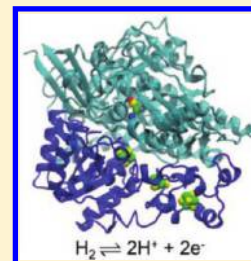


Proton Transport Pathways in [NiFe]-Hydrogenase

Isaiah Sumner and Gregory A. Voth*

Department of Chemistry, Institute for Biophysical Dynamics, James Franck Institute and Computation Institute, University of Chicago, 5735 South Ellis Avenue, Chicago, Illinois 60637, United States

ABSTRACT: Hydrogenases reversibly catalyze the production of molecular hydrogen. Current interest in these enzymes is focused on understanding the catalysis, since this may prove useful for hydrogen-based fuel cell and photosynthetic hydrogen production cell technologies. A key step in the hydrogenase catalytic cycle and the focus of this work is proton transport (PT) to and from the active site. The PT mechanism of the enzyme is studied using reactive molecular dynamics simulations of the full protein and the excess proton transfers via the multistate empirical valence bond (MS-EVB) method. Pathways connecting the bulk and the active site are located that suggest possible participation by several protonatable residues. PT free energy surfaces are calculated to differentiate the pathways.



1. INTRODUCTION

The hydrogenases are a family of biological redox enzymes that can reversibly catalyze the reduction of protons and the oxidation of molecular hydrogen, i.e.,



While these enzymes evolved in a wide variety of microorganisms to facilitate the use of H_2 as an energy source or an electron sink (H_2 production), current interest in these enzymes is focused on understanding the catalysis in order to develop hydrogen-based fuel cells and photosynthetic cells for hydrogen production.^{1,2}

To date, the hydrogenases are classified into three phylogenetically distinct classes that are determined by the metals incorporated into their functional cores, namely, the [NiFe]-, [FeFe]- and [Fe]-only hydrogenases.^{1–5} Both the [NiFe]- and [FeFe]-hydrogenases utilize only hydrogen as the substrate or product, whereas the [Fe]-only hydrogenases do not catalyze the reaction seen in eq 1 but couple the uptake of H_2 with enyltetrahydromethanopterin reduction.^{1,6,7} [NiFe]-hydrogenase is the target of this study.

[NiFe]-hydrogenases are composed of a large (L) and small (S) subunit of ~60 and ~28 kDa each,⁸ as shown in Figure 1. The active site is composed of a Ni atom bound to four cysteine residues, two of which are terminal to the Ni atom and two of which form a bridge to the active site Fe atom. The Fe atom also has two CN^- and one CO ligands. In addition to the active site, [NiFe]-hydrogenases also contain cubane, Fe–S metal clusters that are believed to play a central role in the electron transport chain. The enzyme has two [4Fe4S] clusters, which are proximal and distal to the active site, and one [3Fe4S] cluster, which is mesial to the active site. These clusters are anchored to the protein by cysteine residues coordinated one to each Fe atom.

In order to successfully catalyze the reaction in eq 1, [NiFe]-hydrogenases depend on substrate access to the active site, hydrogen heterolytic cleavage, proton transport (PT) to the solution, and external electron transfer. Generally speaking, they utilize three distinct types of channels/paths to accomplish

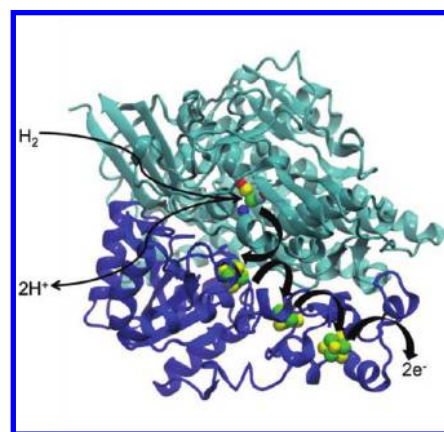


Figure 1. [NiFe]-hydrogenase (PDB 1H2R) with the electron, proton, and molecular hydrogen pathways shown. The large (L) subunit is colored in light blue, whereas the small (S) subunit is dark blue.

these tasks: two that carry H_2 or protons to and from the catalytic active site and a third that allows for electron flow. These are depicted schematically in Figure 1.

The exact nature of the catalysis once H_2 gains access to the active site is an area of intense research.^{8–20} However, there is some consensus on a few salient points. First, the Ni is generally thought to be the H_2 binding site. Second, H_2 is cleaved heterolytically and a Ni-terminal cysteineate (C546^L in the system studied here) is the reactive base (H^+ acceptor). The particular cycle assumed in this work is shown in Figure 2. The top portion of the cycle involves H_2 cleavage, followed by H^+/e^- loss and a hydride transfer. The cycle is renewed after a second H^+/e^- loss. The bottom half of the cycle includes an oxidative addition step of H_2 , which is unique to this cycle.¹⁷ Since the present work is focused on using reactive molecular dynamics (MD) simulations to gain insight into the PT

Received: September 3, 2011

Revised: January 9, 2012

Published: February 7, 2012

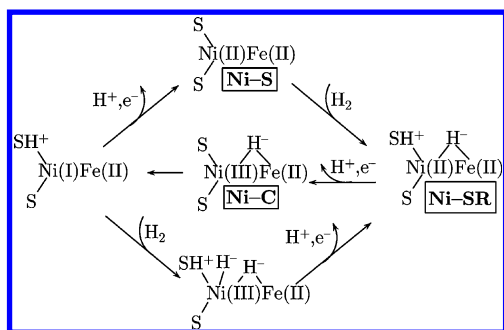


Figure 2. Catalytic cycle as proposed in ref 17. The named intermediates are the geometries used in this publication. At each step, the particular hydrogen moieties (i.e., hydride or proton) found in the active site are noted.

mechanism in [NiFe]-hydrogenase, the relevant portions of the cycle are the states following proton loss. Due to the oxidative addition step, the Ni-SR state is included in this study, whereas other cycles exclude this possibility.

Despite the focus on the hydrogenase catalytic cycle, work on PT pathways in [NiFe]-hydrogenase remains scant.^{8,15,16,21–23} Proposed PT pathways are mostly based on hydrogen bonding networks, gleaned from crystal structures, that lead from the active site to the enzyme surface.^{8,21} There have also been some H^+/D^+ isotope exchange experiments that implicate a glutamic acid residue in the vicinity of the terminal cysteineate mentioned previously as the initial site of the PT pathway.²² In these studies, this glutamic acid was mutated to glutamine and all PT ceased, although there was ortho-para H_2 conversion indicating the continued access to the active site and cleavage of H_2 .

In addition to the experimental studies, there have also been some computational work on the PT process in [NiFe]-hydrogenase.^{15,16} Clearly, the computational study of these events represents a significant challenge. In particular, it is known that PT occurs via the Grotthuss mechanism wherein the electronic charge defect associated with the proton can be delocalized across several molecular species.^{24,25} This implies that excess protons “hop” not only along water wires embedded in the enzyme via Zundel-type ($H_2O-H^+-OH_2$) water clusters, but protonatable amino acids like glutamic acids may also provide proton shuttling sites. The accurate representation of this mechanism is not possible with traditional MD methods, which fix the bonding topology throughout the simulation. Therefore, the authors of ref 15 utilized Poisson–Boltzmann and Monte Carlo simulations as well as a distance-based network analysis to find possible groups that participate in the PT. These simulations kept the protein fixed at its crystal structure geometry but possibly identified several viable PT pathways. In contrast, the authors of ref 16 used minimum energy pathways calculated using QM/MM-hybrid electronic structure techniques, wherein the shuttling proton as well as possible amino acid shuttling sites were included in the QM subsystem, which was represented using electronic density functional theory. The authors used both the crystal structure as well as structures taken from MD simulations. The pathways in this latter study were chosen on the basis of geometric considerations. The results of both papers singled out similar PT networks as the most likely to occur. The present work instead utilizes multi-state empirical valence bond (MS-EVB)^{26–30} reactive MD simulations with an enhanced path sampling methodology to study the PT process.

MS-EVB is a technique that dynamically allows chemical bonds to break and form during MD simulations and can be parametrized for any generic chemical reaction,³¹ although here it exclusively allows proton transfer between water molecules. The PT pathway methodology is coupled with metadynamics,^{32–35} a technique that can be used to find complex, nonlinear minimum free-energy pathways. These two schemes are used to identify possible proton hopping channels which are then characterized by their free energy surfaces using MS-EVB/umbrella sampling.³⁵ These calculations also serve to help identify the participation by key amino acid residues. The use of these techniques to generate unbiased PT paths, i.e., paths that are not chosen *a priori*, in a dynamical protein environment, as well as the calculation of PT free-energy surfaces distinguishes the present work from previously published studies.

2. METHODOLOGY

A. Simulation Setup. The [NiFe]-hydrogenase protein was built from the X-ray crystal structure (PDB 1H2R) of the periplasmic enzyme found in *Desulfovibrio vulgaris* Miyazaki F.³⁶ Protonation states for all but two of the protonatable amino acids were determined from pK_a calculations using the PROPKA method^{37,38} assuming neutral pH. Experiments have shown that the optimal environmental pH with regards to hydrogenase function in *Desulfovibrio* species resides between 6.0 and 8.0.^{39,40} The residue E34^L was placed in its neutral state based on crystal structure data³⁶ taken at pH 7.4 which seems to indicate the presence of a hydrogen bond to C546^{L41} and H88^L was placed in its doubly protonated, positive state based on theoretical considerations.⁴¹ In order to explore any dependence of the PT pathway on the active site and FeS cluster state, several geometries corresponding to different points in the catalytic cycle and electron transfer path were generated. The Ni-S, Ni-SR, and Ni-C active site geometries were chosen, as described previously. The FeS cluster oxidation states used are shown in Table 1 and were chosen based on refs 8, 13, and 15.

Table 1. Metal Cluster (FeS and Active Site) Oxidation States Considered

| FeS-Cluster | Nia-C(PR) | Nia-C(DR) | Nia-S(MO) | Nia-S(MR) | Nia-SR |
|-------------|-----------|-----------|-----------|-----------|--------|
| Proximal | Red | Ox | Ox | Ox | Red |
| Mesial | Red | Red | Ox | Red | Red |
| Distal | Ox | Red | Ox | Ox | Red |

In order to derive charge parameters for the FeS and active site clusters, each initial geometry was minimized using the ONIOM^{42,43} QM/MM scheme at the b3lyp/lanl2dz:am-ber96⁴⁴ level of theory and the protein alpha-carbons were fixed to their crystal-structure positions. The high-level (QM), model system contained the FeS clusters and the active site clusters as well as the R-group portion of the coordinated cysteineate, whereas the low-level (MM), real system contained the entire enzyme including crystallographic waters. The QM/MM boundary was placed between the alpha- and beta-carbons of the cysteineate, and the beta-carbon valences were saturated with a hydrogen link-atom. The Gaussian 03⁴⁵ suite of programs was utilized for these calculations. Furthermore, to ensure that each metal cluster had the correct number of electrons, the geometry optimizations were iterated over the FeS and active site clusters. In other words, the targeted cluster was optimized as part of the high-level system, while the other

clusters were held fixed as part of the low-level system. The remainder of the enzyme was allowed to relax. The geometries of the metal clusters were considered optimized once a loose, geometry-based convergence was reached (rmsd change of <0.01 Å after a full iteration).

The FeS and active site clusters, including the R-group of the coordinated cysteine residues, were then excised from the enzyme and partial atomic charges were calculated using the RESP⁴⁶ method at the b3lyp/lanl2dz level of theory. The beta-carbon valences were saturated with a methyl group placed at the alpha-carbon position, and the methyl hydrogen atoms were optimized. The methyl group charge was restrained to mimic the backbone charge of a standard, deprotonated AMBER CYS residue (CYM). The Lennard-Jones terms for the Ni, Fe, and S atoms were taken from the literature.¹⁷ Finally, the metal clusters including the ligated cysteine residues were held fixed to their ONIOM-optimized geometries during the dynamics simulations using harmonic restraining potentials with a force constant of 1000 kcal mol⁻¹ Å⁻².

As the last setup step, the system, including all 557 crystallographic waters, was solvated in a rectangular box with ~14000 modified TIP3P⁴⁷ waters and counterions were added to neutralize the system. Each structure was then optimized and heated to 298.15 K using the AMBER MD⁴⁸ program. Finally, the structures were equilibrated for 200 ps in a constant NPT ensemble at 1.0 atm and 298.15 K using a Nose–Hoover thermostat followed by 2 ns of equilibration in a Nose–Hoover constant NVT ensemble at 298.15 K. All equilibrations were performed using the LAMMPS⁴⁹ MD software package.

B. MS-EVB. The MS-EVB methodology is described in detail elsewhere^{25–30} and will only be described briefly here. MS-EVB is a reactive molecular-mechanics scheme that allows bond topologies to change (chemical reactions) during a simulation. Each bond topology, or diabatic state $|i\rangle$, is described by a classical force field (AMBER parm99⁵⁰ in this paper) and is a diagonal element of the MS-EVB Hamiltonian

$$H^{\text{EVB}}(\vec{r}) = \sum_{ij} |i\rangle h_{ij}(\vec{r}) \langle j| \quad (2)$$

where \vec{r} represents the nuclear degrees of freedom. At each time-step, this matrix is formed from the available states and diagonalized. The system is then propagated on the ground state, or adiabatic surface, which is a linear combination of all the possible bonding topologies. The off-diagonal elements describe the coupling between the bonding topologies, and the functional form of $h_{ij}(\vec{r})$ is designed and parametrized so that the MS-EVB scheme reproduces experimental properties such as pK_a 's³¹ and *ab initio* potential surfaces.³⁰

The MS-EVB model also provides a convenient coordinate to track the excess charge defect. This variable, the center of excess charge (CEC), is the preferred method of following the protonic electronic charge defect because the excess proton may change identity during the transport process due to the Grotthuss mechanism. The CEC is defined as the sum of the center of charge of each EVB state weighted by the MS-EVB coefficient, i.e.,

$$\vec{r}_{\text{CEC}} = \sum_i^{N_{\text{EVB}}} c_i^2(\vec{r}) \vec{r}_i^{\text{COC}} \quad (3)$$

where

$$\vec{r}_i^{\text{COC}} = \frac{\sum_{k \in i} |q_k| \vec{r}_k}{\sum_{k \in i} |q_k|} \quad (4)$$

The term $c_i^2(\vec{r})$ is the contribution of diabatic state i to the ground state potential surface, the sum is over all atoms, and q_k is the k th partial charge as defined in the i th diabatic state. All MS-EVB simulations were performed using a modified version of the LAMMPS⁴⁹ package.

C. Free Energy Calculation. The free energy calculations in this paper were performed using a combination³⁵ of metadynamics (MTD)^{32–34} and umbrella sampling (US).^{51–54} In MTD, a history-dependent biasing potential is used to lift a collective variable (CV) out of potential wells so that the system can explore a fuller range of the CV. The potential that is generally used is Gaussian in form, i.e.,

$$V_{\text{MTD}}(s, t) = \sum_{t_i} V_0 \exp \left[-\frac{|s - s(t_i)|^2}{2w^2} \right] \quad (6)$$

where V_0 is the height of the Gaussian, w is the width, s is the CV, and t_i represent the (regularly spaced) time intervals at which the Gaussians are added. MTD not only increases the sampling of high energy regions of the CV, but it can also calculate free energy surfaces.^{32,55–59} However, in this publication, MTD is a path-finding tool and functions to locate the PT pathways in [NiFe]-hydrogenase. The free energy of PT is then calculated using US.³⁵

The free-energy difference along the pathway is calculated by discretizing the following expression:^{35,53}

$$F(s(\alpha)) - F(s(0)) = \int_0^\alpha \sum_i \frac{ds_i(\alpha')}{d\alpha'} \frac{\partial F(s_i(\alpha'))}{\partial s_i} d\alpha' \quad (7)$$

where the sum is over N paths ($N = 1$ in this work) and $\alpha \in [0, 1]$ is used to parametrize the pathway $\{s_i\}$ which is defined here by the CV. The mean force in each umbrella window can be used to calculate the derivative of the free energy

$$\frac{\partial F(\beta_0)}{\partial s_j} \approx \frac{k}{T} \int_0^T [s_j(\beta_0) - s_j(\beta(t))] dt \quad (8)$$

where $s_j(\beta_0)$ is the equilibrium value, $s_j(\beta(t))$ is the instantaneous value, and k is the force constant of the umbrella window located at $\alpha' = \beta_0$. The advantage of this scheme³⁵ over other popular methods like WHAM⁵¹ is that good overlap between umbrella windows is not crucial, which allows for arbitrarily curved paths. This method has been successfully used to calculate the free energy of PT through a water wire in a carbon nanotube and a channel protein,³⁵ but we use it here in a globular protein. This represents a more challenging application, since the path is not as well-defined or constrained. US and MTD calculations were performed by interfacing the PLUMED⁶⁰ package with LAMMPS.⁴⁹

3. RESULTS

A. Metadynamics Pathways. For each FeS oxidation state along the catalytic cycle, MTD was utilized in order to find a minimum energy pathway connecting the active site and the bulk. One advantage of using MTD is that it can locate the pathways in an unbiased fashion; i.e., very little about the nature

of these pathways is assumed *a priori*. The chosen CV, represented as $s(\alpha)$ in eqs 6–8, was the distance between the center of mass of the active site and the CEC (see eq 3). Each simulation started with a hydronium near the residue E34^L, which is the assumed initial site in the PT chain,²² and was then re-equilibrated for an additional 20 ps at 298.15 K. Since MTD was used as a path finding and not a free-energy method, the MTD parameters were chosen such that a pathway could be found with minimal computational effort. Thus, a Gaussian with $V_0 = 0.5$ kcal/mol and $w = 0.5$ Å was deposited every 0.5 ps (see eq 6). This technique located three pathways.

The first pathway, referred to as MTD pathway 1 and shown in Figure 3, was found in the Nia-C(DR) and Nia-S (MR and

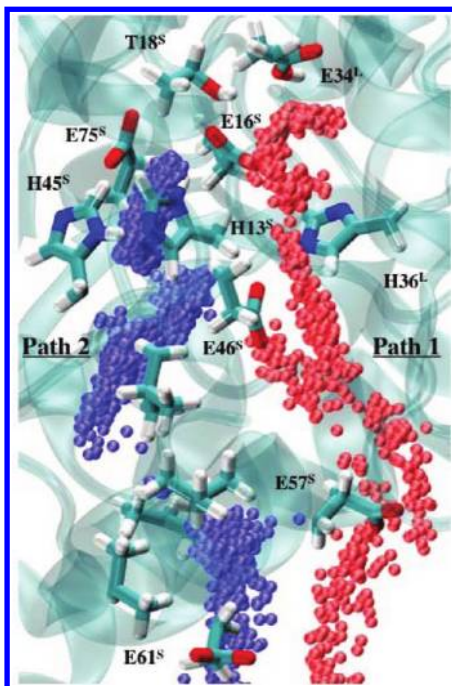


Figure 3. Shown here are MTD pathways 1 (red) and 2 (blue). The spheres that constitute the pathways represent dynamical positions of the excess proton CEC. Several important residues, including the hydrophobic corridor (unlabeled), are shown. The enzyme interior is located toward the top of the figure, and the exterior is at the bottom.

MO states) active site geometries. This pathway is nearly linear and contains hydrophilic residues, several of which are protonatable. Notable among these residues are H13^S, E16^S, T18^S, H36^L, E46^S, E57^S, and E75^S. The residues E16^S, T18^S, and E75^S (or E16^S, T18^S, and E73^S in *D. gigas* numbering) are conserved in all *Desulfovibrio* species.⁸ This pathway was also suggested in previous experimental and computational works.^{8,15,16,21}

A second pathway, referred to as MTD pathway 2 and also seen in Figure 3, was found in the Nia-C(PR) geometry. After re-equilibration, the hydronium leaves E34^L and is stabilized in a pocket of waters behind T18^S, near the active site. Therefore, the excess proton CEC starts in a skewed position in the MTD calculation. Additionally, on its way to the enzyme surface, the CEC hops through a hydrophobic region formed by valine, leucine, and isoleucine residues. Finally, repeated MTD simulations for this structure with the hydronium initially located closer to E34^L resulted in a PT corridor similar to MTD pathway 1.

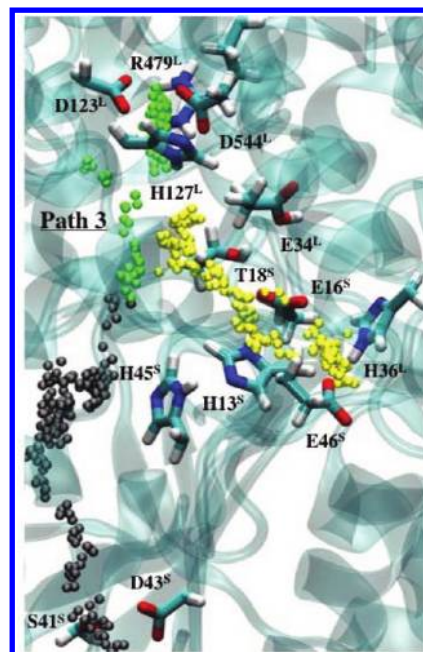


Figure 4. This is MTD pathway 3. For reference, note the placement of residues H13^S, E16^S, T18^S, H36^L, E34^L, and E46^S. Yellow, green, and gray spheres represent the different portions of this path.

A final pathway, referred to as MTD pathway 3, was found for the Nia-SR geometry, and is shown in Figure 4. This pathway is broken into three parts, and it starts along the first portion of MTD pathway 1 (yellow path in Figure 4). It reaches a dead end formed by a hydrogen bond between H13^S and E46^S that does not allow the proton to pass. However, it must be emphasized that these particular MS-EVB simulations do not allow proton transfer to amino acids. If these transfers were allowed, such amino acid networks would not necessarily block the proton and so such studies are targets for future research. Next, the CEC finds another pathway framed in particular by D544^L, H127^L, D123^L, and R479^L (D528^L, H111^L, D107^L, and R463^L *D. gigas* numbering). This portion of the path (shown in green) is also similar to another pathway suggested experimentally and computationally^{8,15} and contains the conserved residues D544^L and R479^L.⁸ Finally, it reaches the bulk via the pathway characterized by H45^S, D43^S, and S41^S. This particular portion of the pathway (shown in gray) has not been explicitly characterized before, although portions have been considered.^{8,15} It is also important to note that an additional MTD simulation for the Nia-SR structure resulted in the CEC traveling along MTD pathway 1.

In summary, the MTD trajectories revealed a network of protonatable and hydrophilic residues that can support several PT chains from the active site to the bulk. Of these, MTD pathway 1 seems to be the most likely. This is supported by the fact that the simulations that located different paths did so due to rare events such as H-bond blocking and nonideal initial structures. In fact, repeated simulations always resulted in MTD pathway 1. Finally, these calculations did not find pathways making use of Mg²⁺ and its water ligands. This structure is found in the enzyme crystal structure and has been proposed to contribute to the PT paths.^{8,15,16,23} Its participation was also questioned in refs 15 and 16.

B. Free Energy Surfaces. After the PT pathways were located with MTD, the free energy surfaces (FES's) were

calculated with the US method outlined in section 2C. These calculations can serve to distinguish the pathways in a more quantitative way. Each free energy calculation consisted of umbrella windows spaced 0.25 Å apart, each of which was sampled for 600–1000 ps (including 100–200 ps equilibrations). The US windows were represented using cylindrical coordinates, and a force constant of 30 kcal mol⁻¹ Å⁻² was applied to the CEC position only in the axial direction. Thus, the CEC was unhindered by the umbrella force in the perpendicular, radial directions, which had two important consequences. First, the CEC could take any pathway parallel to the axial placement of the umbrella windows. Therefore, the US simulations, like the MTD runs, were “path-finding”. This property becomes apparent in the following sections. Second, the reaction coordinate, $s(\alpha)$ in eq 7, was an arbitrary, curvilinear path parametrized by α , the distance between the active site center of mass and the CEC.

The US windows used to calculate the FES along MTD pathway 1 were placed along a straight line located approximately down the center of the MTD pathway. This path is referred to as US pathway 1. The other, more complex MTD channels were also simply approximated. First, dead ends were not considered so the second portion of MTD pathway 3 was not included. (The first portion overlaps with MTD pathway 1.) Second, since MTD pathways 2 and 3 are similar, one set of windows represented both of them. Thus, the approximate pathway, referred to as US pathway 2, contains two straight line segments, the first of which leads from E34^L to the water pocket behind T18^S. The second segment goes to the bulk and is centered roughly between MTD pathway 2 and the last portion of pathway 3.

It is important to note that these MS-EVB simulations, like the MTD trajectories described previously, did not allow proton transfers to amino acids (e.g., shuttling of protons through them). Thus, the FES's produced by this method are missing possibly significant information and the calculated free energies of H⁺ extraction are not necessarily expected to be completely accurate. However, the FES's shown here can serve as a baseline and will also contain information about residues likely to be protonated as well as their possible level of participation in the PT process. Furthermore, since the same errors occur in all FES's (nontitratable residues), qualitative comparisons among the surfaces are expected to be valid.

B.1. US Pathway 1. The real paths taken by the CEC when constrained along the $s(\alpha)$ representation of MTD pathway 1 are consistent across all geometries and oxidation states studied. Figure 5 displays FES's for proton transfer in the Nia-C geometry for both the DR and PR oxidation states. It is clear that the main feature of these FES's is the uphill nature, although one surface contains a significant, initial well. In the DR oxidation state, the initial portion of the FES describes the CEC leaving the influence of E16^S and E34^L. After leaving the initial portion of the pathway, the slope of the FES decreases, since the CEC is affected now by E46^S. Upon its release from E46^S, it travels along the surface due to four residues, D28^L, D54^S, E57^S, and E61^S, which are the last protonatable residues that affect the CEC before it reaches the bulk.

The CEC takes a similar path in the PR oxidation state. After leaving a stable pocket formed by E16^S and E75^S, the CEC is then carried by E46^S to the protein surface (E61^S) before entering the bulk. Despite the similarities in the paths for the DR and PR oxidation states, the initial well is much deeper and the barrier is much higher for the PR state. This difference most

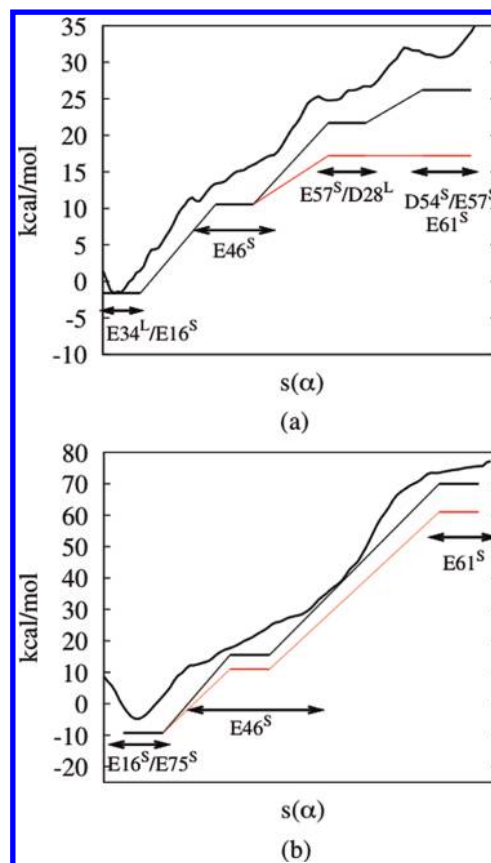


Figure 5. Shown here are the FES's for US pathway 1 in the (a) Nia-C(DR) and (b) Nia-C(PR) states. The rescaled curves estimate the effect titratable residues may have on the calculated FES (black curve) assuming an isolated effect (black lines) and a cooperative effect (red lines). The x-axis is a generic reaction coordinate that leads from the enzyme interior (left) to exterior (right), and the residue labels indicate the protonatable residues that most affect the CEC. Note the differences in the y-axis scale between parts a and b. The average bin errors are (a) ± 1.2 kcal/mol and (b) ± 0.8 kcal/mol.

likely comes from the fact that there are more protonatable residues present along the PT pathway in the DR state, which helps to lower the barrier.

The approximate effect that titratable residues may have on the FES can be seen by examining the straight line segments accompanying the free energy curves in Figure 5. On the basis of ref 31, the free energies of bond formation for glutamate and histidine residues are approximately -4.5 and -9.0 kcal mol⁻¹. Thus, the appropriate energy is added to the FES in each area influenced by a group of protonatable amino acid residues assuming the net effect of each grouping is equal to one residue, E34^L is neutral, and aspartate residues behave like glutamate. The corrections represented by the black lines assume this effect is local, whereas the red lines assume a cooperative effect; i.e., one protonatable residue lowers the entire energy curve, which is again lowered by the next protonatable residue. The actual correction will likely lie between these two limits. These corrections make a significant contribution to the DR curve and may lower the barrier to the bulk by ~ 10 kcal mol⁻¹, whereas the PR curve remains energetically very unfavorable despite the corrections. Again, this is because of the larger participation of protonatable residues along the DR pathway.

Figure 6 shows the FES for the proton channel out of the Nia-S geometries (MO and MR oxidation states). The residues

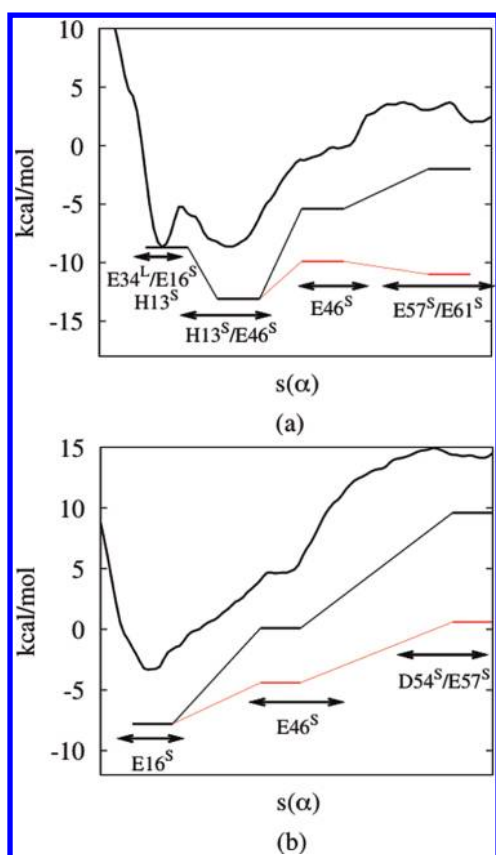


Figure 6. Shown here are the FES's for US pathway 1 in the (a) Nia-S(MR) and (b) Nia-S(MO) states. The rescaled curves estimate the effect titratable residues may have on the calculated FES (black curve) assuming an isolated effect (black lines) and a cooperative effect (red lines). The x -axis is a generic reaction coordinate that leads from the enzyme interior (left) to exterior (right), and the residue labels indicate the protonatable residues that most affect the CEC in a given region. Note the differences in the y -axis scale between parts a and b. The average bin errors are (a) ± 0.5 kcal/mol and (b) ± 0.4 kcal/mol.

that may participate in this path are similar to those seen in the Nia-C geometries, although there is participation by H13^S in the Nia-S(MR) geometry. However, there are marked differences in the FES. The biggest difference is the relatively small barrier required for the CEC to leave the enzyme. Second, there are two local minima in the MR state. The first represents the CEC being trapped between E16^S, E75^S, and H13^S, and the second is the CEC being trapped between E46^S and H13^S. In the MO state, the single energy well corresponds to the attractive interaction between the CEC and E16^S. The CEC then moves on to E46^S in both MO and MR states, after which it leaves the enzyme and moves to the bulk via D54^S, E57^S (MO) and E57^S, E61^S (MR) on the enzyme surface. Clearly, the main difference between these two states is the participation of H13^S, which seems to lower the barrier of the MR state. The energy corrections for the Ni-S FES's serve to exaggerate this difference and also significantly lower the free energy of PT to the bulk environment.

Finally, Figure 7 shows the FES for the Nia-SR geometry. The barrier for the CEC to leave the enzyme is also comparatively lower here. However, there is a sharp well corresponding to the CEC stabilized by E16^S and E75^S. The CEC is then carried through the enzyme mainly by E46^S, and it leaves the enzyme after interacting with D28^L, D54^S, E57^S, and

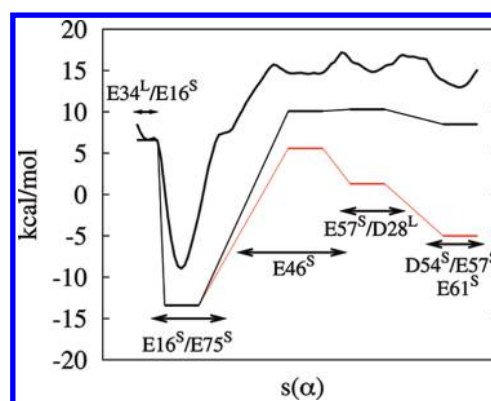


Figure 7. Shown here is the FES for US pathway 1 in the Nia-SR state. The rescaled curves estimate the effect titratable residues may have on the calculated FES (black curve) assuming an isolated effect (black lines) and a cooperative effect (red lines). The x -axis is a generic reaction coordinate that leads from the enzyme interior (left) to exterior (right), and the residue labels indicate the protonatable residues that most affect the CEC in a given region. The average bin error is ± 0.3 kcal/mol.

E61^S. Energy corrections for this pathway are significant and make the PT barrier similar to that seen in the corrected Nia-C(DR) curve.

Several statements can be made about US pathway 1. First, it becomes clear that the oxidation state of the FeS clusters as well as the active site have a significant impact on the free energy required for the CEC to leave the enzyme. Figure 5 illustrates this point by showing that the oxidation states of the proximal and distal FeS cluster can cause free energy changes of ~ 40 – 50 kcal mol⁻¹. Second, E46^S plays a critical role in relaying the CEC from near the active site to the bulk. This residue undergoes significant conformational changes in all pathways as it stabilizes the charge defect and connects the enzyme interior to the surface. Third, the largest factor in barrier height is the level of participation of protonatable residues along the pathway. Allowing PT to the titratable residues may potentially lower many of the barriers to transfer, as indicated by the corrections shown in Figures 5–7. Finally, with the error estimates, many of the free energies become comparable to the energies seen in ref 16, where energy barriers between 9.5–14 kcal mol⁻¹ were calculated for a similar pathway using QM/MM hybrid electronic structure methods.

B.2. US Pathway 2. The initial portion of US pathway 2 overlaps well with the initial portion of US pathway 1 for all geometries and oxidation states. Examining the point of divergence in the pathways reveals interesting insights. In particular, there is a distinct branch-point centered at H13^S. If H13^S is hydrogen-bonded to E75^S, the excess proton can leave via pathway 1. However, if the bond is broken (e.g., if E75^S H-bonds to a water instead), the CEC can take the second pathway. This is depicted in Figure 8, which shows both paths taken by the CEC in the Nia-C(DR) state. Furthermore, US pathway 2 is more varied and has a number of different possibilities, which becomes apparent after examining the differences between the DR and PR states.

The FES's for US pathway 2 in the Nia-C geometries are depicted in Figure 9. In the initial portion of the DR state, there is a well due to a stable structure in which the CEC is located between E16^S and E75^S. After leaving this group of hydrophilic residues, the CEC is affected by H45^S, which follows the CEC until it reaches D43^S. D43^S carries the CEC to near the surface,

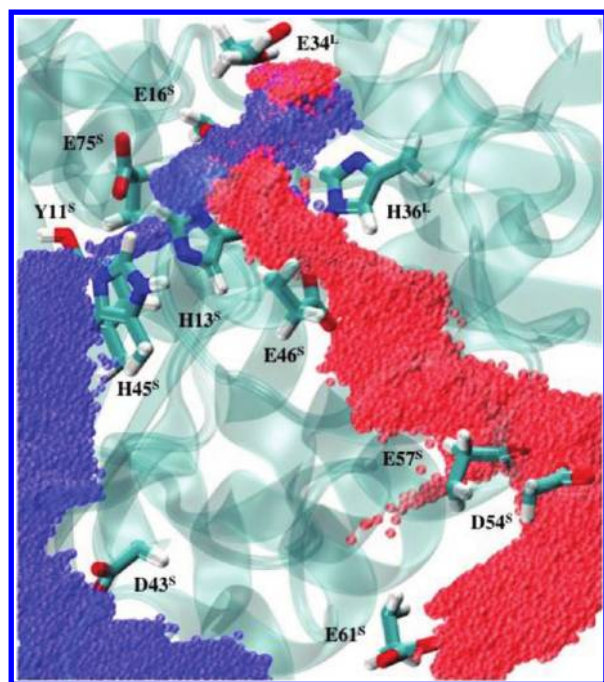


Figure 8. This figure shows US pathways 1 (red) and 2 (blue) in the Nia-C(DR) state. Several important residues including “branch-points” are shown. The enzyme interior is located toward the top of the figure, and the exterior is at the bottom.

after which the CEC enters the bulk via several pathways. In the PR state, the CEC is also initially stabilized by $E16^S$ and $E75^S$. In this pathway, $E75^S$, not $D43^S$ and $H45^S$, carries it through a large portion of the enzyme interior until it is released to the enzyme surface, where it interacts with $D28^L$, $D54^S$, $E57^S$, and $E61^S$.

The transport path taken in the DR state adheres more closely to MTD pathway 3, whereas the path taken in the PR state overlaps better with MTD pathway 2. Again, an amino acid branch-point regulates the difference between the pathways seen in the DR and PR states. If $Y11^S$ H-bonds to $E75^S$, the pathway seen in the DR state is preferred. However, if $Y11^S$ H-bonds to a water molecule, the CEC follows the PR state channel. Both paths are shown in Figure 10.

Next, it is clear upon comparison with the US pathway 1 counterparts seen in Figure 5 that US pathway 2 is not energetically favored. Also, pathway 2 in the PR state has a much larger barrier to the enzyme surface than the DR state. This difference originates not only from the change in electrostatic environment but also because the PT pathways are significantly changed. For example, one contribution to the extremely high PR state barrier is that the excess proton CEC interacts with $E75^S$ via a hydronium ion for much of its path through the enzyme interior. This means that this surface reflects the large free energy required to drag a hydronium ion through a hydrophobic portion of the protein. Protonating $E75^S$ would remove the hydronium in this region and should significantly lower the barrier. The CEC is more solvated in the DR pathway. It is also noteworthy that the CEC moves through the “hydrophobic corridor” in the PR state, as described in section B, which also contributes to the large barrier. Furthermore, corrections to account for PT to amino acids have a small impact on the barriers, since there are generally fewer of these residues.

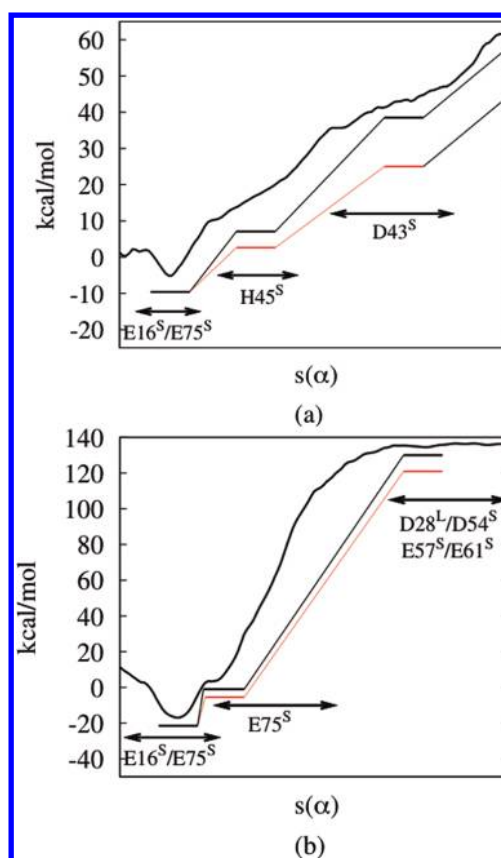


Figure 9. Shown here are the FES's for US pathway 2 in the (a) Nia-C(DR) and (b) Nia-C(PR) states. The rescaled curves estimate the effect titratable residues may have on the calculated FES (black curve) assuming an isolated effect (black lines) and a cooperative effect (red lines). The x-axis is a generic reaction coordinate that leads from the enzyme interior (left) to exterior (right), and the residue labels indicate the protonatable residues that most affect the CEC in a given region. Note the differences in the y-axis scale between parts a and b. The average bin errors are (a) ± 1.0 kcal/mol and (b) ± 0.6 kcal/mol.

Finally, Figure 11 displays the FES for pathway 2 in the Nia-S(MR) geometry. Like its pathway 1 counterpart (see Figure 7), the FES initially has double-well character, which reflects the excess proton CEC being trapped by $E16^S$, $E34^L$, and $H36^L$. This is similar to the previous pathways 2. The second well represents stabilization due to $E16^S$ and $E75^S$, which diverges from the previous paths. Next, the CEC is attracted to $E46^S$ for the remainder of the reaction coordinate, after which it reaches the bulk. Again, US pathway 2 is much higher in energy than the corresponding US pathway 1 and attempts to correct the energetics do not have a significant effect. Attempts to calculate the FES for US pathway 2 in the Nia-S(MO) and Nia-SR state resulted in the CEC adhering closely to pathway 1. These failed attempts as well as the very high energetics involved for other active site geometries all lead to the conclusion that US 2 pathways as explored in this publication are not likely to be utilized during PT to the bulk.

4. CONCLUSIONS

This work utilized the novel MTD/US technique³⁵ combined with MS-EVB²⁵ reactive molecular dynamics simulations to elucidate the proton transfer process in [NiFe]-hydrogenase. In total, three PT pathways were found via the MTD method^{32–35} which allow access to the enzyme active site. These simulations

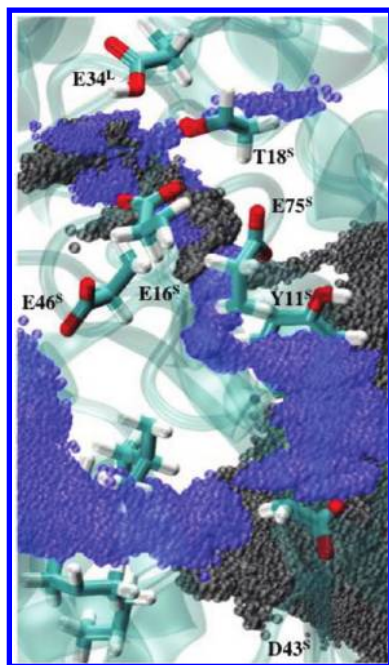


Figure 10. This figure depicts the two divergent US pathways 2 for Nia-C(DR) (gray spheres) and Nia-C(PR) (blue spheres). Note that the blue path goes through the hydrophobic corridor (unlabeled residues) and thus conforms more closely to MTD pathway 2, whereas the gray path corresponds more closely to MTD pathway 3. Both paths were based on the same $s(\alpha)$ representation. The enzyme interior is located toward the top of the figure, and the exterior is at the bottom.

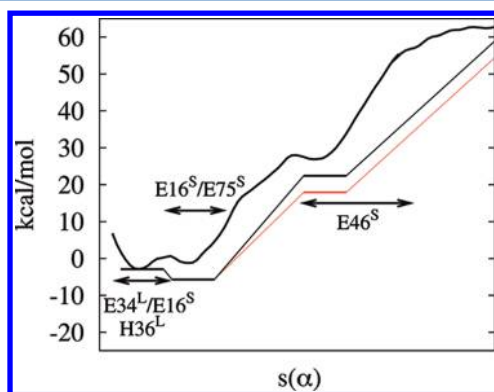


Figure 11. Shown here is the FES for US pathway 1 in the Nia-S(MR) state. The rescaled curves estimate the effect titratable residues may have on the calculated FES (black curve) assuming an isolated effect (black lines) and a cooperative effect (red lines). The x -axis is a generic reaction coordinate that leads from the enzyme interior (left) to exterior (right), and the residue labels indicate the protonatable residues that most affect the CEC in a given region. The average bin error is ± 0.6 kcal/mol.

resulted in a qualitative understanding of the PT process and found that one pathway in particular, characterized by E34^L, E16^S, and E46^S, was dominant (the red paths in Figures 3 and 8). This was based on the frequency with which this pathway was found in all active site geometries and oxidation states under consideration.

Furthermore, the US pathways, which were based on the MTD trajectories, and free energy surfaces elucidated the process further. It was found that the channels based on MTD pathway 1 were consistent for each state under investigation.

The excess proton CEC moved from a stable pocket formed by different combinations of E34^L, E16^S, E75^S, and H13^S to E46^S, which tracked the CEC through the enzyme interior to the surface where it was then stabilized by E57^S, D28^L, D54^S, and E61^S. The channels based on MTD pathways 2 and 3 were more varied across the different structures. The initial, stable regions of the surface as well as the residues on the protein surface overlapped with the previously described pathway. However, the CEC was stabilized by different set of residues as it moved through the interior, i.e., E46^S, E75^S, and H45^S.

The US calculations allow for several conclusions. First, pathway 1 is clearly favored energetically. This agrees well with previous work^{15,16} and with our qualitative MTD results. It is also important to emphasize the presence of key transport residues in all pathways, namely, E46^S, E75^S, and H45^S. These residues are far from static and undergo significant conformational changes to stabilize the CEC as it reaches the surface. Furthermore, these amino acids are attractive targets for mutagenesis experiments due to their seemingly active role in the PT pathways. Second, there are protonatable residues on the surface of the enzyme near the channel entry/exit. These residues represent the final energy barrier before the excess proton reaches the solution, and their participation is likely to significantly impact the PT energetics. Additionally, since these residues reside on the exterior of the hydrogenase, they may play critical roles in the reverse reaction (H_2 production), wherein protons need to enter the enzyme. Third, the oxidation states of the FeS clusters seem to impact the PT free energy, which is likely due to the change in the electrostatic environment. This is seen most dramatically in the Nia-C active site geometry (Figure 5), where changing the oxidation state increases the barrier by ~ 50 kcal mol⁻¹. The role of the FeS clusters in PT¹⁵ and in the catalysis⁶¹ has been previously discussed in the literature. Finally, the state of the active site can have an effect on the PT free energy. This may have an impact on the catalysis and proposed catalytic cycles, and PT should be taken into account when considering possible rates of hydrogenase catalysis. In fact, the calculations in this publication seem to indicate that the top half of the catalytic cycle in Figure 2 is favored over the bottom half if PT is the only consideration.

Moreover, it should be noted that important features of this process are missed when static approaches or approaches based on the crystal structure alone are utilized. As mentioned previously, crucial, protonatable residues are dynamic and change conformations in order to stabilize the excess proton CEC. Also, the fully solvated, equilibrated MD structure has many more water molecules in its interior which results in a more complex proton solvation dynamics. The simulations presented in this paper indicate that not only are Zundel-like structures present, but Eigen-like ($H_9O_4^+$) clusters also occur during PT. Naturally, the presence of more waters in the enzyme impacts not only the PT path but also its free energy.

Although the present work helps to further expand on the previous understanding of the hydrogenase PT process,^{15,16} there are opportunities for future research. For instance, the role the FeS clusters seem to play in the free energy of PT is still unclear. A detailed investigation into this effect is needed. Also, there are several residues that eventually should be included in a more complete MS-EVB Hamiltonian, which will allow them to become protonated and thus explicitly contribute to proton shuttling. This will certainly increase the accuracy of the FES calculations. This effect of residue protonation was

estimated and seen to bring down the free energy barriers in several of the surfaces calculated for US pathway 1 and brings some of them in closer agreement with the energetics seen in other theoretical publications.¹⁶ Furthermore, including these residues in the PT will likely lead to subtle changes to the actual PT pathway taken by the protonic charge defect. For example, amino acid residues would be able to carry the CEC with less presence of water, and thus, the CEC can move through desolvated and hydrophobic areas of the protein at a lower energetic cost. Also, stable, hydrogen-bonded amino acid structures will not necessarily impede the CEC. Finally, the role of the exterior residues in attracting protons could be elucidated with this technique.

AUTHOR INFORMATION

Corresponding Author

*E-mail: gavoth@uchicago.edu.

ACKNOWLEDGMENTS

This research was supported by the Air Force Office of Scientific Research (AFOSR grant FA9550-10-1-0142). Computational resources were provided by the DOD High Performance Computing Modernization Program at the Army Research Laboratory and Maui High Performance Computing DOD Supercomputing Resource Centers. We thank Drs. C. M. Maupin and Y. Zhang for helpful discussions.

REFERENCES

- (1) Vignais, P. M.; Billoud, B. *Chem. Rev.* **2007**, *107*, 4206–4272.
- (2) Lubitz, W.; Reijerse, E. J.; Messinger, J. *Energy Environ. Sci.* **2008**, *1*, 15–31.
- (3) Vignais, P. M.; Billoud, B.; Meyer, J. *FEMS Microbiol. Rev.* **2001**, *25*, 455–501.
- (4) Shima, S.; Thauer, R. K. *Chem. Rev.* **2007**, *7*, 37–46.
- (5) Meyer, J. *Cell. Mol. Life Sci.* **2007**, *64*, 1063–1084.
- (6) Zirnigbl, C.; Hedderich, R.; Thauer, R. K. *FEBS Lett.* **1990**, *261*, 112–116.
- (7) Ghirardi, M. L.; Posewitz, M. C.; Maness, P. C.; Dubini, A.; Yu, J. P.; Seibert, M. *Annu. Rev. Plant Biol.* **2007**, *58*, 71–91.
- (8) Fontecilla-Camps, J. C.; Volbeda, A.; Cavazza, C.; Nicolet, Y. *Chem. Rev.* **2007**, *107*, 4273–4303.
- (9) De Lacey, A. L.; Fernandez, V. M.; Rousset, M.; Cammack, R. *Chem. Rev.* **2007**, *107*, 4304–4330.
- (10) Mulder, D. W.; Boyd, E. S.; Sarma, R.; Lange, R. K.; Endrizzi, J. A.; Broderick, J. B.; Peters, J. W. *Nature* **2010**, *465*, 248–251.
- (11) Pandelia, M. E.; Ogata, H.; Currell, L. J.; Flores, M.; Lubitz, W. *Biochim. Biophys. Acta, Bioenerg.* **2010**, *1797*, 304–313.
- (12) Pandelia, M. E.; Ogata, H.; Lubitz, W. *ChemPhysChem* **2010**, *11*, 1127–1140.
- (13) Teixeira, V. H.; Baptista, A. M.; Soares, C. M. *Biophys. J.* **2006**, *91*, 2035–2045.
- (14) Siegbahn, P. E. M.; Tye, J. W.; Hall, M. B. *Chem. Rev.* **2007**, *107*, 4414–4435.
- (15) Teixeira, V. H.; Soares, C. M.; Baptista, A. M. *Proteins: Struct., Funct., Bioinf.* **2008**, *70*, 1010–1022.
- (16) Galvan, I. F.; Volbeda, A.; Fontecilla-Camps, J. C.; Field, M. J. *Proteins: Struct., Funct., Bioinf.* **2008**, *73*, 195–203.
- (17) Lill, S. O. N.; Siegbahn, P. E. M. *Biochemistry* **2009**, *48*, 1056–1066.
- (18) Keith, J. M.; Hall, M. B. *Inorg. Chem.* **2010**, *49*, 6378–6380.
- (19) Chang, C. H. *J. Phys. Chem. A* **2011**, *115*, 8691–8704.
- (20) Hong, G.; Cornish, A. J.; Hegg, E. L.; Pachter, R. *Biochim. Biophys. Acta, Bioenerg.* **2011**, *1807*, 510–517.
- (21) Volbeda, A.; Charon, M. H.; Piras, C.; Hatchikian, E. C.; Frey, M.; Fontecillacamps, J. C. *Nature* **1995**, *373*, 580–587.
- (22) Dementin, S.; Burlat, B.; De Lacey, A. L.; Pardo, A.; Adryanczyk-Perrier, G.; Guigliarelli, B.; Fernandez, V. M.; Rousset, M. *J. Biol. Chem.* **2004**, *279*, 10508–10513.
- (23) Friedrich, B.; Massanz, C.; Bernhard, M.; Lenz, O. *J. Inorg. Biochem.* **1999**, *74*, 24–24.
- (24) Agmon, N. *Chem. Phys. Lett.* **1995**, *244*, 456–462.
- (25) Swanson, J. M. J.; Maupin, C. M.; Chen, H. N.; Petersen, M. K.; Xu, J. C.; Wu, Y. J.; Voth, G. A. *J. Phys. Chem. B* **2007**, *111*, 4300–4314.
- (26) Schmitt, U. W.; Voth, G. A. *J. Phys. Chem. B* **1998**, *102*, 5547–5551.
- (27) Cuma, M.; Schmitt, U. W.; Voth, G. A. *Chem. Phys.* **2000**, *258*, 187–199.
- (28) Cuma, M.; Schmitt, U. W.; Voth, G. A. *J. Phys. Chem. A* **2001**, *105*, 2814–2823.
- (29) Day, T. J. F.; Soudackov, A. V.; Cuma, M.; Schmitt, U. W.; Voth, G. A. *J. Chem. Phys.* **2002**, *117*, 5839–5849.
- (30) Wu, Y. J.; Chen, H. N.; Wang, F.; Paesani, F.; Voth, G. A. *J. Phys. Chem. B* **2008**, *112*, 467–482.
- (31) Maupin, C. M.; Wong, K. F.; Soudackov, A. V.; Kim, S.; Voth, G. A. *J. Phys. Chem. A* **2006**, *110*, 631–639.
- (32) Laio, A.; Parrinello, M. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 12562–12566.
- (33) Iannuzzi, M.; Laio, A.; Parrinello, M. *Phys. Rev. Lett.* **2003**, *90*, 238301–238302.
- (34) Ensing, B.; Laio, A.; Parrinello, M.; Klein, M. L. *J. Phys. Chem. B* **2005**, *109*, 6676–6687.
- (35) Zhang, Y.; Voth, G. A. *J. Chem. Theory Comput.* **2011**, *7*, 2277–2283.
- (36) Higuchi, Y.; Ogata, H.; Miki, K.; Yasuoka, N.; Yagi, T. *Structure* **1999**, *7*, 549–556.
- (37) Jensen, J. H.; Li, H.; Robertson, A. D. *Proteins: Struct., Funct., Bioinf.* **2005**, *61*, 704–721.
- (38) Jensen, J. H.; Bas, D. C.; Rogers, D. M. *Proteins: Struct., Funct., Bioinf.* **2008**, *73*, 765–783.
- (39) Yagi, T.; Honya, M.; Tamiya, N. *Biochim. Biophys. Acta* **1968**, *153*, 699–705.
- (40) Sakaguchi, S.; Kano, K.; Ikeda, T. *Electroanalysis* **2004**, *16*, 1166–1171.
- (41) Siegbahn, P. E. M. *Adv. Inorg. Chem.* **2004**, *56*, 101–125.
- (42) Maseras, F.; Morokuma, K. *J. Comput. Chem.* **1995**, *16*, 1170–1179.
- (43) Vreven, T.; Byun, K. S.; Komaromi, I.; Dapprich, S.; Montgomery, J. A.; Morokuma, K.; Frisch, M. J. *J. Chem. Theory Comput.* **2006**, *2*, 815–826.
- (44) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. *J. Am. Chem. Soc.* **1995**, *117*, 5179–5197.
- (45) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, J. A., Jr.; Vreven, T.; Kudin, K. N.; Burant, J. C.; et al. *Gaussian 03*, revision C.02; Gaussian, Inc.: Wallingford, CT, 2004.
- (46) Bayly, C. I.; Cieplak, P.; Cornell, W. D.; Kollman, P. A. *J. Phys. Chem.* **1993**, *97*, 10269–10280.
- (47) Schmitt, U. W.; Voth, G. A. *J. Chem. Phys.* **1999**, *111*, 9361–9381.
- (48) Case, D. A.; Darden, T. A.; Cheatham, T. E., III; Simmerling, C. L.; Wang, J.; Duke, R. E.; Luo, R.; Walker, R. C.; Zhang, W.; Merz, K. M.; et al. *AMBER 11*; University of California: San Francisco, CA, 2010.
- (49) Plimpton, S. J. *Comput. Phys.* **1995**, *117*, 1–19 <http://lammps.sandia.gov>.
- (50) Cheatham, T. E.; Cieplak, P.; Kollman, P. A. *J. Biomol. Struct. Dyn.* **1999**, *16*, 845–862.
- (51) Kumar, S.; Bouzida, D.; Swendsen, R. H.; Kollman, P. A.; Rosenberg, J. M. *J. Comput. Chem.* **1992**, *13*, 1011–1021.
- (52) Souaille, M.; Roux, B. *Comput. Phys. Commun.* **2001**, *135*, 40–57.

- (53) Maragliano, L.; Fischer, A.; Vanden-Eijnden, E.; Ciccotti, G. *J. Chem. Phys.* **2006**, *125*, 024101–024106.
- (54) Gan, W.; Yang, S.; Roux, B. *Biophys. J.* **2009**, *97*, L08–L10.
- (55) Gervasio, F. L.; Laio, A.; Parrinello, M. *J. Am. Chem. Soc.* **2005**, *127*, 2600–2607.
- (56) Gervasio, F. L.; Parrinello, M.; Ceccarelli, M.; Klein, M. L. *J. Mol. Biol.* **2006**, *361*, 390–398.
- (57) Domene, C.; Klein, M. L.; Branduardi, D.; Gervasio, F. L.; Parrinello, M. *J. Am. Chem. Soc.* **2008**, *130*, 9474–9480.
- (58) Pfaendtner, J.; Branduardi, D.; Parrinello, M.; Pollard, T. D.; Voth, G. A. *Proc. Natl. Acad. Sci. U.S.A.* **2009**, *106*, 12723–12728.
- (59) Barducci, A.; Bonomi, M.; Parrinello, M. *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **2011**, *1*, 826–843.
- (60) Bonomi, M.; Branduardi, D.; Bussi, G.; Camilloni, C.; Provasi, D.; Raiteri, P.; Donadio, D.; Marinelli, F.; Pietrucci, F.; Broglia, R. A.; Parrinello, M. *Comput. Phys. Commun.* **2009**, *180*, 1961–1972.
- (61) George, S. J.; Kurkin, S.; Thorneley, R. N. F.; Albracht, S. P. J. *Biochemistry* **2004**, *43*, 6808–6819.