in Figure 2) and [(CN)(CO)₂(μ-PDT)Fe₂(CO)₂(CN)]²⁻ (model B in Figure 2). In both cases the formal oxidation state is Fe(I)Fe(I). These models have been used to benchmark theoretical treatments of hydrogenase. For instance, the structure suggested by Schmidt et al.¹³ was used in theoretical studies by Cao and Hall¹⁵ as a way to generate a linear, ad hoc calibration function for the IR frequencies. The computed IR frequencies of the CO and CN ligands were fitted to the experimental ones, and the fitting function was carried over to studies of hydrogenase models at different oxidation states and chemical compositions.

Schmidt et al.¹³ resolved a structural isomer of [(CN)(CO)₂(μ-PDT)Fe₂(CO)₂(CN)]²⁻ by X-ray crystallography (Figure 2, model B1), which is different from those suggested by Lyon et al.¹² and by Le Cloirec et al.¹⁴ (Figure 2, model B3). The difference is in the placement of the CO and CN ligands, as illustrated in Figure 2. Mild as they may seem, such small differences can alter the IR spectrum in a nontrivial way. This has important implications, both for experiment and for theoretical benchmarks, and possibly also for the mechanism.

Several more advanced synthetic models were reported in the past few years. Lawrence et al.¹⁶ synthesized an azadithiolate model, and very recently Tard et al.⁸ reported for the first time a synthetic species that contains both the diiron and the Fe₄S₄ cubane part. All these models are Fe(I)Fe(I) analogues. Razavet et al.^{17–19} are the only authors (to the best of our knowledge) to report a transient FTIR study on a mixed-valence Fe(II)-Fe(I) model [(μ-MeSCH₂C(Me)(CH₂S)₂)Fe₂(CO)₄(CN)₂]²⁻, produced both electrochemically and by a stopped-flow technique, a very important model of the oxidized active state of the enzyme that has not been studied theoretically so far.

In this paper we study several synthetic analogues of the diiron cluster in hydrogenase in two oxidation states, Fe(I)Fe(I) and Fe(II)Fe(I). Our computed structures are in excellent agreement with the available X-ray structure determinations. We show that our predicted infrared spectra are in good agreement with experiment for both frequencies and intensities. We also show that the IR signature is sensitive to subtle structural isomerism, obtaining agreement for model B1 of section II (Figure 2) with the spectrum determined by Schmidt

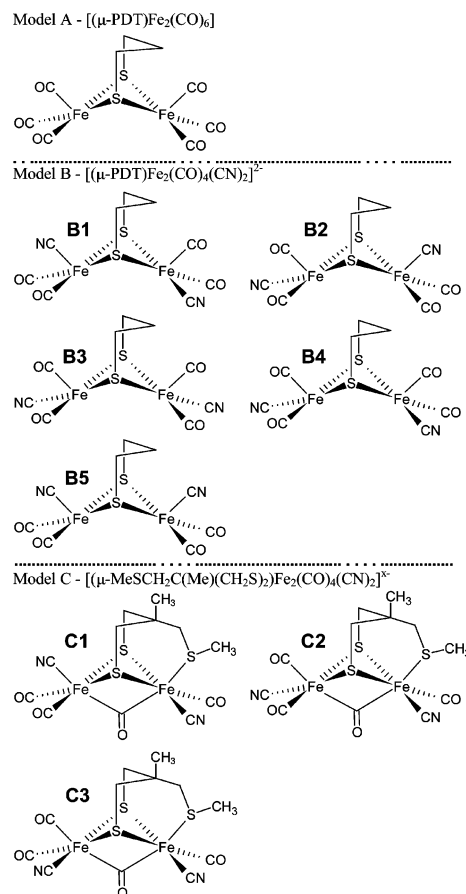


Figure 2. Computational models used here. Models A/B were studied in the Fe(I)Fe(I) oxidation state in spin-restricted calculation. Models C were studied as Fe(I)Fe(I) (spin-restricted), mixed-valence Fe(II)-Fe(I) (spin-unrestricted, doublet), and fully oxidized Fe(II)Fe(II) (spin-restricted).

et al.¹³ Distinguishing among such structural isomers is especially important for the enzyme because of the uncertainty of atomic assignments, given the limited resolution of the reported X-ray structures of [FeFe]-hydrogenase (1.8–1.6 Å). Moreover, one cannot exclude internal rearrangements during the reaction cycle, and it is important to show that computational methods are sensitive enough to track them.

As mentioned above, we are the first to study the Fe(II)Fe(I) synthetic models. The Fe(I)Fe(I) models, on the other hand, were also studied by other groups.^{15,20,21} Our IR spectrum calculations are in better agreement with experiment, both for the absolute and relative positions of the various vibrational bands and for the intensities, underlining the higher accuracy that we have obtained from DFT calculations for these systems.

The paper is organized as follows: in Section 2 we present our model systems, followed by computational details in Section 3. The results of the computations for the various models are presented in Section 4, including a comparison with earlier theoretical work by others. We offer conclusions in Section 5. In the Appendix, we estimate the accuracy of our computed relative energies.

2. Model Systems

2.1. Model A. Model A [(μ-PDT)Fe₂(CO)₆] (Figure 2) was synthesized by several groups^{12–14} and is the best characterized of all hydrogenase models. Here the oxidation state is Fe(I)-Fe(I).

2.2. Model B. Model B ($[(\text{CN})(\text{CO})_2(\mu\text{-PDT})\text{Fe}_2(\text{CO})_2(\text{CN})]^{2-}$) (Figure 2) was synthesized by the same groups as model A,^{12–14} but two structural isomers were proposed: Schmidt et al.¹³ assigned it to be B1 via X-ray crystallography, while Lyon et al.,¹² from their IR analysis, and Le-Cloirec et al.,¹⁴ from their IR and NMR analysis, proposed it to be model B3. It is conceivable that different synthetic pathways may lead to different geometric arrangements of the ligands. However, there is very good agreement among the experimentally reported IR frequencies from these three groups, with peak frequencies differing only by a few reciprocal centimeters, probably reflecting a mild solvent effect. Thus it is likely to be the same isomer in all cases.

Though it is quite certain that B1 is the structure for this system, it serves as an excellent benchmark system to see if computational methods can determine the right arrangement of the CO/CN ligands, in view of the limited resolution of the crystallographic studies of [FeFe]-hydrogenase. For that aim we studied several structural isomers involving different arrangements of the CO and CN ligands, labeled as B1–B5 in Figure 2.

2.3. Model C. Model C ($[(\mu\text{-MeSCH}_2\text{C}(\text{Me})(\text{CH}_2\text{S})_2)\text{Fe}_2(\text{CO})_4(\text{CN})_2]^{x-}$) was synthesized and studied by Razavet et al.^{17–19} in two different oxidation states: Fe(I)Fe(I) (reduced state) and Fe(II)Fe(I) (mixed-valence state). The geometry with a bridging $\mu\text{-CO}$ is stable only at low temperatures for the Fe(I)Fe(I) oxidation state. To the best of our knowledge, this is the only synthetic analogue that has been studied experimentally in the mixed-valence state, similar to the oxidized state of [FeFe]-hydrogenase. It is an important model system for assessing our computational accuracy. We also explored the fully oxidized Fe(II)Fe(II) state, and we show that it is likely to be present in the spectrum of the presumed mixed-valence state. The structure of model C was not determined by X-ray crystallography in any of the oxidation states, and the exact position of the CO/CN ligands is unknown. The C models are presented in Figure 2. Notice that there are three additional isomers (obtained by exchanging the CO and CN ligands on the distal Fe) that were not considered since they are mirror images of the isomers that we considered.

3. About the Computations

We used DFT⁴ with the Perdew–Burke–Ernzerhof (PBE)^{22,23} exchange and correlation functional in the Car–Parrinello plane-wave code, which is included in the Quantum-ESPRESSO²⁴ electronic structure package. We used ultrasoft pseudopotentials²⁵ for all elements, with a 30 Ryd kinetic energy cutoff for the smooth part of the wave functions and a cutoff of 180 Ryd for the augmented charge density. The Fe pseudopotential was generated by a scalar-relativistic DFT calculation²⁶ that includes the semicore 3s and 3p orbitals, as well as the 3d, 4s, and 4p orbitals. Models A and B [cf. Figure 2] are in the Fe(I)Fe(I) oxidation state, which is known from the enzyme to be EPR silent,²⁷ and therefore we used spin-restricted calculations. We tested one case, model B3, in an unrestricted total energy calculation. The energy difference with respect to the restricted case was negligible, less than 10^{-7} hartree. Models C were studied in three oxidation states: reduced [Fe(I)Fe(I)], mixed valence [Fe(II)Fe(I)], and fully oxidized [Fe(II)Fe(II)]. Only the mixed-valence state was treated with spin-unrestricted DFT. The molecules were placed in a cubic periodic cell of dimension 25 atomic units, a cell size large enough for the computed molecular properties to be independent of cell size. The Hessian of the energy with respect to nuclear coordinates was constructed by

finite numerical displacement of each atom in sequence in all three Cartesian directions. The IR cross-section was calculated by evaluating the Born effective charges and projecting them onto the normal modes to obtain the transition dipole.²⁸ Rigid linear and rotational modes of motion are projected out by standard methods,²⁹ and the acoustic sum-rule is imposed on the matrix elements of the Born-charge tensors.³⁰

We stress that *no ad hoc a posteriori empirical corrections were applied to the computed frequencies and intensities*, unless stated otherwise.

Often chemical reactions result in the thermodynamically most stable product, but sometimes the kinetically controlled product may govern. The systems studied here seem to correlate well with the thermodynamic product, and thus it is important to estimate the accuracy of the *relative* total energies of the various structural isomers. In the Appendix we briefly analyze the level of accuracy of our predicted relative energies, estimating that the error is smaller than 1 kcal/mol.

4. Results and Discussion

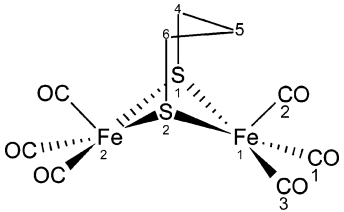
4.1. Model A. We focus first on model A, the best characterized of all the synthetic models studied here. The computed structure (Table 1) agrees extremely well with the reported crystallographic data by Lyon et al.,¹² with typical deviations of 1–2% for the bond length, and 1–2° for the bond angles. Notice that although a typical C–C bond length is about 1.54 Å, similar to what we get in our calculations, the experimentally reported C–C bond lengths for model A are unusually short, about 1.45 Å. Also in the experimental structure unusually short C–H bonds are reported, around 1.0 Å, compared to the typical 1.1 Å, which also comes out from our calculations. We show below that model B, though similar to model A, does not have these unusual short bonds. We therefore believe that these discrepancies indicate an experimental error in resolving the bridge structure of model A.

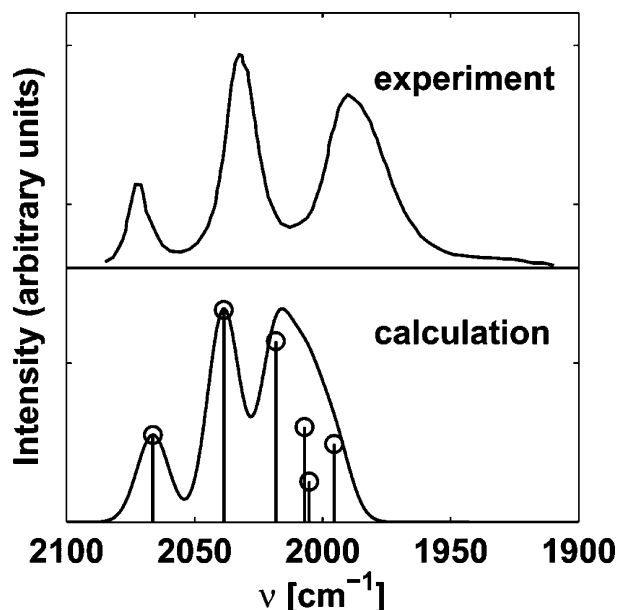
Figure 3 compares the experimental IR spectrum¹² to our simulated spectrum, focusing on the CO bands. The simulated spectrum was computed by broadening the lines by convoluting them with a Gaussian of standard deviation 7 cm^{-1} , chosen to optimize the fit to the experimental data. The agreement between the spectra is very good qualitatively and quantitatively: both for the relative positions of the vibrational frequencies and their intensities (line shapes). The computed spectrum overlaps with the experimental one, and only the low-frequency band is shifted by about 20–25 cm^{-1} toward higher frequencies with respect to experiment. This is presumably due to the absence of solvent and to the limits of the PBE approximation to the DFT exchange-correlation functional.

4.2. Model B. Next we studied models B1–B5 (cf. Figure 2). The computed total energies (Table 2) show that model B1 has the lowest energy, about 3.6 kcal/mol (~ 4 kcal/mol when zero point energy is included) lower than B3, the structure proposed by Lyon et al.¹² and Le-Cloirec et al.¹⁴ The structure of B1 is also in very good agreement with the crystal structure reported by Schmidt et al.¹³ (see Table 3), with typical bond length deviations of 1–2% and angle deviations of 1–3°. Needless to say, the structure we find for all other B models does not agree with the experimental X-ray structure.

Next we turn to the IR spectra of models B, shown in Figure 4. We aim to see if comparing simulated and experimental¹³ spectra will allow us to rank the plausibility of different structures and possibly rule out via their spectra alone some of the isomers as candidates for the structure of model B. This capability is important because of the limited resolution of the

TABLE 1: Selected Bond Lengths and Angles of Model A^a

					
Selected Bond Lengths [Å]					
Fe(1)–Fe(2)	2.541 (2.511) 1.2%	Fe(1)–S(1)	2.271 (2.254) 0.7%	Fe(1)–S(2)	2.273 (2.250) 1.0%
S(1)–C(4)	1.840 (1.824) 0.9%	S(2)–C(6)	1.839 (1.817) 1.2%	C(4)–C(5)	1.520 (1.451) 4.7%
C(6)–C(5)	1.519 (1.469) 3.4%	Fe(1)–C(1)	1.776 (1.800) –1.3%	Fe(1)–C(2)	1.774 (1.803) –1.6%
Fe(1)–C(3)	1.775 (1.797) –1.2%	C(1)–O(1)	1.161 (1.141) 1.7%	C(2)–O(2)	1.164 (1.129) 3.1%
C(3)–O(3)	1.162 (1.136) 2.3%				
Selected Angles [deg]					
S(1)–Fe(1)–S(2)	84.5 (85.3) –0.9%	Fe(1)–S(1)–Fe(2)	68.2 (67.7) 0.7%	Fe(1)–S(2)–Fe(2)	68.2 (67.9) 0.4%
C(1)–Fe(1)–C(2)	98.4 (98.0) 0.4%	C(2)–Fe(1)–C(3)	98.2 (100.0) –1.8%	C(1)–Fe(1)–C(3)	90.7 (91.0) –0.3%
C(1)–Fe(1)–S(2)	157.2 (160.2) –1.9%	C(1)–Fe(1)–S(1)	87.6 (87.8) –0.2%	C(2)–Fe(2)–S(1)	104.4 (101.8) 2.6%
C(2)–Fe(2)–S(2)	106.4 (103.4) 2.9%	C(3)–Fe(1)–S(2)	155.4 (155.9) –0.3%		

^a Experimental data (in parentheses) are from Lyon et al.¹²**Figure 3.** Computed and experimental Infrared spectra of model A [(μ-PDT)Fe₂(CO)₆]. The calculated spectrum is obtained by convoluting the calculated intensities, indicated by the vertical stems, with a Gaussian having a width of 7 cm^{–1}. The experimental spectrum is reproduced with permission from ref 12.**TABLE 2: Internal Energies of Models B1–B5, Relative to the Lowest Energy Model B1^a**

model	total energy [kcal/mol] relative to B1	total energy + ZPE [kcal/mol] relative to B1
B1	0.0	0.0
B2	1.18	1.32
B3	3.64	3.96
B4	1.97	2.47
B5	1.74	1.80

^a The zero-point-energy (ZPE) is included in the right column. The estimated error is smaller than 1 kcal/mol (see the Appendix).

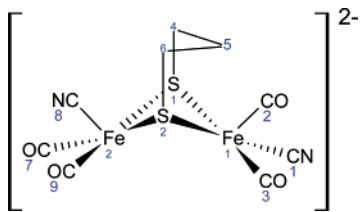
available X-ray data of [FeFe]-hydrogenases where we would have to rely more heavily on the computational and spectroscopic methods. For model systems B only peak positions and intensities have been reported. We assumed the same 7 cm^{–1} Gaussian broadening as for model A. From comparison with

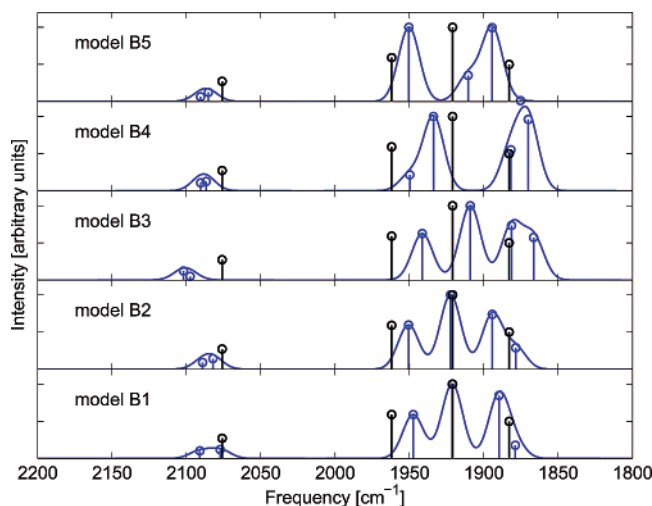
the experimental spectrum, it is clear that models B4 and B5 are in qualitative disagreement with experiment. The positions and the relative spacing of the peaks do not match the experiment, nor does the intensity pattern. In contrast, the agreement for model B1 is excellent, which also agrees uniquely with the crystallographic data by Schmidt et al. The *cis* geometry (B3) that was suggested by Lyon et al. and by Le Cloirec et al. has the right qualitative features but the spacing between the CO and CN bands is too large. Note also that the spectrum of model B2 is almost identical with that of B1, and IR analysis cannot really distinguish between those geometrical isomers. However, the total energy of model B2 is slightly higher than B1 by ~1.3 kcal/mol (Table 2), making it a somewhat less favorable candidate for model B.

We thus conclude that IR analysis alone is very useful in eliminating *some* of the possible B structures, though still leaving us with more than one good candidate (B1, B2, and B3). We can rank the candidates having the best qualitative agreement, with models B1 and B2 ahead of B3. The total energies (Table 2) of the isomers correlate well with this ranking and suggest that B1 is the most stable structure.

An open question is whether the experimental spectrum of model B corresponds to a single isomer or to a superposition of two or more. This is directly related to whether there is fast interconversion among the CO/CN ligands. Several studies^{6,31} have suggested that the reaction mechanism by which CN[–] nucleophiles attack model A to form model B involves a transition state in which one of the terminal CO moves to a bridging position, thus exposing the corresponding Fe atom to nucleophilic attack. The energy of this (rotated) transition state is ~12 kcal/mol at room temperature. The presence of a pathway involving this transition state suggests that some interconversion occurs, at least during the reaction, to form structure B from structure A. So far, however, it is unknown whether once structure B is formed, can it exploit this particular pathway to facilitate fast intraconversions.^{6,31} Experimentally it is well-known that nucleophilic substitution stops after two CO's have been replaced by CN's.^{12–14} If the proposed transition state were accessible in model B, it would have exposed one of the Fe atoms to further nucleophilic attacks to form species with three or more CN groups. This indicates that either there is no fast rearrangements of the CO/CN ligands in model B or that the mechanism is different and does not involve the transition state

TABLE 3: Selected Bond Lengths and Angles of Model B1^a

					
Selected Bond Lengths [Å]					
Fe(1)–Fe(2)	2.558 (2.517) 1.6%	Fe(1)–S(1)	2.276 (2.269) 0.3%	Fe(1)–S(2)	2.272 (2.266) 2.6%
S(1)–C(4)	1.848 (1.820) 1.5%	S(2)–C(6)	1.844 (1.830) 0.8%	C(4)–C(5)	1.523 (1.505) 1.2%
C(6)–C(5)	1.522 (1.540) –1.2%	Fe(1)–C(1)	1.907 (1.929) –1.1%	Fe(1)–C(2)	1.743 (1.769) –1.5%
Fe(1)–C(3)	1.743 (1.742) 0.1%	CN(1)	1.181 (1.150) 2.7%	CO(2)	1.181 (1.158) 2.0%
CO(3)	1.178 (1.167) 0.9%	Fe(2)–C(7)	1.743 (1.749) –0.9%	Fe(2)–C(8)	1.904 (1.933) –1.5%
Fe(2)–C(9)	1.740 (1.728) 0.7%	CO(7)	1.177 (1.162) 1.3%	CN(8)	1.182 (1.158) 2.1%
CO(9)	1.180 (1.176) 0.3%	Selected Angles [deg]			
S(1)–Fe(1)–S(2)	85.1 (85.7) –2.3%	Fe(1)–S(1)–Fe(2)	68.3 (67.2) 1.6%	Fe(1)–S(2)–Fe(2)	68.2 (67.2) 1.5%
C(1)–Fe(1)–C(2)	96.6 (98.2) –1.6%	C(2)–Fe(1)–C(3)	99.8 (98.8) 1.0%	C(1)–Fe(1)–C(3)	89.0 (90.1) 1.2%
C(1)–Fe(1)–S(1)	86.0 (84.0) 2.4%	C(1)–Fe(1)–S(2)	156.3 (158.9) –1.6%	C(2)–Fe(1)–S(1)	110.9 (106.8) 3.2%
C(2)–Fe(1)–S(2)	106.9 (102.6) 4.2%	C(3)–Fe(1)–S(2)	89.2 (89.8) –0.7%		

^a Experimental data (in parentheses) are from Schmidt et al.¹³**Figure 4.** The simulated (blue lines) IR spectra of models B1–B5. Experimental peak frequencies and intensities from ref 13 are shown as black stems, identical in each panel. A Gaussian broadening of 7 cm^{−1} is used, as for model A.

discussed above. If we *assume* nevertheless that fast interconversions occurred then we get a thermodynamic equilibrium distribution, having population fractions for the B isomers of 85%, 9%, 0.1%, 2%, and 4% for B1 to B5, respectively, using our calculated ground state energies. Therefore the spectrum would still be dominated by the lowest energy configurations B1 and B2 and primarily B1.

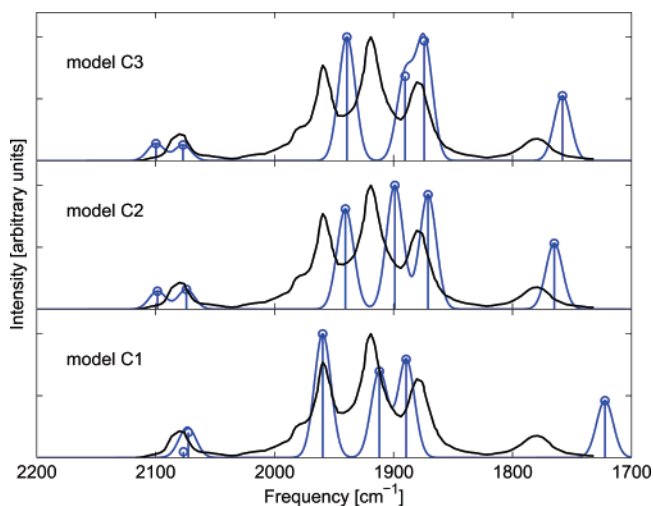
4.3. Model C. In contrast to model B where the correct structure is known, the placement of the CO/CN groups is unknown for models C since this structure was not solved by X-ray crystallography. We apply the same analysis as before to identify the most probable isomer/s of model C in two oxidation states.

We start with the reduced Fe(I)Fe(I) state. C1 is the lowest energy isomer, about 3.4 kcal/mol lower than the second-ranking trans configuration, C2 (Table 4).

The IR spectra of these species are shown in Figure 5. Consistent with the relative total energies, the cis configuration C3 shows the least agreement with the experimental spectrum:¹⁹ the CN bands are split and the CO bands are incorrectly spaced.

TABLE 4: Internal Energies (kcal/mol) of Models C in the Three Oxidation States, Relative to the Lowest Energy Model C1^a

model	Fe(I)Fe(I)	Fe(II)Fe(I)	Fe(II)Fe(II)
C1	0.00	0.00	0.00
C2	3.44	5.40	6.82
C3	4.97	4.41	7.53

^a The estimated error is smaller than 1 kcal/mol (see the Appendix).**Figure 5.** The simulated (blue lines) IR spectra of models C1–C3 in the reduced Fe(I)Fe(I) state. The experimental spectrum (black lines) was adapted from ref 19 (reproduced by permission of The Royal Society of Chemistry). A Gaussian broadening of 7 cm^{−1} is used.

Model C2 (trans configuration) is in better agreement for both frequencies and intensities, but there the splitting in the CN modes, found for C3, is still present. The frequencies of model C1 are in the best agreement for the central CO bands and also capture the near degeneracy of the CN modes. However, the intensity and the position of the low-frequency μ -CO mode are not very well represented.

On the basis of the IR spectra alone we can rule out only model C3 which is the structure originally inferred by Razavet et al.¹⁹ When considering also the total energies, however, we conclude that model C1 is probably the correct structure.

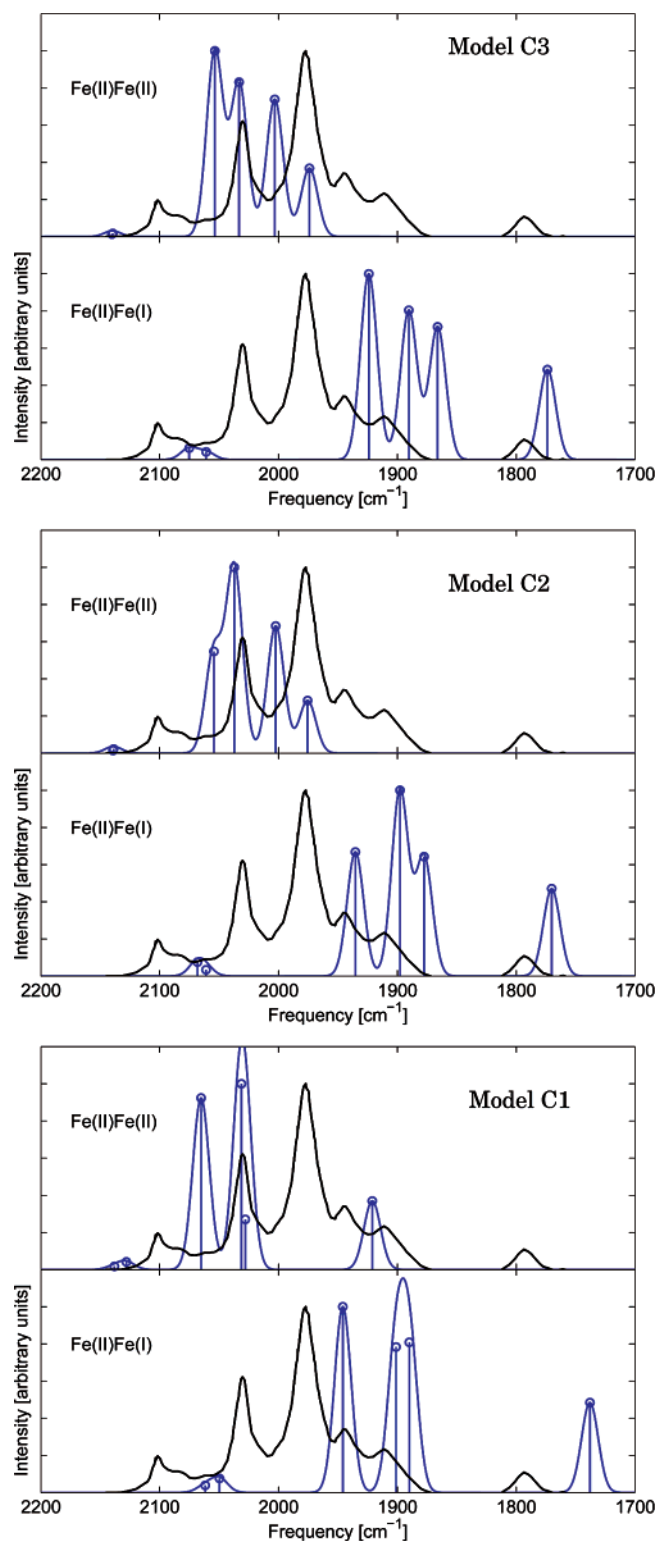


Figure 6. The simulated (blue lines) IR spectra of models C1–C3 in the mixed valence Fe(II)Fe(I) and fully oxidized state Fe(II)Fe(II). The experimental spectrum (black line) of the presumed mixed-valence state is from ref 17 (reproduced by permission of The Royal Society of Chemistry). A Gaussian broadening of 7 cm^{−1} is used.

Turning to models C in the mixed-valence state [Fe(II)Fe(I)], we note in Figure 6 that the experimental data¹⁷ are not as clean as for the Fe(I)Fe(I) case. In particular, the CO bands have five peaks instead of the four expected based on the number of CO groups. This mixed-valence species was not isolated and crystallized; it is a *transient* species obtained by electrochemical methods or chemical oxidation in rapid, stopped-flow experi-

TABLE 5: Fit Parameters for Mixing of All Possible C Isomer Pairs in Two Different Oxidation States^a

	x	$\delta\nu_1$ [cm ^{−1}]	$\delta\nu_2$ [cm ^{−1}]
C1–C1	0.37	43	−55
C2–C2	0.41	39	−19
C3–C2	0.48	51	−16
C1–C2	0.38	29	−19
C2–C1	0.43	36	−48
C1–C3	0.37	28	−23
C3–C1	0.52	49	−43
C3–C3	0.45	49	−21
C2–C3	0.39	37	−23

^a Each row is labeled C α –C β , where α and β are the indexes of the isomer pair. C α is in the (II)(I) oxidation state, having a weight x in the fit, and C β is in the (II)(II) oxidation state with a weight $Y = 1 - x$. Both $\delta\nu_1$ and $\delta\nu_2$ are rigid shift parameters (cf. eq 1).

ments, and it is likely that the experimental spectrum has contributions coming from two or more species, possibly in different oxidation states.

Figure 6 shows the computed IR spectra of models C in the mixed-valence state, as well as in the fully oxidized Fe(II)Fe(II) state. It is evident in all cases that the mixed-valence state samples the lower part of the experimental CO band, whereas the fully oxidized state samples the higher frequency part. Thus it is reasonable to assume that the experimental spectrum has contributions coming from both a mixed-valence and an oxidized state.

We explored this assertion by studying all possible isomer pairs to determine how well a mixture of the mixed-valence and the fully oxidized states can reproduce the experimental spectrum. We used a nonlinear fitting procedure to optimize the theoretically predicted spectrum $S(\nu)$

$$S(\nu) = xS_{\text{Fe(II)Fe(I)}}(\nu + \delta\nu_1) + (1 - x)S_{\text{Fe(II)Fe(II)}}(\nu + \delta\nu_2) \quad (1)$$

where ν is the frequency, $S_{\text{Fe(II)Fe(I)}}$ and $S_{\text{Fe(II)Fe(II)}}$ are the spectra at the two oxidation states, and x and $(1 - x)$ are their relative weights. We vary x , $\delta\nu_1$, and $\delta\nu_2$ to obtain a best least-squares fit to the experimental spectrum. It is important to note that this is the only context in which we have employed empirical frequency shifts. When assessing the quality of the fit, special attention was given to finding, *when possible*, a best fit that reproduces the correct experimental number of peaks by a proper choice of the starting point of the fitting procedure. The other criteria were the deviations with respect to experiment of peak positions and intensity pattern. The results for the fit parameters are presented in Table 5 and for the spectra in Figure 7. It is curious that the shifts are of opposite sign for the two oxidation states. It would be interesting to build a larger database in the future regarding the sensitivity of the shift sign to the oxidation state.

We focus first on superpositions of the same isomer in the two oxidation states. It is clear from Figure 7a,b that C1 alone and C2 alone, each in both oxidation states, are good candidates for the observed species, reproducing the right number of peaks. C1 reproduces the intensities better while C2 reproduces the peak positions better. Model C3 alone, in both oxidation states (Figure 7h), has one extra peak in the simulated spectrum. However, mixed superpositions of two isomers in two different oxidation states may occur if different isomers are more stable at different oxidation states. In our total energy calculations isomer C1 is always the most stable followed by C2 and C3, though the differences between C2 and C3 are very small. Thus an isomerization from C1 to other isomers upon oxidation is less likely and indeed superpositions of the spectrum of C1 in

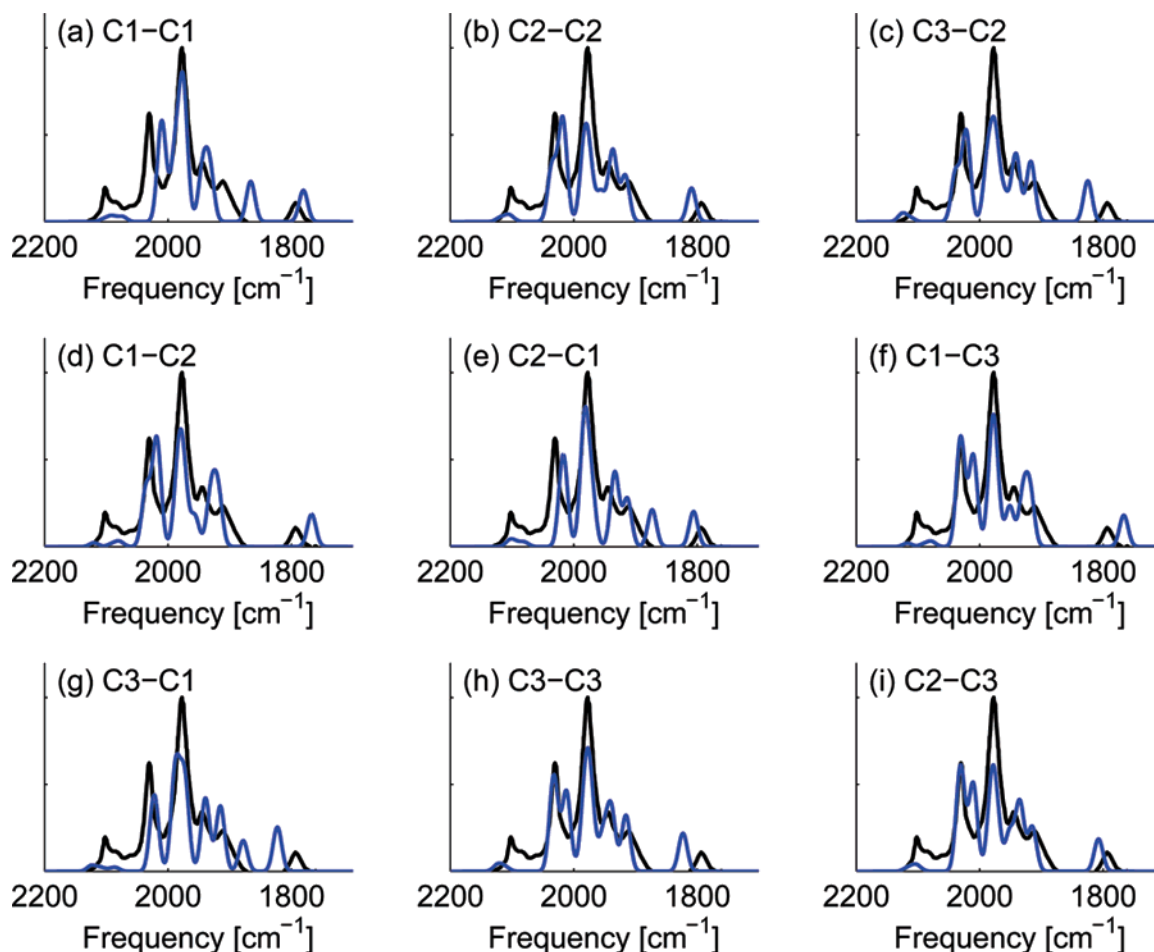


Figure 7. Model C in the mixed-valence state. Simulated IR spectra (blue lines) and experimentally reported IR spectrum (black lines, reproduced from ref 17 by permission of The Royal Society of Chemistry). As explained in the text, the experimental spectrum is most likely a superposition of the mixed-valence Fe(II)Fe(I) and the fully oxidized Fe(II)Fe(II) states. A superposition fit of the nine combinations of pairs of models C in these two oxidation states is shown. The label of each plot is $C\alpha-C\beta$, where α is the index of the isomer in the (I)(II) oxidation state and β is the isomer in the (II)(II) oxidation state. A Gaussian broadening of 7 cm^{-1} is used.

the mixed valence state with any of the other isomers in the fully oxidized state do not agree very well with the experimental spectrum (Figure 7). However, mixing C3 with C2 (Figure 7c) also yields a surprisingly good agreement with experiment, but based on the total energies we believe that these are unlikely to be the correct species.

4.4. Comparison with Earlier Calculations. Several other theoretical explorations of [FeFe]-hydrogenase used models A and B to calibrate and demonstrate the accuracy of their calculations. Cao and Hall,¹⁵ as well as Bruschi et al.,²¹ used combinations of Gaussian-type basis sets with the B3LYP³² (having the Becke exchange and Lee–Yang–Parr correlation functionals) approximation to the exchange correlation functional. B3LYP is known generally, though not always, to perform better than pure generalized gradient (GGA) functionals, such as the PBE²³ used here, within a given computational scheme. However, in electronic-structure calculations the quality of the results is also very sensitive to the particular basis set used in the different computational schemes. Obtaining results converged with respect to the basis set is not a trivial task for atomically localized orbitals especially when it comes to predicting IR intensities. In contrast, a plane-wave basis set has the advantage that the basis functions are naturally delocalized and orthogonal. The energy cutoff of the plane waves serves as a simple tuning parameter. Moreover, the absence of dependence of the plane waves on the atomic positions provides advantages

in computing forces and vibrational frequencies as well as intensities.

These advantages show up in improved accuracy of the IR spectra. In Figures 8 and 9 we compare our calculations with experiment and calculations by Cao and Hall,¹⁵ showing their original *computed spectrum* and their *predicted spectrum* after applying an empirical linearly scaled shift of the frequencies to match the experiments. For model A (Figure 8), in our IR spectrum the two higher frequency peaks overlap the experimental spectrum closely, and the lowest peak is blue-shifted by about $20\text{--}25\text{ cm}^{-1}$ with respect to experiment, whereas in ref 15 the entire spectrum is blue-shifted by $\sim 60\text{ cm}^{-1}$. More important, in our calculations the *spacing* between the peaks and the overall shape of the spectrum are in much better agreement with experiment. Even the linear scaling of the computed spectrum results in poor agreement with experiment, particularly with regard to the number of peaks. Similarly, our calculations for model B are centered on the experimental spectrum, while they are blue-shifted by $\sim 70\text{--}80\text{ cm}^{-1}$ in the calculations of Cao and Hall.¹⁵ After applying their empirical linear shift and scale, the spectrum is of similar quality to ours. However, these two examples, A and B, show that an empirical fitting formula does not guarantee uniform quality of the predicted frequencies.

Liu and Hu²⁰ also reported DFT calculations with a plane wave code, similar to ours. They do not report the intensities

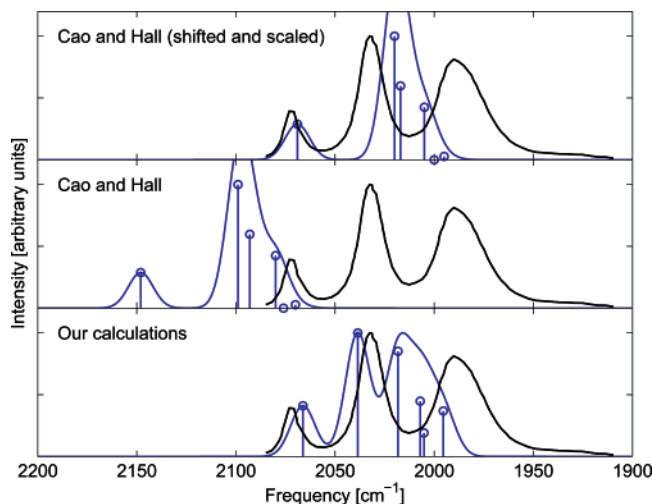


Figure 8. Simulations (blue lines) of the IR spectrum of model A (black line, identical in all panels, reproduced with permission from ref 12). The lower panel contains our data, while the other two panels are reconstructed from the theoretical results of Cao and Hall.¹⁵ The middle panel is their original data, whereas the upper panel is the result of their empirical linear fit to the experimental data. A Gaussian broadening of 7 cm^{-1} is used.

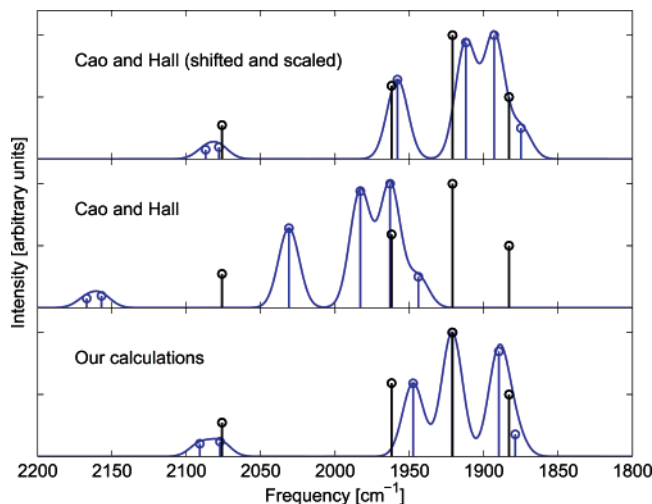


Figure 9. Simulations (blue lines) of the IR spectrum of model B (black stem lines, identical in all panels, reported in ref 13). The lower panel contains our data, while the other two panels are reconstructed from the theoretical results of Cao and Hall.¹⁵ The middle panel is their original data, whereas the upper panel is the result of their empirical linear fit to the experimental data. A Gaussian broadening of 7 cm^{-1} is used.

of the IR transitions, but their computed frequencies of model B deviate from experiment by $\sim 115 \text{ cm}^{-1}$. The differences from our results may arise from the DFT functional (which is not specified there), the Fe pseudopotential (which seems not to include the semicore 3s and 3p states and we found them to be important), or insufficient convergence in the structural optimization.

5. Conclusions and Outlook

DFT calculations were used to study synthetic analogues of the diiron subcluster in the active site of [FeFe]-hydrogenase. We show that well-converged calculations of energy, structure, and vibrations can accurately reproduce the IR spectra of these systems. We are able to distinguish among various structural isomers of models of [FeFe]-hydrogenase which are currently unresolvable in the enzyme itself. In particular, our results for

model system B are in best agreement with those of the experimentally observed isomer B1. We also analyze, for the first time, synthetic models in the mixed-valence state (models C). Our analysis shows that model C1 is consistently the lowest energy isomer, and its IR spectrum is consistent with experiment for all oxidation states. Model C2, though less stable than C1, is also consistent with IR data, and in the absence of X-ray data we cannot rule out this isomer.

Our results for models A and B clearly demonstrate the high accuracy that can be obtained from converged DFT calculations in the study of [FeFe]-hydrogenase, which is better than what one might expect based on earlier theoretical explorations. By combining available IR data with *ab initio* calculations, we can make important structural predictions, as demonstrated for models B and C.

In forthcoming papers we shall exploit this high accuracy to address important unresolved structural uncertainties for [FeFe]-hydrogenase. We shall point out a previously unrecognized inconsistency between the known structure of the diiron cluster and the available IR data. We shall also show that the combination of calculations and experimental measurements of the IR spectrum of [FeFe]-hydrogenase can be used to refine the known structure of the diiron subcluster to resolve that particular uncertainty and to address the uncertainty in the nature of the bridging ligand referred to in the Introduction.

Appendix: Accuracy of Relative Energies of Structural Isomers

An important quantity in our calculations is the relative energy of various structural isomers. Pure generalized gradient (GGA) functionals, such as the PBE,²³ are known to have a typical accuracy of $\sim 10\text{--}30 \text{ kcal/mol}$ for the standard enthalpy of formation,³³ with a systematic underestimation, and even hybrid functionals, such as the B3LYP, have errors of $\sim 5 \text{ kcal/mol}$. However, when calculating *relative energies*, such as proton affinities or hydrogen bond energies, the errors are much smaller, better than 1 kcal/mol for PBE.³³ The energy differences that are relevant for our study are in the range of 1–5 kcal/mol . One can expect that for isomers as similar as those of models B or C (Figure 2) the relative energy is of considerably higher accuracy, due to the cancellation of systematic components in the DFT error. However, there is no general theorem to support that assumption. It is therefore very important to show that at least for one representative [FeFe]-hydrogenase analogue, the relative energies predicted by DFT are meaningful.

We do this by repeating some of our calculations using the more accurate B3LYP functional and comparing the relative energies obtained. Specifically, we studied the total energies of models B1 and B3 using both the PBE and the B3LYP functionals in the Gaussian03³⁴ electronic structure code with a very large (all electron) standard basis set [6-311++G(3df,3pd), SCF=Tight]. These were total energy calculations at the optimized geometry obtained from our plane wave PBE calculations. For both functionals, model B1 has lower energy than B3, by 4.01 kcal/mol for the PBE functional and by 4.77 kcal/mol for the B3LYP. The difference between the results with the two functionals is less than 1 kcal/mol , and is reasonably smaller than the relative energies of the isomers. Though the absolute energies predicted with the two functionals are quite different, the energy difference between the two isomers is very similar and consistent. Therefore, we can reasonably assume that there is a systematic cancellation of errors for a given functional when comparing such similar structures, and the error in the energy differences is about 1

kcal/mol or less. Moreover, our use of pseudopotentials is validated by this comparison with an all-electron computation.

Acknowledgment. We thank M. Y. Darensbourg for permission to reproduce the IR spectrum of model A and C. Pickett for permission to reproduce the spectra of model C in the Fe-(I)Fe(I) and Fe(II)Fe(I) oxidation states. We gratefully acknowledge financial supported from the National Science Foundation grant CHE-012432 (R.C., S.Z., M.H.C.), DARPA grant BAA 04-12 "New technology for Inside-Out Design of novel protein Catalysts" (R.C., S.Z.), and the Princeton Center for Environmental and Bioinorganic chemistry, funded by the National Science Foundation and the Department of Energy (E.I.S.).

References and Notes

- (1) Nicolet, Y.; Piras, C.; Legrand, P.; Hatchikian, C. E.; Fontecilla-Camps, J. C. *Struct. Folding Design* **1999**, 7 (1), 13–23.
- (2) Nicolet, Y.; de Lacey, A. L.; Vernet, X.; Fernandez, V. M.; Hatchikian, E. C.; Fontecilla-Camps, J. C. *J. Am. Chem. Soc.* **2001**, 123 (8), 1596–1601.
- (3) Siegbahn, P. E. M.; Blomberg, M. R. A. *Annu. Rev. Phys. Chem.* **1999**, 50, 221–249.
- (4) Parr, R. G.; Yang, W. *Density-Functional Theory of Atoms and Molecules*; Oxford University Press: Oxford, UK, 1989; Vol. 16.
- (5) Darensbourg, M. Y.; Lyon, E. J.; Smee, J. J. *Coord. Chem. Rev.* **2000**, 206, 533–561. Evans, D. J.; Pickett, C. J. *Chem. Soc. Rev.* **2003**, 32 (5), 268–275. He, C. J.; Wang, M.; Li, M. A.; Sun, L. C. *Prog. Chem.* **2004**, 16, (2), 250–255. Madamwar, D.; Garg, N.; Shah, V. *World J. Microbiol. Biotechnol.* **2000**, 16 (8–9), 757–767. Nandi, R.; Sengupta, S. *Crit. Rev. Microbiol.* **1998**, 24 (1), 61–84. Tamagnini, P.; Axelsson, R.; Lindberg, P.; Oxelfelt, F.; Wunschi, R.; Lindblad, P. *Microbiol. Mol. Biol. Rev.* **2002**, 66 (1), 1+. Vignais, P. M.; Billoud, B.; Meyer, J. *FEMS Microbiol. Rev.* **2001**, 25 (4), 455–501. Adams, M. W. W.; Stiefel, E. I. *Science* **1998**, 282 (5395), 1842–1843. Adams, M. W. W.; Stiefel, E. I. *Curr. Opin. Chem. Biol.* **2000**, 4 (2), 214–220.
- (6) Georgakaki, I. P.; Thomson, L. M.; Lyon, E. J.; Hall, M. B.; Darensbourg, M. Y. *Coord. Chem. Rev.* **2003**, 238, 255–266.
- (7) Tsuda, M.; Dino, W. A.; Kasai, H. *Solid State Commun.* **2005**, 133 (9), 589–591. Karyakin, A. A.; Morozov, S. V.; Karyakina, E. E.; Zorin, N. A.; Pereygin, V. V.; Cosnier, S. *Biochem. Soc. Trans.* **2005**, 33, 73–75. Darensbourg, M. Y. *Nature* **2005**, 433 (7026), 589+. Lamle, S. E.; Halliwell, L. M.; Armstrong, F. A.; Albracht, S. P. J. *Inorg. Biochem.* **2003**, 96 (1), 174–174. Karyakin, A. A.; Morozov, S. V.; Karyakina, E. E.; Varfolomeyev, S. D.; Zorin, N. A.; Cosnier, S. *Electrochem. Commun.* **2002**, 4 (5), 417–420. Tsujimura, S.; Fujita, M.; Tatsumi, H.; Kano, K.; Ikeda, T. *Phys. Chem. Chem. Phys.* **2001**, 3 (7), 1331–1335. Woodward, J.; Cordray, K. A.; Edmonston, R. J.; Blanco-Rivera, M.; Mattingly, S. M.; Evans, B. R. *Energy Fuels* **2000**, 14 (1), 197–201. De Lacey, A. L.; Detcherry, M.; Moiroux, J.; Bourdillon, C. *Biotechnol. Bioeng.* **2000**, 68 (1), 1–10.
- (8) Tard, C.; Liu, X. M.; Ibrahim, S. K.; Bruschi, M.; De Gioia, L.; Davies, S. C.; Yang, X.; Wang, L. S.; Sawers, G.; Pickett, C. J. *Nature* **2005**, 433 (7026), 610–613.
- (9) Peters, J. W. *Science* **1999**, 283 (5410), 2102–2102.
- (10) Peters, J. W.; Lanzilotta, W. N.; Lemon, B. J.; Seefeldt, L. C. *Science* **1998**, 282 (5395), 1853–1858.
- (11) Pierik, A. J.; Hulstein, M.; Hagen, W. R.; Albracht, S. P. J. *Eur. J. Biochem.* **1998**, 258 (2), 572–578. Pierik, A. J.; Roseboom, W.; Happe, R. P.; Bagley, K. A.; Albracht, S. P. J. *J. Biol. Chem.* **1999**, 274 (6), 3331–3337. Lyon, E. J.; Shima, S.; Boecher, R.; Thauer, R. K.; Grevels, F. W.; Bill, E.; Roseboom, W.; Albracht, S. P. J. *J. Am. Chem. Soc.* **2004**, 126 (43), 14239–14248.
- (12) Lyon, E. J.; Georgakaki, I. P.; Reibenspies, J. H.; Darensbourg, M. Y. *Angew. Chem., Int. Ed.* **1999**, 38 (21), 3178–3180.
- (13) Schmidt, M.; Contakes, S. M.; Rauchfuss, T. B. *J. Am. Chem. Soc.* **1999**, 121 (41), 9736–9737.
- (14) Le Cloirec, A.; Best, S. P.; Borg, S.; Davies, S. C.; Evans, D. J.; Hughes, D. L.; Pickett, C. J. *Chem. Commun.* **1999**, (22), 2285–2286.
- (15) Cao, Z. X.; Hall, M. B. *J. Am. Chem. Soc.* **2001**, 123 (16), 3734–3742.
- (16) Lawrence, J. D.; Rauchfuss, T. B.; Wilson, S. R. *Inorg. Chem.* **2002**, 41 (24), 6193–6195.
- (17) Razavet, M.; Borg, S. J.; George, S. J.; Best, S. P.; Fairhurst, S. A.; Pickett, C. J. *Chem. Commun.* **2002**, (7), 700–701.
- (18) Razavet, M.; Le Cloirec, A.; Davies, S. C.; Hughes, D. L.; Pickett, C. J. *J. Chem. Soc., Dalton Trans.* **2001**, (24), 3551–3552.
- (19) Razavet, M.; Davies, S. C.; Hughes, D. L.; Pickett, C. J. *Chem. Commun.* **2001**, (9), 847–848.
- (20) Liu, Z. P.; Hu, P. J. *Am. Chem. Soc.* **2002**, 124 (18), 5175–5182.
- (21) Bruschi, M.; Fabucci, P.; De Gioia, L. *Inorg. Chem.* **2001**, 41 (6), 1421–1429.
- (22) Perdew, J. P.; Burke, K.; Ernzerhof, M. *Phys. Rev. Lett.* **1997**, 78 (7), 1396–1396.
- (23) Perdew, J. P.; Burke, K.; Ernzerhof, M. *Phys. Rev. Lett.* **1996**, 77 (18), 3865–3868.
- (24) *Quantum-ESPRESSO*, 3.0; 2005 (www.pwscf.org).
- (25) Laasonen, K.; Car, R.; Lee, C.; Vanderbilt, D. *Phys. Rev. B* **1991**, 43 (8), 6796–6799. Vanderbilt, D. *Phys. Rev. B* **1990**, 41 (11), 7892–7895.
- (26) Bachelet, G. B.; Schlüter, M. *Phys. Rev. B* **1982**, 25 (4), 2103–2108.
- (27) Adams, M. W. W. *Biochim. Biophys. Acta* **1990**, 1020 (2), 115–145.
- (28) Gonze, X.; Lee, C., *Phys. Rev. B* **1997**, 55 (16), 10355–10368.
- (29) Wilson, E. B. *Molecular vibrations: The theory of infrared and Raman vibrational spectra*; McGraw-Hill: New York, 1955.
- (30) Pick, R. M.; Cohen, M. H.; Martin, R. M. *Phys. Rev. B* **1970**, 1 (2), 910–919.
- (31) Darensbourg, M. Y.; Lyon, E. J.; Zhao, X.; Georgakaki, I. P. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, 100 (7), 3683–3688. Lyon, E. J.; Georgakaki, I. P.; Reibenspies, J. H.; Darensbourg, M. Y. *J. Am. Chem. Soc.* **2001**, 123 (14), 3268–3278.
- (32) Becke, A. D. *J. Chem. Phys.* **1993**, 98 (7), 5648–5652. Lee, C.; Yang, W.; Parr, R. G. *Phys. Rev. B* **1988**, 37 (2), 785–789.
- (33) Staroverov, V. N.; Scuseria, G. E. *J. Chem. Phys.* **2003**, 119 (23), 12129–12137.
- (34) Frisch, M. J., et al. *Gaussian 03*; Gaussian Inc.: Pittsburgh, PA, 2003.