Studying O_2 pathways in [NiFe]- and [NiFeSe]-hydrogenases

Tiago M. Barbosa, Carla S. A. Baltazar, Davide R. Cruz, Diana Lousa and Cláudio M. Soares*

ITQB, Instituto de Tecnologia Química e Biológica António Xavier, Universidade Nova de Lisboa, Av. da República, 2780-157 Oeiras, Portugal

CORRESPONDING AUTHOR: Cláudio M. Soares

EMAIL: claudio@itqb.unl.pt

Abstract

Hydrogenases are efficient biocatalysts for H_2 production and oxidation with various potential biotechnological applications. [NiFe]-class hydrogenases are highly active in both production and oxidation processes but suffer from being sensitive to O_2. [NiFeSe] hydrogenases are a subclass of [NiFe] hydrogenases with an increased tolerance to aerobic environments. In this study we aim to understand the structural causes of the low sensitivity of [NiFeSe]-hydrogenases, when compared with the [NiFe] enzymes, by studying the diffusion of O_2. To unravel the differences between the two enzymes, we used computational methods comprising Molecular Dynamics simulations with explicit O_2 and Implicit Ligand sampling methodologies. With the latter, we were able to map the free energy landscapes for O_2 permeation in both enzymes. We derived pathways from these energy landscapes and selected the kinetically more relevant ones with reactive flux analysis using transition path theory. These studies evidence the existence of quite different pathways in both enzymes and predict a lower permeation efficiency for O_2 in the case of the [NiFeSe]-hydrogenase when compared with the [NiFe] enzyme. These differences can explain the experimentally observed lower inhibition by O_2 on [NiFeSe]-hydrogenases, when compared with [NiFe]-hydrogenases. A comprehensive map of the residues lining the most important O_2 pathways in both enzymes is also presented.

Introduction

Hydrogenases are metalloenzymes that catalyse the reaction of $H_2 \leftrightharpoons 2H^+ + 2e^-$ [1–4]. Functioning at a high turnover frequency, they are considered the most efficient noble-metal free H_2 production and oxidation catalysts, being at least as effective as economically expensive platinum based catalysts [5–7]. Their applications are many, ranging from fuel cells to electro- and photocatalysis [5–7]. Studying their catalytic mechanisms is very important for making H_2 an economically viable, carbon-free alternative to current energy sources. Most hydrogenases are sensitive to O_2, which is one of the major problems for their use in large scale applications [3]. Therefore, studying the behaviour of O_2 inside the structure can be extremely valuable and may open new avenues in their engineering.

The nomenclature and classification of hydrogenases lies on the composition of their bimetallic active centre, with [NiFe] and [FeFe] hydrogenases being the two most common hydrogenases in nature [2]. [FeFe] hydrogenases are generally irreversibly inactivated and damaged by O_2 [8], while the [NiFe]-class shows a more diverse behaviour towards exposure [9]. Standard [NiFe] hydrogenases are generally O_2-sensitive with inactivation occurring by the formation of a mixture of two inactive states (Ni-A and Ni-B) in the active centre [10,11]. While in an inactive state, the Ni ion is in a Ni(III) oxidation state and a bridging peroxo group or hydroxo ligand is present between the Ni and Fe ions; other modifications such may also contribute to the inhibition, as the oxidation of the sulfur ligands [4,9,12–15]. [NiFeSe]-hydrogenases are a subclass of [NiFe]-hydrogenases which are characterized by having a selenocysteine coordinating the Nickel in the active site [16] and have very interesting properties, such as high catalytic activities and lower sensitivity to O_2, when compared to [NiFe]-hydrogenases, making them more suited to biotechnological applications[16,17].

Recovery time from the oxidised states caused by exposure to O_2 is remarkably different in both hydrogenases, as [NiFeSe] is extremely fast, while standard [NiFe] can take several hours [18].

Structurally, both [NiFe]- and [NiFeSe]-hydrogenase enzymes are almost identical, being comprised of a minimum of two subunits. The active site lies deep inside the large subunit, while the small subunit generally contains three iron-sulfur clusters in a wire like formation, forming an electron transfer chain between the active site and the enzyme surface. The exact cluster composition differs in both enzymes: in the O_2 tolerant [NiFeSe]-hydrogenase the iron-sulfur clusters are all [4Fe4S], while in [NiFe]-hydrogenase there are two [4Fe4S] and one [3Fe4S] [4,19–21].

Several structural features of the [NiFeSe] hydrogenase have been evidenced to explain its catalytic prowess: the “cage effect” of the protein structure surrounding the active site [22], differences in residues comprising proton transfer pathways and H_2 channels [23] and the nature of the selenocysteine complex. The complex has been suggested [20] to have a severe influence on the O_2 sensitivity of the [NiFeSe]-hydrogenase, both by promoting the rapid recovery from O_2 damage as well for increasing H_2 production [24–26]. Other factors, such as the access of O_2 to the active site, may also play a role in the unique feature of [NiFeSe] hydrogenases.

Determining the O_2 paths via experimental methods is very challenging as O_2 is very mobile, has a low electron count and weak interaction with amino acids [27]. Computational studies on the subject are then a valuable way to propose pathways by directly observing a representation of the nature of the phenomenon in an atomic level.

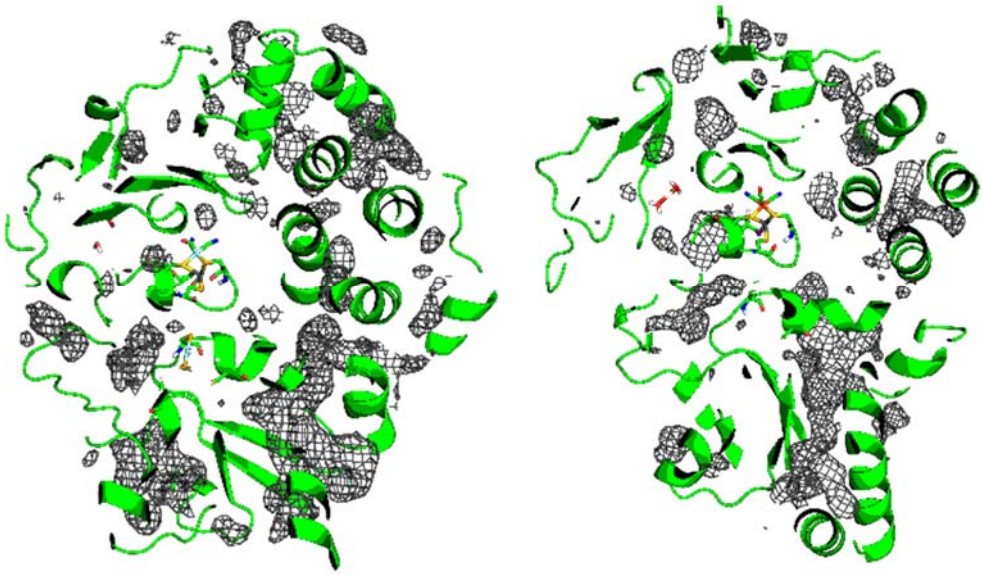
Permeation pathways for entrance of H_2 in [NiFe]- and [NiFeSe]-hydrogenases have been studied using computational methods by us and other authors[23,28,29], but the subject of O_2 permeation has been less studied [27,28], and, to the best of our knowledge, never studied on a [NiFeSe] hydrogenase. Therefore, the aim of the present work is to

study a [NiFe]- and a [NiFeSe]-hydrogenase to compare their differences in O_2 internalization, diffusion and protection inside the protein structures, trying to unravel the structural and dynamic differences that might explain the different O_2-sensitivity. With the present study, we were able to map the free energy landscape for O_2 permeation on both enzymes and found very marked differences. Analysing these landscapes using probabilistic models has shown evidence for a more defined pathway for O_2 internalization in [NiFe]-hydrogenase and a more diffuse and less specific set of pathways in [NiFeSe]-hydrogenase.

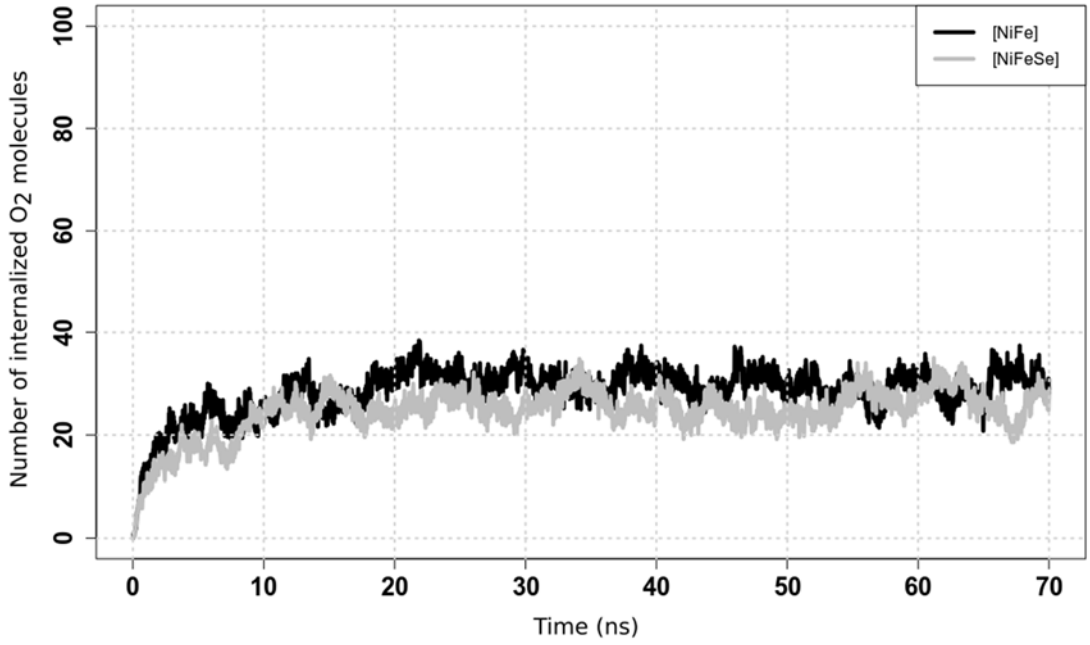
Results and discussion

MD simulations of both enzymes in water show a stability plateau after about 20-30ns, as can be seen in Figure S1 of Supporting Information, displaying the root mean square deviations (RMSD's). Additionally, introducing the O_2 in the system did not compromise this stability.

To illustrate O_2 internalization we calculated average Probability Density Functions (PDFs) from the five trajectories calculated for each hydrogenase (Figure 1). The probability maps show similar patterns of internalization on both enzymes, with a main channel in line with the nickel-iron and iron-sulfur centres. There are also diffuse zones of higher probability all around both enzymes and several zones where the probability is not continuous. There are not enough continuous O_2 zones near the active centres to define pathways. This is likely due to the insufficient sampling provided by the simulated timescale.



A **B**



C

*Figure 1 – Top –Slice of the protein structure with PDF's represented by wireframe meshes at probability 0.002. PDF's were calculated from 35ns to 70ns. The protein is represented by a green cartoon while metal centres are represented using sticks. The [NiFe] and [NiFeSe] centres are visible at the centre of the figure; Panel **A**– [NiFe], Panel **B**– [NiFeSe] hydrogenase. Panel **C**– Average number of internalized O_2 molecules over time for [NiFe]- (black line) and [NiFeSe]-hydrogenase (gray line).*

Figure 1 also contains a plot of O_2 internalization, which shows that the quantity of molecules internalised reaches a plateau at ~30 molecules out of the total 100, for both enzymes, and this process is relatively fast (~10ns) in both cases. From this data, we conclude that, within the simulated time scale, both the [NiFe] and the [NiFeSe]

enzymes, do not show any differences in the capacity to internalise and hold molecular oxygen.

Interesting as these results may be, it is also clear that the sampling obtained in the time scale of these simulations does not allow to adequately find clear paths for molecular oxygen permeation up until the active site zone. This is in contrast with our previous experience with molecular hydrogen in these hydrogenases, which rapidly reaches the active site [22,23,30] and this is certainly due to the significant larger size of molecular oxygen, when compared with molecular hydrogen. We have observed this type of situation before on oxygen metabolising enzymes [31,32] and the solution we resorted was to use ILS, which can explore higher energy zones in the permeation free energy surface. This was the route we decided to follow in the present work, and use the oxygen free trajectories of the enzymes in water to infer about the free energy surface of molecular oxygen placement, in the whole space of the hydrogenases.

By applying the ILS methodology to a trajectory, O_2 was *forced* in the whole space of both enzymes, mapping even the deeper structural layers. This comprehensive analysis allowed a detailed examination of the landscape near the active centre as well as the whole of the enzyme's conformational space. Figure 2 (A and D panels) displays the results of this method applied to the five trajectories of the [NiFe]- and [NiFeSe]-hydrogenases, respectively.

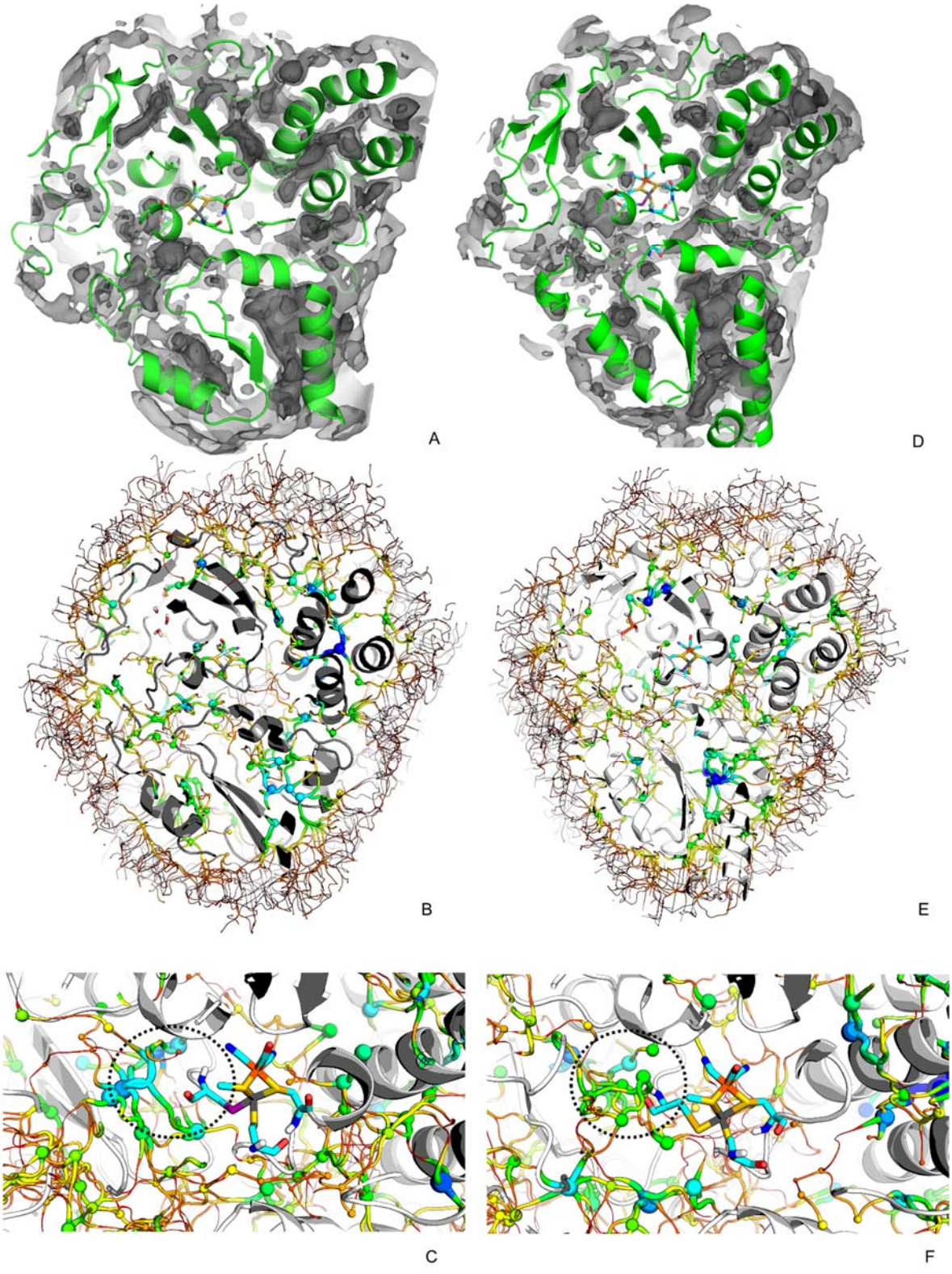


*Figure 2 – Panels **A**,**B**, and **C** – correspond to [NiFe]-hydrogenase; Panels **D**,**E** and **F** – correspond to [NiFeSe]-hydrogenase; Panels **A**, and **D** – contain the ILS isosurfaces for each hydrogenase – Energy cut-offs of -1, -5 and -10 kJ.mol^{-1}, coloured from lighter to darker grey; Panels **B** and **E** show an overview of the pathway tessellation for each hydrogenase - Minima are represented by spheres and the pathways by cartoon traces with thicknesses inversely proportional to the pathway energy. Both spheres and traces are colour coded from lower (-16 kJ.mol^{-1}) to higher (2 kJ.mol^{-1}) energy using blue, yellow, green, orange, and red, respectively; Panels **C** and **F** - Like panels **B** and **E**, with a zoom of the active site of both enzymes, highlighting (using a dotted black circle) the selenocysteine (panel **F**) zone in [NiFeSe]-hydrogenase and the corresponding cysteine (panel **C**) zone in [NiFe]-hydrogenase.*

These results are consistent with the MD simulations with molecular oxygen, as the lower energy zones are roughly similar with the higher PDF's evidenced on Figure 1. However, in the case of ILS results of Figure 2, low probability zones near the active sites are also present, evidencing the higher sampling power of ILS, when compared with the MD simulations with explicit O_2.

Panels B and E of Figure 2 show that the tessellation pathways are extremely intricate; a detailed visual observation (results not shown) evidences that the outline and surface of the [NiFe]-hydrogenase have large zones paved with low energy minima, while the [NiFeSe]-hydrogenase shows fewer and smaller zones being shielded at the surface by higher energy barriers.

Figure 2 also contains (panels C and F) the ILS pathways found near the active centres, represented by the minima (as spheres) and the paths between minima (as cartoon traces). By analysing these two landscapes near the active centres, it is evident that the [NiFe]-hydrogenase contains more low energy basins near the cysteine that is replaced by a selenocysteine (Sec) in the [NiFeSe] enzyme, the latter being relatively empty of basins in the same location (circled zones in panels C and F of Figure 2). This is already an indication for the higher difficulty of placing O_2 near the active site in the [NiFeSe]-hydrogenase, when compared with the [NiFe] counterpart. Therefore, the protein structure and dynamics of the [NiFeSe]-hydrogenase seem to be more tailored to reduce the O_2 access to the active site, when compared with the [NiFe]-hydrogenase, which can be used to explain the lower O_2 sensitivity of the former, when compared with the latter. This is interesting and in contrast with our findings for H_2 permeation [23], where, using MD simulations, we found higher density for H_2 in [NiFeSe]-hydrogenase, when compared with the [NiFe]-hydrogenase. This was the molecular basis that was suggested by us to explain the higher catalytic activity towards H_2 of [NiFeSe]-hydrogenase.

With flux analysis using transition path theory, we can calculate the net flux of O_2 from the exterior of the protein to the active site. From this overall analysis we determined the flux of O_2 to the active site of both hydrogenases, and the values are $5.28x10^{-5}$ and $1.2x10^{-5}$ for the [NiFe]-hydrogenase and the [NiFeSe]-hydrogenase, respectively. With this we put a number on the visual analyses present in Figure 2, clearly showing the higher capacity of [NiFe]-hydrogenase to permeate O_2, when compared with the [NiFeSe]-hydrogenase. As said above, this correlates well with the lower O_2 sensitivity of [NiFeSe]-hydrogenase.

There are a number of pathways contributing to the overall flux towards the active site of both hydrogenases. These are displayed in Figure 3 and quantified on Table 1, where the final energy basins are identified. We decided to highlight sets of pathways instead of individual ones, since these appear in interconnected clusters. Note also that the sum of the fluxes of the pathways on each enzyme does not correspond to the complete flux calculated, since these pathways communicate with each other. Figure 3 shows the paths on the whole protein with inset highlights of the active site zone. We selected the more prevalent reactive pathways for each hydrogenase. Each pathway comprises product basins apparently sharing the same reactive network. Table 1 describes the net flux values and the pathway selection.

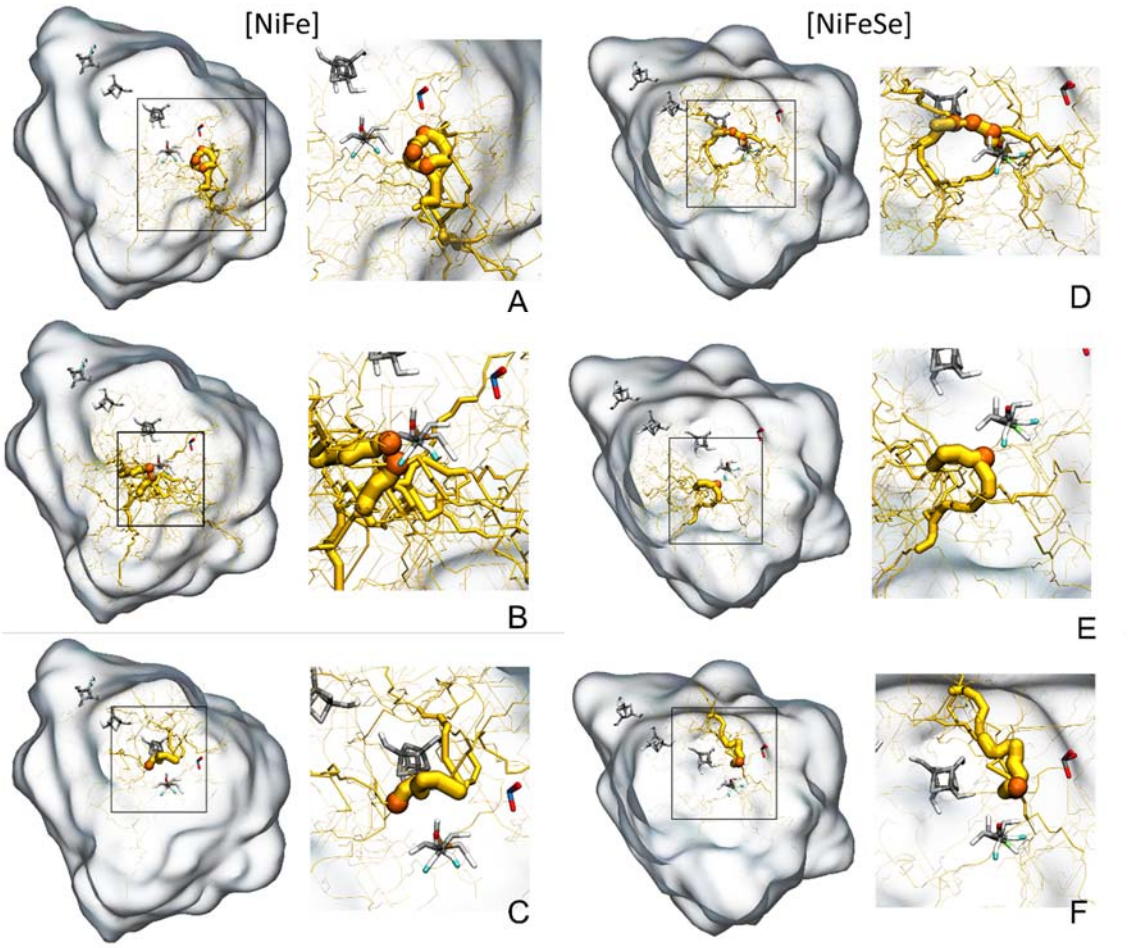


*Figure 3 – Main fluxes, found by Reactive Flux Analysis of the ILS pathways, targeting basins near the active centre. The net flux is represented by the yellow trace (thickness proportional to the normalized flux – non comparable between different paths); Target basins are represented by the orange spheres. Panels **A**, **B** and **C** correspond to [NiFe]-hydrogenase, while Panels **D**, **E** and **F** correspond to [NiFeSe]-hydrogenase. The different panels correspond to the following pathways (same designation as in Table 1): Panel **A** – NF-A; Panel **B** – NF-B; Panel **C** – NF-C; Panel **D** – NFS-A; Panel **E** – NFS–B; Panel **F** – NFS-C.*

Several entrance pathways were found in both hydrogenases, suggesting the presence of multiple entry points on the protein surface. The reactive pathways are remarkably different, converging to different points near the centres, suggesting multiple inactivation mechanisms and kinetics for each.

Table 1 – Grouping of reactive pathways per product basin. Percentages were calculated from the sum of all the fluxes per enzyme.

Enzyme	*Path*	*Product basin(s)*	*Total Flux*	*%of sum*
[NiFe]	NF-A	NF961	4.06×10^{-5}	36.03%
		NF963	3.09×10^{-5}	27.37%
		NF1004	2.64×10^{-5}	23.36%
	NF-B	NF1259	2.11×10^{-6}	1.87%
		NF1329	1.10×10^{-5}	9.76%
	NF-C	NF1430	1.82×10^{-6}	1.61%
[NiFeSe]	NFS-A	NFS1363	2.71×10^{-6}	17.97%
		NFS1510	1.80×10^{-6}	11.98%
		NFS1578	1.88×10^{-6}	12.51%
	NFS-B	NFS1129	1.72×10^{-6}	11.41%
	NFS-C	NFS1291	6.95×10^{-6}	46.12%

Overall, this analysis provides evidence for a main pathway to the active centre in the [NiFe]-hydrogenase (NF-A), which has dominant flux values, contrasting with the several representative pathways in the case of the [NiFeSe]-hydrogenase. Nevertheless, NFS-C is dominant in [NiFeSe]-hydrogenase. The values of the fluxes are considerably higher for most of the [NiFe]-hydrogenase target basins, when compared with the [NiFeSe] enzyme, which correlates well with the higher value of the total flux found for the former.

Path NF-A of [NiFe]-hydrogenase comprises three target basins sharing the same network. The pathway converges directly to the Ni coordinating $Cys530^{L}$, which is replaced by a Sec in [NiFeSe]-hydrogenases. This pathway has no representation in the [NiFeSe] hydrogenase and accounts for most of the flux, suggesting that it is the main inactivation spot. The presence of selenium in the [NiFeSe]-hydrogenase may influence inactivation, as it was suggested in previous research [24].

As for the NF-B pathway of [NiFe]-hydrogenae, its target basins are located near the Fe ion of [NiFe]-hydrogenase, and have similarly located and contiguous basins in the [NiFeSe]-hydrogenase (path NFS-B), suggesting that these two pathways are conserved among the two hydrogenases.

NF-C from [NiFe]-hydrogenase converges near the active centre coordinating $CYS323^{L}$ and has a very low flux. This pathway has correspondence with the NFS-C of the [NiFeSe]-hydrogenase pathway. Similarly to NF-C, pathway NFS-A is also comprised of three product basins, with their respective reactive networks, and converges to an intermediate location between the proximal FeS centre and the active centre in the [NiFeSe]-hydrogenase

These findings suggest that the preferred pathways for O_2 differ in both enzymes, possibly determining the inactivation mechanism, as the active site of the [NiFeSe]-hydrogenases is less exposed to O_2.

The fact that neither basins nor pathways are present near the selenocysteine (as opposed to the same space of the [NiFe] $Cys530^{L}$) suggests that the Sec or the surrounding environment may also have a role in the protection of the [NiFeSe] hydrogenase's centre.

To illustrate the differences on the hydrogenase's O_2 pathways, we identify all residues at a van der Waals distance of the highest fluxes (higher than one half of the maximum flux the pathway) and mapped on Figure 4. The corresponding residue of the other hydrogenase was also selected by aligning the two structures to check for conservation in both hydrogenases (Supporting material – tables S3-S4).

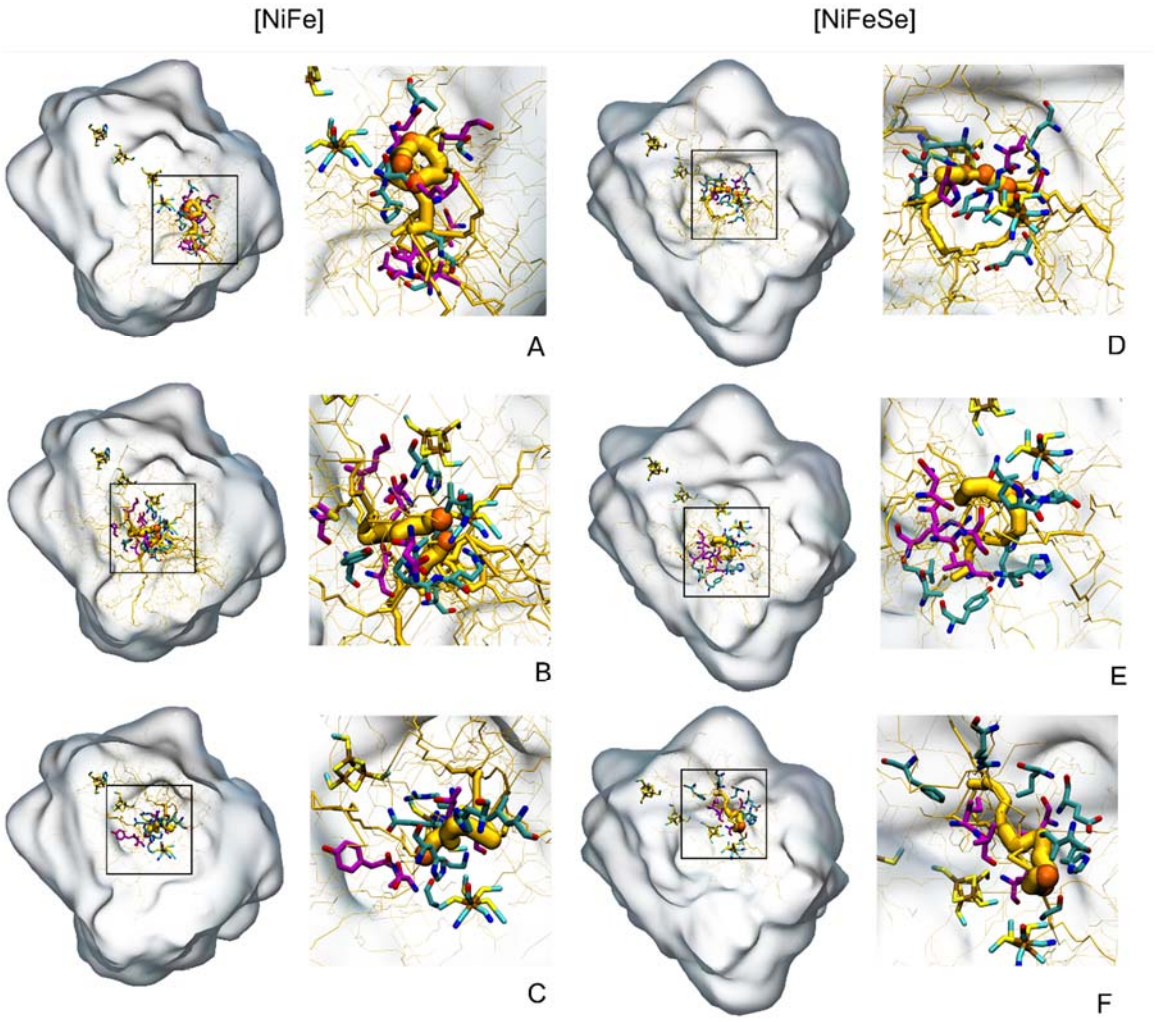


*Figure 4 – Representation of the residue conservation among the two hydrogenases near the O_2 pathways. The residues at van der Waals distance (considering the Se:O_2 distance) from the pathways are selected. Pathways are arranged in the same orientation as in Figure 3 (corresponding to the same Panels). Non-conserved residues are coloured with magenta carbons atoms, while conserved residues have their carbon atoms coloured cyan. Panels **A, B** and **C** correspond to [NiFe]-hydrogenase, while Panels **D, E** and **F** correspond to [NiFeSe]-hydrogenase. The different panels correspond to the following pathways (same designation as in Table 1): Panel **A** – NF-A; Panel **B** – NF-B; Panel **C** – NF-C; Panel **D** – NFS-A; Panel **E** – NFS–B; Panel **F** – NFS-C.*

There are multiple non conserved residues around the reactive pathways (SM-tables S3-S4) which is consistent with the fact that there are distinct pathways in each hydrogenase.

Our present results may suggest residues found in different pathways in both hydrogenases that can be mutated to change their characteristics towards O_2 inhibition. Actually, in a recent study done in the [NiFeSe] hydrogenase, one of the residues of NFS-1 identified here - GLY491 - was successfully mutated experimentally by a bulkier

ALA (the direct [NiFe] counterpart) and by a SER, leading to decreased O_2 inhibition while not affecting H_2 production in comparison to the wildtype [33]. This inhibition pathway does not exist in the [NiFe]-hydrogenase, which, according to our results, is mainly inhibited by the NF-1 path. Placing a bulkier residue in this position on [NiFeSe]-hydrogenase may eliminate or reduce the NFS-1 path, thereby reducing inhibition by O_2 even further. Other residues found by our analysis (listed in tables S3 and S4), may have similar impacts, in both hydrogenases.

Conclusions

Using two different approaches, the pathways of O_2 permeation were comprehensively mapped in two different [NiFe] class hydrogenase structures displaying different O_2 sensitivities. The methods used here consider, not only the structure, but the dynamic behaviour of the protein structures, allowing for a more realistic analysis that can deal with transient pathways for O_2 access. ILS in particular allows for a thermodynamic quantification of the O_2 affinity on the whole protein matrix, which, together with further analysis, allows for predicting the fluxes of O_2 from the exterior towards the active site of the enzymes.

We found marked differences in the diffusion patterns of both enzymes, being the [NiFe]-hydrogenase more prone for O_2 access and potential inactivation, when compared with the [NiFeSe]-hydrogenase. Additionally, there is evidence for different mechanisms for O_2 inactivation of each enzyme, which may help explain the different performances of both in aerobic settings. The pathways for inactivation were also mapped in an atomistic level, which may help understand the structural properties of the focal points of oxygen diffusion. This knowledge may prove useful in future manipulation towards the development of more efficient hydrogen catalysts that are less inhibit by O_2.

Methods

System setup

The X-ray structures of [NiFe] (*D. gigas* PDB ID 3frv) [20] and [NiFeSe] (*D. vulgaris* PDB ID 2wpn) [21] hydrogenases were used in this study. Each system was solvated in a rhombic dodecahedral water box using SPC water [34]. A minimum distance of 8Å between the protein and box walls was imposed. Each system was neutralized with Na^+ ions to counter act its negative charge. Protonation states were determined through a combination of PB/MC calculations/simulations using MEAD version 2.2.9 and PETIT version 1.6.0 respectively [30,35] at pH 7.0. These predicted that all lysine and arginine residues were positively charged, while glutamate and aspartate residues were considered negatively charged (but see details on supplementary material for an exception). Details on the Histidine protonation can be found in the supplementary material (Tables S1 and S2).

As for the O_2 molecule parameters, the model from Cordeiro[36] , which was parameterised to account for the solvation properties of molecular oxygen, both in aqueous as well as non-aqueous environments, was used in this work. As for the oxidation states we considered the Ni-SI_a state [37] for the active [NiFe] centres and the oxidized state for the [4Fe4S] clusters. More details on the parametrization of the metallic centres can be found in Baltazar *et al.* [23] and Teixeira *et al.* [38] for [NiFeSe]-hydrogenase and [NiFe]-hydrogenase, respectively.

Molecular dynamics simulations

The GROMOS 54A7 [39] forcefield and single point charge (SPC) water model [34] were used to describe the systems, and GROMACS version 5.0.7 [40] was used to perform all MD simulations. Five 100ns long simulations in solvent were performed for each system. These simulations were carried out with a constant number of particles, pressure (1 atm – controlled using a semi-isotropic Parrinello-Rahman barostat [41,42]), temperature (300K – controlled by a V-rescale thermostat [43]) and periodic boundary conditions. Different temperature couplings were applied to protein and solvent + O_2 atoms using a coupling constant of 0.1 ps. A pressure coupling constant of 1.6 ps was used. All solute bond lengths were constrained with the P-LINCS algorithm [44] while the SETTLE algorithm [45] was used for solvent. Equations of motion were integrated with a time-step of 2 fs, with neighbour lists being updated every 40 steps. Electrostatic interactions were treated with the Particle mesh Ewald method [46] with a real space cut-off at 10 Å and a Fourier grid spacing of 1.2 Å. The Verlet cut-off scheme was selected.

To remove unfavourable atomic contacts, the systems were energy minimized without positional restraints using a combination of steepest descent and Low memory Broyden-Fletcher-Goldfarb-Shanno algorithms [47]. System initialization comprises four 50 ps MD steps with velocities being generated from a Boltzmann distribution at the defined temperature. At the first step, in the *NVT* ensemble, the Berendsen thermostat [48] was utilized with positional restrains on the C-alpha atoms with force constant of 10000 kJ/mol $Å^2$. Pressure coupling using the Berendsen barostat [48] was added in the subsequent step with a coupling constant of 3ps. In the following step all parameters were kept, but the coupling constant was decreased to 2ps. In the final step all restraints were removed, the pressure coupling constant was reduced to 1.6ps, with the barostat being altered to Parrinello-Rahman and the thermostat to V-rescale.

A protocol was prepared to study O_2 diffusion assuring system stability, conformational variety and statistical accuracy. From the solvent only simulations a snapshot of each replicate was retrieved at the 30ns mark (assuring system stability). 100 water molecules were randomly selected from the outside of the protein structure and substituted by O_2 molecules. The velocities from the removed water oxygen and one of hydrogen atoms were kept and assigned to the inserted O_2. The remaining hydrogen atom and its velocity were discarded. 1 ns of equilibration with a smaller timestep (1 fs) was calculated so as the newly introduced molecules stabilize with the solvent (avoiding clashes). The simulations with O_2 were kept for a further 70ns amounting to a total of 350ns of simulation with explicit O_2 per system.

MD – O_2 distribution analysis

The VMD volmap plugin [49] was utilized calculate probability density functions (PDF's) of the O_2 distribution along the MD trajectories with explicit O_2. A total of 175000 frames per enzyme, corresponding to the final 35ns of each trajectory (of the five 70ns trajectories per enzyme) were used for this calculation, with a grid resolution of 1 Å. We calculated the internalization of O_2 using a previously implemented and described method [38]. Maps were visualized and images rendered using Pymol (The PyMOL Molecular Graphics System, Version 1.8, Schrödinger, LLC) and VMD [49].

Implicit ligand Sampling

The implicit ligand sampling [50] (ILS) method was used to calculate the free energy of transferring O_2 from pure water to anywhere inside both hydrogenases and surrounding

environment. This method allows for studying the whole landscape of molecular oxygen placement, even regions such as the deep lying hydrogenase active site, where explicit molecules of O_2 have difficulties in reaching within the time scale of the simulation. This methodology uses molecular dynamics simulations of the system without molecular oxygen, in contrast with the previously described simulations.

From the ILS method the potential mean force ($PMF(\boldsymbol{r})$) of having a diatomic ligand at a position **r** is given by:

$$PMF(\boldsymbol{r}) = -k_b T \ln \sum_{m=1}^{M} \sum_{k=1}^{C} \frac{e^{-(k_b T)^{-1} \Delta E(\boldsymbol{r}, \boldsymbol{q_m}, \Omega_k)}}{MC} \quad (1)$$

where M is the number of utilized protein-solvent configurations, C is the number of random orientations of the ligand and $\Delta E(\boldsymbol{r}, \boldsymbol{q_m}, \Omega_k)$ is the protein-solvent interaction energy in the configuration $\boldsymbol{q_m}$ with the diatomic ligand located at $\boldsymbol{r}$ with an orientation Ω_k. Non-bonded interactions (electrostatic and van der Waals) are accounted by $\Delta E(\boldsymbol{r}, \boldsymbol{q_m}, \Omega_k)$. In the O_2 model used [36], given that it has no partial charges, only van der Waals interactions were considered. For performing these calculations, a modified version [51] of the GROMACS 4.5.4 Widom TPI algorithm was used to perform ILS [51]. The last 10ns of the five MD trajectories in water were used (accounting in total for ~25000 configurations for each enzyme), with the configurations being fitted to the C-alpha atoms of the energy minimized structure. Grids of 58x62x61 Å and 62x62x63 Å dimensions wad used in the calculations for the [NiFe] and [NiFeSe] structures, respectively. For each grid point, 400 insertions in random positions and orientations (*C* in equation 1) per grid cube were made. The results of all calculations were averaged for each system resulting in two discretized scalar fields (3D energy landscapes). These landscapes detail the Gibbs free energy of moving O_2 from vacuum to a given position of the system, $\Delta G_{vac \rightarrow prot}(O_2)$. Finally, as our interest is to study a landscape of the Gibbs-free energy of moving O_2 from a position in water to a position in the

system, $\Delta G_{wat \rightarrow prot}(O_2)$, we made additional simulations to calculate the free energy of moving O_2 from the vacuum to water, $\Delta G_{vac \rightarrow wat}(O_2)$ and subtracting it to every grid point of $\Delta G_{vac \rightarrow prot}(O_2)$.

To calculate $\Delta G_{vac \rightarrow wat}(O_2)$ we adopted a method [51] , which takes 10ns pure water simulations in the NPT ensemble and applies the ILS method to the final 2000 conformations (2 ns). The resulting 3D landscape of this calculation was then averaged over all the grid points resulting in the final $\Delta G_{vac \rightarrow wat}(O_2)$. The calculated value was of 8.30 kJ/mol for the O_2 model [36] used.

ILS - free energy landscape analysis

ILS details extensively the free energy landscape of both enzymes. Using that information, it is possible to infer low energy pathways (more thermodynamically favourable) of O_2 inside the structures. To achieve this, a previously implemented method [51] extending on another previous approach [52] was adopted. This method starts by linking each grid point to the neighbour grid point of lowest energy (neighbours are defined as the adjacent 26 grid points forming a 3x3x3 cube around it) until a local minimum is found. All grid points 'falling' to the same minima are grouped into sets and classified as basins. After the classification, the algorithm identifies the lowest energy points within the boundaries between each pair of neighbouring basins – the saddle points. A network of paths between all energy minima of the landscape can then be constructed using the steepest-descent paths from the saddle points to the minima.

O_2 Diffusion Kinetics Modelling

ILS provides an exhaustive sampling over the energy landscape of the whole system (including high-energy regions) representing a suitable model for a kinetic analysis. In addition, classifying the energy landscape into basins provides a division of the landscape into macrostates. Considering these basins as belonging to the state space of O_2 diffusion inside the two hydrogenases a Markov process describing the time-discrete evolution of the system in the state space can be constructed. The construction of the representative model relies on calculating a transition probability matrix where each element $T_{ij}(\Delta t)$ corresponds to the probability of transition to basin/state j after a time Δt when being in a basin i at an arbitrary time. As ILS does not provide statistics of these dynamics in the state space the matrix was inferred from the energy landscape using Metropolis sampling for the jumps between neighbour grid points. Following Kramer's assumption (assuming the grid-point probability distribution within any state i at time t can be approximated by the steady state of state i) the transition probability from two different states (i, j) can be calculated using the following method [53] :

$$T_{ij}(\Delta t) = \frac{1}{Z_i(3^D - 1)} \sum_{x \in i} \sum_{\substack{y \in j \\ y \sim j}} \{e^{-\beta E(x)}, e^{-\beta E(y)}\} \qquad (2)$$

Where Z_i is the partition function of state i given by $\sum_{x \in i} e^{-\beta E(x)}$, D is number of dimensions of the landscape, x and y are the neighbour grid points that belong to the border, $\beta = 1/k_b T$ representing k_b as the Boltzmann constant and T the absolute temperature of the system and finally $E(x)$ representing the energy at the grid point x.

The self-transition probabilities $T_{ii}(\Delta t)$ were calculated as $1 - \sum_{i \neq j} T_{ii}(\Delta t)$. Using this method, a Markov model was constructed for each ILS 3D energy landscape for all

transitions with a cut-off for saddle pair energy of < 40 kj.mol^{-1}. Therefore, this model excludes very low probability transitions and very hard to reach states. As the solvent states were not crucial in the model building, they were coarse grained into a single state. Denoting the probability of a state i at a time t as $p_i(t)$, the time discrete evolution for the Markov chain can be inferred by:

$$p_j(t + \Delta t) = \sum_i p_i(t) T_{ij}(\Delta t) \tag{3}$$

Iterating this Markov chain for $t \rightarrow \infty$ gives the equilibrium of the stationary probability distribution $\pi_i = p_i(\infty)$, obeying to the invariance relation $\pi_j = \Sigma_i \pi_i T_{ij}(\Delta t)$. The iteration process from any starting probability distribution, $p_i(0) \neq \pi_i$, corresponds to a relaxation process toward π_i, where $T_{ij}(\Delta t)$ is calculated from the above method (see equation 2). As the border is the same for any given states pair ij the detailed balance relation $\pi_i T_{ij}(\Delta t) = \pi_j T_{ji}(\Delta t)$ is also verified. The iteration of the Markov chain (equation 3) utilized a probability distribution of

$$p_i(0) = \begin{cases} 1 & i = solvent \\ 0 & i \neq solvent \end{cases} \tag{4}$$

Reactive Flux Analysis – Transition Path Theory

We applied transition path theory (TPT)[54] to the resultant Markov model in order to characterize the transition pathways and calculate reactive fluxes between the solvent state and the product state. Our approach is based on finding the subsets in the whole ensemble of transitions, which we can consider trajectories of molecular oxygen, leaving the solvent state (reagent) and continue until reaching the catalytic [NiFe] and [NiFeSe] centres (product states), and consider them reactive trajectories. As we cannot still pinpoint the exact place of the inactivation inside of both hydrogenases, all

basins in contact (we considered the Selenium-oxygen van der Waals radius as the contact distance) with the most distant atom of the cysteines connected to the Nickel-Iron centre were considered product states and trajectories leading to those basins were considered reactive trajectories.

Using TPT the reactive trajectories were statistically characterized using committors (forward and backward). In our case the forward committor is defined as the probability that a process will reach first the product state than the solvent state, being the backwards committor the inverse. TPT also allows for the calculation of the effective flux, the net average number of reactive trajectories per time unit that transition from state i to state j while converging to the product states. Each basin was considered as a state and the pathways reactive trajectories. These calculations were performed using the PyEmma software [55]. Details on the use of this methodology to a similar system can be found in Damas *et al.* work [32].

Author Contributions Statement

T.M.B. and C.M.S. designed the experiments that were carried out by T.M.B., C.S.A.B., D.R.C. and D.L. T.M.B. prepared figures. T.M.B. and C.M.S. wrote the manuscript. All authors reviewed the manuscript.

Acknowledgments

We would like to acknowledge helpful discussions with João M. Damas, Pedro R. Magalhães, António M. Baptista, Pedro M. Matias and Inês C. Pereira. This work was financially supported by FCT - Fundação para a Ciência e a Tecnologia, Portugal, through project PTDC/BBB-BEP/2885/2014, and fellowships SFRH/BD/73369/2010 (to C.S.A.B.), SFRH/BD/52205/2013 (to D.R.C.) and SFRH/BPD/92537/2013 (to D.L.).

This work was also financially supported by Project LISBOA-01-0145-FEDER-007660 (Microbiologia Molecular, Estrutural e Celular) funded by FEDER funds through COMPETE2020 - Programa Operacional Competitividade e Internacionalização (POCI) and by national funds through FCT - Fundação para a Ciência e a Tecnologia.

Ethics declarations

Competing interests

The authors declare no competing interests.

References

1. Stephenson, M. & Stickland, L. H. Hydrogenase: a bacterial enzyme activating molecular hydrogen: The properties of the enzyme. *Biochem. J.* **25**, 205–214 (1931).

2. Vignais, P. M. & Billoud, B. Occurrence, Classification, and Biological Function of Hydrogenases: An Overview. *Chem. Rev.* **107**, 4206–4272 (2007).

3. Lubitz, W., Ogata, H., Rdiger, O. & Reijerse, E. Hydrogenases. *Chemical Reviews* **114**, 4081–4148 (2014).

4. Fontecilla-Camps, J. C., Volbeda, A., Cavazza, C. & Nicolet, Y. Structure/Function Relationships of [NiFe]- and [FeFe]-Hydrogenases. *Chem. Rev.* **107**, 4273–4303 (2007).

5. Vincent, K. A. *et al.* Investigating and Exploiting the Electrocatalytic Properties of Hydrogenases. *Chem. Rev.* **107**, 4366–4413 (2007).

6. Karyakin, A. A. *et al.* Hydrogenase electrodes for fuel cells. *Biochem. Soc. Trans.* **33**, 73–75 (2005).

7. Jones, A. K., Sillery, E., Albracht, S. P. J. J. & Armstrong, F. A. Direct comparison of the electrocatalytic oxidation of hydrogen by an enzyme and a platinum catalyst. *Chem. Commun. (Camb).* **2**, 866–867 (2002).

8. De Lacey, A. L., Fernández, V. M., Rousset, M. & Cammack, R. Activation and inactivation of hydrogenase function and the catalytic cycle: Spectroelectrochemical studies. *Chemical Reviews* **107**, 4304–4330 (2007).

9. Vincent, K. A. *et al.* Electrochemical definitions of O2 sensitivity and oxidative inactivation in hydrogenases. *J. Am. Chem. Soc.* **127**, 18179–18189 (2005).

10. Marques, M. C., Coelho, R., Pereira, I. A. C. & Matias, P. M. Redox state-dependent changes in the crystal structure of [NiFeSe] hydrogenase from Desulfovibrio vulgaris Hildenborough. *Int. J. Hydrogen Energy* **38**, 8664–8682 (2013).

11. Ogata, H., Lubitz, W. & Higuchi, Y. Structure and function of [NiFe] hydrogenases. *Journal of Biochemistry* **160**, 251–258 (2016).

12. Ogata, H. *et al.* Activation process of [NiFe] hydrogenase elucidated by high-resolution X-ray analyses: Conversion of the ready to the unready state. *Structure* **13**, 1635–1642 (2005).

13. Volbeda, A. *et al.* Structural differences between the ready and unready oxidized states of [NiFe] hydrogenases. *J. Biol. Inorg. Chem.* **10**, 239–249 (2005).

14. Cracknell, J. A., Wait, A. F., Lenz, O., Friedrich, B. & Armstrong, F. A. A kinetic and thermodynamic understanding of O2 tolerance in [NiFe]-hydrogenases. *Proc. Natl. Acad. Sci.* **106**, 20681–20686 (2009).

15. Fernandez, V. M., Hatchikian, E. C. & Cammack, R. Properties and reactivation of two different deactivated forms of Desulfovibrio gigas hydrogenase. *Biochim. Biophys. Acta - Protein Struct. Mol. Enzymol.* **832**, 69–79 (1985).

16. Baltazar, C. S. A. A. *et al.* Nickel-iron-selenium hydrogenases - An overview. *Eur. J. Inorg. Chem.* 948–962 (2011). doi:10.1002/ejic.201001127

17. Valente, F. M. A. *et al.* Hydrogenases in Desulfovibrio vulgaris Hildenborough: Structural and physiologic characterisation of the membrane-bound [NiFeSe]

hydrogenase. *J. Biol. Inorg. Chem.* **10**, 667–682 (2005).

18. Wombwell, C., Caputo, C. A. & Reisner, E. [NiFeSe]-Hydrogenase Chemistry. *Acc. Chem. Res.* 2858–2865 (2015). doi:10.1021/acs.accounts.5b00326

19. Garcin, E. *et al.* The crystal structure of a reduced [NiFeSe] hydrogenase provides an image of the activated catalytic center. *Structure* **7**, 557–566 (1999).

20. Marques, M. C., Coelho, R., De Lacey, A. L., Pereira, I. A. C. & Matias, P. M. The three-dimensional structure of [nifese] hydrogenase from desulfovibrio vulgaris hildenborough: A hydrogenase without a bridging ligand in the active site in its oxidised, 'as-isolated' state. *J. Mol. Biol.* **396**, 893–907 (2010).

21. Volbeda, A. *et al.* Structure of the [NiFe] hydrogenase active site: Evidence for biologically uncommon Fe ligands. *J. Am. Chem. Soc.* **118**, 12989–12996 (1996).

22. Gutiérrez-Sanz, O. *et al.* Influence of the protein structure surrounding the active site on the catalytic activity of [NiFeSe] hydrogenases. *J. Biol. Inorg. Chem.* **18**, 419–427 (2013).

23. Baltazar, C. S. A. A., Teixeira, V. H. & Soares, C. M. Structural features of [NiFeSe] and [NiFe] hydrogenases determining their different properties: A computational approach. *J. Biol. Inorg. Chem.* **17**, 543–555 (2012).

24. Marques, M. C. *et al.* The direct role of selenocysteine in [NiFeSe] hydrogenase maturation and catalysis. *Nat. Chem. Biol.* **13**, 544–550 (2017).

25. Parkin, A., Goldet, G., Cavazza, C., Fontecilla-Camps, J. C. & Armstrong, F. A. The difference a Se makes? Oxygen-tolerant hydrogen production by the [NiFeSe]-hydrogenase from Desulfomicrobium baculatum. *J. Am. Chem. Soc.* **130**, 13410–13416 (2008).

26. Ceccaldi, P., Marques, M. C., Fourmond, V., Pereira, I. C. & Léger, C. Oxidative inactivation of NiFeSe hydrogenase. *Chem. Commun.* **51**, 14223–14226 (2015).

27. Kalms, J. *et al.* Tracking the route of molecular oxygen in O2-tolerant membrane-bound [NiFe] hydrogenase. *Proc. Natl. Acad. Sci.* **115**, E2229–E2237 (2018).

28. Wang, P. H., Best, R. B. & Blumberger, J. Multiscale simulation reveals multiple pathways for H2 and O2 transport in a [NiFe]-hydrogenase. *J. Am. Chem. Soc.* **133**, 3548–3556 (2011).

29. Teixeira, V. H., Baptista, A. M. & Soares, C. M. Pathways of H2 toward the active site of [NiFe]-hydrogenase. *Biophys. J.* **91**, 2035–2045 (2006).

30. Teixeira, V. H., Soares, C. M. & Baptista, A. M. Studies of the reduction and protonation behavior of tetraheme cytochromes using atomic detail. *J. Biol. Inorg. Chem.* **7**, 200–216 (2002).

31. Oliveira, A. S. F., Damas, J. M., Baptista, A. M. & Soares, C. M. Exploring O 2 Diffusion in A-Type Cytochrome c Oxidases: Molecular Dynamics Simulations Uncover Two Alternative Channels towards the Binuclear Site. *PLoS Comput. Biol.* **10**, e1004010 (2014).

32. Damas, J. M., Baptista, A. M. & Soares, C. M. The pathway for O2 diffusion inside CotA laccase and possible implications on the multicopper oxidases family. *J. Chem. Theory Comput.* **10**, 3525–3531 (2014).

33. Zacarias, S. *et al.* A Hydrophilic Channel Is Involved in Oxidative Inactivation of

a [NiFeSe] Hydrogenase. *ACS Catal.* **9**, 8509–8519 (2019).

34. Hermans, J., Berendsen, H. J. C., Van Gunsteren, W. F. & Postma, J. P. M. A consistent empirical potential for water–protein interactions. *Biopolymers* **23**, 1513–1518 (1984).

35. Baptista, A. M. & Soares, C. M. Some Theoretical and Computational Aspects of the Inclusion of Proton Isomerism in the Protonation Equilibrium of Proteins. *J. Phys. Chem. B* **105**, 293–309 (2002).

36. Cordeiro, R. M. Reactive oxygen species at phospholipid bilayers: Distribution, mobility and permeation. *Biochim. Biophys. Acta - Biomembr.* **1838**, 438–444 (2014).

37. Niu, S., Thomson, L. M. & Hall, M. B. Theoretical characterization of the reaction intermediates in a model of the nickel-iron hydrogenase of Desulfovibrio gigas. *J. Am. Chem. Soc.* **121**, 4000–4007 (1999).

38. Teixeira, V. H., Baptista, A. M. & Soares, C. M. Pathways of H2 toward the active site of [NiFe]-hydrogenase. *Biophys. J.* **91**, 2035–2045 (2006).

39. Schmid, N. *et al.* Definition and testing of the GROMOS force-field versions 54A7 and 54B7. *Eur. Biophys. J.* **40**, 843–856 (2011).

40. Berendsen, H. J. C., van der Spoel, D. & van Drunen, R. GROMACS: A message-passing parallel molecular dynamics implementation. *Comput. Phys. Commun.* **91**, 43–56 (1995).

41. Nosé, S. & Klein, M. L. Constant pressure molecular dynamics for molecular systems. *Mol. Phys.* **50**, 1055–1076 (1983).

42. Parrinello, M. & Rahman, A. Polymorphic transitions in single crystals: A new molecular dynamics method. *J. Appl. Phys.* **52**, 7182–7190 (1981).

43. Bussi, G., Donadio, D. & Parrinello, M. Canonical sampling through velocity rescaling. *J. Chem. Phys.* **126**, 014101 (2007).

44. Hess, B., Bekker, H., Berendsen, H. J. C. & Fraaije, J. G. E. M. LINCS: A Linear Constraint Solver for molecular simulations. *J. Comput. Chem.* **18**, 1463–1472 (1997).

45. Miyamoto, S. & Kollman, P. A. Settle: An analytical version of the SHAKE and RATTLE algorithm for rigid water models. *J. Comput. Chem.* **13**, 952–962 (1992).

46. Essmann, U. *et al.* A smooth particle mesh Ewald method. *J. Chem. Phys.* **103**, 8577–8593 (1995).

47. Press, W. H., Teukolsky, S. A., Vetterling, W. T. & Flannery, B. P. *Numerical Recipes in C (2nd Ed.): The Art of Scientific Computing*. (Cambridge University Press, 1992).

48. Berendsen, H. J. C., Postma, J. P. M., Van Gunsteren, W. F., Dinola, A. & Haak, J. R. Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* **81**, 3684–3690 (1984).

49. Humphrey, W., Dalke, A. & Schulten, K. VMD: Visual molecular dynamics. *J. Mol. Graph.* **14**, 33–38 (1996).

50. Cohen, J., Arkhipov, A., Braun, R. & Schulten, K. Imaging the migration pathways for O 2 , CO, NO, and Xe inside myoglobin. *Biophys. J.* **91**, 1844–

1857 (2006).

51. Damas, J. M., Baptista, A. M. & Soares, C. M. The pathway for O2 diffusion inside CotA laccase and possible implications on the multicopper oxidases family. *J. Chem. Theory Comput.* **10**, 3525–3531 (2014).

52. Campos, S. R. R. & Baptista, A. M. Conformational analysis in a multidimensional energy landscape: Study of an arginylglutamate repeat. *J. Phys. Chem. B* **113**, 15989–16001 (2009).

53. Mann, M. & Klemm, K. Efficient exploration of discrete energy landscapes. *Phys. Rev. E - Stat. Nonlinear, Soft Matter Phys.* **83**, 011113 (2011).

54. Metzner, P., Schütte, C. & Vanden-Eijnden, E. Transition Path Theory for Markov Jump Processes. *Multiscale Model. Simul.* **7**, 1192–1219 (2009).

55. Scherer, M. K. *et al.* PyEMMA 2: A Software Package for Estimation, Validation, and Analysis of Markov Models. *J. Chem. Theory Comput.* **11**, 5525–5542 (2015).