H_2 Activation by Hydrogenases

Deutsche Ausgabe: DOI: 10.1002/ange.201501737
Internationale Ausgabe: DOI: 10.1002/anie.201501737

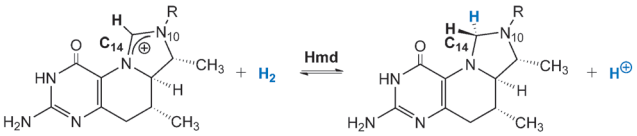
Multiscale Modeling of the Active Site of [Fe] Hydrogenase: The H_2 Binding Site in Open and Closed Protein Conformations**

Erik Donovan Hedegård, Jacob Kongsted, and Ulf Ryde*

Abstract: *A series of QM/MM optimizations of the full protein of [Fe] hydrogenase were performed. The FeGP cofactor has been optimized in the water-bound resting state (**1**), with a side-on bound dihydrogen (**2**), or as a hydride intermediate (**3**). For inclusion of H_4MPT in the closed structure, advanced multiscale modeling appears to be necessary, especially to obtain reliable distances between CH-H_4MPT$^+$ and the dihydrogen (H_2) or hydride (H^-) ligand in the FeGP cofactor. Inclusion of the full protein is further important for the relative energies of the two intermediates **2** and **3**. We finally find that hydride transfer from **3** has a significantly higher barrier than found in previous studies neglecting the full protein environment.*

Large-scale industrial generation of H_2 is today performed by the steam reforming process, which requires high temperature, even in presence of metal catalysts.[1] Meanwhile, a class of enzymes termed hydrogenases mediates the reversible conversion of H_2 into hydride and protons at ambient temperature and pressure; this is indeed remarkable considering that the pK_a value of H_2 is as high as 35.[1] This reversible generation of hydrogen holds great potential in relation to energy storage and as green alternative to fossil fuel.

To date, three different kinds of hydrogenases have been characterized.[2] The two first classes have been known for decades and are denoted [NiFe] and [FeFe] hydrogenases. They have bimetallic active sites with iron and nickel or two iron ions, respectively. In 1990, a third class was discovered, namely the H_2-forming methylene–tetrahydromethanopterin dehydrogenases (Hmd).[3a] Distinct for this class is that it is only active in presence of the methenyl–tetrahydromethanopterin (CH-H_4MPT$^+$) substrate for which it catalyzes the proton/H_2 exchange (Scheme 1). Hydrogenases of this class



Scheme 1. The reaction catalyzed by [Fe] hydrogenase (Hmd), in which the CH-H_4MPT$^+$ substrate (left) is converted into methylene–tetrahydromethanopterin (CH_2-H_4MPT).

are also denoted [Fe] hydrogenase, referring to their monoatomic active site, which is another feature that makes this third class distinct from the two other classes.

The [Fe] hydrogenase active site consists of an iron-containing cofactor (the FeGP cofactor; Figure 1). It was first identified by Thauer and co-workers[3a,b] and its nature was subsequently revealed,[3] ultimately with the full crystal structure of the protein.[4] It consists of an iron ion bound to two CO groups and a large pyridinol–guanine ribonucleotide. In the original crystal structure (Figure 1 left), the protein is in an open conformation. This structure has a large cleft between the central globular unit (dashed circle in Figure 1, left) and the peripheral units, where the active site is located (solid circle). It is believed that the substrate enters this cleft (that is, at the crossing point of the two circles) and the protein subsequently undergoes large conformational changes into a closed conformation (Figure 1, right). A crystal structure of the protein in the closed conformation has been obtained for the apoenzyme that is, without both the FeGP cofactor and the substrate.[5] Recently, a Cys176Ala mutated holoenzyme-CH_2-H_4MPT complex was also resolved by X-ray crystallography,[10] but in the open conformation where CH_2-H_4MPT is too far from the active site to study the mechanism in detail. Inhibited forms[6b] have also been studied by X-ray crystallography.

Thus, the nature of the H_2 binding and the reaction mechanism are still speculative, and intermediates with H_2 bound to iron have not been identified. Although the crystallographic work has been indispensable for unraveling the mechanism of the hydrogenases, it is hampered by two factors that have proved necessary for proper crystallization.[6a] First, a Cys176Ala mutant has been used. Second, small amounts of dithiothreitol (DTT) were added to stabilize cofactor and protein. Both factors alter the first coordination sphere of the active site iron ion (Cys-176 is a ligand of Fe in the native enzyme that is replaced by DTT in the mutant), rendering the protein inactive. Theory can complement the experimental data by providing optimized structures of the wild-type protein without DTT coordinated and additional insight can be gained by also considering possible reaction intermediates. In this regard, interesting theoretical work on

[*] Dr. E. D. Hedegård, Prof. J. Kongsted
University of Southern Denmark
Department of Physics, Chemistry and Pharmacy
Campusvej 55, 5230 Odense M (Denmark)
E-mail: edh@sdu.dk
Prof. U. Ryde
Department of Theoretical Chemistry, Lund University
Getingevägen 60, P.O.Box 124, 221 00 Lund (Sweden)

[**] E.D.H. thanks the Villum, OTICON, and Augustines funding agencies for stipends. This investigation has been supported by grants from the Swedish research council (project 2010-5025). The computations were performed on computer resources provided by the Swedish National Infrastructure for Computing (SNIC) at Lunarc at Lund University. The authors are grateful to Prof. Segio Shima for providing the coordinates of the modelled closed structure of the protein.

Supporting information for this article is available on the WWW under http://dx.doi.org/10.1002/anie.201501737.

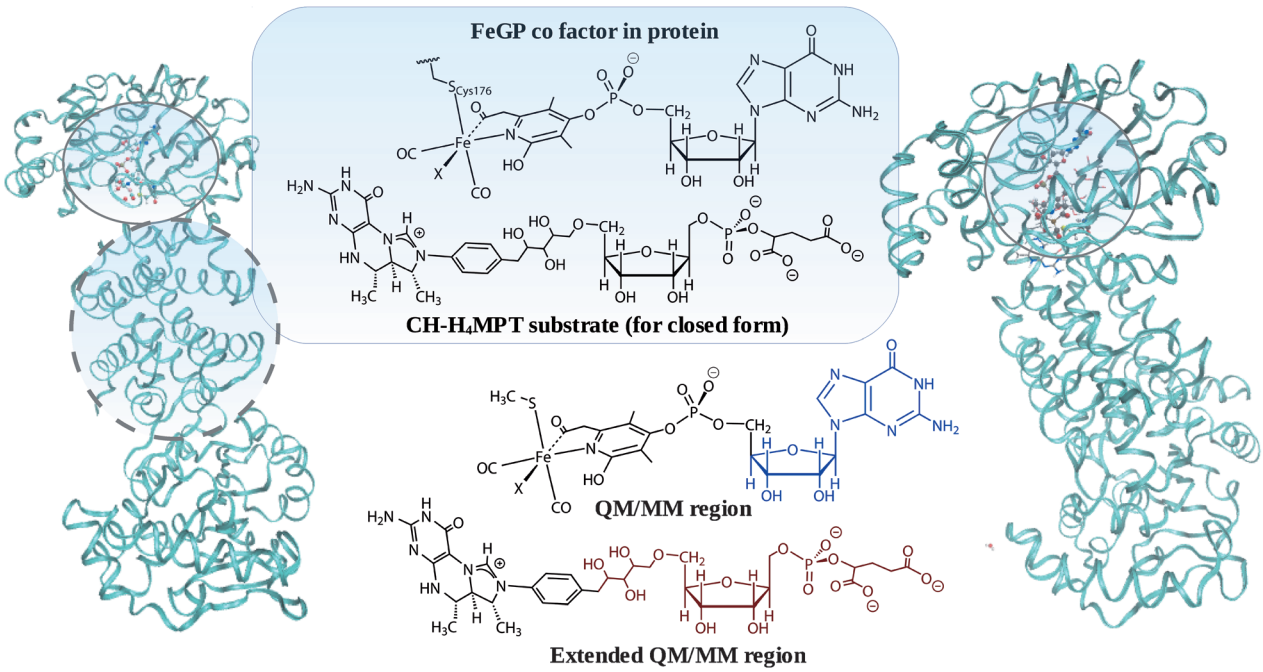


Figure 1. Open (left) and closed (right) conformations of [Fe] hydrogenase. The FeGP cofactor and the region used in QM/MM geometry optimizations are also shown (the blue and red parts are treated by MM). X is either H_2 or H^-.

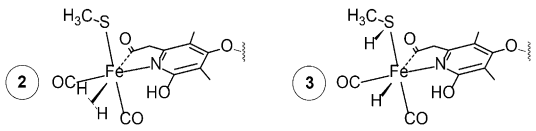


Figure 2. [Fe] hydrogenase intermediates considered herein.

small models without the protein environment has previously been performed,[7a–d] revealing two putative reaction intermediates (Figure 2). We denote them **2** and **3** here, while the water-bound state obtained from X-ray crystallography is denoted **1**. As can be seen in Figure 2, the reaction intermediate **2** has H_2 bound side-on to the Fe ion, *trans* to the acyl group,[6b] replacing the water molecule in the resting state (**1**). In the second intermediate (**3**), the H_2 molecule has been cleaved heterolytically, giving a hydride ion coordinated to Fe and a proton that has moved to the adjacent Cys-176 group.

In this study, these models have been brought to the full protein scale.[8] This is not a straightforward task, since the protein is a dimer and both monomers need to be included because of the exposed location of the active site. The inclusion of water solvent surrounding the full dimer gives a very large system that is beyond reach for any electronic-structure method. Accordingly, we use here a multiscale model approach, namely a hybrid between density functional theory (DFT) and the AMBER molecular mechanics (MM) force field, combined in the QM/MM program ComQum.[9] We investigated several systems: The resting structure known from X-ray crystallography with a water molecule bound at the proposed H_2 binding site (**1**), as well as the two intermediates, **2** and **3**. Furthermore, we studied the hydride abstraction reaction from intermediate **3** by the CH-H_4MPT$^+$ substrate. We first studied the open conformation of the protein, followed by studies of the closed conformation, including H_4MPT. These latter studies pose several problems as there is no starting structure available from X-ray studies: Therefore we started from a superimposition of crystal structures of the protein in the closed conformation and a protein-CH_2-H_4MPT complex[5,6a,9] as carried out previously.[9] Direct (rigid) superimposing renders parts of the protein and especially H_4MPT severely distorted. Therefore, substantial work-up was required to obtain a reliable structure (see the Experimental Section and the Supporting Information); in particular a 100 ns molecular dynamics (MD) simulation turned out to be crucial. The crystal structure and the closed-structure model contained the CH_2-H_4MPT product, which we kept during the MD simulations so that the obtained structures can be compared to experimental investigations with CH_2-H_4MPT (see the Supporting Information). Finally, QM/MM structures were also obtained with the CH-H_4MPT$^+$ substrate and the hydride-transfer reaction was studied.

In Table 1, we show hydrogen–iron and hydrogen–C_{14} distances in the calculations in the open conformation without H_4MPT and in the closed conformation with CH_2-H_4MPT or CH-H_4MPT$^+$. For CH-H_4MPT$^+$ we also carried out calculations, including the reactive part of CH-H_4MPT$^+$ in the QM region during the QM/MM optimization (the extended QM/MM region in Figure 1). Further data are provided in the Supporting Information. The Fe–H bond lengths are similar to those found in the QM-vacuum optimizations by Yang and Hall,[7a] while our optimized Fe–C_{14} and H–C_{14} distances (determining the relative location of H_4MPT and FeGP) differ more compared the previous calculations without the protein. These differences emphasize the importance of the

Table 1: Optimized bond distances [Å] for intermediates **2** and **3**.[a]

	No H_4MPT	With CH_2-H_4MPT	With CH-H_4MPT$^+$	Refs. [7a, 8][b]
Bonds (**2**)				
Fe–H_1	1.90	1.88	1.86 (1.82)	1.81/1.71
Fe–H_2	1.90	1.90	1.91 (1.86)	1.84/1.73
H_1–H_2	0.79	0.79	0.79 (0.80)	0.79/0.80
H_1–C_{14}	N.A.	3.25	3.00 (3.23)	N.A./–
H_2–C_{14}	N.A.	2.92	2.63 (2.88)	N.A./2.16
Fe–C_{14}	N.A.	4.75	4.46 (4.52)	N.A./–
Bonds (**3**)				
Fe–H	1.60	1.61	1.61 (1.62)	1.62/1.61
H–C_{14}	N.A.	3.33	2.95 (2.88)	3.07/1.93
Fe–C_{14}	N.A.	4.93	4.50 (4.48)	4.61/–

[a] Optimizations with the extended QM region, including the reactive part of CH-H_4MPT$^+$, are shown in brackets (see also Figure 3).
[b] Ref. [7a] used a truncated model of CH-H_4MPT$^+$.

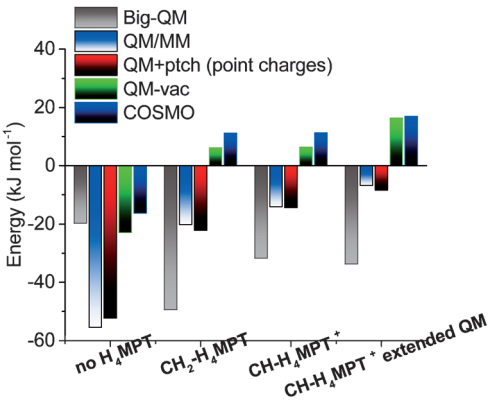


Figure 4. Relative energies of compounds **2** and **3** (E_2–E_3) calculated with the TPSS functional and def2-TZVP basis set. All of the energies are calculated for geometries optimized in the protein.

multiscale modeling approach. Our optimized H–C_{14} distances are also somewhat different from those obtained in a previous QM/MM study.[8] Figure 3 shows selected structural parameters for the QM/MM optimized intermediates **2** and **3** with CH-H_4MPT$^+$ (extended QM regions).

The QM/MM calculations also give estimates of the relative energies of the two intermediates **2** and **3** (Figure 4). The QM/MM energies in all of the calculations (without

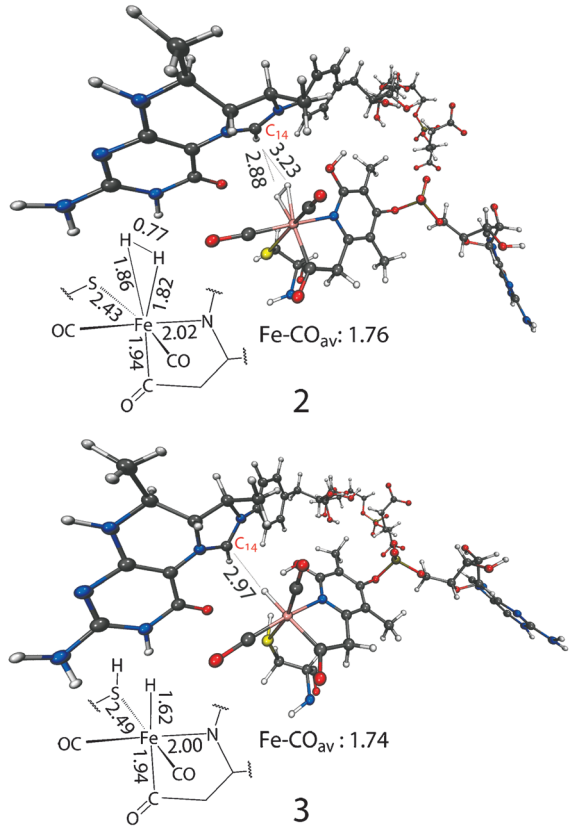


Figure 3. QM/MM optimized structures and selected distances with CH-H_4MPT$^+$ (100 ns). C black, H white, Fe pink, O red, N blue, S yellow.

H_4MPT, with CH_2-H_4MPT, or with CH-H_4MPT$^+$) indicate that the H_2 intermediate (**2**) is 6–20 kJ mol^{-1} more stable than the hydride intermediate (**3**). The same qualitative trend was found using other snapshots from the MD simulations. Interestingly, this is opposite to what was found in previous vacuum calculations[7a] in which **3** was 14 kJ mol^{-1} more stable than **2**. To understand this difference, we carried out vacuum and continuum-solvent (COSMO) calculations on the protein-optimized structures, as well as on the vacuum-optimized structure of the isolated QM region (Supporting Information, Table S5). With our slightly larger QM system and basis set we obtain a much smaller E_2–E_3 energy difference (2 kJ mol^{-1}) for structures optimized in vacuum (this discrepancy is caused mainly by a difference in the hydrogen-bond patter, as is explained in the Supporting Information). Without the COSMO solvation, intermediate **2** is more stable by 6 kJ mol^{-1}.

If we instead use the QM/MM geometries of the two intermediates, but still calculate single-point vacuum DFT energies, intermediate **3** is 6–16 kJ mol^{-1} more stable than **2** for the closed structures (with CH_2-H_4MPT or CH-H_4MPT$^+$), whereas **2** is 22 kJ mol^{-1} more stable in the open structure (QM-vac in Figure 4). Thus, there is a rather large geometric effect from the surrounding enzyme in the latter case. If we add a point-charge model of the protein to the QM calculation, intermediate **2** is stabilized by about 30 kJ mol^{-1} for all four systems (QM + ptch in Figure 4). If we finally add the MM energies of the surroundings, we end up in final QM/MM energies mentioned above. These MM energy corrections are small for all systems (0–3 kJ mol^{-1}). Thus, we can conclude that the difference between QM-cluster and QM/MM results is mainly caused by electrostatic effects in the protein. However, as shown in Figure 4, COSMO calculations on the protein structures do not reproduce this stabilization of intermediate **2**, but instead favor intermediate **3**, showing that the protein effects are specific in a way that cannot be modelled by a continuum solvent model.

Apparently, the stability of the intermediates **2** and **3** strongly depend on the size of the QM system and how the surroundings are modeled, as has been observed previously for other systems.[11] To check our QM/MM results, we also

performed single-point calculations with very large QM systems, including all residues with atoms within 6 Å of the original QM system, the back-bone of three residues on each side of the Cys-176 ligand, and all buried charged residues in the protein (in total about 700 atoms; see Figure S3 in the Supporting Information). This choice of QM system follows rules established for other systems[12] and uses a point-charge description of the remaining protein. Although somewhat different energies were obtained, these calculations confirmed that intermediate **2** is the most stable state by 20–50 kJ mol^{-1} (Big-QM in Figure 4); this is our best estimate of the energy difference. The electrostatic stabilization of intermediate **3** can be further understood at the MM level (using fitted QM charges for the QM system), by calculating residue contributions to the energy difference. Such an analysis shows that the backbone of Cys-176, as well as four charged residues close to the active site (Arg-12, His-14, Lys-151, and Asp-251) give the largest contributions to the energy difference. The effect of the protein is counteracted by water molecules near the active site, especially in the open conformation (in which the active site is solvent exposed).

Finally, we have investigated the hydride abstraction reaction by the CH-H_4MPT$^+$ substrate from intermediate **3**. The results in Table 2 show that the reaction is downhill by 46–76 kJ mol^{-1}. This is opposite to what was found before (uphill by 33 kJ mol^{-1}).[7a] Also, the barriers are significantly higher (69–88 kJ mol^{-1} versus 36 kJ mol^{-1}).[7a] An analysis similar to that in Figure 4 shows that the COSMO solvation strongly disfavors the product (Supporting Information, Table S7). Thus, both the instability of intermediate **3** and the high activation barrier speak against a hydride as the reactive species. Instead, a more likely mechanism seems to be a concerted H_2 splitting in which the FeGP pyridinol OH group is deprotonated and acts as the proton acceptor, as has recently been suggested.[8]

To summarize, we have performed a series of QM/MM optimizations of the full (dimeric) protein of [Fe] hydrogenase starting from either a substrate-free crystal structure in the open conformation or a modelled and relaxed closed conformation including CH-H_4MPT$^+$ or CH_2-H_4MPT. The FeGP cofactor has been optimized in the water-bound resting state (**1**), with a side-on bound H_2 (**2**), or as a hydride intermediate (**3**). In all forms, the coordination environment of the iron ion is rather similar and it is also similar to what has been obtained in previous vacuum optimizations.[7] However, for inclusion of H_4MPT in the closed structure, advanced multiscale modeling appears to be necessary, especially to obtain reliable distances between CH-H_4MPT$^+$ and the H_2 or H^- ligand in the FeGP cofactor. Inclusion of the full protein is further important for the relative energies of the hydride and H_2 intermediates, and we find that the side-on bound H_2 intermediate is most stable, thus supporting a previously outlined mechanism.[8] We also find that hydride transfer from **3** has a significantly higher barrier than found in previous vacuum studies,[7a] mainly owing to differences in the structures in vacuum and in the protein. Consequently, it is important to experimentally validate the QM/MM structures by spectroscopic methods, for example, using Mössbauer parameters.[13] Therefore, full coordinates of our structures are provided in the Supporting Information.

Table 2: The hydride abstraction reaction from **3** to product [kJ mol^{-1}].

Hydride abstraction	QM/MM	Big-QM	COSMO	Vacuum	Ref. [7a][a]
$\Delta E_{Barrier}$	87.7	68.6	95.4	83.4	36.1
$\Delta E_{Reaction}$	−46.4	−76.0	−3.9	−51.5	33.1

[a] Using a continuum model.

Experimental Section

For the open conformation, the starting coordinates were obtained from the re-interpreted crystal structure (PDB: 3F47) in which the carbon atom from the FeGP acyl group coordinates to the iron ion.[4,6] The closed conformation was started from the model constructed by Hiromoto et al.[5,9] In the QM/MM and MM calculations, the protein as well as the guanidine part of the FeGP cofactor were treated by the AMBER 99SB force field,[14a,b] whereas the remainder of the FeGP cofactor and H_4MPT were treated by the all-atom GAFF force field.[14c] For both H_4MPT and FeGP, a few new MM parameters had to be determined (see the Supporting Information).

The proteins were simulated in a sphere of water molecules with a radius of 60 Å and were equilibrated by simulated annealing. Subsequently, structures for the three states (**1**, **2**, and **3**) were obtained for both the open and closed conformations (with either CH_2-H_4MPT or CH-H_4MPT$^+$) by QM/MM geometry optimization using the ComQum[9] software. An MD run of 100 ns was performed using **1** and CH_2-H_4MPT to relax unfavorable interactions in the closed-structure model. QM/MM optimizations were performed after 10, 50, and 100 ns (all data shown herein are for the 100 ns snapshot). The QM region is shown in Figure 1 (with X = H_2O for **1**, H_2 for **2**, and H^- for **3**. For **3**, Cys-176 is protonated. In calculations with CH_2-H_4MPT, the substrate was not included in the QM system for the geometry optimizations, while for structures including CH-H_4MPT$^+$, additional optimizations were carried out with the reactive part of CH-H_4MPT$^+$ in the QM system (extended QM region in Figure 1). All of the studies of the hydride transfer were performed with this extended QM region. Accurate energies were also estimated with the Big-QM approach[15] as detailed in the Supporting Information.

All QM calculations were performed with the Turbomole package[14d] using DFT as the QM method. All reported data were obtained with the TPSS functional[14e] (results with other functionals are presented in the Supporting Information), using the def2-SV(P) basis set for geometry optimizations, followed by single-point energy calculations with the def2-TZVP basis set. The MM part of the QM/MM calculations used the AMBER 10 software, whereas MD simulations were performed with AMBER 12.[14f] More detailed descriptions of the individual steps are given in the Supporting Information.

Keywords: [Fe] hydrogenase · hydrogen activation · molecular mechanics · multiscale modeling · quantum mechanics

How to cite: *Angew. Chem. Int. Ed.* **2015**, *54*, 6246–6250
Angew. Chem. **2015**, *127*, 6344–6348

[1] a) R. Zimmermann, *Science* **1999**, *285*, 687–689; b) E. Buncel, B. Menon, *J. Am. Chem. Soc.* **1977**, *99*, 4457–4461.
[2] a) W. Lubitz, E. Reijerse, M. van Gastel, *Chem. Rev.* **2007**, *107*, 4331–4365; b) J. C. Fontecilla-Camps, A. Volbeda, C. Cavazza,

Y. Nicolet, *Chem. Rev.* **2007**, *107*, 4273 – 4303; c) S. Shima, U. Ermler, *Eur. J. Inorg. Chem.* **2011**, 963 – 972.
[3] a) C. Zirngibl, R. Hedderich, R. K. Thauer, *FEBS Lett.* **1990**, *261*, 112 – 116; b) G. Buurman, S. Shima, R. K. Thauer, *FEBS Lett.* **2000**, *485*, 200 – 204; c) X. Wang, Z. Li, X. Zeng, Q. Luo, D. J. Evans, C. J. Pickett, X. Liu, *Chem. Commun.* **2008**, 3555 – 3557; d) J. A. Wright, P. J. Turrell, C. J. Pickett, *Organometallics* **2010**, *29*, 6146 – 6156.
[4] S. Shima, S. Pilak, O. Vogt, M. Schick, M. S. Stagni, W. Meyer-Klaucke, E. Warkentin, R. K. Thauer, U. Ermler, *Science* **2008**, *321*, 572 – 575.
[5] O. Pilak, B. Mamat, S. Vogt, C. H. Hagemeier, R. K. Thauer, S. Shima, C. Vonrhein, E. Warkentin, U. Ermler, *J. Mol. Biol.* **2006**, *358*, 798 – 809.
[6] a) T. Hiromoto, K. Ataka, O. Pilak, S. Vogt, M. Salomone-Stagni, W. Meyer-Klaucke, E. Warkentin, R. K. Thauer, S. Shima, U. Ermler, *FEBS Lett.* **2009**, *583*, 585 – 590; b) H. Tamura, M. Salomone-Stagni, T. Fujishiro, E. Warkentin, W. Meyer-Klaucke, U. Ermler, S. Shima, *Angew. Chem. Int. Ed.* **2013**, *52*, 9656 – 9659; *Angew. Chem.* **2013**, *125*, 9838 – 9841.
[7] a) X. Yang, M. B. Hall, *J. Am. Chem. Soc.* **2009**, *131*, 10901 – 10908; b) M. T. Stiebritz, M. Reiher, *Inorg. Chem.* **2010**, *49*, 5818 – 5823; c) A. R. Finkelmann, M. T. Stiebritz, M. Reiher, *J. Phys. Chem. B* **2013**, *117*, 4806 – 4817; d) A. Dey, *J. Am. Chem. Soc.* **2010**, *132*, 13892 – 13901.
[8] A. R. Finkelmann, H. M. Senn, M. Reiher, *Chem. Sci.* **2014**, *5*, 4474 – 4482.
[9] a) U. Ryde, M. H. M. Olsson, *Int. J. Quantum Chem.* **2001**, *81*, 335 – 347; b) U. Ryde, L. Olsen, K. Nilson, *J. Comput. Chem.* **2002**, *23*, 1058 – 1070.
[10] T. Hiromoto, E. Warkentin, J. Moll, U. Ermler, S. Shima, *Angew. Chem. Int. Ed.* **2009**, *48*, 6457 – 6460; *Angew. Chem.* **2009**, *121*, 6579 – 6582.
[11] a) L. Hu, J. Eliasson, J. Heimdal, U. Ryde, *J. Phys. Chem. A* **2009**, *113*, 11793 – 11800; b) L. Hu, P. Söderhjelm, U. Ryde, *J. Chem. Theory Comput.* **2011**, *7*, 761 – 777.
[12] a) L. Hu, P. Söderhjelm, U. Ryde, *J. Chem. Theory Comput.* **2013**, *9*, 640 – 649; b) S. Sumner, P. Söderhjelm, U. Ryde, *J. Chem. Theory Comput.* **2013**, *9*, 4205 – 4214.
[13] a) E. D. Hedegård, S. Knecht, U. Ryde, J. Kongsted, T. Saue, *Phys. Chem. Chem. Phys.* **2014**, *16*, 4853 – 4863; b) J. Gubler, A. R. Finkelmann, M. Reiher, *Inorg. Chem.* **2013**, *52*, 14205.
[14] a) V. Hornak, R. Abel, A. Okur, B. Strockbine, A. Roitberg, C. Simmerling, *Proteins Struct. Funct. Bioinf.* **2006**, *65*, 712 – 725; b) K. Lindorff-Larsen, S. Piana, K. Palmo, P. Maragakis, J. L. Klepeis, R. O. Dror, D. E. Shaw, *Proteins Struct. Funct. Bioinf.* **2010**, *78*, 1950 – 1958; c) S. J. Weiner, P. A. Kollman, D. T. Nguyen, D. A. Case, *J. Comput. Chem.* **1986**, *7*, 230 – 252; d) R. Ahlrichs, M. Bär, M. Häser, H. Horn, C. Kölmel, *Chem. Phys. Lett.* **1989**, *162*, 165 – 169; e) J. M. Tao, J. P. Perdew, V. N. Staroverov, G. E. Scuseria, *Phys. Rev. Lett.* **2003**, *91*, 146401; f) D. Case et al., AMBER 10 and AMBER 12, University of California, San Francisco.
[15] S. P. de Visser, M. G. Quesne, B. Martin, P. Comba, U. Ryde, *Chem. Commun.* **2014**, *50*, 262 – 282.

Received: February 23, 2015
Published online: April 13, 2015

