Targeting Intermediates of [FeFe]-Hydrogenase by CO and CN Vibrational Signatures

Yu, Lian; Greco, Claudio; Bruschi, Maurizio; Ryde, Ulf; De Gioia, Luca; Reiheet, Markus

Published in:
Inorganic Chemistry

DOI:
10.1021/ic102039z

2011

Targeting Intermediates of [FeFe]-Hydrogenase by CO and CN Vibrational Signatures

Lian Yu[a], Claudio Greco[b], Maurizio Bruschi[b,1], Ulf Ryde[c], Luca De Gioia[b] and Markus Reiher[a,1]

[a] Laboratorium für Physikalische Chemie, ETH Zurich, Wolfgang-Pauli-Strasse 10, CH-8093 Zürich, Switzerland

[b] Department of Biotechnology and Biosciences, and Department of Environmental Science, University of Milano-Bicocca, Piazza della Scienza 1-2, 20126 Milan, Italy

[c] Department of Theoretical Chemistry, Lund University, Chemical Center, P.O.B. 124, Lund 22100, Sweden

Abstract

In this work we employ density functional theory to assign vibrational signatures of [FeFe]-hydrogenase intermediates to molecular structures. For this purpose, we perform an exhaustive analysis of structures and harmonic vibrations of a series of CN and CO containing model clusters of the [FeFe]-hydrogenase enzyme active site considering also different charges, counter ions and solvents. The pure density functional BP86 in combination with a triple-zeta polarized basis set produce reliable molecular structures as well as harmonic vibrations. Calculated CN and CO stretching vibrations are analyzed separately. Our results are discussed in close comparison to previous studies. Finally, we apply the scaled vibrational frequencies to assign intermediates in [FeFe]-hydrogenase's reaction cycle.

Date: 17th of December, 2010
Keywords: DFT calculations; infrared spectroscopy; Hydrogenases

[1]Authors to whom correspondence should be addressed: M. Bruschi (email: maurizio.bruschi@unimib.it), M. Reiher (email: markus.reiher@phys.chem.ethz.ch)

1 Introduction

In the global biological hydrogen metabolism three types of metal containing enzymes, [NiFe]-hydrogenases, [FeFe]-hydrogenases, and [Fe]-hydrogenases play a key role [1–8]. Generally [FeFe]-hydrogenases catalyze the reduction of protons to yield H_2 [9–11]. The active site of [FeFe]-hydrogenase, referred to as the H-cluster, contains a dinuclear iron cluster with CO and CN^- ligands, known as $[2Fe]_H$ cluster, to which a Fe_4S_4 cubane is attached by the sulphur atom of a cysteine residue (Figure 1). In the last decade, many experimental and quantum chemical studies focussed on their structures and redox properties, and much progress has been made regarding the elucidation of the reaction mechanism [11–23]. For example, mechanisms for heterolytic cleavage of bound H_2 have been studied, based either on a transient formation of a terminal or on a bridging hydride isomer [11, 24–29]. Of course, interest into the mechanism of these enzymes is triggered by the importance of the catalyzed reaction for society: it is most desiderable to fully understand the individual steps of the catalytic cycle to eventually create biomimetic catalysts for hydrogen production.

[Figure 1 about here.]

However, spectroscopic evidence on reactive intermediates often does not allow one to directly draw a conclusion on the molecular structure that produces the spectroscopic signals. Then, theoretical studies may close the gap by producing spectral signatures of well characterized structures to be ultimately compared with the experimental signals. In such a way, especially when thoroughly calibrated, quantum chemical studies can provide reliable insight into structure and reactivity of intermediates from the catalytic cycle.

Infrared (IR) spectroscopy can be a very useful tool for studying the intermediate states of the catalytic cycle by monitoring the shifts in wave number of the CO and CN stretching vibrations. The vibrational frequencies of these diatomic ligands might be a sensitive indicator for changes in the environment [9, 30]. In 2002, Liu and Hu showed that there is a linear relationship between DFT calculated CO stretching frequencies and the experimental data obtained for mononuclear Fe complexes [15].

Later, Darensbourg, Hall and co-workers have employed DFT — with the B3LYP density functional and a double-zeta basis set plus polarization functions — to predict the solution-phase IR spectra for 15 model clusters for the active site of [FeFe]-hydrogenase. These authors confirmed that a simple linear scaling of the computed stretching fre-

quencies yields accurate predictions of the experimentally determined values [23, 31]. Darensbourg, Hall and collaborators then applied the scaling factors obtained to elaborate on the IR signals of the potential intermediates [23, 31].

However, the scatter of the calculated frequencies around the experimentally measured frequencies is still fairly large in Refs. [23, 31], and hence the uncertainty of the prediction might be too large. Therefore, we attempt to re-calibrate calculated vibrational frequencies employing a different density functional, namely BP86 [32, 33], and a larger triple-zeta basis set, which are known to produce quite accurate vibrational frequencies [34–37]. Still, it needs to be shown whether the scatter of the calculated frequencies around their experimental references and, in particular, the scatter of the *scaled* frequencies can be reduced. If so, we may be able to more reliably connect the experimental vibrational signatures of [FeFe]-hydrogenases to molecular structures optimized in quantum chemical calculations.

In order to base our calibration on [FeFe]-hydrogenase model clusters in many different conditions, such as in different environments and as neutral, positively or negatively charged species, we selected a variety of model clusters whose structures have all been experimentally known and no ambiguity regarding the first-shell ligand environment is present. Also, of course, their infrared spectra must have been measured.

The goal of this study is to develop and test a computational methodology for a sufficiently accurate prediction of the IR spectra of the active site of [FeFe]-hydrogenase. It is decisive to understand that even after scaling only little individual deviation can be tolerated in order to unambiguously assign intermediates.

2 Computational Methods

All-electron Kohn–Sham DFT calculations were performed with the quantum chemical program package TURBOMOLE [38, 39]. We applied the pure density functional BP86 [32, 33] (in combination with the resolution-of-the-identity ('RI') density fitting technique as implemented in TURBOMOLE) as well as the hybrid functional B3LYP [40–42]. The TZVP basis set by Schäfer *et al.* [43] was employed. It is well known that this combination of a high quality basis set and the BP86 density functional produces accurate molecular structures (see, e.g., Refs. [34, 44–46]) as well as harmonic vibrational frequencies (see, e.g., Refs. [34–36]). The latter is due to a fortunate error compensation [35] that allows us to directly compare harmonic BP86 frequencies with

experimental fundamental ones. All frequency calculations were carried out with the program package SNF [47]. The calculations of solvation effects have been carried out with the default COSMO parameters of the TURBOMOLE/COSMO implementation and with optimized atomic COSMO radii taken from Refs. [48, 49]. In particular, for the Fe atom we applied the COSMO radius of 1.393 Å from Ref. [50]. The molecular structures were visualized with the program MOLDEN [51].

All structures of the selected model clusters were optimized in an unrestricted low-spin state, in order to check for possible anti-ferromagnetic couplings between Fe atoms. In fact, in the single-determinant approximation of DFT, anti-ferromagnetic interactions are modeled according to the broken-symmetry approach introduced by Noodleman [52, 53], which produce Kohn–Sham determinants of spin-unrestricted type, often corresponding to the lowest energy states (ground states) for metal clusters (see, e.g., Refs. [54, 55]). However, for all model clusters considered in the calibration set, a with vanishing total spin, S=0, the unrestricted (open-shell) calculation always converge to the restricted (closed-shell) solution.

3 Model Clusters and Calibration of the Calculated CO and CN Frequencies

There are altogether 32 model clusters that have been considered for this study (figures of all BP86-optimized structures are provided in the Supporting Information). They can be classified in four groups: eleven of them are neutral clusters and their IR spectra were measured in solid phase (e.g. KBr); the IR spectra of the other 21 clusters, which include eight neutral, five negatively and eight positively charged clusters, were measured in solution (e.g. in THF). Furthermore, we selected not only dinuclear iron clusters, but also a few tri- and tetranuclear iron clusters (for example, complex **23** is a trinuclear iron cluster and complexes **19**, **21** and **22** are tetranuclear iron clusters; see Supporting Information for the numbering scheme). Table 1 lists the references to the original literature, solvent and counter-ions of all model clusters selected.

[Table 1 about here.]

The DFT-optimized molecular structures are usually in good agreement with the experimentally determined crystal structures. For a statistical analysis of the structural data calculated, we refer to Table 2 which provides the mean signed deviation (MSD),

the mean unsigned deviation (MUD), and the maximum deviation (MaxDev) for all Fe–Fe, Fe–S. Fe–P, Fe–CO, and Fe–CN distances for which X-ray crystallographic data are available. Obviously, the agreement of optimized and experimental clusters structures is very good and the maximum MUD is only 0.035 Å as found for the Fe–S bonds.

[Table 2 about here.]

A detailed comparison between experimental and DFT calculated structures of four clusters, which represent the four possible situations found in the 32 model clusters discussed above, is reported in the Supporting Information.

3.1 Solvent Effects and IR Intensities

In the BP86/RI/TZVP optimizations all cluster structures are obtained for the isolated species at zero Kelvin, while experimentally the IR-spectra were determined in condensed phase. Hence, it is important to assess the magnitude and possible significance of solvent effects. For this purpose, we selected complex **28**, whose IR-spectrum has been measured in CH_3CN as solvent. The molecular structure of complex **28** has been reoptimized within the COSMO dielectric *continuum* model as well as within a micro-solvated model with explicit CH_3CN solvent molecules placed at appropriate positions which has then been subjected to structure optimization (see Figure 2).

[Figure 2 about here.]

In general, the comparison with experimentally determined frequencies [19] shows that the COSMO and micro-solvated model calculations always yield about 45 cm^{-1} and 25 cm^{-1} lower wave numbers, respectively, while the isolated cluster model leads to wave numbers that are within 15 cm^{-1} of the experimental result. According to the correlation coefficients, the data from the COSMO and micro-solvation model calculations are more linearly related to the measured frequencies than the isolated model calculations. However, since micro-solvation is very demanding in terms of computing time and also not free of model artefacts, while the improvement is still comparatively small, we rely on the standard *isolated-cluster* calculations for all model clusters and hope to capture environment effects solely in the calibration parameters. Hence, we

circumvent the solvation problem by absorbing these effects into the linear regression parameters.

The calculated relative intensities obtained from isolated-cluster calculations also resemble the experimental results (see Supporting information). Therefore, we may use isolated models to obtain the CO and CN frequency raw data for our calibration.

3.2 Calibration of the Calculated CO and CN Frequencies

Calculated CN and CO frequencies for all isolated model clusters, as well as details on the assignment of the experimentally observed ν_{CX} (X = C, N) stretching frequencies are given in the Supporting Information.

Complexes **8**, **11**, **12** and **24**–**27** are CN^--containing iron clusters (see Supporting Information). Because of the different nature of the CN^- vs. the CO ligands, we analyze the CN stretching vibrations separately in order to eventually arrive at more accurate linear regression parameters.

As shown in Figure 3a, we employed three different functional forms for the regression in order to derive scaling factors for the calculated frequencies. Two of them have only one parameter, either intercept or slope, i.e., they add a constant or scale all values by a constant, respectively. Only the third one is a standard linear regression and features both parameters. In order to identify the best functional form a simple statistical analysis has been carried out on the three regression models. The results are shown in Table 3 (for details on the statistical measures see the Supporting Information). In the case of CN frequencies, the two-parameter model turned out to have the largest r^2 value. However, this model is inferior in predictive terms, as witnessed by the comparison of q^2 values reported in Table 3 forthe three different fittings. This observation allows us to restrict our analysis on the two one-parameter models. Notably, they feature very similar q^2, MSD, MUD and MaxDev values. Our preference of the form

$$\nu_{CN}(\text{Scal.}) = 0.990\ \nu_{CN}(\text{Cal.}) \tag{1}$$

is the result of a careful evaluation of significance of the parameters: the confidence intervals with 90% confidence level for the two regression parameters

$$\nu_{CN}(\text{Scal.}) = b \cdot \nu_{CN}(\text{Cal.}) + a \tag{2}$$

have been determined with $a \in [-832, 215]$ and $b \in [0.888, 1.385]$. Since the zero point falls well within the interval of intercept a, we selected the slope-based one parameter model to produce the scaled frequencies (Scal.) for all CN stretching vibrations.

The BP86 calculations yield overall higher frequencies (about 21 cm^{-1} on average) compared to the corresponding measured CN frequencies. Because of the polar nature of the CN ligand, which dissociates as a CN^- anion, larger effects of the environment, such as a solvent, can be expected. By contrast, for the less polar CO ligand calculated frequencies match the observed frequencies (*vide infra*). Hence, it is justified to scale the calculated CO frequencies separately from the CN ones.

[Figure 3 about here.]

It can be seen from Figure 3b that the intercept of the first one-parameter regression is small (-3 cm^{-1}, i.e. the calculated frequencies are only about 3 cm^{-1} larger than measured frequencies, on average). In the second one-parameter regression the slope is 0.999, thus very close to 1. These observations allow us to conclude that BP86 CO stretching vibrations are close to the observed ones. As far as a comparison between the one-parameter models and the two-parameter model is concerned, both r^2 and q^2 values show negligible differences; on the other hand, MSD, MUD and MaxDev values of the two-parameter model are globally better than those of the one parameter models. We also calculated the confidence intervals for the two regression parameters a and b, which are $a \in [-102, 4]$ and $b \in [0.997, 1.050]$. Even though the interval of intercept a includes zero, the upper limit is so close to zero that we decided to rely on the comparison among MUD, MSD and MaxDev values for the choice of the best regression approach. We thus propose to use

$$\nu_{\mathrm{CO}}(\mathrm{Scal.}) = 1.023 \cdot \nu_{\mathrm{CO}}(\mathrm{Cal.}) - 49. \tag{3}$$

to fit the BP86/RI/TZVP calculated CO frequencies. Furthermore, we have tested separate linear regressions for neutral, positively and negatively charged clusters, but found basically no difference so that a distinction with respect to electric charge is not necessary.

[Table 3 about here.]

There is only one scaled frequency with an error larger than 25 cm^{-1} compared to the measured frequencies. This is the bridging CO stretching frequency in cluster **9** (measured in solution). To assess the magnitude of solvent effects, DFT calculations with the COSMO dielectric *continuum* model were carried out for complex **9**. The COSMO calculations yield 1987, 1930, and 1878 cm^{-1} giving a much better agreement

for the bridging CO, but also a much larger error (about 40 cm^{-1}) for the other two CO frequencies. It can be seen that the changes due to the COSMO model do not improve on average the agreement between calculated and experimental frequencies for complex **9**. This might point to counter ions in solution that may play a decisive role here.

Because of the lack of experimental knowledge on the position of counter ions in clusters **9** and **24**, we could only optimize their structures without the counter ions (otherwise the structures might be too different). Since the true position of the counter ion $[Et_4N]^+$ in complex **8** is found in the X-ray determined structure in Ref. [56], the BP86/RI/TZVP optimization of complex **8** including the counter-ion has been be carried out (structure and calculated CN and CO frequencies are given in the Supporting Information). The BP86/RI/TZVP frequencies without counter ion are 2119 cm^{-1} (CN), 2021 cm^{-1}, 1981 cm^{-1}, 1958 cm^{-1}, 1950 cm^{-1}, and 1943 cm^{-1}. Upon comparison with the isolated-cluster results, two pairs of calculated stretching frequencies show significant differences: the CN frequencies, 2119 cm^{-1} and 2083 cm^{-1}, and the asymmetric CO frequencies, 1943 cm^{-1} and 1918 cm^{-1}. Thus, the data indicate that the counter ions in solution can have an influence on CN and CO stretching frequencies.

3.3 Comparison to B3LYP Results

The density functional B3LYP has been widely used in previous quantum chemical investigations of [FeFe]-hydrogenases. In order to provide a comparison to the present study, we scaled the vibrational frequencies of the same 15 clusters as those calculated with the B3LYP functional in Refs. [23, 31]. Since the CO and CN frequencies have not been analyzed separately in those B3LYP calculations, we also determined a *single* scaling factor for the BP86 calculated CO and CN frequencies of the set of complexes reported in Refs. [23, 31].

[Figure 4 about here.]

As shown by the data in Figure 4, the intercept of the first regression obtained for BP86 calculations is very small, namely -2, while for B3LYP it is -100, i.e., the averaged difference between the calculated and measured frequencies obtained for B3LYP is about 100 cm^{-1}, much larger than for the BP86 calculations. Compared to the B3LYP calculations of Refs. [23, 31] the slope of the second regression obtained for BP86 is much closer to one. The coefficient of determination (r^2) of the best (third) linear

regression of the BP86 data is 0.972, while for the B3LYP calculations [23] it is 0.948. In view of this comparison we can conclude that the BP86/RI/TZVP computational scheme is more reliable in predicting CO and CN vibrational frequencies.

However, the decisive point is whether the BP86 data are still more reliable than the B3LYP results *after* scaling. For this, the square root of the error $s^2_{y|x}$ has been reconstructed from the literature data, since only r^2 was given, while s^2_y has to be calculated from the experimental data (see the Supporting Information for the definition of these quantities). The square root of the error $s^2_{y|x}$ for the BP86 linear regression $y = 0.955x + 88$ results in about 10 cm^{-1} while it is for the B3LYP linear regression $y = 0.9319 + 43.246$ about 14 cm^{-1}. These results indicate that scaled BP86 data produce a more accurate estimate of the CO and CN stretching frequencies than the scaled B3LYP data.

4 Structural Assignment of [FeFe]-Hydrogenase Intermediates

The possibility to compute reliable scaled stretching vibrations prompted us to analyze a series of plausible H-cluster models, corresponding to characterized states of the enzyme, with the aim of assigning their structures by comparison with experimental IR spectra. The states investigated are referred to as $\mathbf{H_{ox}}$**-CO**, $\mathbf{H_{ox}}$, and $\mathbf{H_{red}}$, corresponding to the CO-inhibited, the paramagnetic *'oxidized'*, and the diamagnetic *'reduced'* forms of the enzyme, respectively. In particular, our effort was aimed at complementing and extending previous works, which used computed IR frequencies to unravel the details of the molecular structures of the IR active enzyme forms. [23,57–59]

[Figure 5 about here.]

As for the atomic composition of ligands different from cyanide and carbonyl groups in our models, notice that several lines of evidence [60,61] converge on the conclusion that the bidentate ligand bridging the Fe_p and Fe_d ions corresponds to a dithiomethylamine (dtma) residue. However, it has been possible neither to give further support nor to falsify this conclusion on the basis of IR data analysis to date. Thus, here we considered not only dtma containing models of the H-cluster, but also their propandithiolate (pdt) and dithiomethylether (dtme) containing counterparts [12, 24, 62, 63], hoping that the

new approach for IR frequencies scaling can unravel crucial details on the H-cluster atomic composition.

Since the explicit modelling of the Fe_4S_4 cluster was outside the scope of the present investigation, the CH_3SH ligand has been used to simply model the cysteine residue bridging the $[2Fe]_H$ and the Fe_4S_4 subclusters, following previous DFT investigations [11, 26, 64–66]. In order to consider the effects of possible changes in the redox state of the Fe_4S_4 cluster, the CH_3S^- ligand has also been used to model the cysteine residue. However, the comparative analysis of DFT and experimental spectra shows that all CH_3S^- containing models feature ν(CX) vibrational frequencies shifted to significantly lower wavenumber with respect to those containing the CH_3SH fragments, consequently leading to a large systematic deviation from the experimental spectra (see Tables 4, 5, 6, and 7 in Supporting Information). In fact, comparison of atomic charges for a few relevant CH_3SH and CH_3S^- model complexes shows that the additional charge in the CH_3S^- complexes is mainly localized on the π^* orbitals of the CO ligands leading to longer C-O distances and explaining the shift of the ν(CX) vibrational frequencies. In this respect, deprotonation of CH_3SH may lead to an enhancement of the σ donor character that has no parallel even in the most reduced form of the naturally occurring cubane cluster. In conclusion, the CH_3S^- ligand does not appear to be an adequate model of the cubane moiety in the H cluster for all of the redox states investigated, confirming the results of previous works [11, 65].

A schematic representation of the complexes investigated as models of the $\mathbf{H_{ox}}$**-CO**, $\mathbf{H_{ox}}$, and $\mathbf{H_{red}}$ states is reported in Figure 5. The ν(CO) and ν(CN) vibrational frequencies, and the corresponding intensities for all of the model complexes investigated are reported in Tables 5, 6, 7, and 8 in the Supporting Information. Hereafter, only the model complexes which resemble the experimental IR spectra best will be discussed in detail (a statistical analysis of the deviation from the experimental spectra is also reported in the Supplementary Information).

4.1 The $\mathbf{H_{ox}}$-CO Redox State

The first set of diiron species investigated was designed to include all plausible models of the $\mathbf{H_{ox}}$**-CO** state, i.e., the CO inhibited form of the enzyme, which has been characterized as an Fe^IFe^{II} species by Mössbauer [67–69] and EPR [70,71] spectroscopy. Scaled vibrational frequencies of CO and CN^- ligands in the models of $\mathbf{H_{ox}}$**-CO** under investigation are listed in Table 5 in the Supporting Information and compared to

the corresponding IR spectral features collected experimentally investigating [FeFe]-hydrogenases [72, 73]. It must be noted that all models of the **H_{ox}-CO** state taken into account correspond to minimum energy structures in which a CO ligand occupies a bridging position, being simultaneously coordinated to both iron ions, and a CO or a CN^- ligand occupy the apical position of the Fe_d atom, in *trans* position to the bridging CO.

The experimental spectra corresponding to the **H_{ox}-CO** state, as reported in Ref. [72] and the corresponding best fitting simulated spectra are shown in Figure 6. Experimental spectra are characterized by a ν(CO) band at 1811 cm^{-1}, assigned to the bridging CO, two ν(CO) bands at 1972 and 2016 cm^{-1} assigned to the COs coordinated to the Fe_d atom, one ν(CO) band at 1963 cm^{-1} assigned to the CO coordinated to the Fe_p atom, and two ν(CN) bands at 2096 and 2089 cm^{-1}.

The comparison between DFT and experimental spectra shows that the complexes featuring the protonated chelating dtma and dtme ligands (referred to as dtmaH and dtmeH, respectively) can be excluded due to the non satisfactory agreement between simulated and experimental spectra. On the other hand, complexes with pdt, neutral dtma or dtme match fairly well the experimental spectra, considering both vibrational frequencies and intensity ratio. Therefore, the strict similarity of the calculated spectra did not allow the assignment of the hetero-atom of the dithiol ligand by the analysis of the IR-spectra. As an example, in Figure 6 the simulated spectra of the neutral dtma complexes with a CN^- ligand (complex **1-CN**; panel B), or the CO ligand (complex **1-CO**; panel C) at the apical position of the Fe_d atom, are reported.

It can be noted that the spectrum computed for the former model fairly well matches the experimental bands, in agreement with previous calculations [58]: the maximum error for CO frequencies is observed in the case of the bridging-carbonyl group. Similar conclusions hold true for the cyanide groups frequencies. The simulated spectrum of **1-CO** features larger deviations instead, with the μ-CO again showing the larger difference between theoretical and experimental values. In this context, it is worth underlining that it is extremely difficult to reliably evaluate the quality of computed μ-CO vibrational frequencies, because very few complexes featuring μ-CO have been experimentally characterized and therefore could be included into the training sets used to calibrate DFT vibrational frequencies. Moreover, in the **1-CO** $\rightarrow$ **1-CN** isomerization the protein environment is expected to play a crucial role: the strong hydrogen bond between CN^- and Lys237 in the protein [74] would be disrupted as a consequence of such conformational rearrangement, while the hydrogen bond between the Fe_d-bound cyanide group and the dtma amine group in **1-CO** would be influenced by the presence

of another hydrogen bond donor/acceptor in the protein - namely Cys178 thiol group - which is never included in our QM models. Such effects from the protein matrix have already proved sufficient to invert the stability order between **1-CO** and **1-CN**, thus favoring the former in QM/MM models of the whole enzyme [74]; similar relevant influences are likely to be exerted on the $[2Fe]_H$ vibrational properties as well. Thus we propose that a definitive assignment of the orientation of CN^- and CO ligands could be made only in the context of QM/MM representation of the whole enzyme, which we shall tackle in subsequent work.

[Figure 6 about here.]

4.2 The $\mathbf{H_{ox}}$ Redox State

The $\mathbf{H_{ox}}$ state of the enzyme has been extensively characterized using several spectroscopic techniques, and has been proposed to be a key intermediate in the catalytic cycle of the enzyme. $\mathbf{H_{ox}}$ is paramagnetic and the $[2Fe]_H$ subcluster features a $Fe^{II}Fe^{I}$ redox state in which a CO ligand should bridge the two iron ions. The experimental spectrum of the $\mathbf{H_{ox}}$ state reproduced from Ref. [73] and Ref. [62] as well as the simulated spectra of selected model complexes are shown in Figure 7. The two experimental spectra are nearly identical and feature a weak band at 1802 cm^{-1} assigned to the bridging CO, two intense bands at 1940 and 1965 cm^{-1} assigned to the terminal COs, and two weak ν(CN) bands at 2079 and 2093 cm^{-1}. In addition, a weak band at 2017 cm^{-1} is visible in the spectrum reported both by Albracht *et al.* [73] and by Nicolet *et al.* [62]. This band, which is also present with the same intensity in the spectra of $\mathbf{H_{red}}$ (*vide infra*), exactly corresponds to the most intense band observed in the CO-inhibited form of the enzyme, and could therefore be assigned to traces of $\mathbf{H_{ox}}$**-CO** formed in the experimental condition.

The comparison between computed vibrational frequencies of plausible models of the $\mathbf{H_{ox}}$ state and corresponding experimental IR frequencies (Table 6 in Supporting Information) reveals that bimetallic model complexes with pdt, dtma or dtme chelating ligands give very similar simulated spectra, which are all in remarkable agreement with the experimental spectra. As an example, the spectrum reported in panel C of Figure 7 for the dtma complex **2** is characterized by ν(CO) bands at 1833, 1944 and 1969 cm^{-1} and ν(CN) bands at 2076 and 2085 cm^{-1}, which very well match the experimental spectra.

Since it has been proposed that the $\mathbf{H_{ox}}$ state in the enzyme can bind H_2, the vibrational

spectra of bimetallic models, in which H_2 is terminally bound to the Fe_d atom, was also computed. The simulated spectra of the H_2-coordinated model complexes with CH_3SH and pdt, dtma or dtme chelating ligands are very similar to those obtained for the H_2-free complexes. As an example, panel D of Figure 7 reports the simulated spectrum of **2-H$_2$** which corresponds to the H_2-bound form of **2**. Therefore, it can be concluded that on the basis of the IR-spectra it is not possible to discriminate between H_2-bound or H_2-free models of the $\mathbf{H_{ox}}$ form.

Finally, we investigated complexes in which the nitrogen and oxygen atoms of dtma and dtme are protonated. In this cases, however, the simulated spectra do not match the experimental spectra (panel E of Figure 7 for the $\mathbf{2H_N}^+$ complex) indicating that $\mathbf{H_{ox}}$ models in which the dithiol ligand is protonated can be excluded.

[Figure 7 about here.]

4.3 The $\mathbf{H_{red}}$ redox state

The $\mathbf{H_{red}}$ redox state has been experimentally characterized as a Fe^IFe^I species, and the elucidation of its structural features is crucial to understand the enzyme activity, as well as for the design of efficient bio-inspired synthetic catalysts, since $\mathbf{H_{red}}$ should correspond to the enzyme form that initially binds a proton in the reaction eventually leading to the conversion of two protons and two electrons to dihydrogen. However, the structural features of $\mathbf{H_{red}}$ have been particularly elusive.

In fact, X-ray structures and IR analysis of the protein suggest that the bridging CO observed in the $\mathbf{H_{ox}}$ form moves to a semi-bridged position in $\mathbf{H_{red}}$ [62, 73], a structural feature which is very difficult to reproduce by DFT calculations. Another complication arises from the large number of ν(CO) bands observed in the IR spectra of $\mathbf{H_{red}}$, strongly suggesting that the investigated samples contain different forms of the enzyme active site. Furthermore, the IR spectrum of $\mathbf{H_{red}}$ from *Chlamydomonas reinhardtii* green alga reported very recently by Lubitz et al. [75] differs significantly from those previously reported [62, 73], and it is compatible with a CO ligand fully bridged between the two iron ions.

DFT optimization of the bimetallic models of $\mathbf{H_{red}}$ generally leads to two isomers: one in which a CO ligand is fully bridged, and one in which all CO ligands are coordinated in terminal positions. In view of the above mentioned experimental data on $\mathbf{H_{red}}$, it is reasonable to compare the IR frequencies of the CO-bridged isomers with the

experimental spectrum reported by Lubitz et al. (hereafter the corresponding form will be labelled $\mathbf{H_{red}(b)}$) [75], and the IR frequencies of the all-terminal CO isomers with the spectra reported in Ref. [73], and Ref. [62] (hereafter the corresponding form will be labelled $\mathbf{H_{red}(t)}$). Possible differences between $\mathbf{H_{red}(t)}$ and $\mathbf{H_{red}(b)}$ forms also lie in the redox state of the Fe_4S_4 cluster, which however, is not included in our model complexes.

The experimental spectrum of the $\mathbf{H_{red}(t)}$ state reproduced from Ref. [73] and Ref. [62], as well as simulated spectra of selected bimetallic models of $\mathbf{H_{red}(t)}$ are shown in Figure 8. A very intense ν(CO) band at 1894 cm^{-1}, weak ν(CO) bands at 1916, 1940, 1965 cm^{-1}, ν(CN) bands at 2079 and 2093 cm^{-1}, and one band at 2040 cm^{-1} which could either correspond to a ν(CO) or a ν(CN) band, are present in both experimental spectra. In the experimental spectrum reported in Ref. [73] a weak band at 2015 cm^{-1} was also identified, which however, as discussed in the previous section, should be assigned to traces of $\mathbf{H_{ox}}$**-CO**. The ν(CO) bands at 1940 and 1965 cm^{-1} correspond to those observed in the IR spectrum of the $\mathbf{H_{ox}}$ state, even though they are much less intense. In particular, the band at 1940 cm^{-1}, which is the most intense one in $\mathbf{H_{ox}}$, is still visible in the $\mathbf{H_{red}(t)}$ spectrum of Ref. [73], but it has almost disappeared in the spectrum reported in Ref. [62], suggesting that this band can in fact be assigned to traces of $\mathbf{H_{ox}}$. On the other hand, the band at 1965 cm^{-1} is visible in both $\mathbf{H_{red}(t)}$ spectra and the 1965/1940 intensity ratio is reversed with respect to that observed in $\mathbf{H_{ox}}$, suggesting that the 1965 cm^{-1} band does not correspond to an $\mathbf{H_{ox}}$ band only, but that it is also a true band of the $\mathbf{H_{red}(t)}$ state. The ν(CN) bands at 2079 and 2093 cm^{-1} closely correspond to those observed in the $\mathbf{H_{ox}}$ form, as well. In this case, however, either both, or only one band should be assigned to the $\mathbf{H_{red}(t)}$ form, depending on the plausible assignments of the 2040 cm^{-1} band which could correspond to either a ν(CO) or a ν(CN) vibrational frequency. The above assignments are supported by results of Albracht *et al.*, who assigned the bands at 1894, 1916, 1965, 2040, and 2079 cm^{-1} to the $\mathbf{H_{red}(t)}$ state on the basis of difference spectra [73].

The bimetallic complexes with neutral dtma or dtme fairly match the 1894, 1916, 2040 and 2079 cm^{-1} bands of the experimental spectra (see Table 7 in Supplementary Information). The simulated IR spectrum of the complex with dtma **3t** is reported in panel C of Figure 8. The two lowest energy bands are very close to each other, and overlap in the simulated spectrum, explaining the large intensity of the experimental band at 1894 cm^{-1}. These bands are computed to be about 30 cm^{-1} lower than the experimental one. However, as discussed previously, the X-ray geometry is not well reproduced by DFT calculations, as in the DFT model the semi-bridged CO moves

to a terminal position, an observation that can account for the differences between experimental and calculated spectra. It is important to note that in complex **3t** the ν(CN) vibrational frequency calculated at 2032 cm^{-1} is assigned to the experimental band at 2040 cm^{-1}. For the same complex, using the B3LYP functional, Hall *et al.* [23] calculated the ν(CN) frequencies at 2080 and 2063 cm^{-1} (values significantly higher than those calculated in this work), and assigned these frequencies to the 2065 and 2079 cm^{-1} experimental bands. As a consequence, the 2040 cm^{-1} band was attributed in Ref. [23] to the presence of the μ-hydride species.

Protonation of the heteroatom of dtma or dtme results in model complexes which may also match experimental data. In the case of the dtmaH ligand (complex $\mathbf{3tH_N}^+$; panel D in Figure 8), in fact, the 1878 and 1925 cm^{-1} ν(CO) bands of the protonated complex fairly overlap with those of the non-protonated complex, while the band at 1971 cm^{-1} can correspond to the 1965 cm^{-1} band observed in the experimental spectra. The calculated ν(CN) bands at 2069 and 2091 cm^{-1} overlap with the 2065 cm^{-1} band calculated for **3t** (2069 cm^{-1} band), or can be assigned to the 2093 cm^{-1} band of the experimental spectra (2091 cm^{-1} band). As shown in the panel E of Figure 8, the experimental spectra of the $\mathbf{H_{red}(t)}$ state can be fairly well predicted by considering a 50-50 mixture of **3t** and $\mathbf{3t_NH}^+$ complexes.

Differently from the proposal of Ref. [23], the μ-hydride species with non-protonated dtma (or dtme) should not contribute to the spectrum of $\mathbf{H_{red}(t)}$. Indeed, this species, in our calculations, is characterized by an intense ν(CO) band at 2004 cm^{-1} in a region free of bands in the experimental spectra. The experimental spectra of $\mathbf{H_{red}(t)}$ can also be reproduced by considering a mixture of **3t** and the terminal hydride complex $\mathbf{3\text{-}tH_{Fe}}^+$ (see panel C of Figure 9), with the assumption that the 1940 cm^{-1} band observed in the experimental spectra is a true band of $\mathbf{H_{red}}$, and not due to traces of $\mathbf{H_{ox}}$. In this case, the calculated 1920, 1950 and 1986 cm^{-1} ν(CO) bands are assigned to the 1916, 1940, and 1965 cm^{-1} experimental bands, respectively, while the 2085 and 2112 cm^{-1} ν(CN) bands are assigned to the 2079, and 2093 cm^{-1} experimental bands, respectively. The assignment of $\mathbf{3tH_N}^+$ or $\mathbf{3\text{-}tH_{Fe}}^+$, or both isomers to the mixture can hardly be accounted for only on the basis of the IR analysis. However, it should be noted that, as reported in previous computational studies [74], the terminal hydride species $\mathbf{3\text{-}tH_{Fe}}^+$ is significanlty more stable than the protonated dtma isomer $\mathbf{3tH_N}^+$, clearly suggesting that the former is the protonated species actually present in the $\mathbf{H_{red}(t)}$ mixture.

[Figure 8 about here.]

The spectrum assigned to $\mathbf{H_{red}(b)}$ by Lubitz et al. [75], and the corresponding best fitting simulated spectrum are shown in Figure 9. The experimental spectrum is characterized by three ν(CO) bands at 1793, 1891, and 1935 cm^{-1}, and two ν(CN) bands at 2070 and 2083 cm^{-1}. The first ν(CO) band is shifted by about 90 cm^{-1} to lower wave numbers with respect to the spectra of $\mathbf{H_{red}(b)}$, suggesting that the CO ligand is fully bridged between the two iron ions.

The vibrational spectra of the model complexes with pdt, dtma, or dtme ligands are very similar to each other, and reproduce fairly well the experimental CO-bridged band at 1793 cm^{-1}, and the ν(CO) band at 1891 cm^{-1} (Table 8 in the Supporting Information). As an example, in the panel B of Figure 9 is shown the spectrum for the complex with dtma **3b**. In the simulated spectrum of **3b** the third ν(CO) band is close to the second one, and can also be assigned to the 1891 cm^{-1} experimental band. Indeed, a careful inspection of the experimental spectrum shows that the 1891 cm^{-1} band features a shoulder at lower wave numbers, which is consistent with the simulated spectrum. On the other hand, the 1935 cm^{-1} band is not reproduced in the simulated spectrum of **3b**, and cannot be assigned. This band is very close to the 1940 cm^{-1} band of $\mathbf{H_{ox}}$ and could therefore be assigned to traces of $\mathbf{H_{ox}}$ still present in the experimental conditions, or more probably to a mixture of different forms as in $\mathbf{H_{red}(t)}$. In fact, as above mentioned, the simulated spectrum of the terminal hydride complex (panel C of Figure 9 for the complex with dtma;) shows an intense band at 1950 cm^{-1} and a weaker band at 1986 cm^{-1}, which could be assigned to the experimental 1935 cm^{-1} band and to another band visible on the spectrum at higher wave numbers, respectively. The panel D of Figure 9 shows that the experimental spectrum of the $\mathbf{H_{red}(b)}$ form is well reproduced by a 50-50 mixture of **3b** and **3-t**$\mathbf{H_{Fe}}^+$. The experimental ν(CN) bands at 2083 and 2070 cm^{-1} in the spectrum of the mixture can be assigned to the ν(CN) bands of **3-t**$\mathbf{H_{Fe}}^+$ (2112 and 2085 cm^{-1}), while the ν(CN) bands of the **3b** complex (2047 and 2043 cm^{-1}) can be assigned to a very weak band visible in the spectrum at lower wave numbers. However, it is worth underlining that the assignment of CN bands cannot be more than a tentative one, given the weakness of the experimental bands and the noise that characterize the spectrum baseline above 2000 cm^{-1}. The experimental CO bands present a much larger signal/noise ratio, and it is thus significant to discuss the maximum deviation between theory and experiment for carbonyl groups; notably, such maximum deviation is smaller than 25 cm^{-1} when the 50% **3b** + 50% **3-t**$\mathbf{H_{Fe}}^+$ mixture is taken into account, thus supporting the assignment here proposed for the $\mathbf{H_{red}(t)}$ enzyme form.

[Figure 9 about here.]

In summary, we report in Table 4 a statistical analysis of the deviation of the scaled ν(CO) and ν(CN) vibrational frequencies with respect to the experimental ones for the H_{ox}-CO, H_{ox}, and H_{red} model complexes. Clearly, this analysis is carried out on a few frequencies for each complex, therefore a single large deviation significantly affects the average values. In the case of the CO frequencies the MUD values are larger than those calculated for the model complexes used in the calibration (see Table 3). In this respect we note that the largest deviation always occurs for the bridging CO vibration. If the analysis is carried out by excluding the bridging CO vibrational frequency (values given in parenthesis) the MUD values are significantly lower, and, in the case of models that allowed a confident assignment of the protein active site, they are smaller than the MUD value caluculated for the complexes used for the calibration. The large deviation of the bridging CO vibrational frequencies has two reasons. Apparently, they are more sensitive to the protein environment, the effects of which will be the specific object of a future QM/MM study on the enzyme. Secondly, in the set of about 120 CO vibrational frequencies used for the parametrization of fitting equations, the weight of bridging CO is much smaller (in other words, only a few of those 120 CO vibrations correspond to a bridging CO). As for the CN frequencies, the MUD values are only slightly larger than that calculated for the complexes used for the calibration. Obviously, it is important to note that the CN ligands should be more sensitive to the protein environment than the terminal CO ligands, due, for example, to the formation of H-bonds with nearby residues. When considered from the above point of view, the deviations calculated for CN^- groups in our models of the isolated binuclear portion of the H-cluster can be considered quite satisfactory. Finally, it is interesting to note that complex $\mathbf{2H_N}^+$, which is not considered to be a good model for H_{ox}, as discussed in the main text, is indeed characterized by values of MUD, MSD and MaxDev much larger than those of the other model complexes.

[Table 4 about here.]

5 Conclusions

CO and CN stretching vibrations of 32 [FeFe]-hydrogenase model clusters have been investigated. To our knowledge this is the most extensive investigation in which computed vibrational frequencies are used to assign the structure of reaction intermediates characterized by IR spectroscopy. It nicely complements the studies of Darensbourg and Hall [23, 31] and Zilberman *et al.* [57–59]. CN and CO stretching vibrations were

analyzed separately.

For all CO stretching vibrations the mean unsigned error between the measured and the calculated frequencies is 7.5 cm^{-1}, and becomes 7.4 cm^{-1} after scaling, which indicates that already the unscaled BP86/RI/TZVP CO stretching vibrations are very reliable. In terms of mean error, the effect of scaling is, however, larger for the CN stretching vibrations, where the mean error between the measured and the calculated frequencies is 20.5 cm^{-1}, and becomes as low as 5.1 cm^{-1} after scaling. On the basis of these results we suggest to use Eq. (1) and Eq. (3) for scaling the calculated CN and CO vibrational frequencies, respectively. In comparison to the hybrid density functional B3LYP employed in Refs. [23,31] the BP86/RI/TZVP scheme yields vibrational frequencies that are in better agreement with the experimental ones, both before and after scaling. In particular, the separation in the scaling procedure of the CN vibrational frequencies from the CO vibrational frequencies gives scaled CN frequencies which closely reproduce the experimental ones.

The possibility of computing reliable scaled IR frequencies allowed us to investigate a large set of bimetallic model complexes, designed as plausible models of the $\mathbf{H_{ox}}$-**CO**, $\mathbf{H_{ox}}$, and $\mathbf{H_{red}}$ states of the enzyme. It turned out that the protonation of the thiol group coordinated to the Fe_p atom is necessary to model the Fe_4S_4 moiety of the H-cluster. In addition, the presence of the dithiolate chelating ligand pdt, dtma, and dtme results in very similar spectra. Therefore, the comparison between calculated and experimental spectra does not allow for the assignment of the heteroatom of the dithiolate ligand. With these remarks in mind, we showed that calculated spectra of the complexes $[(CH_3SH)(CO)(CN)Fe(\mu\text{-}CO)(\mu\text{-}dtma)Fe(CO_2)(CN)]^{-1}$ **1-CN**, and $[(CH_3SH)(CO)(CN)Fe(\mu\text{-}CO)(\mu\text{-}dtma)Fe(CO)(CN)]^{-1}$ **2** (or the corresponding complexes with pdt and dtme) are in remarkable agreement with the measured spectra of $\mathbf{H_{ox}}$-**CO** and $\mathbf{H_{ox}}$, respectively. In the latter case, the H_2 adduct also gives a spectrum in good agreement with the experiment. Among the forms of the enzyme investigated, $\mathbf{H_{red}}$ is the most elusive and hard to be studied. On the basis of the most recent spectroelectrochemical IR investigations, two forms of $\mathbf{H_{red}}$ have been proposed, one in which a CO ligand is semi-bridged between the two iron ions ($\mathbf{H_{red}(t)}$), and one in which the CO ligand is fully bridged ($\mathbf{H_{red}(b)}$). Possibly, the two forms also differ for the redox state of the Fe_4S_4 subcluster. The spectra of the $\mathbf{H_{red}(t)}$ and $\mathbf{H_{red}(b)}$ forms have been compared with those calculated for the all-terminal COs and CO-bridged isomers of the bimetallic model complexes, respectively. For both forms the calculated spectra are consistent with a mixture of complexes: a mixture of $[(CH_3SH)(CO)(CN)Fe(\mu\text{-}dtma)Fe(CO)_2(CN)]^{-2}$ **3t**, $[(CH_3SH)(CO)(CN)Fe(\mu\text{-}dtmaH)Fe(CO)_2(CN)]^{-1}$ $\mathbf{3tH_N}^+$,

and $[(CH_3SH)(CO)(CN)Fe(\mu\text{-dtma})Fe(H)(CO)_2(CN)]^{-1}$ **3-tH$_{Fe}^+$** for **H$_{red}$(t)**, and a mixture of $[(CH_3SH)(CO)(CN)Fe(\mu\text{-dtma})(\mu\text{-CO})Fe(CO)(CN)]^{-2}$ **3b** and $[(CH_3SH)$-$(CO)(CN)$-$Fe(\mu\text{-dtma})Fe(H)(CO)_2(CN)]^{-1}$ **3-tH$_{Fe}^+$** for **H$_{red}$(b)**.

It should be noted that for all bimetallic complexes that closely fit the **H$_{ox}$-CO** and **H$_{ox}$**, or **H$_{red}$** experimental spectra, the lowest ν(CO) frequency corresponding to the bridging (or semi-bridging) CO gives the largest deviation with respect to the experiment. The inclusion of the vicinal Fe_4S_4 cluster, as well as the protein environment in the models, could be crucial for an even more accurate calculation of the IR spectra. In a subsequent study, we plan to apply the results obtained here on improved models that take also into account the protein environment. Such a QM/MM model has already been developed [76] and now requires the calculation of vibrational frequencies for which we can employ the efficient mode-tracking approach [77, 78]. Work along these lines is now in progress in our laboratories.

6 Acknowledgments

This work has been supported by the Schweizer Nationalfonds SNF (project 200020-121870). C.G. thanks the Humboldt Foundation for financial support.

References

[1] Cammack, R. *Nature* **1999**, *397*, 214.

[2] Holm, R. H.; Kennepohl, P.; Solomon, E. I. *Chem. Rev.* **1996**, *96*, 2239.

[3] Frey, M. *ChemBioChem* **2002**, *3*, 153.

[4] Thauer, R. K.; Klein, A. R.; Hartmann, G. C. *Chem. Rev.* **1996**, *96*, 3031.

[5] Volbeda, A.; Charon, M. H.; Piras, C.; Hatchikian, E. C.; Frey, M.; Fontecilla-Camps, J. C. *Nature* **1995**, *373*, 580.

[6] Canaguier, S.; Artero, V.; Fontecave, M. *Dalton Trans.* **2008**, 315.

[7] Albracht, S. P. J. *Biochim. Biophys. Acta* **1994**, *1178*, 167.

[8] Korbas, M.; Vogt, S.; Meyer-Klaucke, W.; Bill, E.; Lyon, E. J.; Thauer, R. K.; Shima, S. *J. Biol. Chem.* **2006**, 30804–30813.

[9] Nicolet, Y.; Cavazza, C.; Fontecilla-Camps, J. C. *J. Inorg. Biochem.* **2002**, *91*, 1.

[10] Peters, J. W. *Curr. Opin. Struct. Biol* **1999**, *6*, 670.

[11] Cao, Z.; Hall, M. B. *J. Am. Chem. Soc.* **2001**, *123*, 3734–3742.

[12] Peters, J. W.; Lanzilotta, W. N.; Lemon, B. J.; Seefeldt, L. C. *Science* **1998**, *282*, 1853.

[13] Nicolet, Y.; Piras, C.; Legrand, P.; Hatchikian, E. C.; Fontecilla-Camps, J. C. *Structure* **1999**, *7*, 13.

[14] Pierik, A. J.; Hagen, W. R.; Redeker, J. S.; Wolbert, R. B. G.; Boersma, M.; Verhagen, M. F. J. M.; Grandee, H. J.; Veeger, C.; Mutsaers, P. H. E.; Dunhan, R. H. S. W. R. *Eur. J. Biochem.* **1992**, *209*, 63.

[15] Liu, Z.-P.; Hu, P. *J. Am. Chem. Soc.* **2002**, *124*, 5175–5182.

[16] Le Cloirec, A.; Davies, S. C.; Evans, D. J.; Hughes, D. L.; Pickett, C. J.; Best, S. P.; Borg, S. *Chem. Commum.* **1999**, *22*, 2285.

[17] Razavet, M.; Davies, S. C.; Hughes, D. L.; Barclay, J. E.; Evans, D. J.; Fairhurst, S. A.; Liu, X.; Pickett, C. J. *Dalton Trans.* **2003**, 586–595.

[18] Gloaguen, F.; Lawrence, J. D.; Schmidt, M.; Wilson, S. R.; Rauchfuss, T. B. *J. Am. Chem. Soc.* **2001**, *123*, 12518–12527.

[19] Zhao, X.; Georgakaki, I. P.; Miller, M. L.; Mejia-Rodriguez, R.; Chiang, C. Y.; Darensbourg, M. Y. *Inorg. Chem.* **2002**, *41*, 3917–3928.

[20] Hasan, M. M.; Hursthouse, M. B.; Kabir, S. E.; Malik, K. A. *Polyhedron* **2001**, *20*, 97–101.

[21] Mejia-Rodriguez, R.; Chong, D.; Reibenspies, J. H.; Soriaga, M. P.; Darensbourg, M. Y. *J. Am. Chem. Soc.* **2004**, *126*, 12004–12014.

[22] Gloaguen, F.; Lawrence, J. D.; Rauchfuss, T. B.; Bénard, M.; Rohmer, M.-M. *Inorg. Chem.* **2002**, *41*, 6573–6582.

[23] Tye, J. W.; Darensbourg, M. Y.; Hall, M. B. *Inorg. Chem.* **2008**, *47*, 2380.

[24] Fan, H. J.; Hall, M. B. *J. Am. Chem. Soc.* **2001**, 3828–3829.

[25] Bruschi, M.; Fantucci, P.; De Gioia, L. *Inorg. Chem.* **2002**, *41*, 1421–1429.

[26] Bruschi, M.; Fantucci, P.; De Gioia, L. *Inorg. Chem.* **2003**, *42*, 4773–4781.

[27] Zhou, T.; Mo, Y.; Liu, A.; Zhou, Z.; Tsai, K. R. *Inorg. Chem.* **2004**, *43*, 923–930.

[28] Zhou, T.; Mo, Y.; Zhou, Z.; Tsai, K. R. *Inorg. Chem.* **2005**, *44*, 4941–4946.

[29] Zampella, G.; Greco, C.; Fantucci, P.; De Gioia, L. *Inorg. Chem.* **2006**, *45*, 4109–4118.

[30] Lukin, J. A.; Ho, C. *Chem. Rev.* **2004**, *104*, 1219.

[31] Tye, J. W.; Darensbourg, M. Y.; Hall, M. B. *J. Comput. Chem.* **2006**, *27*, 1454–1462.

[32] Becke, A. D. *Phys. Rev. A* **1988**, *38*, 3098–3100.

[33] Perdew, J. P. *Phys. Rev. B* **1986**, *33*, 8822–8824.

[34] Brehm, G.; Reiher, M.; Schneider, S. *J. Phys. Chem. A* **2002**, *106*, 12024–12034.

[35] Neugebauer, J.; Hess, B. A. *J. Chem. Phys* **2003**, *118*, 7215–7225.

[36] Reiher, M.; Brehm, G.; ; Schneider, S. *J. Phys. Chem. A* **2004**, *108*, 734–742.

[37] Galinato, M. G. I.; Whaley, C. M.; Lehnert, N. *Inorganic Chemistry* **2010**, *49*, 3201–3215.

[38] R. Ahlrichs *et al.*; TURBOMOLE, http://www.cosmologic.de/turbomole.html.

[39] R. Ahlrichs, M. Bär, M. Häser, H. Horn, C. Kölmel *Chem. Phys. Lett.* **1989**, *162*, 165.

[40] Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 5648–5652.

[41] Lee, C.; Yang, W.; Parr, R. G. *Phys. Rev. B* **1988**, *37*, 785–789.

[42] Stephens, P. J.; Devlin, F. J.; Chabalowski, C. F.; Frisch, M. J. *J. Phys. Chem.* **1994**, 11623.

[43] Schäfer, A.; Huber, C.; Ahlrichs, R. *J. Chem. Phys* **1994**, *100*, 5829.

[44] Reiher, M.; Salomon, O.; Sellmann, D.; Hess, B. A. *Chem. Eur. J.* **2001**, *7*, 5195–5202.

[45] Reiher, M.; Neugebauer, J.; Hess, B. A. *Z. Phys. Chem* **2003**, *217*, 91–103.

[46] Koch, W.; Holthausen, M. C. *A Chemist's Guide to Density Functional Theory*; Wiley-VCH: Weinheim, 2001.

[47] Neugebauer, J.; Reiher, M.; Kind, C.; Hess, B. A. *J. Comput. Chem.* **2002**, *23*, 895–910.

[48] Klamt, A.; Eckert, F. *Fluid Phase Equilibria* **2000**, *172*, 43–72.

[49] Schäfer, A.; Klamt, A.; Sattel, D.; Lohrenz, J. C.; Eckert, F. *Phys. Chem. Chem. Phys.* **2000**, *2*, 2187–2193.

[50] Sillanpää, A. J.; Aksela, R.; Laasonen, K. *Phys. Chem. Chem. Phys.* **2003**, *5*, 3382–3393.

[51] Schaftenaar, G.; Noordik, J. H. *J. Comput.-Aided Mol. Design* **2000**, *14*, 123–134.

[52] Noodleman, L.; J. G. Norman, J. *J. Chem. Phys.* **1979**, *70*, 4903.

[53] Noodleman, L. *J. Chem. Phys.* **1981**, *74*, 5737–5743.

[54] Noodleman, L.; Davidson, E. R. *Chem. Phys.* **1986**, *109*, 131–143.

[55] Herrmann, C.; Yu, L.; Reiher, M. *J. Comput. Chem.* **2006**, *27*, 1223–1239.

[56] Windhager, J.; Görls, H.; Petzold, H.; Mloston, G.; Linti, G.; Weigand, W. *Eur. J. Inorg. Chem.* **2007**, 4462–4471.

[57] Zilberman, S.; Stiefel, E. I.; Cohen, M. H.; Car, R. *Inorg. Chem.* **2006**, 5715–5717.

[58] Zilberman, S.; Stiefel, E. I.; Cohen, M. H.; Car, R. *Inorg. Chem.* **2007**, 1153–1161.

[59] Zilberman, S.; Stiefel, E. I.; Cohen, M. H.; Car, R. *J. Phys. Chem. B* **2006**, 7049–7057.

[60] Silakov, A.; Wenk, B.; Reijerse, E.; Lubitz, W. *Phys. Chem. Chem. Phys.* **2009**, *11*, 6592–6599.

[61] Ryde, U.; Greco, C.; Gioia, L. D. *J Am. Chem. Soc.* **2010**, *132*, 4512–4513.

[62] Nicolet, Y.; de Lacey, A. L.; Vernede, X.; Fernandez, V. M.; Hatchikian, E. C.; Fontecilla-Camps, J. C. *J. Am. Chem. Soc.* **2001**, 1596–1601.

[63] Pandey, A. S.; Harris, T. V.; Giles, L. J.; Peters, J. W.; Szilagyi, R. K. *J. Am. Chem. Soc.* **2008**, 4533–4540.

[64] Bruschi, M.; Greco, C.; Fantucci, P.; De Gioia, L. *Inorg. Chem.* **2008**, *47*, 6056–6071.

[65] Fiedler, A. T.; Brunold, T. C. *Inorganic Chemistry* **2005**, *44*, 9322–9334.

[66] Bruschi, M.; Fantucci, P.; De Gioia, L. *Inorg. Chem.* **2004**, *43*, 3733–3741.

[67] Rusnak, F. M.; Adams, M. W. W.; Mortenson, L. E.; Munck, E. *J. Biol. Chem.* **1987**, 38–41.

[68] Popescu, C. V.; Munck, E. *J. Am. Chem. Soc.* **1999**, 7877–7884.

[69] Pereira, A. S.; Tavares, P.; Moura, I.; Moura, J. J. G.; Huynh, B. H. *J. Am. Chem. Soc.* **2001**, 2771–2782.

[70] Bennet, B.; Lemon, B. J.; Peters, J. W. *Biochemistry* **2000**, 7455–7460.

[71] Albracht, S. P. J.; Roseboom, W.; Hatchikian, E. C. *J. Biol. Inorg. Chem.* **2006**, 88–101.

[72] Lacey, A. L. D.; Stadler, C.; Cavazza, C.; Hatchikian, E. C.; Fernandez, V. M. *J. Am. Chem. Soc.* **2000**, *122*, 11232–11233.

[73] Roseboom, W.; Lacey, A. L. D.; Fernandez, V. M.; Hatchikian, E. C.; Albracht, S. P. J. *J. Biol. Inorg. Chem.* **2006**, 102–118.

[74] Greco, C.; Bruschi, M.; Heimdal, J.; Fantucci, P.; Gioia, L. D.; Ryde, U. *Inorg. Chem.* **2007**, 7256–7258.

[75] Silakov, A.; Kamp, C.; Reijerse, E.; Happe, T.; Lubitz, W. *Biochemistry* **2009**, *48*, 7780–7786.

[76] Greco, C.; Bruschi, M.; De Gioia, L.; Ryde, U. *Inorg. Chem.* **2007**, *46*, 5911–5921.

[77] Reiher, M.; Neugebauer, J. *J. Chem. Phys.* **2003**, *118*, 1634–1641.

[78] Herrmann, C.; Neugebauer, J.; Reiher, M. *J. Comput. Chem.* **2008**, *29*, 2460–2470.

[79] Justice, A. K.; Rauchfuss, T. B.; Wilson, S. R. *Angew. Chem. Int. Ed.* **2007**, *46*, 6152–6154.

[80] Boyke, C. A.; van der Vlugt, J. I.; Rauchfuss, T. B.; Wilson, S. R.; Zampella, G.; De Gioia, L. *J. Am. Chem. Soc.* **2005**, *127*, 11010–11018.

[81] Schmidt, M.; Contakes, S.; Rauchfuss, T. B. *J. Am. Chem. Soc.* **1999**, *121*, 9736–9737.

[82] Georgakaki, I. P.; Miller, M. L.; Darensbourg, M. Y. *Inorg. Chem.* **2003**, *42*, 2489–2494.

[83] Singleton, M. L.; Bhuvanesh, N.; Reibenspies, J. H.; Darensbourg, M. Y. *Angew. Chem. Int. Ed.* **2008**, *47*, 9492–9495.

[84] Lyon, E. J.; Georgakaki, I. P.; Reibenspies, J. H.; Darensbourg, M. Y. *Angew. Chem. Int. Ed.* **1999**, *38*, 3178–3180.

[85] Li, H.; Rauchfuss, T. B. *J. Am. Chem. Soc.* **2002**, *124*, 726–727.

[86] Lawrence, J. D.; Li, H.; Rauchfuss, T. B.; Bénard, M.; Rohmer, M.-M. *Angew. Chem. Int. Ed.* **2001**, *40*, 1768–1771.

List of Figures

Figure 1: Schematic structure of the H-cluster found in the active site of [FeFe]-hydrogenases. The binuclear iron cluster is usually referred to as the $[2Fe]_H$ subcluster.

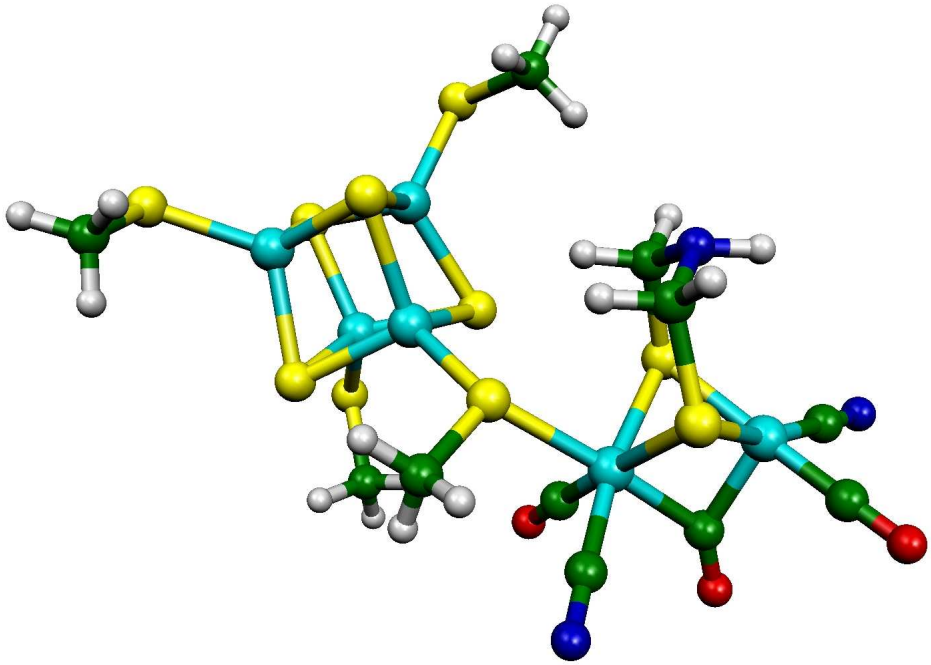


Figure 2: The BP86/RI/TZVP optimized structure of complex **28** micro-solvated by eight CH_3CN molecules. The labels $r1$ to $r4$ denote local bond-stretching vibrations.

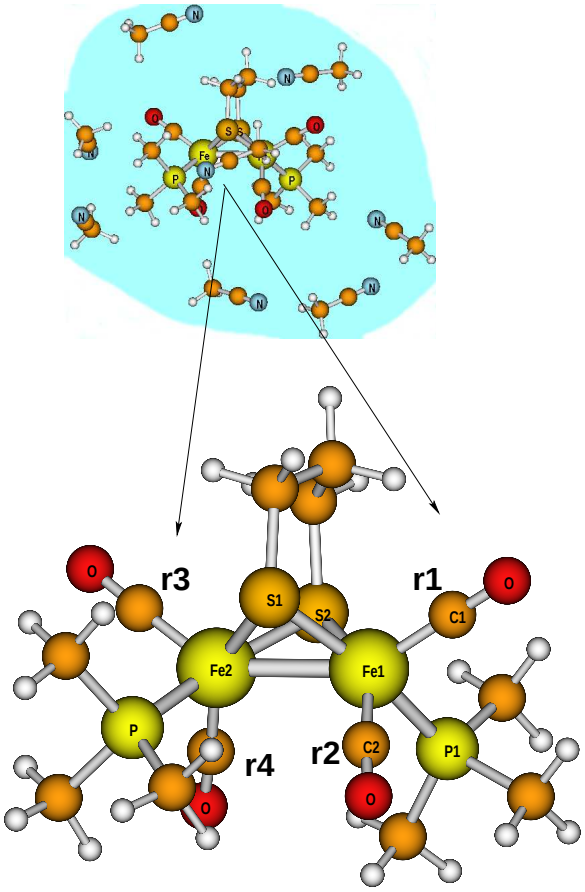


Figure 3: a) Correlation of the calculated (Calc.) and experimentally determined (Exp.) CN frequencies (in units of cm^{-1}) of selected clusters. b) Correlation of the calculated and observed CO frequencies (cm^{-1}) of all 32 clusters. Here R stands for the correlation coefficient and the r^2 value is the coefficient of determination (for details on the statistical measures see the Supporting Information).

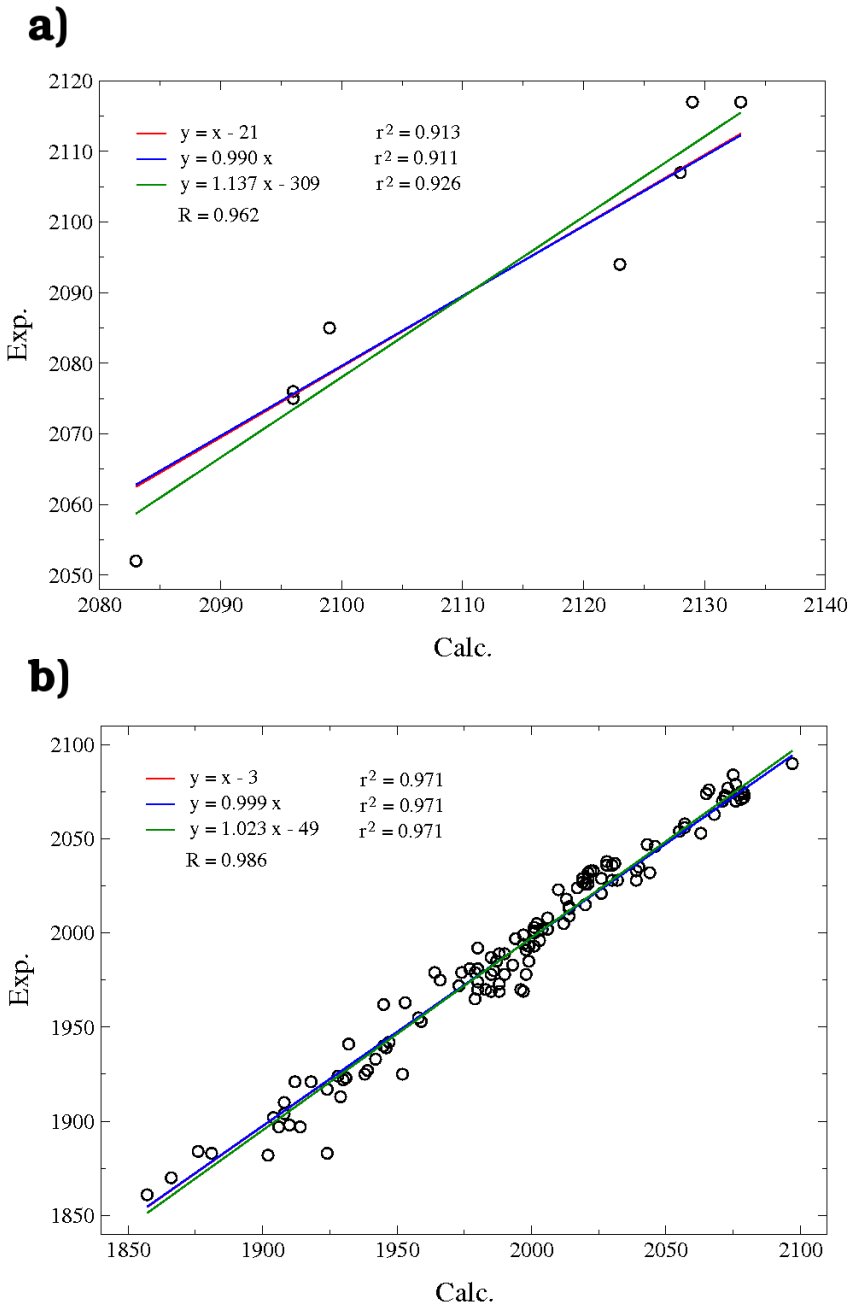


Figure 4: Correlation of calculated and observed CO and CN frequencies (cm^{-1}) of clusters **12**–**15**, **18**, **20**, and **24**–**32**, which have been selected according to the choice in Ref. [23,31] The presentation contains our BP86/RI/TZVP data. The B3LYP results of Darensbourg and Hall [23, 31], to which our results shall be compared, have been given in their work as $y = 0.9538x$ with $r^2 = 0.9477$. Here R denotes the correlation coefficient and r^2 stands for the coefficient of determination.

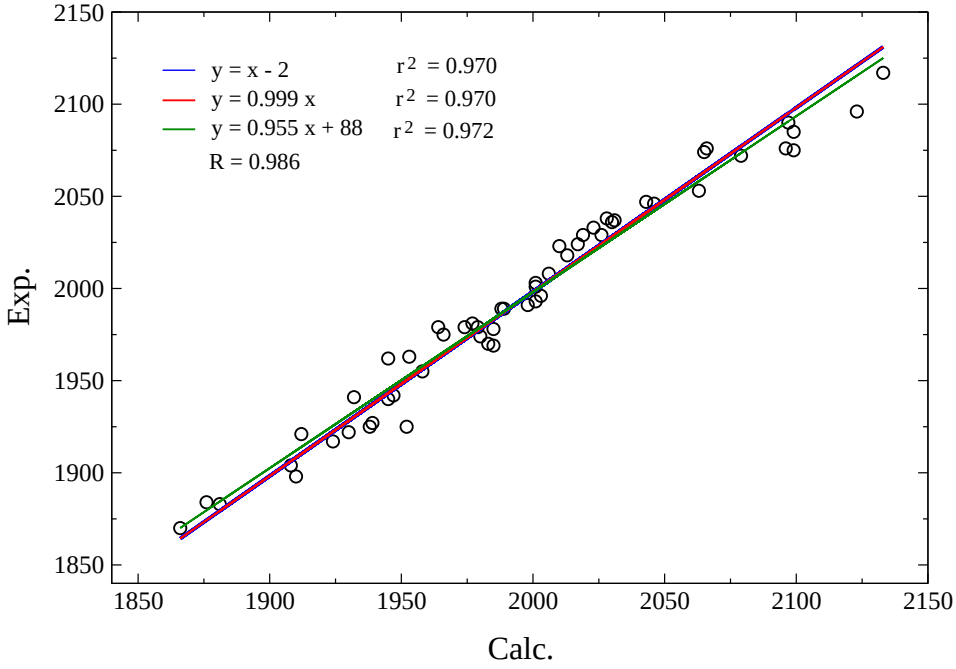


Figure 5: Schematic representation of the investigated models of the $[2Fe]_H$ subcluster for the $\mathbf{H_{ox}}$**-CO**, $\mathbf{H_{ox}}$, and $\mathbf{H_{red}}$ redox states of [FeFe]-hydrogenases. The thiol ligand coordinated to the Fe_p atom is either CH_3S^- or CH_3SH. The X group of the dithiol bridging ligand is CH_2, NH, NH_2, O, or OH. As for the $\mathbf{H_{ox}}$**-CO** models, the two isomers, with either a CO or CN^- ligand in *trans* to the μ-CO ligand have been investigated. The $\mathbf{H_{ox}}$ models investigated also include the complexes with H_2 coordinated to the Fe_d centre. In the case of the $\mathbf{H_{red}}$ state the μ-CO, and all-terminal CO isomers, and the corresponding protonated species have been included in the set of models.

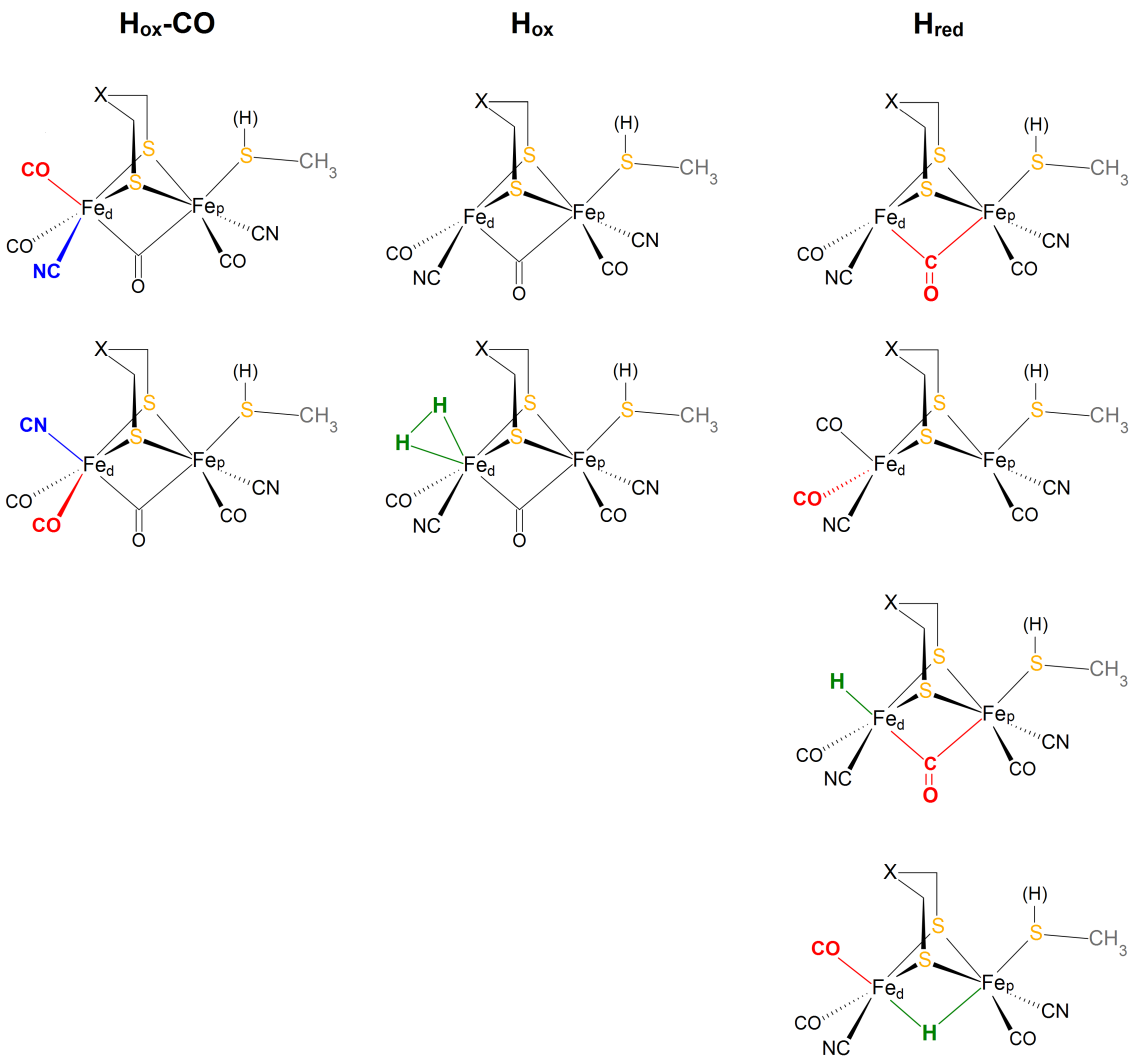


Figure 6: Measured IR spectrum of the $\mathbf{H_{ox}}$**-CO** form of [FeFe]-hydrogenase enzyme reproduced from Ref. [72] (panel A), and BP86/RI/TZVP calculated IR spectra for the model complexes **1-CN** (panel B), and **1-CO** (panel C). The wave numbers are given in cm^{-1}. All computed frequencies are scaled according to Eq. (1) and Eq. (3). The simulated spectra are presented with a half-width of 15 cm^{-1}.

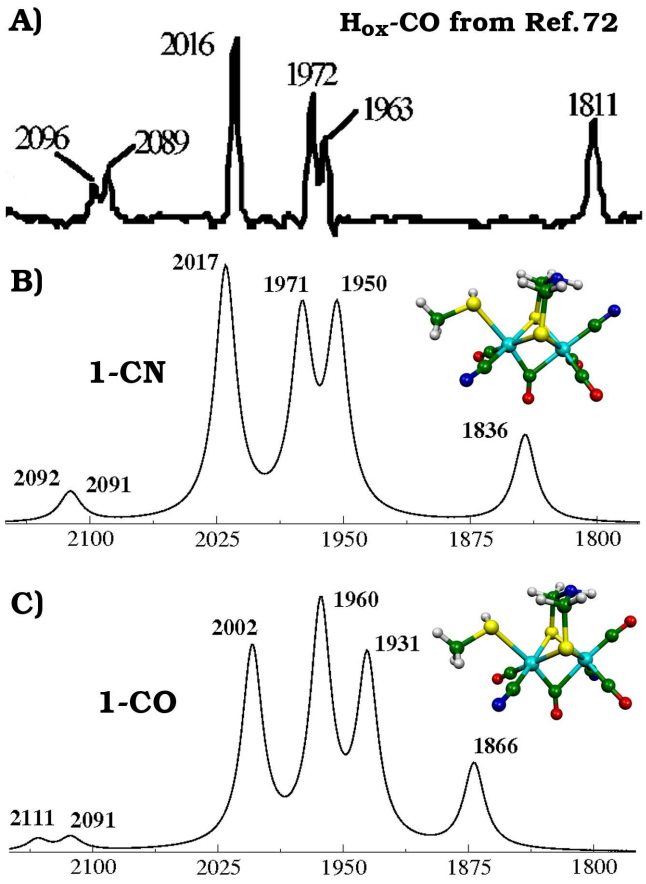


Figure 7: Measured IR spectrum of the $\mathbf{H_{ox}}$ form of [FeFe]-hydrogenase enzyme reproduced from Ref. [73] (panel A) and [62] (panel B), and BP86/RI/TZVP calculated IR spectra of the model complexes **2** (panel C), **2-H**$_2$ (panel D), and $\mathbf{2H_N}^+$ (panel E). The wave numbers are given in cm^{-1}. All computed frequencies are scaled according to Eq. (1) and Eq. (3). The simulated spectra are presented with a half-width of 15 cm^{-1}.

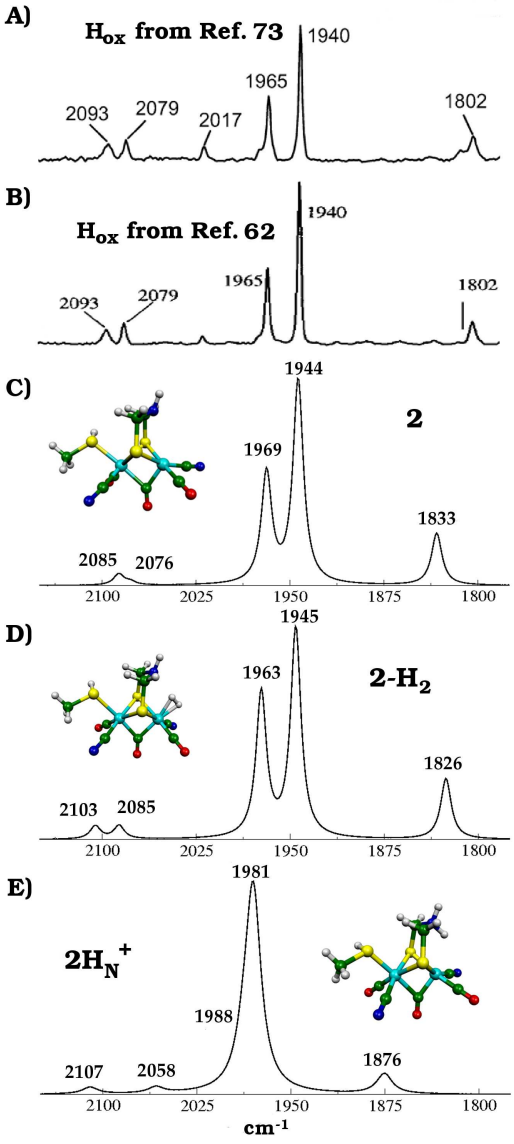


Figure 8: Measured IR spectrum of the $\mathbf{H_{red}(t)}$ form of [FeFe]-hydrogenase enzyme reproduced from Ref. [73] (panel A) and [62] (panel B), and BP86/-RI/TZVP calculated IR spectra for the model complexes **3t** (panel C), and $\mathbf{3tH_N}^+$ (panel D). In panel E is reported the simulated spectrum of 50-50 mixture of **3t** and $\mathbf{3tH_N}^+$. The wave numbers are given in cm^{-1}. All computed frequencies are scaled according to Eq. (1) and Eq. (3). The simulated spectra are presented with a half-width of 15 cm^{-1}. The lowest wave number frequency is presented as a single band at frequency given by the arithmetic mean of the 1859, 1865 and 1878 cm^{-1} calculated for **3t** and $\mathbf{3tH_N}^+$.

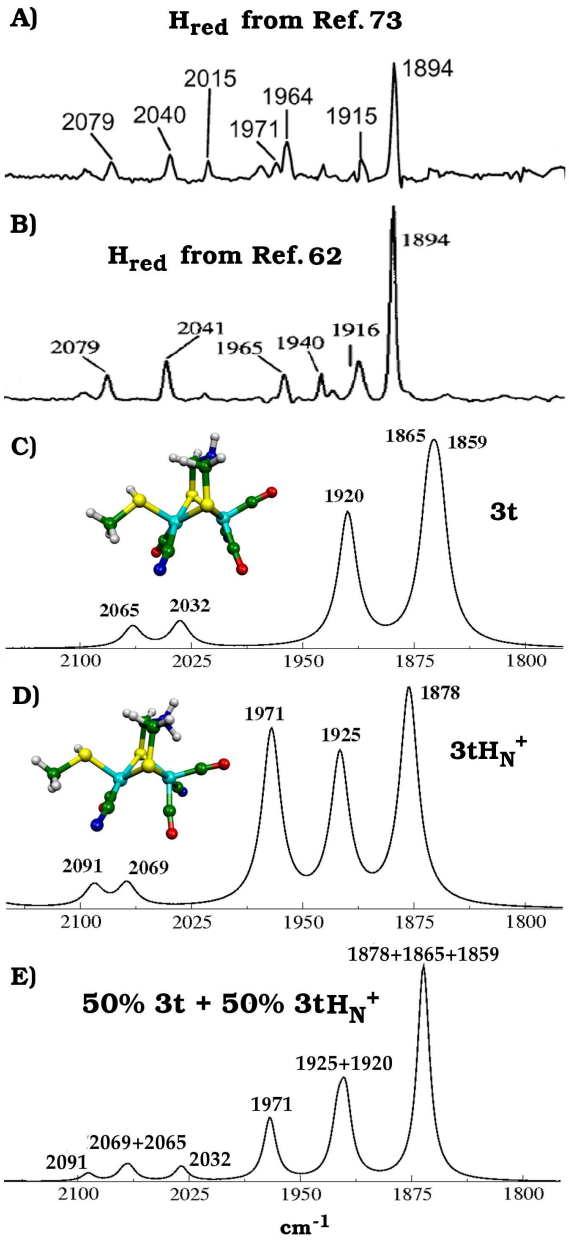
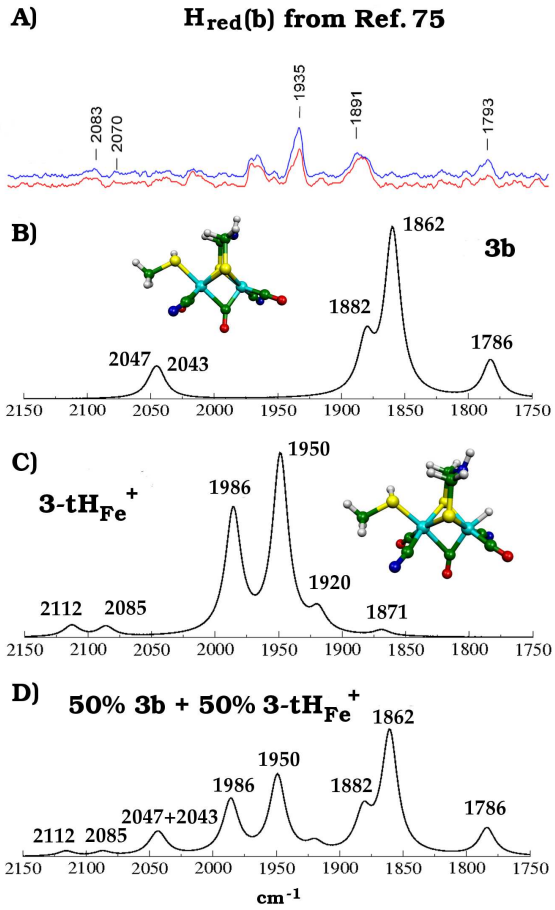


Figure 9: Measured IR spectrum of the $\mathbf{H_{red}(b)}$ form of [FeFe]-hydrogenase enzyme reproduced from Ref. [75] (panel A), and BP86/RI/TZVP calculated IR spectra for the model complexes **3b** (panel B), and **3-tH**$_{\mathbf{Fe}}^{+}$ (panel C). In panel D is reported the simulated spectrum of 50-50 mixture of **3b** and **3-tH**$_{\mathbf{Fe}}^{+}$. The wave numbers are given in cm^{-1}. All computed frequencies are scaled according to Eq. (1) and Eq. (3). The simulated spectra are presented with a half-width of 15 cm^{-1}.



List of Tables

Table 1: Counter ions, solvents and references to the IR spectra of the 32 clusters considered for the calibration of the calculated CO and CN frequencies (abbreviations in the formulae are: ad = adamantyl; IMes = 1,3-bis(2,4,6-trimethylphenyl)imidazol-2-ylidene; PTA-H = protonated 1,3,5-triaza-7-phosphaadamantane; PTA-Me = methylated 1,3,5-triaza-7-phosphaadamantane; edt = ethane-1,2-dithiolato; pdt = propane-1,3-dithiolate; dtma = di(thiomethyl)amine. For structural representations of these clusters see the Supporting Information).

Complex		Counter ion	Solvent	Ref.
1	$[Fe_2(CO)_6\{SCH_2SCH_2S\}]$	—	—	[56]
2	$[Fe_2(CO)_6\{SC(Me)_2SC(Me)_2S\}]$	—	—	[56]
3	$[Fe_2(CO)_6\{SC(Et)_2SC(Et)_2S\}]$	—	—	[56]
4	$[Fe_2(CO)_6\{SC(C_5H_{10})SC(C_5H_{10})S\}]$	—	—	[56]
5	$[Fe_2(CO)_6\{SC(ad)_2SC(ad)_2S\}]$	—	—	[56]
6	$[Fe_2(CO)_6\{SC(ad)S\}]$	—	—	[56]
7	$[Fe_2(CO)_6\{SC(C_5H_{10})S\}]$	—	—	[56]
8	$[Fe_2(CO)_5(CN)\{SCH_2SCH_2S\}][Et_4N]$	$[Et_4N]^+$	THF	[56]
9	$[Fe_2(CO)_3(PMe_3)(dppe)(edt)]^{+1}$	$[BF_4]^-$	CH_2Cl_2	[79]
10	$[Fe_2(CO)_4(PMe_3)(dppe)(edt)]^{+2}$	$[BF_4]^-$	CH_2Cl_2	[79]
11	$[Fe_2(CO)_3(CN)_2(PMe_3)_2(edt)]$	—	CH_3CN	[80]
12	$[Fe_2(CO)_4(CN)_2(pdt)]^{-2}$	[18-crown-6-K]$^+$	CH_3CN	[16, 81]
13	$[Fe_2(CO)_4(PMe_3)_2(pdt)(\mu\text{-H})]^{+1}$	$[PF_6]^-$	CH_3CN	[19]
14	$[Fe_2(CO)_4(PMe_3)_2(pdt)(\mu\text{-}SCH_3)]^{+1}$	$[BF_4]^-$	CH_2Cl_2	[82]
15	$[Fe_2(CO)_5\{MeSCH_2C(Me)(CH_2S)_2\}]$	—	CH_3CN	[17]
16	$[Fe_2(CO)_4(PMe_3)(IMes)(pdt)]$	—	toluene	[83]
17	$[Fe_2(CO)_4(PMe_3)(IMes)(pdt)]^{+1}$	$[PF_6]^-$	CH_2Cl_2	[83]
18	$[Fe_2(CO)_6(pdt)]$	—	THF	[20, 84]
19	$[Fe_4\{SC(Et)_2SC(Et)_2S\}(CO)_{11}(S)]$	—	—	[56]
20	$[Fe_2(CO)_2(PTA\text{-}Me)_2(pdt)]^{+2}$	$[PF_6]^-$	CH_3CN	[21]
21	$[Fe_4\{SCMe_2SCMe_2S\}(CO)_{11}(S)]$	—	—	[56]
22	$[Fe_4\{SC(C_5H_{10})SC(C_5H_{10})S\}(CO)_{11}(S)]$	—	—	[56]
23	$[Fe_3(CO)_9(ad)(S)]$	—	—	[56]
24	$[Fe_2(CO)_4(CN)(PMe_3)(pdt)(\mu\text{-H})]$	—	THF	[22]
25	$[Fe_2(CO)_4(CN)\{MeSCH_2C(Me)(CH_2S)_2\}]^{-1}$	$[Et_4N]^+$	CH_3CN	[17]
26	$[Fe_2(CO)_4(CN)_2\{MeSCH_2C(Me)(CH_2S)_2\}]^{-2}$	[18-crown-6-K]$^+$	CH_3CN	[17]
27	$[Fe_2(CO)_5(CN)(pdt)]^{-1}$	$[Et_4N]^+$	THF	[18, 84]
28	$[Fe_2(CO)_4(PMe_3)_2(pdt)]$	—	CH_3CN	[19]
29	$[Fe_2(CO)_6(dtma)]$	—	hexane	[85]
30	$[Fe_2(CO)_6\{SCH_2N(Me)CH_2S\}]$	—	CH_3CN	[86]
31	$[Fe_2(CO)_6\{SCH_2N(Me)_2CH_2S\}]^{+1}$	unknown	CH_3CN	[86]
32	$[Fe_2(CO)_2(PTA\text{-}H)_2(pdt)]^{+2}$	$[F_3CSO_3]^-$	CH_3CN	[21]

Table 2: Mean signed deviation (MSD), mean unsigned deviation (MUD) and maximum deviation (MaxDev) for all Fe-CO, Fe-S, Fe-CN, Fe-P, and Fe-Fe distances in Å based on models **1**–**9**, **12**–**15**, **18**, **19**, **21**–**24**, **27**–**30**, **32**

bond length	MSD	MUD	MaxDev
Fe–CO	0.009	0.012	0.056
Fe–S	−0.008	0.035	−0.062
Fe–CN	0.017	0.019	0.042
Fe–P	−0.033	0.033	−0.052
Fe–Fe	−0.033	0.033	−0.092

Table 3: Statistical analysis of the CN and CO calibration models. Here, r^2 is the coefficient of determination, q^2 is the predictive squared correlation coefficient, CI(90%) is the confidence interval, MSD is the mean signed deviation, MUD is the mean unsigned deviation, and MaxDev is the maximum deviation

	r^2	q^2	CI(90%)	MSD	MUD	MaxDev
			CN			
ν_{CN}(Scal.)=ν_{CN}(Cal.)−21	0.913	0.886	a=[−832, 215]	0.5	5.0	10.0
ν_{CN}(Scal.)=0.990·ν_{CN}(Cal.)	0.911	0.884	b=[0.888, 1.385]	0.6	5.1	−10.2
ν_{CN}(Scal.)=1.137·ν_{CN}(Cal.)−309	0.926	0.868		−0.7	4.7	−10.9
			CO			
ν_{CO}(Scal.)=ν_{CO}(Cal.)−3	0.971	0.970	a=[−102, 4]	−0.6	7.4	−39.0
ν_{CO}(Scal.)=0.999·ν_{CO}(Cal.)	0.971	0.970	b=[0.997, 1.050]	−0.6	7.4	−39.1
ν_{CO}(Scal.)=1.023·ν_{CO}(Cal.)−49	0.971	0.970		0.5	7.4	−36.3

Table 4: Mean signed deviation (MSD), mean unsigned deviation (MUD) and maximum deviation (MaxDev) for the ν(CO) and ν(CN) vibrational frequencies of the relevant H_{ox}-CO, H_{ox}, and H_{red} model complexes discussed in the main text. The scaled frequecies of the **3t** complex are compared with the 1894, 1916, 2041 and 2079 cm^{-1} experimental bands, while the scaled frequecies of the $\mathbf{3tH_N}^+$ complex are compared with the 1894, 1916, 1965, 2079 and 2093 cm^{-1} experimental bands. For the CO vibrational frequencies the values calculated by excluding the bridging CO vibration from the analysis are given in parenthesis.

	MSD	MUD	MaxDev	MSD	MUD	MaxDev
		CO			CN	
1-CO	0.5 (19.0)	28.0 (19.0)	-55 (31)	-9.5	9.5	-16
1-CN	-3.0 (4.3)	10.0 (5.0)	-25 (13)	0.5	2.5	4
2	-13.0 (-7.5)	13.0 (7.5)	-31 (-4)	5.5	5.5	8
$\mathbf{2H_N}^+$	-46.0 (-21.3)	46.0 (21.3)	-74 (-41)	3.5	17.5	21
2-H_2	-12.3 (-6.5)	12.3 (6.5)	-24 (-8)	-8.0	8.0	-10
3t	12.5 (-4.0)	16.5 (4.0)	29 (-4)	11.5	11.5	14
$\mathbf{3tH_N}^+$	0.3 (-7.5)	10.3 (7.5)	16 (-9)	6.0	6.0	10