

Finding Gas Diffusion Pathways in Proteins: Application to O₂ and H₂ Transport in Cpl [FeFe]-Hydrogenase and the Role of Packing Defects

Jordi Cohen,¹ Kwiseon Kim,² Paul King,² Michael Seibert,² and Klaus Schulten^{1,*}

¹Department of Physics and Beckman Institute
University of Illinois
Urbana, Illinois 61801

²National Renewable Energy Laboratory
Golden, Colorado 80401

Summary

We report on a computational investigation of the passive transport of H₂ and O₂ between the external solution and the hydrogen-producing active site of Cpl [FeFe]-hydrogenase from *Clostridium pasteurianum*. Two distinct methodologies for studying gas access are discussed and applied: (1) temperature-controlled locally enhanced sampling, and (2) volumetric solvent accessibility maps, providing consistent results. Both methodologies confirm the existence and function of a previously hypothesized pathway and reveal a second major pathway that had not been detected by previous analyses of Cpl's static crystal structure. Our results suggest that small hydrophobic molecules, such as H₂ and O₂, diffusing inside Cpl, take advantage of well-defined preexisting packing defects that are not always apparent from the protein's static structure, but that can be predicted from the protein's dynamical motion. Finally, we describe two contrasting modes of intraprotein transport for H₂ and O₂, which in our model are differentiated only by their size.

Introduction

Hydrogenases are enzymes that catalyze the reversible oxidation or production of molecular hydrogen according to the reaction



Hydrogenases can be found in a wide variety of unicellular organisms and usually come in one of two flavors: [NiFe]-hydrogenases, which are usually associated with H₂ uptake, and “iron-only” [FeFe]-hydrogenases (Adams, 1990), which are generally involved in H₂ production. In a majority of microorganisms, [FeFe]-hydrogenases (Nicolet et al., 2002) function in anaerobic metabolism to oxidize overly reduced electron carriers. Much of the recent scientific interest in [FeFe]-hydrogenases, however, concerns a different role entirely: the H₂ production properties of [FeFe]-hydrogenases offer the promise of a means for affordable large-scale production of H₂ as a source of renewable energy (Ghirardi et al., 2000; Mertens and Liese, 2004).

In this paper, we investigate the structural properties

of Cpl [FeFe]-hydrogenase from *Clostridium pasteurianum*. Hydrogen production in Cpl happens at the H cluster, a metallic cluster bound to and embedded inside the Cpl protein matrix, and is achieved by the reduction of H⁺ ions from the external solution through the use of electrons acquired from a reduced carrier such as ferredoxin (Adams, 1990; Peters, 1999). The H⁺ ions (or “protons”) probably reach the H cluster by means of a putative, but yet unverified, proton pathway contained in the protein (Peters et al., 1998). The electrons are transferred to the embedded H cluster through a series of accessory iron-sulfur clusters aligned in a chain between the H cluster and one end of Cpl. The Cpl enzyme, along with its embedded iron-sulfur clusters and H cluster, is displayed in Figure 1.

In addition to electron and H⁺ transport, Cpl must allow for gas transport between the buried H cluster and the external solution. Cpl must allow the H₂ product to exit the protein, but it also allows small gas molecules such as O₂ and CO to transit through the enzyme and reach the H cluster, which becomes inactivated upon their binding. In the case of O₂, this process is irreversible. While the O₂-mediated deactivation of hydrogenase is in some cases beneficial for the host organism, it severely limits the practicality of using hydrogenase to produce H₂ as a carrier of consumable energy (Flynn et al., 2002). As of the writing of this paper, a few studies have investigated gas access in hydrogenases by using either a molecular dynamics (MD) approach to sample the possible paths for H₂ permeation (Cohen et al., 2005; Montet et al., 1997), hydrophobic cavity searches on static structures (Nicolet et al., 2002, 1999; Montet et al., 1997), or crystallography of xenon-saturated structures (Montet et al., 1997). Although some progress has been made, aside from the certain but incomplete predictions from X-ray crystallography on proteins in the presence of xenon, no other method has been able to comprehensively predict all of the gas pathways in hydrogenases or in any other protein.

In this paper, we investigate the possible pathways taken by H₂ and O₂ across Cpl hydrogenase. We notice a significant difference between the dynamics of H₂ and of O₂ gas permeation through the protein matrix, which is caused only by differences in ligand size. This behavior prompted us to examine the effect of the protein's internal dynamics in regulating gas access to its buried active site. For this purpose, we introduce a new, to our knowledge, approach for mapping transient cavities based on the dynamics of the lone protein (i.e., without the gas ligand) extracted from computer simulation and compare the resulting maps to trajectories of the intraprotein gas diffusion process. An almost perfect match between the two computations prompts us to suggest that the protein-wide pathways taken by hydrophobic gases in Cpl hydrogenase, and possibly in any gas transport protein, are fully determined by density fluctuations within the protein. Furthermore, in the case of hydrogenase, two major gas transport pathways are fully characterized by the protein's motion on

*Correspondence: kschulte@ks.uiuc.edu

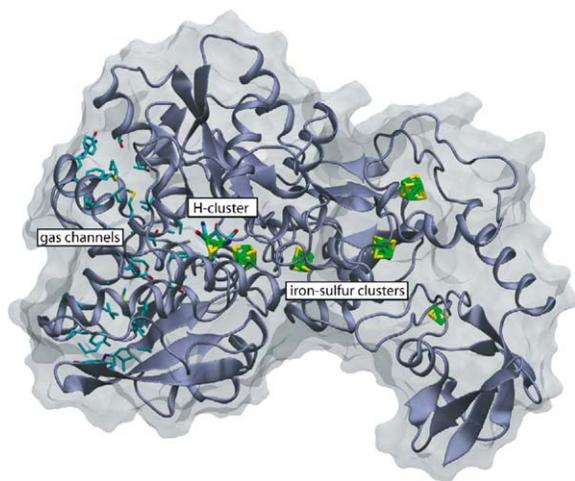


Figure 1. Structure of Cpl Hydrogenase

Cpl hydrogenase showing the enzyme with its embedded H cluster and iron-sulfur clusters. Also shown are the residues lining the two principal gas pathways connecting the external solution to the buried H cluster.

the nanosecond timescale, despite the fact that the actual diffusion of gases such as O_2 may take much longer. In addition to the physical insight gained, the cavity mapping method that we introduce holds the promise of being able to predict protein-wide gas transport pathways inside macromolecules. Our results provide insights into the gas transport mechanisms for H_2 and O_2 in Cpl hydrogenase, as well as a different way of looking at gas transport pathways that is immediately applicable to other proteins and structures.

Results

In this section, we describe and compare MD trajectories of H_2 and O_2 gas diffusion inside Cpl. We then compare our results with a dynamic mapping of the protein cavities in the absence of gas, and we find a strong correlation between the locations of the protein's natural cavities and the locations of the diffusing gas molecules.

Simulations of Gas Diffusion Reveal Different Transport Mechanisms for H_2 and O_2

While the diffusion of O_2 or H_2 molecules inside a protein is not a particularly slow process (a transport event can take from 100 ps to hundreds of nanoseconds), it is a stochastic process, and it cannot be completely described by simply examining a few nanosecond-long trajectories. In order to drastically speed up the exploration of O_2 and H_2 entry/exit pathways, it becomes necessary to use certain approximations. One such approximation, known as locally enhanced sampling (LES) (Elber and Karplus, 1990) and based on the time-dependent Hartree approximation (Gerber et al., 1982), allows a certain subset of particles in the simulated system to be replicated many times, where each replicated subset is simulated independently. In such a

scheme, each set of replicated particles interacts with a common environment consisting of the unreplicated particles, but the replicated copies do not interact with each other at all. In the present paper, we make use of a variant of LES: temperature-controlled locally enhanced sampling (TLES), which is described in the [Experimental Procedures](#).

We have run separate TLES simulations for H_2 and O_2 , by using 1000 copies of the diffusing gas molecule, in order to determine the pathways taken by H_2 and O_2 while transiting between the active site of Cpl and the external solution. In our simulations, we have used a heavy version of H_2 , “heavy dihydrogen” hH_2 , in which the hydrogen atoms' masses were set to that of oxygen. The reason for this is that we were mainly interested in investigating the accessibility of the protein to gas molecules, as a function of the gas' molecule size. Since we use the same mass for both O_2 and hH_2 , we remove the variation in behavior between the two gases that is caused by their difference in mass (and consequently velocity), and we instead focus purely on the variation in behavior that is due to their size difference. Changing the mass of the diffusing gas does not affect the kinetic energy, momentum, or energy profile perceived by the gas or experienced by the protein. The system will explore the same gas-protein conformational space as it would otherwise, except that the actual velocities of the hH_2 molecules will be slowed down with respect to H_2 . However, the rate of diffusion of hH_2 is not expected to be significantly different from that of real H_2 , due to friction effects. The larger mass also circumvents the use of much smaller simulation time steps, which are necessary when dealing with the very sharp Lennard-Jones potentials of TLES-replicated H_2 .

In the case of H_2 , we performed four simulations of 4 ns each, in which the hH_2 molecules were initially placed at the active site (at which hydrogen production takes place), based on the location of the active-site bound carbon monoxide in the X-ray structure of CO-saturated Cpl described by [Lemon and Peters \(1999\)](#). In all cases, hH_2 exited predominantly through two major diffusion pathways, the first one (pathway A) having been previously proposed as an H_2 channel candidate ([Nicolet et al., 2002](#); [Peters, 1999](#)), and the second one (pathway B) being newly discovered ([Cohen et al., 2005](#)). Both pathways meet at a large cavity right next to the H cluster binding site. Aside from just the main cavity and the two pathways, the hH_2 molecules from all of the simulations taken together consistently explored similar regions of the protein, as displayed in [Figure 2](#). Despite the fact that hH_2 exited simultaneously through both pathways in each simulation and that the shape of these pathways explored by hH_2 was the same for all simulations, the proportion of hH_2 exiting through one pathway or the other and the exit rates of hH_2 out of the protein varied significantly from one simulation to the next, as detailed in [Table 1](#). In all cases, the average hH_2 time of first exit out of the protein was very fast and was on the order of nanoseconds (with our simulation suggesting that the first hH_2 molecules will have found the exit within 200 ps and that roughly half will have exited after 4 ns). We expect that

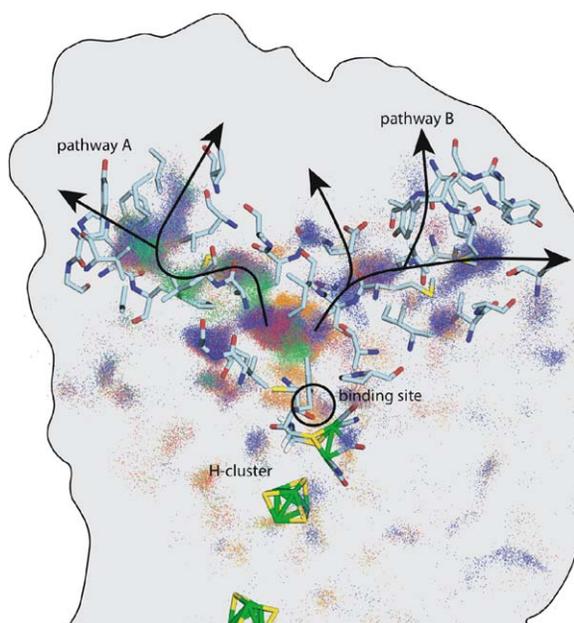


Figure 2. TLES Simulations of hH₂ Diffusion

Four 4 ns TLES simulations of 1000 copies of hH₂ diffusing out from the H cluster. Each independent simulation is shown in a different color, highlighting the fact that the space explored by hH₂ was consistent between simulations. Frames taken from every 100 ps of the trajectories of the hH₂ molecules located inside the protein are superimposed as a cloud. Shown in licorice are the iron-sulfur clusters and H cluster, as well as the residues that line the two major exit pathways. Possible exits, based on the proximity of the external solution, are highlighted with arrows.

real H₂ would exhibit exit times very similar to those of hH₂.

In the case of O₂ diffusion from the H cluster, five independent 3.5–4 ns simulations were performed, some of which are represented in Figure 3. In these simulations, 1000 TLES copies of O₂ were placed at the H cluster binding site and were allowed to diffuse. In only one of these five simulations did we observe O₂

Table 1. Proportion of hH₂ Exits by Pathway

Simulation	Pathway A	Pathway B	Other Pathways	Total Exits
#1	21%	22%	4%	47%
#2	72%	4%	3%	79%
#3	10%	3%	6%	19%
#4	4%	25%	7%	36%

The percentages (rounded) of the total replicated hH₂ molecules that have exited Cpl during each of four different 4 ns simulations of the diffusion of hH₂ starting at the H cluster binding site, sorted by pathways, are shown. Pathways “A,” “B,” and “Other,” respectively, refer to the previously suggested pathway identified by visual inspection of the Cpl structure (Pathway A), to the additional pathway identified by means of the TLES and VSAM methods (Pathway B), or to neither (Other). “Total” refers to the total percentage of molecules that found their way out of the protein during the 4 ns time period. A molecule is considered to have exited when one of its atoms comes into proximity to a water molecule located outside of the protein.

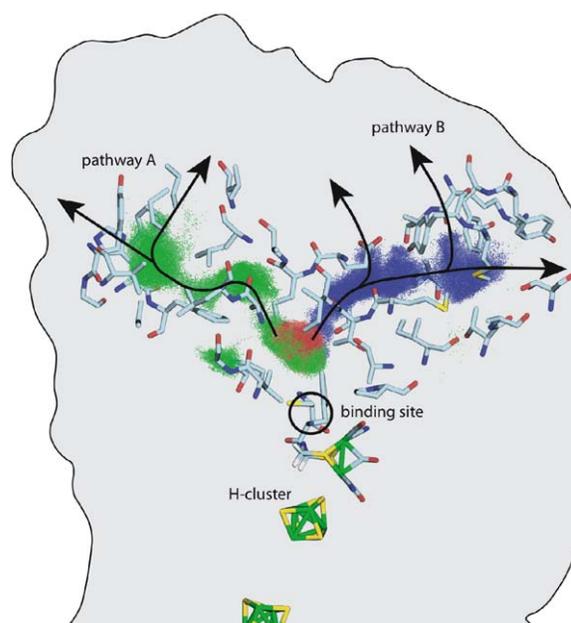


Figure 3. TLES Simulations of O₂ Diffusion

Three representative TLES simulations of 1000 copies of O₂ diffusing out from the H cluster or from the middle of a previously identified H₂ channel. Each independent simulation is shown in a different color, and one can see that, contrary to H₂ diffusion, the O₂ molecules move collectively through the same pathway for a given simulation, though they may employ different pathways for different independent simulations. Overall, the set of pathways explored by O₂ matches the dominant pathways explored by hH₂. Snapshots were taken every 50 ps. The representation of the protein is the same as in Figure 2.

leaving the central cavity through the newly discovered pathway B (see Figure 3 [blue]). In the four other cases, O₂ remained in the central cavity near the binding site (Figure 3 [red]) for the duration of the simulation. Since the other pathway (pathway A) through which we observed hH₂ to diffuse appears to be a narrow but unambiguous hydrophobic channel in the crystal structure, it is suspected that O₂'s failure to exit through it in our simulations is simply due to lack of sufficient sampling of the protein's degrees of freedom. With this in mind, additional TLES simulations of O₂ were performed, by using as starting positions various locations at which large densities of hH₂ were observed in the previous hH₂ diffusion simulations. When we placed O₂ inside the originally proposed hydrophobic channel (pathway A), one cavity away from the central cavity, we were able to successfully observe O₂ diffusion along that channel in all three of the independent 3.5 ns simulations. A fraction of the O₂ molecules placed in this cavity even diffused inward to the central cavity and back in one of the three 3.5 ns simulations performed (one is shown in Figure 3 [green]). In no case was O₂ observed to completely exit the protein and partition into the water solution. We suspect that the hydrophobicity of O₂ causes it to prefer the protein environment to that of the water solution; however, the influence of our O₂ model parameters and of the TLES dynamics might

also be playing a role. The rate of diffusion of O_2 inside hydrogenase appears to be entirely determined by the protein's dynamical fluctuations. O_2 was observed to be able to diffuse to the surface of the protein very quickly (in as little as 1.2 ns) when the protein explored a set of ideal conformations. However, in most simulations, we did not observe any significant travel of O_2 (in four instances out of five), reflecting the fact that the protein conformations are usually unfavorable to O_2 diffusion.

For O_2 , we simulated the reverse of the natural process of O_2 diffusion from the bulk solvent toward the active site. This was done because, at the beginning of our investigation, it was not known where O_2 could enter the protein, and the active site was the only location of Cpl that was a priori known with certainty to be accessible to O_2 , based on previous studies of O_2 inactivation (Adams, 1990). Since the transport mechanism of O_2 inside Cpl is almost undoubtedly passive, it does not matter in which direction we simulate the diffusion, unless there is a strong overall energy gradient between the outside and inside. In our simulations, we have observed back-and-forth motion of O_2 and H_2 , suggesting that the energy profile is approximately flat between solvent and active site (excluding the intermediary energy barriers). However, two aspects of the gas diffusion process may affect the function of the pathways that we describe, but they are not addressed by our current methodology: (1) the free energy profile experienced by the diffusing gas molecules has not been fully characterized, and there may in fact be a slight bias favoring inward or outward diffusion; and (2) the accessibility of the identified pathway entrances to gas entering Cpl from the external solution has not been determined.

From our TLES results, we see that both O_2 and hH_2 can diffuse across the protein and exit through two common pathways. However, we noticed that hH_2 can penetrate a broader region of the protein, and on shorter timescales, whereas O_2 in our simulations was strictly limited to the above-mentioned pathways. Aside from exploring very similar regions of Cpl, the TLES trajectories for O_2 were qualitatively very different from those of hH_2 in terms of the collective dynamics. While the different copies of hH_2 spread out with time into a diffuse cloud covering the entire protein-water system, the O_2 molecules typically clustered together as a single cloud (which occasionally could also split into more clouds on a ~ 2 – 4 ns timescale). The fact that the O_2 molecules cluster cannot simply be attributed to the fact that they all experience similar conditions: they all have different initial velocities, Langevin random forces, and interactions with the protein, and, in addition, the dramatic clustering behavior is not observed at all for the smaller hH_2 molecules. Instead, the behavior of the collective O_2 motion suggests that O_2 does not diffuse in and out of Cpl through a permanent channel, contrary to what was previously assumed. As will be shown later, O_2 fills up small cavities inside the protein that are themselves dynamic. Through the protein's natural motion, combined with the disruptive effect of the O_2 , these cavities dynamically fluctuate in size and in their connections with neighboring cavities at certain favorable locations. The transport of O_2 seems to be

guided much more by the protein's random peristaltic motion than by simple diffusion through a static complex medium (Nadler and Stein, 1996). For H_2 on the other hand, the protein appears much more porous, and, at any given time, there are many more cavities and partial channels that are accessible to H_2 than to O_2 . Because, in our simulations, O_2 and H_2 have the same mass, the differences in behavior between the two gases is not due to differences in diffusion speeds but is solely caused by their different Lennard-Jones parameters.

The major problem encountered with TLES was insufficient sampling, despite the 1000-fold increase in sampling as compared to regular MD, especially since the effect of the single protein trajectory on ligand diffusion appears to be of significant importance. Taken together, the TLES simulations do in fact confirm the existence of two major gas transport pathways, and the calculated trajectories are both realistic and consistent. But, for the case of O_2 , we had difficulty reproducing the same pathways from one simulation run to another. While we observed several very likely permeation events from the active site to the external solution, we could not determine, by using TLES alone, with certainty whether there exist other pathways through which O_2 could exit on the simulated timescale. To obtain a picture of the pathway topology for H_2 and O_2 inside Cpl, we clearly need a better method. We believe that the maximum volumetric solvent accessibility map (VSAM) methods described in the [Experimental Procedures](#) section provide an excellent tool by which to acquire the needed information.

Predicting Hydrophobic Gas Accessibility

from the Equilibrium Dynamics of the Protein Alone
In the previous section, we stated that O_2 molecules permeating inside Cpl moved as if they were trapped in small dynamic pockets of empty space. Almost every copy of O_2 followed exactly the same trajectory in a given TLES simulation, yet these trajectories varied widely from one simulation to the next. Since TLES has a single copy of the protein interacting with many replicated gas molecules, our results suggest that transient conformations of the protein have a huge impact on the pathways taken by gas molecules diffusing inside it. In this section, we specify and confirm this hypothesis by monitoring the transient cavities that are intrinsically present inside Cpl in the absence of gas. We map out, by means of VSAMs, which regions of the entire protein would be accessible to a "ghost" solvent of a given radius, assuming that the solvent does not interact with the protein: it can only "occupy" free space if such space is ever spontaneously available in the protein. Surprisingly, we find that the set of possible trajectories, taken by both O_2 and H_2 gas (and very likely by any other spherically shaped hydrophobic ligand), can be predicted remarkably well. These results complement other related investigations of the influence of protein conformations on ligand diffusion inside the myoglobin distal pocket (Carlson et al., 1996, Gibson et al., 1992) and catalase (Amara et al., 2001). Our results differ from most of these previous studies by the methodologies that we have used (TLES and VSAMs), as

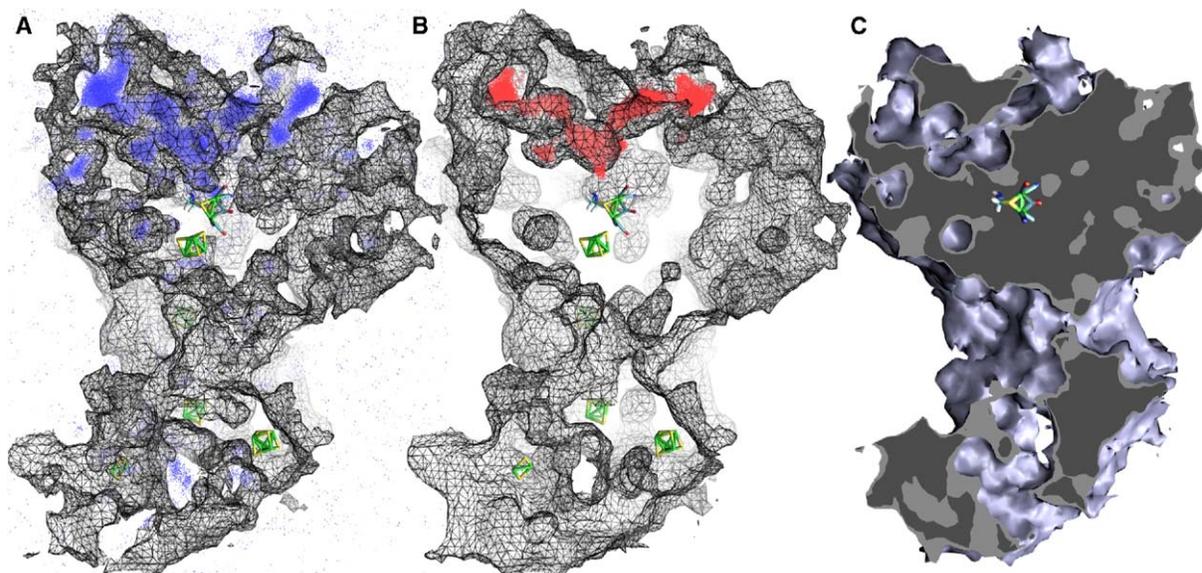


Figure 4. Map of Internal Gas Pathways

(A and B) Comparison of the surface delimiting the maximum VSAM predicted from the equilibrium simulation of Cpl in the absence of gas for a particle the size of (A) H₂ and (B) O₂. The locations explored by the centers of the (A) hH₂ or (B) O₂ atoms during the TLES simulations are also shown.

(C) A slice through the computed gas-accessible surfaces for O₂ (light-gray internal volume) and H₂ (dark-gray internal volume), highlighting the differences between the two.

well as through the, in many cases, significantly longer timescales and larger areas of the protein probed, and by the fact that we looked at the protein's accessible volume in the absence of ligands.

We have calculated a static 3D map of the largest ghost solvent spheres that could be placed at any given time inside Cpl, based on a 2 ns equilibrium simulation, according to the VSAM protocol. VSAMs that were calculated based on either the first or second nanosecond of the computed equilibration trajectory showed little variation and strong reproducibility, as opposed to the TLES trajectories. Figure 4A shows the isosurface contour representing the area accessible to a solvent of radius 1.35 Å (which characterizes the overall radius of an H₂ molecule) along with the TLES trajectories of hH₂ diffusion. Visually, the isosurface accurately describes the regions of space that had been explored by hH₂ during the four 4 ns TLES simulations of hH₂ inside Cpl, even though the VSAM was calculated based on a trajectory that did not contain any gas molecules. Almost every predicted cavity throughout the Cpl structure was observed to have been visited by H₂, and almost all areas visited by H₂ corresponded to regions in which cavities had been predicted, including areas away from where the H₂ diffusion originated as well as, surprisingly, internal cavities disconnected from the surface. The same excellent match was observed in the case of O₂ and the 1.6 Å isovalue contour of the same VSAM (Figure 4B), though in our TLES simulations, the O₂ molecules only had time to explore the cavities directly adjacent to the H cluster active site; thus, the comparison was only performed there. A comparison of the area of the protein accessible to H₂- and

O₂-sized particles is shown in Figure 4C, and one can see that both pathways A and B are clearly identified. There were very few exceptions in which the TLES simulations did not match the cavity predictions: (1) exactly at the binding site at which the 1000 copies were placed, no cavity was predicted there (but gas was observed there during TLES because this was the starting position), (2) in the case of O₂, there was one single region of disagreement, which corresponds to a region occupied by water during equilibration, but in which O₂ managed to go during the TLES simulation (if we consider the space occupied by water to be accessible to O₂, then O₂ was never observed in any other unpredicted cavity), and (3) hH₂ was occasionally observed in regions not predicted by the 1.35 Å contour (but this happened for less than 1% of all hH₂ positions explored). Figure 5 shows the cumulative occupancy of O₂ and H₂ based on the value of the underlying maximum VSAM grid points in the region around the active site. The figure displays definite thresholds for the occupancy of H₂ and O₂ as a function of the predicted maximum radius of solvent, below which no TLES-simulated gas has been found to go. This shows that the maximum VSAM correctly predicts the accessibility of both H₂ and O₂ (in the sense that gas does not enter regions not predicted, the converse appears to be true according to visual inspection for the case of H₂ and cannot be proven for the case of O₂ due to lack of sampling). Only 0.9% of the TLES hH₂ molecules were found in hollow regions with maximum predicted radii below 1.35 Å, and only 0.1% of the O₂ was found in regions with a radius of less than 1.6 Å.

It is important to note that the gas-transport path-

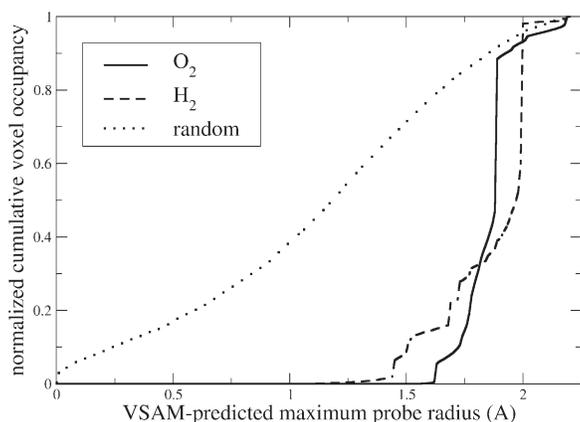


Figure 5. Comparison of TLES- and VSAM-Predicted Gas Pathways Histogram of the values of the VSAM-predicted maximum probe radius for each of the positions explored by both O_2 and hH_2 in the TLES simulations. The abscissas represent the interpolated value of a 0.5 Å resolution maximum VSAM grid at the location of the center of each O_2 or H_2 atom from the TLES simulations, and the ordinates indicate the number of times (normalized and cumulative) that these values have been explored by O_2 or hH_2 .

ways described above could not be identified by simple analysis of static crystal structures. In the case of pathway A, a solvent-accessible surface for a solvent the size of H_2 (radius ~ 1.35 Å) can indeed be detected this way, though it becomes disconnected if one includes equilibrated hydrogen atoms or uses larger probe molecules (such as O_2 with a radius ~ 1.6 Å). Pathway B, however, could not be detected for either H_2 or O_2 by using this type of analysis. If one compares the iso-value contours of our maximum and average VSAMs, for H_2 - and O_2 -sized solvent particles, one can understand the difference in dynamics observed between O_2 and hH_2 diffusion during the TLES simulations. The VSAM containing the average value of the solvent molecule does not exhibit any channels of sufficient size to allow O_2 to access the active site. Only very few cavities along the two main diffusion pathways are large enough to accommodate an O_2 molecule on average, but these do not form a continuous channel from the external solution to the active site (see Figure 6). Only for the case of H_2 is one of the two major diffusion pathways at least partially visible in either the average or instantaneous VSAMs (a continuous channel is in fact observed for a probe radius of approximately 1 Å, which is smaller than H_2). This is the “hydrogen channel” that was originally proposed from an analysis of the X-ray crystal structure of Cpl (Peters et al., 1998). As suggested by our TLES simulations, it does appear that O_2 moves from cavity to cavity as the cavities fluctuate into existence inside the protein, and there is no permanent “channel” to speak of. In the case of H_2 , the same also holds true, but H_2 is sufficiently smaller in size, as compared to O_2 , such that more open space is accessible to it at any given time. The instantaneous H_2 -sized cavities connect in more places as well as more often, allowing for easier diffusion. A quick analysis of the probabilities at which different parts of the pathways are large enough to accommodate gas molecules reveal that, over the course

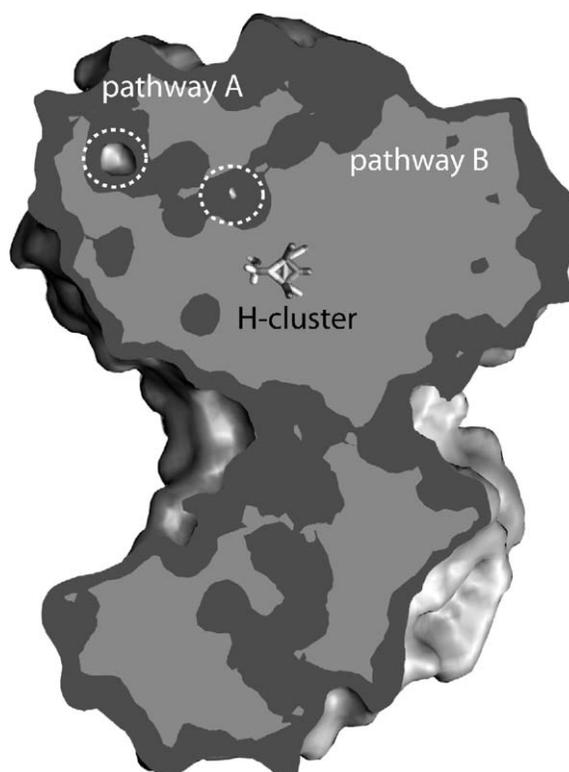


Figure 6. Comparison of Average and Maximum VSAMs Average and maximum cavities for the case of O_2 -sized particles. The outer contour represents the average VSAM. A solid slice through the surface shows which areas of the average VSAMs are accessible to the gas, including the two main diffusion pathways (dark gray), and which areas are excluded (light gray) according to the maximum VSAM. The dotted circles indicate the only two discernable O_2 -sized cavities in the average VSAM (as well as in the static crystal structure).

of the 2 ns equilibration, most regions of pathways A and B were open 5%–8% of the time for O_2 - and 30%–35% of the time for H_2 -sized particles. Also, each pathway had one constricted region, right outside the central cavity, which was only open for about 2% of the time for O_2 and 20% of the time for H_2 , thus limiting the rate of exit of the gas from the central cavity.

Finally, we comment on the approximations made in the maximum VSAM method. We have shown that we can predict with excellent accuracy where both H_2 and O_2 gas can go in a hydrogenase protein, based solely on an analysis of the space accessible inside the protein during an equilibrium simulation in the absence of gas. This statement appears to imply that gas and protein do not interact strongly. However, other studies of gas transport in protein cavities have suggested that the presence of a gas can strongly influence the internal conformations of the protein near the gas (Bossa et al., 2004). To test this suggestion, we have performed our VSAM analysis on the trajectories that did contain O_2 and hH_2 , and we clearly see that the proteins that contain gas exhibit significantly larger cavities (not shown) compared to the gas-less trajectories. What the present paper intends to demonstrate is not that the gas diffuses as if it were a noninteracting ghost particle,

but rather that, though the gas can strongly bias the openness and shape of nearby cavities, it does not create new cavities that would not spontaneously appear by themselves inside the protein. The presence of gas does not create new diffusion pathways. The gas molecules merely insert themselves into packing defects that arise spontaneously with or without gas in the protein and then alter the defects' sizes and "open" probabilities. In this respect, the lone protein approximation is a very good indicator of which areas of the protein are accessible to hydrophobic gases.

Conclusion

There has been a steadily increasing body of evidence suggesting that packing defects play a major role in gas transport inside many proteins (Gibson et al., 1992; Brunori et al., 2004; Bossa et al., 2004; Carlson et al., 1996; Amara et al., 2001; Brunori and Gibson, 2001). Our results further confirm previous indications that a permanent channel is not needed to allow gas from a protein's exterior to reach a buried active site. Transient cavities, arising from the protein's natural equilibrium dynamic motion on the nanosecond timescale, can define predetermined pathways for hydrophobic gas transport. Such observations have been stated before, but we also demonstrate that the location of the pathways taken by diffusing hydrophobic gases (in this case, H₂ and O₂) can be fully described on a protein-wide scale, by simply analyzing the motion of the protein in the absence of internally diffusing gas, and that the presence of the gas is not absolutely needed to open or activate these pathways. We do not expect this to be the case for polar ligands: strong protein-ligand electrostatic interactions might make accessible pathways that would otherwise remain tightly shut during the protein's equilibrium motion (Kocher et al., 1996).

Comparing all of our TLES simulations for a given gas molecule (O₂ or H₂), we notice that even though we could not reproduce the same trajectories and gas exit rates from one multianosecond run to another, all of our runs had in common the fact that they were exploring the same maximum cavity predicted by our VSAM, which itself was reproducible with very good agreement between runs on the nanosecond timescale. This observation highlights the important possibility that all of the necessary protein conformations that enable gas permeation across Cpl can occur on the nanosecond timescale. Though the essential dynamics needed for the understanding of gas permeation in globular proteins (namely, the formation of transient cavities) occur on a short timescale, results obtained by simulating the diffusion of individual particles were never observed to converge during that timescale. This is due to the fact that, if we ignore for now the effects of the gas-protein interactions, gas diffusion, as probed by MD, relies on the temporal and spatial coupling of three simple random processes, namely, the transient formation of cavities, the transient formation of passages between these cavities, and the random hopping of gas molecules across these passages. Combined together, these three effects give the appearance of a very complex gas diffusion process that cannot be fully characterized by using multianosecond MD diffusion studies alone. Our

results demonstrate that the very slow diffusion of O₂ and H₂ inside Cpl can be characterized by sampling the dynamics of a protein on a much shorter timescale. We cannot exclude the effect of rare protein conformations not sampled in a 1 ns run on gas diffusion, but we can assume that their effect is probably a very minor one, since, in order to be effective, these rare protein conformations must also coincide with the presence of O₂ molecules at just the right place and time.

Experimental Procedures

Volumetric Solvent Accessibility Maps

We introduce volumetric solvent accessibility maps (VSAMs) as a means of representing the solvent-accessible surface of a protein. Given a set of hard atomic spheres representing a macromolecule, existing methods typically calculate a closed polyhedron that represents the boundary between space that is penetrable and impenetrable to a model solvent molecule represented by a hard sphere of specified radius (e.g., Richards' Smooth Molecular Surface [Richards, 1977] and Connolly's method [Connolly, 1983]). While traditional techniques can produce almost exact results, there are significant advantages to be gained in representing solvent maps as a volumetric data set (i.e., a 3D grid of scalar values). VSAMs can contain information about not just one size of solvent molecule, but about all sizes simultaneously, enabling the interactive visualization of how the intermolecular cavities change as a function of probe radius. Most importantly, however, many VSAMs can be combined together by using average or maximum rules. When performed over a set of trajectory frames, rather than for a static structure, they can reveal information that cannot be gotten from conventional methods, such as that presented in this paper. However, this comes at the cost of limited resolution, precision, and increased computer memory requirements.

In practice, the solvent accessibility of a macromolecule is stored in a 3D volumetric grid that overlaps all or part of the macromolecule in coordinate space. The value of each voxel (i.e., grid point) of the grid is set to be the radius of the largest possible solvent sphere that (1) contains the spatial coordinates of the center of that voxel, and (2) does not overlap with any of the van der Waals spheres that represent the molecule's atoms. Voxels whose coordinates lie inside the molecule's van der Waals spheres are set to zero, since no solvent can be present there. Voxels in small interstices will have very small radius values, whereas voxels in large cavities or voxels outside the protein can have very large radius values. To retrieve a close approximation of the conventional solvent-accessible surface, one needs only to display the isocontours of the VSAMs for a given value corresponding to the desired probe radius (e.g., such as a 1.4 Å radius for a water solvent). All of the voxels with values smaller than the probe radius will lie outside the contour, and those with larger values will lie inside, because if a voxel can be contained in a large solvent sphere, it will automatically be accessible to smaller solvent spheres.

In this paper, we report on *average* and *maximum* VSAMs, calculated from a set of frames from MD simulation trajectories. One must imagine that a separate VSAM is calculated for each simulation frame; that all of the volumetric grids are aligned and are of the same size; and that, for all frames, the macromolecule has been repositioned such that its C_α atoms are aligned by using a best fit. The average solvent map is then simply a map in which the voxel values have been averaged over all frames, and it provides a description of the average size of the solvent that can reach a specific position in space over the course of the trajectory. The maximum solvent map, then, contains at each voxel the maximum value encountered during the course of the trajectory. The maximum map describes the maximum solvent size that can ever be spontaneously placed inside the macromolecule during any frame of the trajectory. If a channel is only transiently open, or if its different sections are open at different times, then the maximum solvent map will show what the channel would look like in its maximally opened state.

Temperature-Controlled Locally Enhanced Sampling

LES defines an algorithm for simulating multiple copies of a certain number of particles (in this case, gas ligands), which interact in a mean-field way, with a single copy of the environment (here, the protein). One of the features of LES, as described by [Elber and Karplus \(1990\)](#), is that the effective temperature of the replicated particles' dynamics (with N replicas) is increased by a factor of N . This enhanced temperature, while often used as a means to cross barriers, also has disadvantages. For one, the resulting trajectories can be unphysical (i.e., assuming 10 replicated copies, the behavior of a particle at 3000 K is very different from that of one at 300 K), and the copies' traveling speed and energy landscape is dramatically altered compared to reality. Because of this, the extra factor of N in the temperature of the replicated particles effectively limits the amount of copies possible. If one could control the dynamics of the replicated particles such that they all act like 300 K particles, then it would become practical to use much more than 10 simultaneous copies and to gain a greater amount of information from a single simulation (we used 1000 copies for the present case).

In LES, the kinetic energy K and potential energy V are defined as follows:

$$K = \frac{1}{2} \sum_{a \text{ in } A} m_a \dot{q}_a^2 + \frac{1}{2N} \sum_{i=1}^N \sum_{x \text{ in } X} m_x \dot{q}_{x,i}^2 \quad (2)$$

$$V = V_{AA}(\mathbf{q}_A) + \frac{1}{N} \sum_{i=1}^N [V_{XX}(\mathbf{q}_{X,i}) + V_{AX}(\mathbf{q}_A, \mathbf{q}_{X,i})], \quad (3)$$

where \mathbf{q}_n , $\dot{\mathbf{q}}_n$, and m_n are the position, velocity, and mass, respectively, of particle n . A and X represent the sets of unreplicated and replicated particles, respectively, and N is the number of LES copies. $V_{AA}(\mathbf{q}_A)$ represents the total interaction energy between particles within A , $V_{AX}(\mathbf{q}_A, \mathbf{q}_X)$ is the interaction between particles in A and X , etc.

As outlined in [Straub and Karplus \(1991\)](#), [Equations 2 and 3](#) describe a non-Newtonian system (Newton's second law is not satisfied), with the consequence that, at equilibrium, they describe a system in which each replicated particle will acquire a kinetic energy that is N times greater than that of an unreplicated particle (for an LES particle, $\langle K_{LES} \rangle = 3/2Nk_B T$, assuming free particles for the sake of simplicity). While their formal temperatures (defined according to the zeroth law of thermodynamics) are the same, the "effective" temperature (defined from the average kinetic energy per particle $\langle K \rangle = (3/2) k_B T$) of the LES particles is larger by a factor of N . A commonly used method for dealing with the divergent kinetic energy of the replicated particles has been to increase the mass of these particles by a factor of N ([Straub and Karplus, 1991](#); [Roitberg and Elber, 1991](#)). This slows down the copied particles so that their motion can be calculated efficiently, but does nothing to address the divergent kinetic energy itself or to avoid the effective reduction in energy barriers for the copied particles. The increased ligand energy results in significantly skewed simulations, and one is unable to reproduce results from a limited set of single copy simulations, such as a preference for certain cavities and pathways, by using straight LES with as few as ten ligand copies (data not shown). To overcome these problems, improvements to the LES algorithm have been suggested and tested by [Ulitsky and Elber \(1993\)](#). Comparison of the original LES and its variant with single-copy dynamics highlights the shortcomings of straight LES, as well as the success of the LES variant, in reproducing the correct dynamics ([Ulitsky and Elber, 1994](#)).

In our own attempt to improve the LES method, we sample a constant temperature ensemble by coupling the actual protein and the replicated gas ligands to different Langevin heat baths. The idea of controlling the replicated particles' temperature was originally suggested in [Straub and Karplus \(1991\)](#) and was also applied in [Czerminski and Elber \(1991\)](#). For the regular particles, we use a temperature target of $T_A = 310$ K, while for the N replicated gas particles, the target must be set to T_A/N or lower, such that the resulting "effective" temperature of the replicated particles, as measured in the simulation, is close to 310 K. As a result, the average kinetic energy is the same for all particles, and the sam-

pling of phase space corresponds to a 310 K constant temperature ensemble. Maintaining parts of the system at two different formal temperatures keeps the system out of equilibrium, and, as such, the LES copies will have a tendency to heat up, whereas the protein matrix will want to cool down. In practice, we found that in order to keep the LES copies at a stable temperature, a larger Langevin friction coefficient was needed for the replicated gas particles (since they are surrounded by a much larger bath of protein). We used a Langevin damping term of 5 ps^{-1} for the unreplicated particles, and a term of 10 ps^{-1} for the replicated particles. The resulting temperature distribution of the TLES-replicated particles is broader than that of the unreplicated system (we measure a normalized standard deviation for the temperature fluctuations of 10.0 K for the replicated particles and of 1.2 K for the equivalent non-LES simulation). Using TLES, the only approximation made is that instead of feeling one gas molecule, the environment feels a delocalized cloud of gas. Finally, in order to drastically speed up the simulation, we ignored the contribution of the gas molecule to the system's charge distribution when computing PME electrostatics. This gives exact results in the present case, since our models for O_2 and H_2 contain no partial charges. The increased sampling is thus achieved predominantly by increasing the number of copies (reducing entropic barriers), and not by altering the energy landscape, as is done in the original LES method.

Simulation Parameters

O_2 and H_2 gas access was investigated by all-atom MD simulations of the diffusion of O_2 and H_2 molecules inside Cpl, originating at the active site. We used a model for the gas molecules that did not include any partial charges, and for which O_2 and H_2 differed only in their Lennard-Jones parameters (H_2 : $\epsilon = -0.1521$ kcal/mol, $R_{\text{min}}/2 = 1.7682$ Å; O_2 : $\epsilon = -0.022$ kcal/mol, $R_{\text{min}}/2 = 1.32$ Å; according to the CHARMM form of the Lennard-Jones formula), bond spring constants and bond spring lengths (H_2 : 0.72 Å; O_2 : 1.16 Å). As stated before, we used a heavy version of dihydrogen hH_2 instead of H_2 , so that we could compare the diffusion properties of O_2 and H_2 based on their size differences alone.

Our model of hydrogenase was based on the X-ray structure of Cpl [FeFe]-hydrogenase ([Peters et al., 1998](#)) (PDB accession code 1feh). A series of atoms in the active site, or H cluster, were missing from the Protein Database structure and have been modeled here as a di(thiomethyl)amine bridge between the two H cluster sulfur atoms, as suggested by later studies ([Nicolet et al., 2002](#); [Fan and Hall, 2001](#)). The partial charges for the rest of the H cluster were based on a density functional theory calculation on the 2-oxidation state ([Torres et al., 2003](#)), and individual charges were adjusted by a maximum of $\pm 0.02e$ to guarantee the system's charge neutrality. The structure was embedded in a water box, resulting in a 57,000 atom system consisting of 9,000 hydrogenase atoms, 16,000 water molecules, and 15 sodium ions to cancel the excess integer charge.

The system was then equilibrated at a constant temperature (310 K) and pressure (1 atm) for a duration of 1 ns. The last frame of this equilibration was used as a starting point for all subsequent simulations. Aside from the initial equilibration, all simulations were performed at constant volume and temperature (310 K). In all cases, periodic boundary conditions were used. Temperature was regulated within the TLES approach by using Langevin dynamics with damping constants of 5 ps^{-1} for unreplicated atoms and 10 ps^{-1} for replicated gas atoms, respectively. Multiple time stepping was used, with integration time steps of 1 fs, 2 fs, and 4 fs, respectively, for bonded, nonbonded, and long-range electrostatic interactions (a nonbonded time step of 1 fs was used for the case of hH_2 TLES simulations because of energy stability issues when replicated hH_2 molecules enter the water solution environment). Particle Mesh Ewald with a grid resolution of better than 1 Å was used for long-range electrostatics, and all other nonbonded interactions were calculated by using a cutoff of 12 Å. The CHARMM22 force field ([MacKerell et al., 1992, 1998](#)) was employed for all protein interactions, and simulations were performed by using the NAMD ([Kalé et al., 1999](#)) MD software, modified by the present authors to allow for TLES.

Appendix

For the interested reader, we list the residues of Cpl that form the gas pathways A and B according to our study. The common central

cavity is defined by Cpl residues 272, 275, 276, 279, 297, 299, 324, 417, and 423; the central cavity extends into pathway A, defined by residues 274, 280, 283, 284, 287, 293, 295, 424, 427, 428, 431, 459, 461, 466, 467, 468, 492, and 493 and into pathway B, defined by residues 275, 278, 279, 321, 327, 330, 331, 334, 551, 552, 555, 556, 563, 564, 567, 568.

Acknowledgments

We wish to thank Maria Ghirardi and Wesley B. Jones at the National Renewable Energy Laboratory for helpful discussions. This work was supported by the National Institutes of Health Research grant P41-RR05969, the National Science Foundation supercomputer time grant NRAC MCA93S028, the DOE Hydrogen Fuel Cells and Infrastructure Technologies Program (K.K., P.K., and M.S.), the Office of Basic Energy Sciences, DOE (K.K., P.K., and M.S.) and by a DOE Hydrogen Fuel Initiative Award (J.C. and R.S.). Three-dimensional molecular graphics were generated by using the VMD visualization software (Humphrey et al., 1996).

Received: March 18, 2005

Revised: May 17, 2005

Accepted: May 23, 2005

Published: September 13, 2005

References

- Adams, M.W.M. (1990). The structure and function of iron-hydrogenase. *Biochim. Biophys. Acta* 1020, 115–145.
- Amara, P., Andreoletti, P., Jouve, H.M., and Field, M.J. (2001). Ligand diffusion in the catalase from *Proteus mirabilis*: a molecular dynamics study. *Protein Sci.* 10, 1927–1935.
- Bossa, C., Anselmi, M., Roccatano, D., Amadei, A., Vallone, B., Brunori, M., and Nola, A.D. (2004). Extended molecular dynamics simulation of the carbon monoxide migration in sperm whale myoglobin. *Biophys. J.* 86, 3855–3862.
- Brunori, M., and Gibson, Q.H. (2001). Cavities and packing defects in the structural dynamics of myoglobin. *EMBO Rep.* 2, 676–679.
- Brunori, M., Bourgeois, D., and Vallone, B. (2004). The structural dynamics of myoglobin. *J. Struct. Biol.* 147, 223–234.
- Carlson, M.L., Regan, R.M., and Gibson, Q.H. (1996). Distal cavity fluctuations in myoglobin: protein motion and ligand diffusion. *Biochemistry* 35, 1125–1136.
- Cohen, J., Kim, K., Posewitz, M., Ghirardi, M.L., Schulten, K., Seibert, M., and King, P. (2005). Molecular dynamics and experimental investigation of H₂ and O₂ diffusion in [Fe]-hydrogenase. *Biochem. Soc. Trans.* 33, 80–82.
- Connolly, M.L. (1983). Solvent-accessible surfaces of proteins and nucleic acids. *Science* 221, 709–713.
- Czerminski, R., and Elber, R. (1991). Computational studies of ligand diffusion in globins: I. Leghemoglobin. *Proteins* 10, 70–80.
- Elber, R., and Karplus, M. (1990). Enhanced sampling in molecular dynamics: use of the time-dependent Hartree approximation for a simulation of carbon monoxide diffusion through myoglobin. *J. Am. Chem. Soc.* 112, 9161–9175.
- Fan, H.-J., and Hall, M.B. (2001). A capable bridging ligand for Fe-only hydrogenase: density functional calculations of a low-energy route for heterolytic cleavage and formation of dihydrogen. *J. Am. Chem. Soc.* 123, 3828–3829.
- Flynn, T., Ghirardi, M.L., and Seibert, M. (2002). Accumulation of O₂-tolerant phenotypes in H₂-producing strains of *Chlamydomonas reinhardtii* by sequential applications of chemical mutagenesis and selection. *Int. J. Hydr. Ener.* 27, 1421–1430.
- Gerber, R., Buch, V., and Ratner, M. (1982). Time-dependent self-consistent field approximation for intramolecular energy transfer. I. Formulation and application to dissociation of van der Waals molecules. *J. Chem. Phys.* 94, 3022–3030.
- Ghirardi, M.L., Zhang, L., Lee, J.W., Flynn, T., Seibert, M., Greenbaum, E., and Melis, A. (2000). Microalgae: a green source of renewable H₂. *Trends Biotechnol.* 18, 506–511.

Gibson, Q.H., Regan, R., Elber, R., Olson, J.S., and Carver, T.E. (1992). Distal pocket residues affect picosecond ligand recombination in myoglobin. *J. Biol. Chem.* 267, 22022–22034.

Humphrey, W., Dalke, A., and Schulten, K. (1996). VMD – visual molecular dynamics. *J. Mol. Graph.* 14, 33–38.

Kalé, L., Skeel, R., Bhandarkar, M., Brunner, R., Gursoy, A., Krawetz, N., Phillips, J., Shinzaki, A., Varadara jan, K., and Schulten, K. (1999). NAMD2: greater scalability for parallel molecular dynamics. *J. Comp. Phys.* 151, 283–312.

Kocher, J.-P., Prevost, M., Wodak, S.J., and Lee, B. (1996). Properties of the protein matrix revealed by the free energy of cavity formation. *Structure* 4, 1517–1529.

Lemon, B.J., and Peters, J.W. (1999). Binding of exogenously added carbon monoxide at the active site of the iron-only hydrogenase (Cpl) from *Clostridium pasteurianum*. *Biochemistry* 38, 12969–12973.

MacKerell, A.D., Jr., Bashford, D., Bellott, M., Dunbrack, R.L., Field, M.J., Fischer, S., Gao, J., Guo, H., Ha, S., Joseph, D., et al. (1992). Self-consistent parameterization of biomolecules for molecular modeling and condensed phase simulations. *FASEB J.* 6, A143.

MacKerell, A.D., Jr., Bashford, D., Bellott, M., Dunbrack, R.L., Evanseck, J.D., Field, M.J., Fischer, S., Gao, J., Guo, H., Ha, S., et al. (1998). All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J. Phys. Chem. B* 102, 3586–3616.

Mertens, R., and Liese, A. (2004). Biotechnological applications of hydrogenases. *Curr. Opin. Biotechnol.* 15, 343–348.

Montet, Y., Amara, P., Volbeda, A., Vernece, X., Hatchikian, E.C., Field, M.J., Frey, M., and Fontecilla-Camps, J.C. (1997). Gas access to the active site of Ni-Fe hydrogenase probed by X-ray crystallography and molecular dynamics. *Nat. Struct. Biol.* 4, 523–526.

Nadler, W., and Stein, D.L. (1996). Reaction-diffusion description of biological transport processes in general dimension. *J. Chem. Phys.* 104, 1918–1936.

Nicolet, Y., Piras, C., Legrand, P., Hatchikian, C.E., and Fontecilla-Camps, J.C. (1999). *Desulfovibrio desulfuricans* iron hydrogenase: the structure shows unusual coordination to an active site Fe binuclear center. *Structure* 7, 13–23.

Nicolet, Y., Cavazza, C., and Fontecilla-Camps, J.C. (2002). Fe-only hydrogenases: structure, function and evolution. *J. Inorg. Biochem.* 91, 1–8.

Peters, J.W. (1999). Structure and mechanism of iron-only hydrogenases. *Curr. Opin. Struct. Biol.* 9, 670–676.

Peters, J.W., Lanzilotta, W.N., Lemon, B.J., and Seefeldt, L.C. (1998). X-ray crystal structure of the Fe-only hydrogenase (Cpl) from *Clostridium pasteurianum* to 1.8 Å resolution. *Science* 282, 1853–1858.

Richards, F.M. (1977). Areas, volumes, packing, and protein structure. *Annu. Rev. Biophys. Bioeng.* 6, 151–176.

Roitberg, A., and Elber, R. (1991). Modeling side chains in peptides and proteins: application of the locally enhanced sampling technique and the simulated annealing methods to find minimum energy conformations. *J. Chem. Phys.* 95, 9277–9287.

Straub, J.E., and Karplus, M. (1991). Energy equipartitioning in the classical time-dependent Hartree approximation. *J. Chem. Phys.* 94, 6737–6739.

Torres, R.A., Lovell, T., Noodleman, L., and Case, D.A. (2003). Density functional and reduction potential calculations of Fe₂S₄ clusters. *J. Am. Chem. Soc.* 125, 1923–1936.

Ulitsky, A., and Elber, R. (1993). The thermal equilibrium aspects of the time dependent Hartree and the locally enhanced sampling approximations: formal properties, a correction, and computational examples for rare gas clusters. *J. Phys. Chem.* 98, 3380–3388.

Ulitsky, A., and Elber, R. (1994). Application of the locally enhanced sampling (LES) and a mean field with a binary collision correction (cLES) to the simulation of Ar diffusion and NO recombination in myoglobin. *J. Phys. Chem.* 98, 1034–1043.