Supporting Information

Hydrogen-Activation Mechanism of [Fe] Hydrogenase Revealed by Multi-Scale Modeling

Arndt Robert Finkelmann[a], Hans Martin Senn[b]*, Markus Reiher[a]†

[a] ETH Zürich, Laboratorium für Physikalische Chemie, Vladimir-Prelog-Weg 2 8093 Zürich, Switzerland

[b] WestCHEM and School of Chemistry, University of Glasgow, Glasgow G12 8QQ, UK

May 31, 2014

*Corresponding author; e-mail: hans.senn@glasgow.ac.uk

†Corresponding author; e-mail: markus.reiher@phys.chem.ethz.ch

1 Computational details

1.1 Parametrization of methylene-tetrahydromethanopterin and iron-guanylylpyridinol

To be able to study [Fe] hydrogenase by means of molecular-dynamics (MD) simulations and quantum-mechanics/molecular-mechanics (QM/MM) calculations a force-field description of the whole system is necessary. We chose the Amber force field (version ff03[1,2]) for the molecular-mechanics (MM) part. This force field can be combined with the general Amber force field (GAFF),[3] which allows for an easy parametrization of organic molecules. We followed the GAFF parametrization procedure to derive parameters for the iron-guanylylpyridinol (FeGP) (see Fig. 1 in the main paper) cofactor and the substrate methylene-tetrahydromethanopterin (methylene-H_4MPT, see Fig. 1 in the main paper). The parametrization can be divided into two steps. As first step, partial charges derived from the electrostatic potential (ESP) for all atoms were calculated. Then, GAFF topology files for the organic parts were created. The parameters used are given in Figs. 1, 2 and Tables 1, 2.

1.1.1 Calculation of partial charges

For the calculation of partial charges, we chose a procedure similar to that applied in Ref. 1. Structures were optimized with the Gaussian09 program package,[4] utilizing density functional theory with the B3LYP exchange-correlation functional,[5–7] a 6-31+G* basis set[8,9] and the integral equation formalism polarizable continuum model (IEFPCM)[10,11] with $\epsilon = 4$ to account for electrostatic screening. Note that in Ref. 1 structures were optimized with the Hartree–Fock method. Partial charges were calculated with the Turbomole program package (version 6.3.1)[12,13] according to a scheme related to that suggested by Kollman.[14] For the Turbomole calculations, density functional theory with the B3LYP exchange-correlation functional,[5–7] the conductor-like screening model ($\epsilon = 4$)[15] and an aug-cc-pVTZ basis set[16–18] on all atoms was employed.

Unconstrained structure optimization of methenyl-H_4MPT (starting structure from PDB file 3H65[19]) leads to the formation of internal hydrogen bonds of the tailing carboxylate groups to hydroxo groups of either the furanose ring or the glycerin-derived part. These internal hydrogen bonds are not found in the protein structure because the carboxylates can form intermolecular hydrogen bonds. Hence, intramolecular hydrogen bonds do not resemble the bonding situation in the crystal structure and need to be avoided. This was achieved by constraining several dihedral angles during optimization.

1.1.2 Generation of GAFF topologies

The generation of GAFF topologies for methylene-H_4MPT was straightforward. Topology files were created with the program ACPYPE,[20] which interfaces Antechamber[21] of AmberTools 13[22] to create Amber and Gromacs topology files. The whole methylene-H_4MPT molecule was treated as one MM-residue.

Generation of suitable parameters for the FeGP cofactor was more involved.[23] Since metal complexes are not parametrizable with GAFF the cofactor was split up into the guanylylpyridinol ligand, 2 CO molecules, the iron ion (each treated

as one MM-residue) and the cysteine rest. For the iron ion a new ion type was introduced. The two CO molecules were parametrized with GAFF. The guanylylpyridinol ligand could be parametrized with GAFF, however, the parameters of the acyl part coordinating the iron atom had to be adjusted (equilibrium value of the acyl O–C–C angle) (see Fig. 1 in the main paper for the structure). In the crystal structure, the open coordination site is coordinated by water. For this water molecule, the standard parameters of the TIP3P water model[24] were used as the ESP-calculated partial charges differed insignificantly from the standard charges. For the cysteinate rest, a modified cysteinate residue was defined, which had the same bonded and van der Waals parameters as a standard cysteinate, but ESP-derived charges. With a view on terminating the QM region at the cysteinate C^{β}–C^{α} bond in the subsequent QM/MM calculations, while maintaining integer charges for the QM and MM regions, a charge of $+0.069588e$ was distributed equally among the $C^{\beta}H_2$ atoms of the new cysteinate residue to obtain a total integer charge of $-1e$ for FeGP. Since the Fe atom, both CO molecules, the cysteine sulfur atom, and the CO atoms of the acyl-ligand were positionally restrained during the classical MD simulations, no metal–ligand bonded parameters had to be derived.

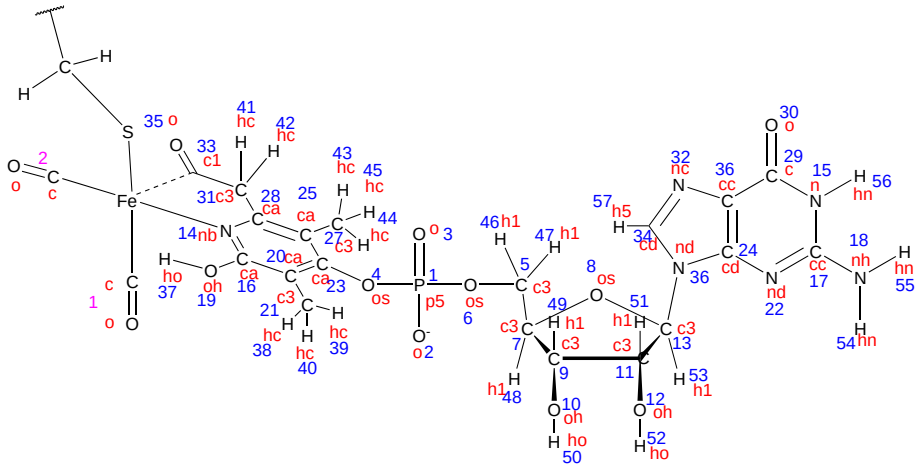

Figure 1: Atom types and atom numbers for the FeGP cofactor. GAFF atom types are given in red, atom numbers in blue. Pink numbers indicate the first and the second CO ligand.

1.2 Preparation of the protein structures for MD simulations

1.2.1 *Open* conformation

The simulations are based on the crystal structure published by Hiromoto *et al.*[19] (PDB code: 3H65). They achieved to crystallize a C176A mutant that harbors the cofactor, ligated by dithiothreitol (buffer molecule) instead of Cys176, in complex with the substrate methylene-H_4MPT. The crystal structure is in the *open* conformation. In the crystal structure, the C-terminal residues 346–358

Table 1: Atom types and partial charges of the FeGP cofactor for the MD simulations and for the QM/MM calculations. The structure with the corresponding atom numbers is given in Fig. 1.

atom	MD charge	QMMM charge	atom	MD charge	QMMM charge
$Fe(CO)_2SC^{\beta}H_2$					
Fe	0.7416375	0.741585	S	−0.4603	−0.460350
C1	0.074899	0.074849	C^{β}	−0.186503	−0.186553
O1	−0.223267	−0.223317	H^{β}	0.126296	0.126246
C2	0.143396	0.143346	H^{β}	0.126296	0.126246
O2	−0.243743	−0.243793			
guanylylpyridinol					
1 p5	1.173428	1.173379	30 o	−0.638592	−0.638525
2 o	−0.804905	−0.804954	31 c3	−0.290532	−0.290582
3 o	−0.758151	−0.758200	32 nc	−0.659167	−0.659100
4 os	−0.248012	−0.248061	33 c1	0.510845	0.510795
5 c3	−0.232453	−0.232502	34 cd	0.214577	0.214644
6 os	−0.451749	−0.451798	35 o	−0.634698	−0.634748
7 c3	0.336367	0.336435	36 na	−0.102300	−0.102233
8 os	−0.579760	−0.579692	37 ho	0.456286	0.456236
9 c3	−0.019723	−0.019655	38 hc	0.127078	0.127028
10 oh	−0.646053	−0.645985	39 hc	0.164999	0.164949
11 c3	0.161410	0.161478	40 hc	0.168804	0.168754
12 oh	−0.687887	−0.687819	41 hc	0.148313	0.148263
13 c3	0.430294	0.430362	42 hc	0.124692	0.124642
14 nb	−0.296088	−0.296137	43 hc	0.100120	0.100070
15 n	−0.657113	−0.657045	44 hc	0.147854	0.147804
16 ca	0.364888	0.364839	45 hc	0.130894	0.130844
17 cc	0.770287	0.770355	46 h1	0.120443	0.120393
18 nh	−0.851084	−0.851016	47 h1	0.172266	0.172216
19 oh	−0.601450	−0.601499	48 h1	0.088891	0.088958
20 ca	0.244988	0.244939	49 h1	0.235349	0.235416
21 c3	−0.501514	−0.501563	50 ho	0.408312	0.408379
22 nd	−0.690035	−0.689967	51 h1	0.093079	0.093146
23 ca	−0.222549	−0.222598	52 ho	0.439216	0.439283
24 cd	0.291468	0.291536	53 h2	0.035636	0.035703
25 ca	0.385152	0.385101	54 hn	0.400592	0.400659
26 cc	0.184629	0.184696	55 hn	0.389254	0.389321
27 c3	−0.432668	−0.432718	56 hn	0.415823	0.415890
28 ca	−0.163103	−0.163153	57 h5	0.138101	0.138168
29 c	0.595243	0.595310			
H_2O					
O	−0.834000	−0.834000	H	0.417000	0.417000
H	0.417000	0.417000			

Table 2: Atom types and partial charges of methylene-H_4MPT for the MD simulations and for the QM/MM calculations. The structure with the corresponding atom numbers is given in Fig. 2.

atom	MD charge	QMMM charge	atom	MD charge	QMMM charge
1 nc	−0.839624	−0.839569	49 o	−0.918099	−0.918141
2 cd	0.866549	0.866604	50 c3	−0.462739	−0.462781
3 nh	−0.869435	−0.869380	51 c3	0.054124	0.054082
4 n	−0.711463	−0.711408	52 c	0.877807	0.877765
5 c	0.716700	0.716755	53 o	−0.876948	−0.876990
6 o	−0.679835	−0.679780	54 o	−0.944601	−0.944643
7 cd	−0.258910	−0.258855	55 h1	0.107622	0.107580
8 nh	−0.378333	−0.378278	56 h1	0.049829	0.049787
9 c3	0.049297	0.049352	57 hc	0.025634	0.025592
10 c3	0.358361	0.358415	58 hc	0.086049	0.086007
11 c3	−0.632035	−0.631981	59 hc	−0.001997	−0.002039
12 nh	−0.584658	−0.584604	60 hc	0.017822	0.017780
13 cc	0.633917	0.633971	61 hc	0.121244	0.121202
14 c3	0.474230	0.474284	62 hc	0.094270	0.094228
15 c3	−0.473377	−0.473323	63 ho	0.417555	0.417513
16 c3	−0.006172	−0.006118	64 ho	0.402850	0.402808
17 nh	−0.348532	−0.348478	65 ho	0.364890	0.364848
18 ca	0.224733	0.224787	66 ho	0.434832	0.434790
19 ca	−0.327418	−0.327364	67 ho	0.440181	0.440139
20 ca	−0.277733	−0.277679	68 hn	0.384391	0.384445
21 ca	0.305265	0.305319	69 hn	0.392875	0.392929
22 ca	−0.231917	−0.231863	70 hc	0.088181	0.088235
23 ca	−0.320717	−0.320663	71 hc	0.118277	0.118331
24 c3	−0.214567	−0.214610	72 hc	0.129347	0.129401
25 c3	0.490123	0.490080	73 hc	0.152460	0.152514
26 c3	0.131906	0.131863	74 hc	0.165743	0.165797
27 c3	0.268650	0.268607	75 hc	0.170784	0.170838
28 c3	0.007070	0.007027	76 h1	0.075703	0.075661
29 oh	−0.639264	−0.639307	77 h1	0.107458	0.107416
30 oh	−0.670502	−0.670545	78 h2	0.164069	0.164123
31 oh	−0.677646	−0.677689	79 h2	0.045246	0.045300
32 os	−0.562346	−0.562389	80 h1	0.070626	0.070584
33 os	−0.440433	−0.440475	81 h1	0.013312	0.013270
34 c3	−0.048424	−0.048466	82 h1	−0.018889	−0.018931
35 c3	0.161302	0.161260	83 h1	0.008358	0.008316
36 os	−0.512516	−0.512558	84 hn	0.348778	0.348832
37 c3	0.653326	0.653284	85 hn	0.414010	0.414064
38 oh	−0.750097	−0.750139	86 h1	−0.062031	−0.061977
39 c3	−0.038881	−0.038923	87 h1	0.039663	0.039717
40 oh	−0.748423	−0.748465	88 h1	0.081657	0.081711
41 c3	0.577317	0.577275	89 h1	0.018267	0.018225
42 p5	1.257785	1.257743	90 h1	−0.094589	−0.094631
43 o	−0.893733	−0.893775	91 h1	0.095968	0.095926
44 o	−0.855874	−0.855916	92 h2	0.054208	0.054166
45 os	−0.423072	−0.423114	93 ha	0.180798	0.180852
46 c3	0.299822	0.299780	94 ha	0.146830	0.146884
47 c	0.923333	0.923291	95 ha	0.165934	0.165988
48 o	−0.913326	−0.913368	96 ha	0.181818	0.181872

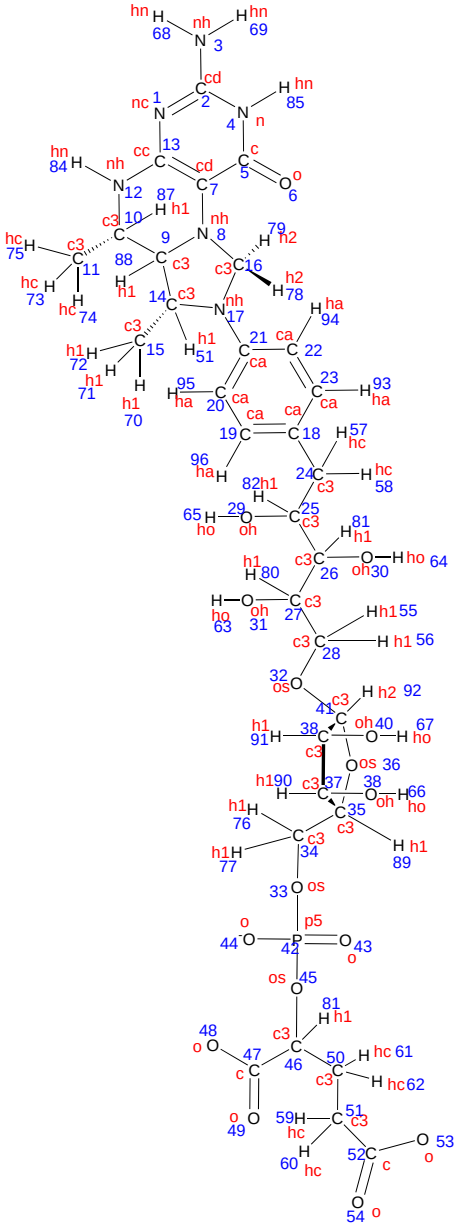

Figure 2: Atom types and atom numbers of methylene-H_4MPT. GAFF atom types are given in red, atom numbers in blue.

could not be refined. They form an unordered tail, far away from the reactive centers, and thus were simply omitted. To prepare the structure for the simulations, several modifications had to be made. FeGP is ligated by dithiothreitol in the crystal structure, which leads to a slight displacement of the whole cofactor compared to its position in the wild-type structure (without methylene-H_4MPT, PDB code: 3F47[25]). To obtain a wild-type-like structure Ala176 was mutated

Table 3: Partial charges of His14 for the QM/MM calculations. Atom names are according to the Amber ff03.

atom	QM/MM charge
N	−0.509660
H	0.348160
CA	0.116205
HA	0.134900
CB	−0.121078
HB2	0.087889
HB3	0.087889
CG	0.000013
ND1	−0.204225
HD1	0.319833
CE1	0.148833
HE1	0.123742
NE2	−0.599922
CD2	0.045304
HD2	0.111717
C	0.513086
O	−0.602692

Table 4: Protonation states of His residues of the modified 3H65 crystal structure for MD simulations, as determined with the program REDUCE.[27] δ indicates a proton at N^{δ} and ϵ indicates a proton at N^{ϵ}. pKa values were calculated with PROPKA.[28]

residue	H position	pKa value
HIS 14	δ	6.57
HIS 41	δ	0.60
HIS 52	δ	5.68
HIS 88	δ	5.24
HIS 120	ϵ	3.99
HIS 123	ϵ	5.82
HIS 174	δ	4.08
HIS 201	ϵ	1.68
HIS 294	ϵ	5.82
HIS 340	δ	7.22

back to cysteinate and the dithiothreitol ligand was removed from the structure. To place the cofactor in the correct (wild-type) position, the structure of the wild type (3F47) was aligned with the structure of the C176A mutant (3H65) using PyMOL.[26] The coordinates of the FeGP moiety (including the cysteine S atom) in the aligned wild-type structure were then inserted into the modified mutant structure, replacing those of the misplaced FeGP. With this procedure, the position of FeGP in the modified mutant structure resembled the position in the wild-type structure. Only the C^{β}–S bond of Cys176 (formerly Ala176) was stretched from 1.79 Å in the wild-type structure to 2.21 Å in the modified mutant structure. The correct bond length is restored in the energy-minimization step at the start of the MD simulations. The protonation states of titratable residues were determined with the program REDUCE,[27] which can also correct flipped Asn/Gln/His residues (none were in this case), and verified with PROPKA.[28] All titratable residues were in their standard protonation states; His protonation states are summarized in Table 4. Finally, the whole

protein dimer was created from the monomer chain with PyMOL.[26]

1.2.2 *Closed* conformation

The *closed* conformation was generated from the modified crystal structure of the *open* conformation (see previous section), which we call starting structure here. The only crystal structure available in the *closed* conformation is for the wild-type apoenzyme (PDB code: 2B0J[29]). To build a model of the complete enzyme in the *closed* conformation, the central subunit (residues 253–345) of 2B0J was first aligned to the central subunit of the starting structure. In the second step, the peripheral subunit (residues 1–241) of the starting structure was aligned to the peripheral subunit of the now aligned 2B0J. The FeGP cofactor was aligned together with the peripheral subunit. The structure thus obtained for the *closed* conformation of the holoenzyme–substrate complex had no significant atom overlaps, except for the tail part of methylene-H_4MPT (after the ribitol part, see Fig. 1 of the main paper). This tail was rotated with the help of the UCSF Chimera program[30] to remove atom overlaps. Given the high flexibility and mobility of the tail, any memory of the initial conformation will be lost during the MD sampling. In the structure thus generated, there is a gap in the backbone between residues 241 and 242, between the hinge region (residues 242–252) and peripheral subunit (see Fig. 3), which will be closed during energy minimization. Finally, the water molecule coordinated to Fe was removed because it would prevent the hydride transfer reaction that we intend to study. With the water molecule absent, the structure is exactly the product of the hydride transfer. The full dimer was again created with PyMOL.[26] A superposition of the starting structure, the final structure, and the apoenzyme in the *open* conformation is presented in Fig. 3. In the generated structure of the *closed* conformation, the distance between Fe and the hydride accepting carbon atom (C14a) is 3.23 Å, which compares well to the distance of 3 Å found by Hiromoto *et al.*, who modelled the structure of the *closed* conformation in a similar fashion.[19]

1.3 MD simulation protocol

All MD preparation steps and simulations were performed with the Gromacs molecular dynamics package version 4.5.5.[31–34] The protein was centered in a triclinic box with a minimal distance of 1 nm between solute and box border. The box was solvated and ions (Na^+ and Cl^-) were added to neutralize the system and to obtain an ion concentration of 0.15 mmol/L. The Fe atom, both CO ligands, the cysteinate S atom, and the CO group of the Fe-coordinating acyl ligand in the FeGP cofactor were kept frozen or positionally restrained at their positions in the prepared crystal structure. Positional restraints, rather than constraints were necessary for the pressure equilibration because pressure scaling with constrained (frozen) atoms leads to technical difficulties in the Gromacs implementation.

The integration time step was 2 fs. The linear constraint solver (LINCS) algorithm to 4th order with 1 iteration was invoked to enforce constraints (all bonds after energy minimization). Water molecules were kept rigid with the SETTLE algorithm in all steps after energy minimization. For neighbor searching, a grid-based group cut-off scheme was used with a cut-off distance of 1 nm

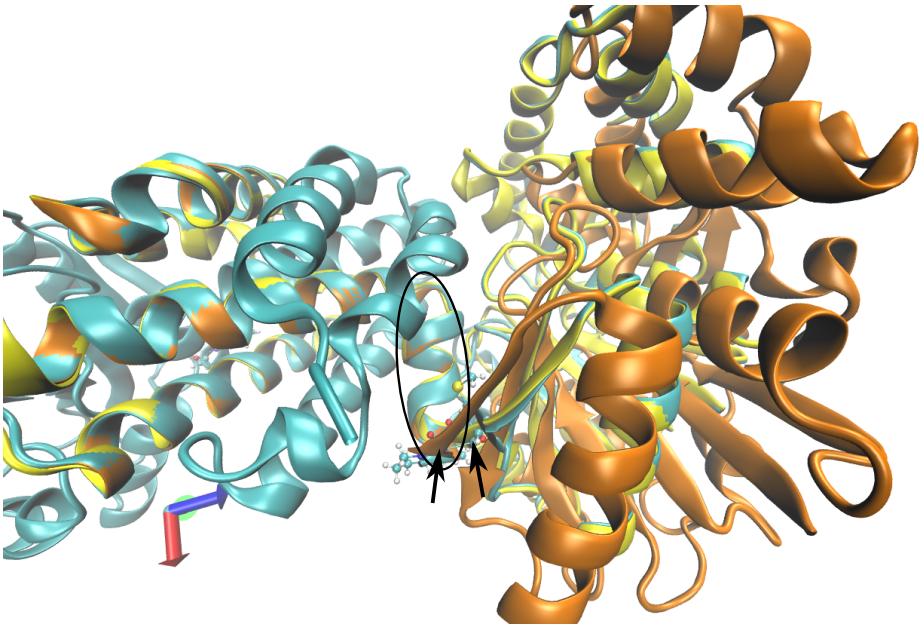

Figure 3: Superposition of the generated *closed* conformation (cyan), the apoenzyme in *closed* conformation (PDB code: 2B0J, yellow) and holoenzyme-substrate complex in *open conformation* (PDB code: 3H65, orange). The central (left side) units are aligned. The peripheral unit of the generated *closed* conformation cyan overlaps with the peripheral unit of the apoenzyme in *closed* conformation. The hinge region is marked by a black oval. The black arrows mark the gap between backbone atoms of residues 241 and 242 (see text).

for short-range interactions and the neighbor list was updated every 10 steps. Coulomb interactions were calculated with a smooth particle–mesh Ewald algorithm with interpolation order of 4, a Coulomb cut-off of 1 nm, Fourier spacing of 0.12 nm, tolerance of 10^{-5} and optimized Fourier transforms. Van der Waals interactions were calculated with a cut-off scheme (radius = 1 nm). Energy and pressure were corrected for long-range dispersion effects. Initial velocities were generated according to a Maxwell–Boltzmann distribution at 293 K. For temperature scaling the system was coupled to a v-rescale thermostat.[35] During equilibration, several subsystems were coupled to their own thermostats. For production, the entire system was coupled to one thermostat with a relaxation time of 2 ps and reference temperature of 293 K. Box equilibration was achieved with isotropic box rescaling by coupling the system to a Berendsen barostat with 1 ps relaxation time, compressibility of $4.5 \cdot 10^{-5}$ bar^{-1}, center-of-mass scaling of reference coordinates, and a target pressure of 1 bar.

The simulation stages are summarized in Table 5. The system in the *open* conformation was energy minimized in vacuum and in solvent. During 200 ps heating in an NVT ensemble, the protein, the cofactor, the substrate and the solvent were coupled to a temperature bath while all heavy atoms of the protein were positionally restrained. Thereafter, the box was equilibrated in an NPT simulation where the restraints on protein atoms were reduced in three

steps (total 900 ps). In the first step, protein, cofactor, substrate and solvent were coupled to three thermostats. In the second and third steps, only two thermostats were utilized (solvent, rest of the system). After adjustment of the box vectors (see Table 6), the system was equilibrated in an NVT ensemble for 400 ps with two thermostats (solvent, rest of the system) and finally 2 ns with one thermostat. The production trajectory was 100 ns. Coordinates were saved to disk every 20 ps.

The simulation procedure for the *closed* conformation was similar. Differences are a reduced force threshold for the vacuum minimization, position restraints for the heavy protein atoms during solvent minimization and only two thermostats (solvent, rest of the system) in all box-equilibration simulations. The equilibrated box parameters are collected in Table 6. The final equilibration simulation (NVT) had to be elongated for an additional 2 ns with a reduced temperature coupling constant and reduced restraints for Fe and its first-shell ligands (see Table 6) to avoid instabilities in the production run. Still, the production run terminated at 94.6 ns. We trace this back to the initial strain put into the system during crystal structure manipulation, which could not fully relax because the Fe atoms of the two FeGP cofactors were positionally restrained. Furthermore, the parameters for FeGP and methylene-H_4MPT might be not optimal for long simulations. However, the trajectory of 94.6 ns is already longer than trajectories normally produced to prepare QM/MM calculations[36] and is sufficiently long to allow us to analyse visited conformations, structural behavior, and protein dynamics around the active site. Trajectories were analysed with the VMD program.[37]

1.4 QM/MM setup

For the QM/MM calculations, several snasphots representing important conformations were selected, as described in the main paper. As QM/MM calculations under periodic boundary conditions are not supported by ChemShell, water molecules and ions outside a shell of 18 Å around the quantum-mechanics (QM) region were therefore discarded to create a finite system. The entire protein dimer was retained. QM/MM optimizations were carried out with the ChemShell program[38–40] (version 3.5.0) with Turbomole (version 6.5)[12,41] as the QM back-end. The optimizations were performed in hybrid delocalized internal coordinates (HDLCs) using the HDLCOpt module.[42] The scaling factor for Cartesian coordinates when constructing HDLCs was set to 0.8; the interval to update the pair list and regenerate HDLCs was set to 100 steps; the convergence threshold was set to 0.001 a.u.

Several regions were defined for the QM/MM optimizations. The *QM region* contained all quantum-mechanically described atoms (52 to 84 atoms). The *MM region* contained all other atoms of the system (approximately 12350 to 12900 atoms). Only the atoms in the *active region* (around 670 to 1700 atoms) were allowed to move in structure optimizations. We used a QM/MM microiterative optimization scheme 43, in which the *inner region* (around 61 to 93 atoms) contained the QM atoms, the MM boundary atoms, and the MM atoms bonded to the them.

The QM part of the calculations was treated with the TPSS exchange—correlation functional[44–46] plus Grimme's DFT-D3 dispersion correction.[47] Structures were optimized with the def2-TZVP basis set[48] on iron and the def2-SVP

Table 5: Simulation stages for simulation of [Fe] hydrogenase in the *open* conformation (top part) and in the *closed* conformation (bottom part). See text for more details.

	Minimization vac.	Minimization solv.	Heating	p equilibration	T equilibration	Production
No. of steps	$F_{\mathrm{tol}} = 2000$ kJmol^{-1}nm^{-1}	$F_{\mathrm{tol}} = 1000$ kJmol^{-1}nm^{-1}	100000	200000 + 50000 +200000	200000 +1000000	50000000
Propagation	Steepest descent	Steepest descent	Leap-frog Verlet			
Ensemble	–	–	NVT	NPT	NVT	NVT
τ_{T}	–	–	0.1/0.1/0.1	0.1/0.1/0.1–0.1/0.1	0.1/0.1–2	2
τ_{p}	–	–	–	1.0	–	–
Solute constraints	None	All bonds				
Restraint force (kJ mol^{-1} nm^{-1})	None	None	1000	1000–100–0	0	0
Fe center restraint	frozen	frozen	99999	99999	99999	99999
	Minimization vac.	Minimization solv.	Heating	p equilibration	T equilibration	Production
No. of steps	$F_{\mathrm{tol}} = 1$ (200 steps) kJmol^{-1}nm^{-1}	$F_{\mathrm{tol}} = 1000$ kJmol^{-1}nm^{-1}	100000	200000 + 50000 +200000	200000 +1000000 +1000000*	47305000
Propagation	Steepest descent	Steepest descent	Leap-frog Verlet			
Ensemble	–	–	NVT	NPT	NVT	NVT
τ_{T}	–	–	0.1/0.1/0.1	0.1/0.1	0.1/0.1–2–0.1*	2
τ_{p}	–	–	–	1.0	–	–
Solute constraints	None	All bonds				
Restraint force (kJ mol^{-1} nm^{-1})	None	1000	1000	1000–100–0	0	0
Fe center restraint	frozen	frozen	99999	99999	99999–50000*	50000

* Run was necessary to avoid system collapse at the production stage

Table 6: Compositions of the simulation systems

	open conformation	*closed* conformation
No. of solute atoms	10852	10846
No. of Na^+/Cl^- ions	128/96	99/67
No. of water molecules	31800	20889
Initial box size		
$(a, b/\text{nm}, V/\text{nm}^3)$	$9.277, 12.273, 1056.34$	$8.056, 11.390, 739.201$
Equilibrated box		
$(a, b/\text{nm}, V/\text{nm}^3)$	$9.276, 12.271, 1055.85$	$8.007, 11.321, 725.755$

basis set [49] on all other atoms. The resolution of the identity approximation with corresponding auxiliary basis sets [50] was invoked to speed up the calculations. [51] The same force-field setup as in the MD simulations was used for the MM region. QM–MM electrostatic interactions were calculated fully (no cut-off), and the QM and MM systems were coupled with the charge-shift scheme. [39,52–54] Where the QM–MM boundary cut through a covalent bond, the QM region was saturated with a H link atom. MM partial charges of residues cut by a QM–MM boundary were corrected such that the QM and MM parts had integer charges. The correction charge (which is usually of the order of $0.01e$) was distributed equally over the entire residue involved, leading to very minor differences between the MM and QM/MM partial charges for these residues (see Tables 1, 2, 3).

Two different QM regions were defined (see Fig. 7 of the main paper). The first QM region contains the Fe center, both CO ligands, the side chain of the iron-coordinating cysteine (Cys176), the gunaylylpyridinol ligand up to the phosphate linker and the hydride-acceptor up to the phenyl part. This first region was used to study H_2 activation reactions at the Fe center and hydride transfer to methenyl-H_4MPT$^+$. The second region did not contain the substrate, but instead the side chain of His14. This region was used for the investigation of proton transfer from the pyridinol hydroxyl group to His14.

2 Supplementary results

2.1 Effect of the basis set

The H_2 cleavage reaction in the *closed* conformation with the snapshot at 11 ns was recalculated with a def2-TZVP basis set [48] on all atoms. Also with the larger basis, no stable hydride species could be located, and structure optimizations converged to the reduced substrate. The reaction energy was -16.0 kcal/mol compared to -18.7 kcal/mol with the smaller basis set. Hence, a larger basis set has no significant quantitative or qualitative effect on structures; the small effect on energies leaves conclusions and interpretation unaffected.

2.2 Potential energy surface scans

To estimate the reaction barriers for the H_2 cleavage reaction with proton transfer to oxypyridine (snapshot at 11 ns) and of the proton transfer from thy pyridi-

nol OH group to His14 in the *open* conformation (snapshot at 10.78 ns), we calculated energy profiles along scan coordinates that closely resemble the reaction coordinates; the profiles are plotted in Fig. 4. The scans provide an upper bound for the reaction barrier. It was not possible to locate transition states for the two reactions. It is likely that the flat profile of the two reactions caused difficulties in numerical hessian calculations.

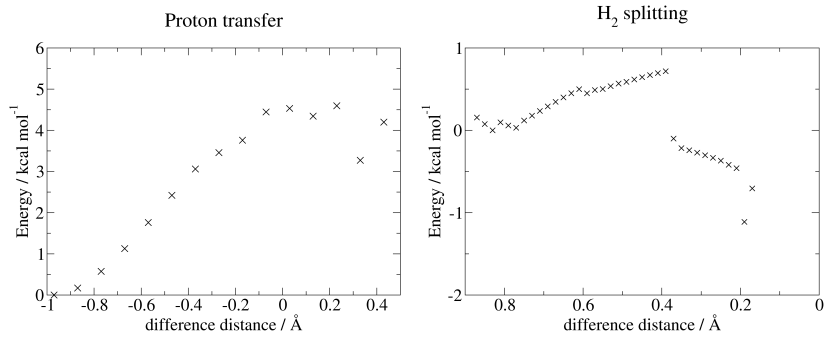

Figure 4: Plots of the potential energy surface scans. Left: Proton transfer from pyridinol OH to His14; the scan coordinate is the difference between the O–H and H–N^{ε} bond lengths. Right: H_2 cleavage with proton transfer to the oxypyridine O.; the scan coordinate is the difference between the O–H and H–H bond lengths.

References

[1] Duan, Y.; Wu, C.; Chowdhury, S.; Lee, M. C.; Xiong, G.; Zhang, W.; Yang, R.; Cieplak, P.; Luo, R.; Lee, T.; Caldwell, J.; Wang, J.; Kollman, P. *J. Comput. Chem.* **2003**, *24*, 1999–2012.

[2] Lee, M. C.; Duan, Y. *Proteins* **2004**, *55*, 620–634.

[3] Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. *J. Comput. Chem.* **2004**, *25*, 1157–1174.

[4] Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnenberg, J. L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Montgomery, J. A., Jr.; Peralta, J. E.; Ogliaro, F.; Bearpark, M.; Heyd, J. J.; Brothers, E.; Kudin, K. N.; Staroverov, V. N.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, J. M.; Klene, M.; Knox, J. E.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Zakrzewski, V. G.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Dapprich, S.; Daniels, A. D.; Farkas, O.; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J. *Gaussian 09 Revision C.01*, Gaussian Inc. Wallingford CT 2009.

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

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

[7] Stephens, P. J.; Devlin, F. J.; Chabalowski, C. F.; Frisch, M. J. *J. Phys. Chem.* **1994**, *98*, 11623–11627.

[8] Hariharan, P.; Pople, J. *Theoret. Chim. Acta* **1973**, *28*, 213–222.

[9] Clark, T.; Chandrasekhar, J.; Spitznagel, G. W.; von Ragueé Schleyer, P. *J. Comput. Chem.* **1983**, *4*, 294–301.

[10] Tomasi, J.; Mennucci, B.; Cammi, R. *Chem. Rev.* **2005**, *105*, 2999–3094.

[11] Scalmani, G.; Frisch, M. J. *J. Chem. Phys.* **2010**, *132*, 114110.

[12] Ahlrichs, R.; Bär, M.; Häser, M.; Horn, H.; Kölmel, C. *Chem. Phys. Lett.* **1989**, *162*, 165–169.

[13] *TURBOMOLE V6.3 2011, a development of University of Karlsruhe and Forschungszentrum Karlsruhe GmbH, 1989-2007, TURBOMOLE GmbH, since 2007; available from* `http://www.turbomole.com`.

[14] Singh, U. C.; Kollman, P. A. *J. Comput. Chem.* **1984**, *5*, 129–145.

[15] Klamt, A.; Schüürmann, G. *J. Chem. Soc. Perk. Trans. 2* **1993**, 799–805.

[16] Dunning, T. H. *J. Chem. Phys.* **1989**, *90*, 1007–1023.

[17] Woon, D. E.; Dunning, T. H. *J. Chem. Phys.* **1995**, *103*, 4572–4585.

[18] Balabanov, N. B.; Peterson, K. A. *J. Chem. Phys.* **2005**, *123*, 064107.

[19] Hiromoto, T.; Warkentin, E.; Moll, J.; Ermler, U.; Shima, S. *Angew. Chem. Int. Ed.* **2009**, *48*, 6457–6460.

[20] Sousa da Silva, A.; Vranken, W. *BMC Research Notes* **2012**, *5*, 367.

[21] Wang, J.; Wang, W.; Kollman, P. A.; Case, D. A. *J. Mol. Graph. Model.* **2006**, *25*, 247–260.

[22] Case, D. A.; Darden, T. A.; III, T. E. C.; Simmerling, C. L.; Wang, J.; Duke, R. E.; Luo, R.; Walker, R. C.; Zhang, W.; Merz, K. M.; Roberts, B.; Hayik, S.; Roitberg, A.; Seabra, G.; Swails, J.; Götz, A. W.; Kolossváry, I.; Wong, K. F.; Paesani, F.; Vanicek, J.; Wolf, R. M.; Liu, J.; Wu, X.; Brozell, S. R.; Steinbrecher, T.; Gohlke, H.; Cai, Q.; Ye, X.; Wang, J.; Hsieh, M.-J.; Cui, G.; Roe, D. R.; Mathews, D. H.; Seetin, M. G.; Salomon-Ferrer, R.; Sagui, C.; Babin, V.; Luchko, T.; Gusarov, S.; Kovalenko, A.; Kollman, P. A. *AMBER13*, University of California, San Francisco 2012.

[23] Carvalho, A. T. P.; Swart, M. *J. Chem. Inf. Model.* **2014**, *54*, 613–620.

[24] Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926–935.

[25] Hiromoto, T.; Ataka, K.; Pilak, O.; Vogt, S.; Stagni, M. S.; Meyer-Klaucke, W.; Warkentin, E.; Thauer, R. K.; Shima, S.; Ermler, U. *FEBS Lett.* **2009**, *583*, 585–590.

[26] DeLano, W. L. *The PyMOL Molecular Graphics System, Version 1.6*, 2002.

[27] Word, J. M.; Lovell, S. C.; Richardson, J. S.; Richardson, D. C. *J. Mol. Biol.* **1999**, *285*, 1735–1747.

[28] Olsson, M. H. M.; Søndergaard, C. R.; Rostkowski, M.; Jensen, J. H. *J. Chem. Theory Comput.* **2011**, *7*, 525–537.

[29] Pilak, O.; Mamat, B.; Vogt, S.; Hagemeier, C. H.; Thauer, R. K.; Shima, S.; Vonrhein, C.; Warkentin, E.; Ermler, U. *J. Mol. Biol.* **2006**, *358*, 798–809.

[30] Pettersen, E. F.; Goddard, T. D.; Huang, C. C.; Couch, G. S.; Greenblatt, D. M.; Meng, E. C.; Ferrin, T. E. *J. Comput. Chem.* **2004**, *25*, 1605–1612.

[31] Berendsen, H. J. C.; van der Spoel, D.; van Drunen, R. *Comput. Phys. Commun.* **1995**, *91*, 43–56.

[32] Van Der Spoel, D.; Lindahl, E.; Hess, B.; Groenhof, G.; Mark, A. E.; Berendsen, H. J. C. *J. Comput. Chem.* **2005**, *26*, 1701–1718.

[33] Hess, B.; Kutzner, C.; van der Spoel, D.; Lindahl, E. *J. Chem. Theor. Comput.* **2008**, *4*, 435–447.

[34] Pronk, S.; Pll, S.; Schulz, R.; Larsson, P.; Bjelkmar, P.; Apostolov, R.; Shirts, M. R.; Smith, J. C.; Kasson, P. M.; van der Spoel, D.; Hess, B.; Lindahl, E. *Bioinformatics* **2013**, *29*, 845–854.

[35] Bussi, G.; Donadio, D.; Parrinello, M. *J. Chem. Phys.* **2007**, *126*, 014101.

[36] Senn, H. M.; Thiel, W. *Angew. Chem. Int. Ed.* **2009**, *48*, 1198–1229.

[37] Humphrey, W.; Dalke, A.; Schulten, K. *J. Mol. Graphics* **1996**, *14*, 33–38.

[38] *ChemShell, a Computational Chemistry Shell*, see www.chemshell.org.

[39] Sherwood, P.; de Vries, A. H.; Guest, M. F.; Schreckenbach, G.; Catlow, C. A.; French, S. A.; Sokol, A. A.; Bromley, S. T.; Thiel, W.; Turner, A. J.; Billeter, S.; Terstegen, F.; Thiel, S.; Kendrick, J.; Rogers, S. C.; Casci, J.; Watson, M.; King, F.; Karlsen, E.; Sjøvoll, M.; Fahmi, A.; Schäfer, A.; Lennartz, C. *Theochem* **2003**, *632*, 1–28.

[40] Metz, S.; Kästner, J.; Sokol, A. A.; Keal, T. W.; Sherwood, P. *WIRS Comput. Mol. Sci.* **2014**, *4*, 101–110.

[41] *TURBOMOLE V6.5 2013, a development of University of Karlsruhe and Forschungszentrum Karlsruhe GmbH, 1989-2007, TURBOMOLE GmbH, since 2007; available from* `http://www.turbomole.com`.

[42] Billeter, S. R.; Turner, A. J.; Thiel, W. *Phys. Chem. Chem. Phys.* **2000**, *2*, 2177–2186.

[43] Kästner, J.; Thiel, S.; Senn, H. M.; Sherwood, P.; Thiel, W. *J. Chem. Theor. Comput.* **2007**, *3*, 1064–1072.

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

[45] Becke, A. D. *Phys. Rev. A* **1988**, *38*, 3098–3010.

[46] Tao, J.; Perdew, J. P.; Staroverov, V. N.; Scuseria, G. E. *Phys. Rev. Lett.* **2003**, *91*, 146401.

[47] Grimme, S.; Antony, J.; Ehrlich, S.; Krieg, H. *J. Chem. Phys.* **2010**, *132*, 154104.

[48] Weigend, F.; Ahlrichs, R. *Phys. Chem. Chem. Phys.* **2005**, *7*, 3297–3305.

[49] Schäfer, A.; Horn, H.; Ahlrichs, R. *J. Chem. Phys.* **1992**, *97*, 2571.

[50] Weigend, F. *Phys. Chem. Chem. Phys.* **2006**, *8*, 1057–1065.

[51] Eichkorn, K.; Weigend, F.; Treutler, O.; Ahlrichs, R. *Theor. Chem. Acc.* **1997**, *97*, 119–124.

[52] Sherwood, P.; de Vries, A. H.; Collins, S. J.; Greatbanks, S. P.; Burton, N. A.; Vincent, M. A.; Hillier, I. H. *Faraday Discuss.* **1997**, *106*, 79–92.

[53] König, P. H.; Hoffmann, M.; Frauenheim, T.; Cui, Q. *J. Phys. Chem. B* **2005**, *109*, 9082–9095.

[54] Lin, H.; Truhlar, D. G. *J. Phys. Chem. A* **2005**, *109*, 3991–4004.