BBA - Bioenergetics 1860 (2019) 325–340

Contents lists available at ScienceDirect

BBA - Bioenergetics

journal homepage: www.elsevier.com/locate/bbabio

Computational analyses, molecular dynamics, and mutagenesis studies of unprocessed form of [NiFe] hydrogenase reveal the role of disorder for efficient enzyme maturation

Marta Albareda, Luis F. Pacios*, Jose M. Palacios*

Centro de Biotecnología y Genómica de Plantas (C.B.G.P.) UPM-INIA, Universidad Politécnica de Madrid, Campus de Montegancedo, 28223 Pozuelo de Alarcón, Spain
Departamento de Biotecnología-Biología Vegetal, Escuela Técnica Superior de Ingeniería Agronómica, Alimentaria y de Biosistemas, Universidad Politécnica de Madrid, Madrid, Spain

ARTICLE INFO

Keywords:
Metalloenzyme
Nitrogen fixation
Hydrogenase
Molecular modelling
C-terminal tail
[NiFe] cofactor

ABSTRACT

Biological production and oxidation of hydrogen is mediated by hydrogenases, key enzymes for these energy-relevant reactions. Synthesis of [NiFe] hydrogenases involves a complex series of biochemical reactions to assemble protein subunits and metallic cofactors required for enzyme function. A final step in this biosynthetic pathway is the processing of a C-terminal tail (CTT) from its large subunit, thus allowing proper insertion of nickel in the unique $NiFe(CN)_2CO$ cofactor present in these enzymes. In silico modelling and Molecular Dynamics (MD) analyses of processed vs. unprocessed forms of *Rhizobium leguminosarum* bv. viciae (*Rlv*) hydrogenase large subunit HupL showed that its CTT (residues 582-596) is an intrinsically disordered region (IDR) that likely provides the required flexibility to the protein for the final steps of proteolytic maturation. Prediction of pKa values of ionizable side chains in both forms of the enzyme's large subunit also revealed that the presence of the CTT strongly modify the protonation state of some key residues around the active site. Furthermore, MD simulations and mutant analyses revealed that two glutamate residues (E27 in the N-terminal region and E589 inside the CTT) likely contribute to the process of nickel incorporation into the enzyme. Computational analysis also revealed structural details on the interaction of *Rlv* hydrogenase LSU with the endoprotease HupD responsible for the removal of CTT.

1. Introduction

Hydrogen is a key element for energy budgets in natural ecosystems, acting either as an oxidizable fuel or as a valve for disposing reducing equivalents in fermentative conditions. Hydrogen oxidation and hydrogen evolution are mediated by hydrogenases, metalloenzymes that catalyse the reversible activation of hydrogen. In addition to these metabolic roles, hydrogenases are also relevant due to their biotechnological applications related to the use of hydrogen as an energy vector [1]. Several hydrogenase classes have been defined depending on the type of metallocenter at the active site (see [2] for a recent review).

[NiFe] hydrogenases are a quite diverse class of hydrogenases widespread in both *Bacteria* and *Archaea*. These enzymes contain at least two subunits: a larger (ca. 60 kDa) catalytic subunit (LSU), and a smaller (ca. 30 kDa) subunit (SSU) with FeS clusters involved in electron transfer. Some hydrogenases present structural variations in the SSU, or additional subunits with different functions such as coupling to specific electron acceptors (NADP, F_{420}; reviewed in ref. [2]). The catalytic subunit harbours an heterobimetallic $NiFe(CN)_2CO$ cofactor at the active site of the enzyme [3,4]. Nickel in this cofactor is coordinated via four conserved cysteinyl residues, two of which also coordinate the iron atom, which is bound to two cyanide and one carbonyl groups; in some hydrogenases, one of the two Ni-coordinating cysteines is replaced by a selenocysteine [2]. Assembly of the complex [NiFe] cofactor is a key process for hydrogenase synthesis, and involves at least six proteins (HypABFCDE; [5,6]). Four of these proteins (HypCDEF) participate in the synthesis of a $Fe(CN)_2CO$ cofactor precursor and its transfer to the LSU. Other auxiliary proteins (HypA and HypB, and in some cases SlyD) are required for the incorporation of Ni^{2+} ions once

Abbreviations: IDR, intrinsically disordered region; LSU, large subunit; SSU, small subunit; CTT, C-terminal tail from hydrogenase LSU; MD, molecular dynamics; RMSD, root mean square deviation; RMSF, root mean square fluctuation

* Corresponding authors at: Centro de Biotecnología y Genómica de Plantas (C.B.G.P.) UPM-INIA, Universidad Politécnica de Madrid, Campus de Montegancedo, 28223 Pozuelo de Alarcón, Madrid, Spain.

E-mail addresses: luis.fpacios@upm.es (L.F. Pacios), jose.palacios@upm.es (J.M. Palacios).

https://doi.org/10.1016/j.bbabio.2019.01.001
Received 24 July 2018; Received in revised form 10 December 2018; Accepted 25 January 2019
Available online 29 January 2019
0005-2728/

the cofactor precursor is associated to the LSU [7,8]. A final step in the maturation of this large catalytic subunit involves the removal of a C-terminal tail (CTT) spanning typically ca. 15 residues on the LSU C-terminus in most bacteria [2], although it is longer in some cases such as *Escherichia coli* hydrogenase-3 [9]. Preceding this C-terminal peptide there is a DPCxxCxxH/R consensus sequence motif (Fig. S1) that (i) begins with an aspartate which participates in a putative proton transfer pathway [10–12], (ii) includes two of the four cysteines which ligate nickel in the mature enzyme, and (iii) ends with a histidine or arginine. The C-terminal peptide always starts with a non-polar residue M/I/V/A and its removal is performed by specific endopeptidases that recognize the consensus motif and cleave between its end basic H/R residue and the first non-polar residue of the CTT [2,13–15] thus leaving a terminal histidine in which multiple structural studies have revealed the presence of a magnesium atom [16–18].

Nickel promotes the recognition by proteases so that cleavage of the CTT only occurs in the presence of Ni^{2+} bound to the LSU [19,20]. Previous studies have shown that truncated *E. coli* hydrogenase-3 versions lacking 20 out of the 33 CTT residues still developed hydrogenase activity [14]. In the case of *E. coli* hydrogenase-2 complete removal of the CTT, or addition of extra residues to it [13], originated relevant effects on hydrogenase processing and activity, suggesting that these alterations on the CTT impair the interaction with Hyp maturation machinery [13,15]. It is accepted that, upon CTT removal, a change of conformation occurs in the neighborhood of the active site somehow burying the $NiFe(CN)_2CO$ metal cluster, thus locking the site in place [6]. This indicates that the presence of the CTT controls the sequence of events that constitute the hydrogenase maturation process [15]. Once processed, the mature LSU associates to the SSU and, in the cases of membrane-bound and periplasmic NiFe hydrogenases, the dimeric enzyme is exported via the Twin-Arginine Translocation (TAT) system [21].

Together with experimental studies, computational methods provide a valuable way (in some cases, the only way) to gain insight into dynamic features of protein structures as well as into mechanistic details of enzymes. In particular, Molecular Dynamics (MD) simulations have been used to investigate aspects like gas and proton transfer in [NiFe] hydrogenases at the atomic level [22–26].

Rhizobium leguminosarum bv viciae (*Rlv*) UPM791 is an α-proteobacterium able to establish nitrogen-fixing symbiosis with pea and other legume plants [27,28]. It contains a [NiFe] hydrogenase system encoded in an 18-gene cluster *hup*SLCDEFGHIJK*hyp*ABFCDEX [29] where *hup*S and *hup*L encode hydrogenase SSU and LSU, respectively. The enzyme is coupled in vivo to the microaerobic respiratory chain in bacteroids [30]. *Rlv* HupL contains a typical CTT that is processed by endoprotease HupD during hydrogenase synthesis [31]. This hydrogenase system, naturally adapted in *Rlv* UPM791 for the removal of hydrogen produced by nitrogenase in nitrogen fixation [32], has been modified by placing the microaerobically-inducible promoter of *Rlv fixN* (P_{fixN}) upstream of the first gene of the *hup/hyp* gene cluster to allow expression in free-living culture [33]. Multiple aspects of the genetics, physiology, and regulation of *Rlv* UPM791 hydrogenase system have been studied, including the requirement of nickel for efficient hydrogenase processing [34,35], and the role of HupE as a nickel transporter for hydrogenase [36]. The fact that this bacterium contains a single hydrogenase system, as opposed to up to 4 systems in other bacteria such as *E. coli* [37], makes it a good model system for functional analysis.

It is known that unprocessed LSU incorporates the $Fe(CN)_2CO$ cofactor precursor from Hyp auxiliary proteins, and also that a nickel atom must be present to induce proteolytic cleavage [5]. Previous work with *E. coli* hydrogenases has shown the essentiality of the removable C-terminal tail for hydrogenase synthesis [13,14]. However, no detailed structure/function analysis of this protein region has been carried out to date. In this study we apply a number of computational analyses of structures and molecular properties together with MD simulations and site-directed mutagenesis to investigate the features of this CTT peptide in the LSU of *Rlv* UPM791 hydrogenase. Our results indicate that this CTT behaves as an intrinsically disordered region with a role in maintaining an appropriate environment around the active site. In addition, we have identified a glutamate in the CTT (E589 in *Rlv* HupL) that, in combination with a highly conserved glutamate in the N-terminal region (E27 in *Rlv* HupL), might play a role in nickel incorporation into the enzyme. Further computational analyses revealed structural details on the interaction of *Rlv* HupL hydrogenase LSU with the endoprotease HupD responsible for the removal of CTT.

2. Results

2.1. C-terminal tails of unprocessed forms of [NiFe] hydrogenase large subunits behave as intrinsically disordered regions

Before proceeding further, we make clear that “unprocessed” and “processed” terms applied to hydrogenase LSUs are not used in a methodological sense (i.e. referring to different stages of computational processing) but they are used throughout this report in a biological sense to mean “unprocessed = CTT still present” and “processed = CTT already removed”.

Computational structural models of [NiFe] hydrogenase LSUs in their unprocessed form were obtained by using I-TASSER [38,39] with their complete sequences. This hybrid method predicts protein structure without relying solely on homology modelling but it considers several experimental structures at a threading stage and predicts ab initio models for regions without experimental backup. This enables to construct 3D models for structural segments not present in available structures as it is the case for the CTT segment. Unprocessed forms of the following LSUs were modelled this way: *Hydrogenovibrio marinus* HoxH, *Ralstonia eutropha* HoxG, *E. coli* Hyd1, and *R. leguminosarum* HupL. Initial structures (further used as starting geometries in molecular dynamics simulations: see below) of complexes that include the NiFe cofactor in the active site were prepared as follows. First, the model structures of unprocessed forms were superimposed with the crystal structures of their corresponding processed forms that contain the cofactors. Then, the location of the NiFe clusters resulting from the superposition was locally optimized in the modelled structures. Except for *Rhizobium* HupL, these crystal structures were those used by I-TASSER as main templates in threading: *H. marinus* HoxH (PDB id: 3AYX) [40], *R. eutropha* HoxG (PDB id: 3RGW; [41]), and *E. coli* Hyd1 (PDB id: 3UQY; [42]). The case of *R. leguminosarum* HupL is different as no experimental structure for its processed form is available. However, nine out of the top ten threading templates actually corresponded to crystal structures of *H. marinus* (PDB id: 3AYX; [40]) and *Allochromatium vinosum* (PDB id: 3MYR; [43]) LSUs. This way, the NiFe cofactor present in these two crystal structures was inserted into HupL in the superposition. However, given that those crystal structures correspond to mature forms of the LSUs, their cofactors have Fe and Ni atoms coordinated by the four cysteines in the binding site and include an OH^- bridge between both metals [40,43]. As our main goal was to explore the structural effects associated to the change between unprocessed and processed forms of HupL, initial structures for our subsequent computational analyses were prepared modifying accordingly the cofactor by removing the OH^- bridge and metal-cysteine bonds. This way, the cofactors in these initial structures consist of the $Fe(CN)_2CO$ molecular moiety and a Ni^{2+} ion because we aimed to explore through MD calculations unprocessed states of LSUs in which no final forms of the NiFe cofactors are still assembled. After these modifications, the local geometries of the binding site were optimized as explained in Methods Sections 4.5 and 4.7. The structural models of all unprocessed forms share a virtually identical architecture (Fig. 1A) and a highly similar location of NiFe cofactors (Fig. 1B). A representative detailed view of the metal binding site for the *R. leguminosarum* HupL case (Fig. 1C) shows that the local optimization locates $Fe(CN)_2CO$, Ni^{2+} and the four

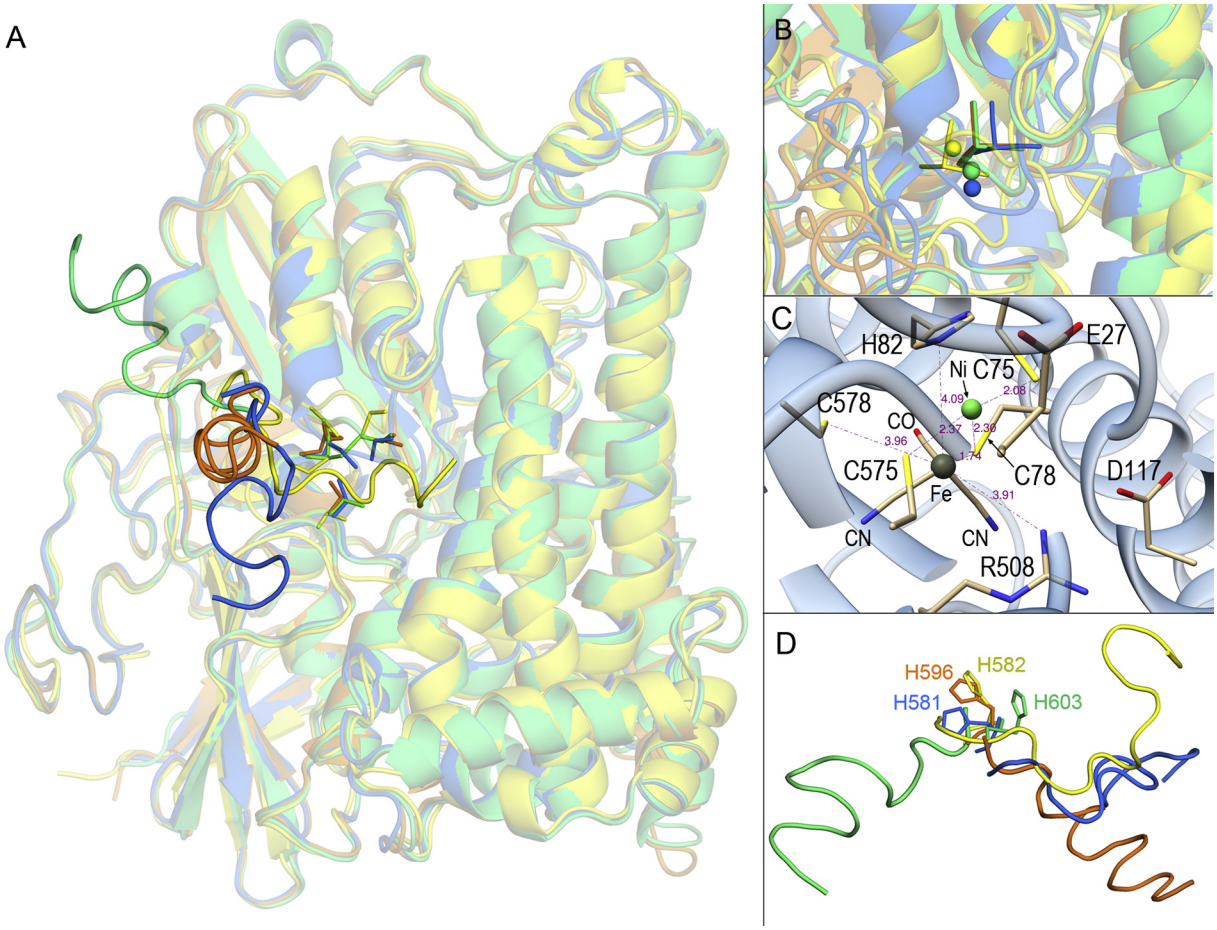


Fig. 1. Structural alignment obtained with the CE algorithm of unprocessed forms of the following hydrogenase LSUs: *R. leguminosarum* HupL (blue), *H. marinus* HoxH (orange), *R. eutropha* HoxG (green), and *E. coli* Hyd1 (yellow). (A) Ribbon diagram of superimposed backbones in semitransparent mode except CTT segments that are shown in non-transparent mode. The four cysteines in the metal-binding site are shown as non-transparent sticks. (B) Iron cofactors (sticks) and nickel atoms (spheres) at locations resulting from the structural alignment in A. (C) Metal-binding site in unprocessed *R. leguminosarum* HupL. Distances from Fe and Ni atoms to the nearest protein atoms are indicated. (D) C-terminal tails in the structural alignment in A. Histidines just preceding the CTTs are displayed as sticks.

Table 1
Disorder prediction for C-terminal peptides cleaved at hydrogenase maturation.

Hydrogenase	Tail[a]	Tail sequence[b]	Disordered regions predicted by		
			DisEMBL	PONDR	IUPred
HupL *Rhizobium*	582-596	**VMSP*D*GQEMARVQVR**	581-588	582-596	586-596
HupL *Rhiz. E589Q mutant*	582-596	**VMSP*D*GQQMARVQVR**	580-588	581-596	586-596
Hyd1 *E. coli*	583-597	**VLGD*D*GSELISVQVR**	582-597	586-597	None
HoxG *Ralstonia*	604-618	**VMSA*E*GQELTTVKVR**	604-618	607-618	609-618
HoxH *Hydrogenovibrio*	597-611	**VMSE*E*GEEMATVKVR**	596-611	597-611	601-611

[a] Sequence numbering defining the C-terminal peptide.
[b] Conserved residues are in bold type and similar residues are underlined and in italics.

cysteines at initial geometries in the site (defined by a 5 Å distance from cofactor atoms). Besides the cysteines, this region close to metal atoms also includes the presence of E27, D117, and R508, residues that are addressed below with regard to their anomalous predicted pK_a's (see Section 2.2). As expected, all models differ greatly in their CTTs (Fig. 1D). These four C-terminal peptides have 15 residues, from which 6 are identical in all four cases (Table 1) and also quite conserved when comparing a set of 40 CTTs from different *Proteobacteria* (Fig. S1). The four LSUs studied have exactly the same DPCLACSTH consensus sequence motif that immediately precedes the CTT so that a conserved His is the last residue retained in the processed form after removal of the peptide.

Modelled structures predict unstructured coil segments for these tails. In fact, the superposition of the four LSUs shows that whereas the location of histidines agrees reasonably well, CTT segments do not fit (Fig. 1D). Prompted by this finding, we used three well-known predictors of disorder from sequence that employ different strategies to predict intrinsically disordered regions (IDRs), i.e. regions with amino acid sequences intrinsically involving disorder. While different small sequence segments were predicted as disordered by these three methods (data not shown), the only sequence region in which the three predictors agreed is the CTT (Table 1). An exception is Hyd1 from *E. coli* for which IUPred (a method that recognizes unstructured regions from sequence based on the estimated amino acid pairwise energy content) predicts no disordered region. Interestingly, our MD data suggest that

Table 2
Predicted pK_a values of residues with ionizable side chains in the initial model structures of processed and unprocessed forms of *Rlv* HupL LSU that show significant deviations from reference pK_a^0 values in water.[a]

Residue	pK_a (processed)	pK_a (unprocessed)	Closest distance to Ni atom (Å)[b]
D129	6.2	8.8	16.9
D132	5.5	5.2	27.7
D277	7.6	4.4	22.4
D311	1.8	3.1	29.0
D539	2.0	3.2	25.6
D573	7.5	7.0	7.9
E27	9.2	7.2	4.7
E56	2.3	6.4	19.0
E72	7.9	9.9	10.7
E505	8.3	7.5	14.9
E548	6.8	7.6	21.1
H82	3.4	1.3	4.5
H118	1.6	1.2	11.9
H121	−0.8	−1.4	10.2
H126	1.7	1.0	15.6
H219	3.5	4.2	12.8
H337	3.8	1.9	19.6
H349	3.1	3.5	24.6
H513	1.4	−0.1	11.5
H570	0.8	3.4	14.5
H581	−2.6	−2.5	12.7
R508	18.9	16.3	4.8
R538	15.0	15.1	25.2

[a] A deviation is considered "significant" if the predicted pK_a differs from the reference value in water, pK_a^0, in ± 1.5 units for aspartate and ± 2 units for the remaining amino acids. pK_a^0 values: aspartate = 3.8, glutamate = 4.5, histidine = 6.5, lysine = 10.5, and arginine = 12.5.

[b] Closer distance from any atom in the side chain to nickel ion.

the CTT in unprocessed Hyd1 lacks the great mobility shown by the other three LSUs (see below and Fig. S4) and, besides, has in its second and third position Leu and Gly instead of Met and Ser present in HupL, HoxG, and HoxH (Table 1). As for the E589Q mutant of HupL studied below, two out of the three predictors increase the length of the segment predicted as disordered in a single residue (Table 1), a scarcely significant result.

The I-TASSER model structure of the processed form of LSU of *R. leguminosarum* HupL was generated from the same threading templates mentioned above selected by this method to construct the unprocessed LSU. As no CTTs are now present, processed forms exhibit a higher similarity than unprocessed forms. The structural comparison of modelled processed LSU of *R. leguminosarum* (581 residues) with these template crystal structures gave the following results (for each LSU, the number of residues included in the superposition / total number of residues and the RMSD computed in the superposition with backbone atoms are indicated): *H. marinus* HoxH (PDB id: 3AYX), 574/595, 0.54 Å; *R. eutropha* HoxG (PDB id: 3RGW) 571/602, 0.70 Å, and *E. coli* Hyd1 (PDB id: 3UQY), 570/582, 0.85 Å. It is thus apparent that the structure of the modelled processed LSU changes little with respect to known structures that in all cases correspond to the complexes with the SSU.

2.2. Some ionizable side chains show pK_a with large deviations from standard values in water

Protonation states of a few key residues may provide clues on structural effects associated to changes underlying the unprocessed to processed transformation of [NiFe] hydrogenase. Experimental details on protonation of residues in hydrogenases can be obtained from high resolution (1.50 Å) crystal structure of *Desulfovibrio vulgaris* hydrogenase (PBD id: 1WUL) [44] as it contains positions of structural water molecules. More recently, a new crystal structure at ultra-high 0.89 Å resolution (PBD id: 4U9H) of this hydrogenase has permitted the detection of hydrogen atoms thus allowing to propose new features of its catalytic cycle [18]. Neither of these *D. vulgaris* hydrogenase structures was used by I-TASSER in predicting our model structure of *R. leguminosarum* LSU.

Analyses on protonation states can also be performed computationally, albeit in an approximate manner. We used Propka 3.1 [45,46] to estimate pK_a values of ionizable side chains for the two crystal structures of *D. vulgaris* LSU cited above and for the model structures of both unprocessed and processed forms of *R. leguminosarum* LSUs as well. Twenty-one residues of the *D. vulgaris* LSU were found to have pK_a's that show large deviations from standard reference values in water pK_a^0 (Table S1). Nineteen out of these twenty-one *D. vulgaris* residues are conserved in processed HupL (Table S1; crystallized hydrogenases are obviously processed forms). Moreover, the structural alignment of *D. vulgaris* and *R. leguminosarum* processed LSUs reveals not only that their structures are rather similar (497 out of the total 534 residues of *D. vulgaris* are included in the superposition with a RMSD of backbone atoms = 0.95 Å), but also that the 3D locations of these conserved residues are highly coincident (Fig. S2).

Propka 3.1 estimates pK_a of an ionizable side chain from empirical calculations of ΔG in the interaction of side chains with their local environment [45,46]. Its 3.1 version implements a new algorithm for modelling non-covalently coupled residues which could influence the titration of each other because of their close proximity [46]. As demonstrated in a test of proteins, this version achieved an increase of reliability with respect to previous Propka 2 results by not only improving the rules and parameters of the method but also by utilizing cancelation of errors efficiently [45]. Thus, only 9% of the pK_a values examined had an error larger than 1 pH unit whereas 55% of the data had an error of 0.4 pH units [45]. Nonetheless, the lack of sufficient experimental values of pK_a's in proteins makes difficult to assess the accuracy of pK_a predictors. In any event, Propka (like other predictors) must be viewed just as a practical tool to obtain merely approximate pK_a's. However, its application to protein structures that only differ in local details could give useful information for comparison purposes. Considering this and in view of the agreement between our model structure of HupL and the experimental structure of *D. vulgaris* hydrogenase not used by I-TASSER in modelling as mentioned at the beginning of this section, we compared pK_a's of unprocessed and processed forms of *R. leguminosarum* HupL to detect effects arising from structural changes provoked by CTT removal. Table 2 lists residues that in processed HupL show large deviations from pK_a^0 values together with their values in unprocessed HupL and the closest distance from atoms in every side chain to nickel given as a parameter to locate the relative position of the residue in the structure. Since Propka estimates for the pK_a values of ionizable side chains strongly depend on their environment, the local geometries around some selected residues that show large deviations from standard values are examined in the superposition of the initial structures of unprocessed and processed HupL LSU (Fig. 2).

To analyze these results, let us recall that the [protonated form] / [deprotonated form] ratio for an ionizable side chain is given by $10^{(pKa-pH)}$ and that the protonated state is neutral in Asp and Glu and has charge +1 in His, Lys, and Arg. Data in Table 2 suggest that all histidines in HupL remain neutral and all basic residues remain positively charged at pH 7, the common behaviour in proteins. For example, three histidines with anomalous pK_a values, H513, H570, and H581, show crowded local environments very different from aqueous medium at which they would have $pK_a \sim 6.5$. Thus, H513 (Fig. 2A) in unprocessed HupL is at a short distance from S86 and E589, residue in the CTT and therefore present only in the unprocessed form (more on this glutamate below). This local environment also includes M583, other CTT residue. In the processed form H513 is close to Y525 and H82. It is thus apparent that the CTT modifies noticeably the local environment of H513 in unprocessed HupL.

H570 (Fig. 2B) in unprocessed HupL has D129 and T24 in its close proximity whereas in processed HupL, this histidine is near R25, T24

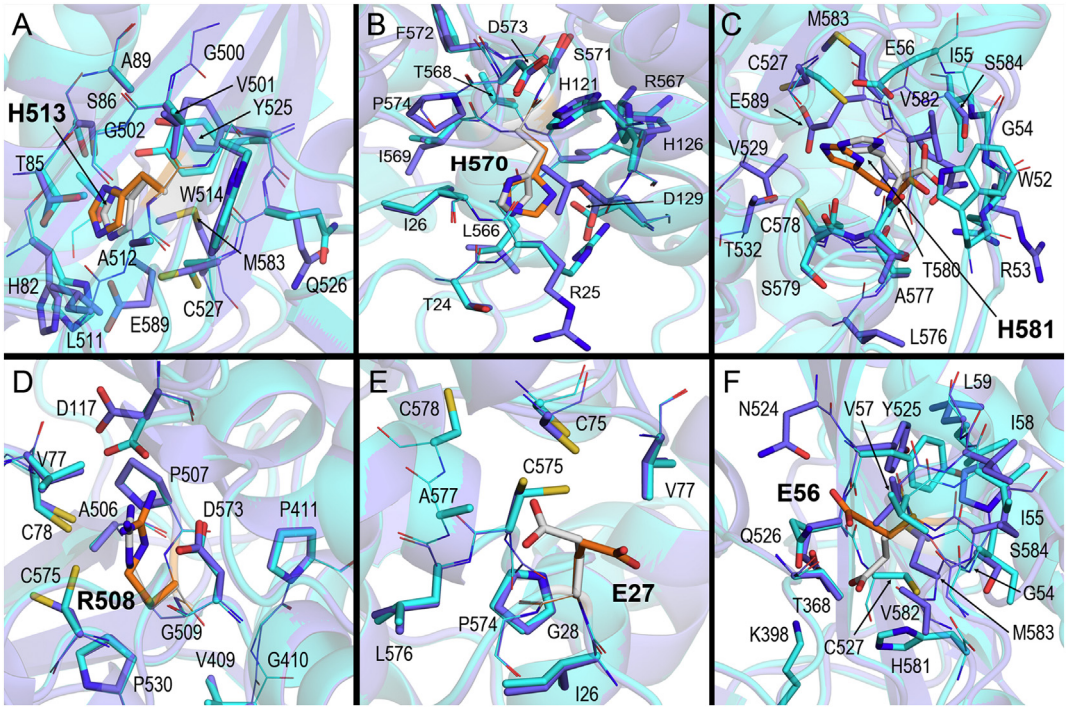


Fig. 2. Local environment of selected residues showing pK_a's that significantly deviate from their standard values (Table 2) in the structural superposition of unprocessed (semi-transparent cartoon and carbon sticks in slate blue) and processed (semi-transparent cartoon and carbon sticks in cyan) forms of *R. leguminosarum* HupL. (A) His513. (B) His 570. (C) His581. (D) R508. (E) E27. (F) E56. In all cases, amino acids having atoms within a 4.0 Å distance from atoms of the residue with anomalous pK_a are shown as sticks and labeled. Carbon sticks of the residue being examined are colored orange in unprocessed HupL and white in processed HupL.

(just as in the unprocessed form), D129 and L566. Note, besides, that the side chain of R25 has a different orientation in both HupL forms (Fig. 2B). Although H570 is located near the C-terminal region in sequence, no CTT residues are present in its local environment.

His581 (Fig. 2C), the residue preceding the CTT, shows the largest pK_a deviations of all histidines with virtually the same value in both LSU forms. In unprocessed HupL, a nitrogen atom of His581 is at short distance from backbone atoms of V582 (residue in the CTT) and of S579 while the other nitrogen atom is close to C527 and E56. In processed HupL, this H581 has in its close proximity backbone atoms of T580 and of C527 as well as one oxygen of E56 (see pK_a features of this glutamate below). The local environment of H581 includes the presence of CTT residues V582, M583, S584 and E589 in unprocessed HupL while it includes W52 in processed HupL as well as a different orientation of side chains of R53 and C527 (Fig. 2C). Note also that E56 is present in the processed but not in the unprocessed form whereas the opposite happens with E589. All in all, these differences in the local environment of H581 could compensate yielding a similar pK_a in both forms of the LSU (Table 2).

Only two arginines out of a total of 56 basic residues in HupL (33 Lys + 23 Arg) have pK_a's that differ in > 2 units from their pK_a^0 values. R508 (Fig. 2D) is special in that it locates in the immediate vicinity of the metal-binding site and shows an anomalously high $pK_a = 18.9$ increased with respect to its value in unprocessed HupL (16.3) and close to the 19.0 value predicted for the homologous R479 in *D. vulgaris* LSU (Table S1). The main difference in its local environment regards the much closer location of D117 and C575 side chains in the processed form with respect to unprocessed form (Fig. 2D). In fact, the shortest distance between N atoms of R508 and O atoms of D117 is 4.0 Å in unprocessed HupL while it is 2.70 Å in processed HupL. The shortest distance between N atoms of R508 and S atom of C575 is 3.68 Å in unprocessed HupL and 3.11 Å in processed HupL.

As for acidic amino acids, the pK_a's of 11 residues (6 Asp + 5 Glu) out of a total of 65 (35 Asp + 30 Glu) in HupL deviate significantly from their pK_a^0 values (Table 2). There is now a different behaviour regarding protonated/deprotonated ratios: at physiological pH, a deprotonated state with charge −1 is favoured for acidic residues with $pK_a < 7$, whereas those with $pK_a > 7$ should be mainly in a neutral state. Yet only two residues (E27 and E505) have pK_a's that imply a predominance of the protonated state by more than one order of magnitude (one pK_a unit). If one considers the pK_a values in unprocessed (7.2) and processed (9.2) forms and its close distance to nickel (< 5 Å), it seems apparent that the behaviour of E27 should be linked to hydrogenase activity. Since the only difference between the two initial structural models of *Rlv* LSU is the presence or absence of the CTT, one could reasonably assume that the spatial proximity of this tail to E27 should be the cause of that pK_a shift. In fact, its local environment shows a dramatic change of the spatial orientation of its carboxylate group away from the cofactor site in unprocessed HupL and directed towards C75, C575, and C78 in processed HupL (Fig. 2E).

As for E56, it again shows a change of orientation of its carboxylate group (Fig. 2F). Thus, while E56 remains close to T368 in both LSU forms, its carboxylate group moves from pointing towards N524 in unprocessed HupL to pointing towards Q526 and H581 in processed HupL. Together with the presence of V582, M583, and S584 (CTT amino acids) only in unprocessed HupL, those conformational differences determine the distinct local environment of E56 in both LSU forms.

Finally, pK_a's for residues in the CTT of unprocessed HupL (not included in Table 2) show disparate behaviour. D586 and R596 have normal pK_a values (4.9 and 12.3, respectively) whereas E589 has $pK_a = 10.5$, indeed the largest value for glutamates in these initial model structures of the LSU of *R. leguminosarum* HupL. Given the large mobility associated to this IDR, one could expect large variations in the local environments of ionizable side chains and consequently large variations of their pK_a's if that mobility implies spatial localization of CTT residues close to amino acids in the protein at a given time. Since the initial model structure of unprocessed HupL places the CTT very close to residues near the cofactor site (Figs. 1A, 2A, and C), the local environment of E589 is very different from the aqueous medium and its

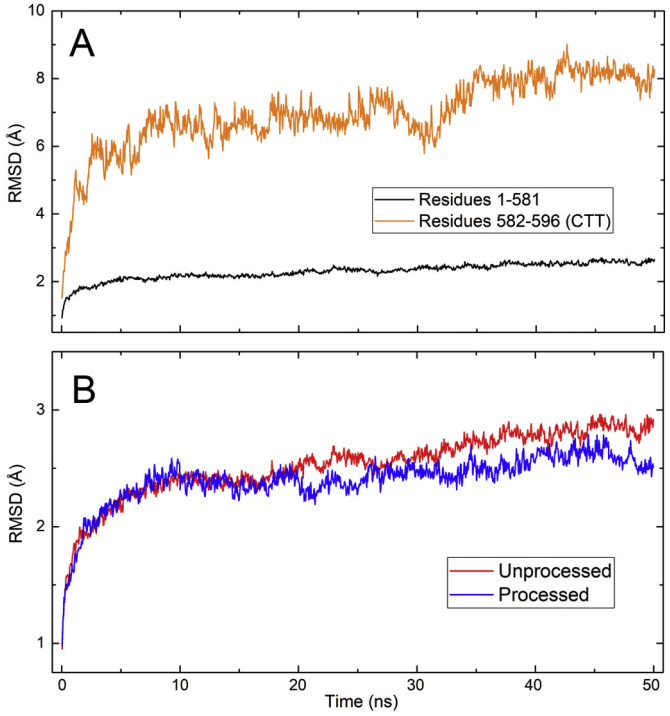


Fig. 3. Average RMSD values of backbone atoms in 50-ns MD simulations of HupL. (A) Averages of four trajectories in unprocessed HupL computed for the structure excluding the CTT ("Residues 1-581") and for the CTT segment ("Residues 582-596"). (B) Average of four trajectories in unprocessed HupL (complete structure) compared with average of three trajectories in processed HupL.

pK_a is very far from the standard value. However, after four different 50-ns MD simulations (see below), the dynamical motion shifts the CTT away from the protein and the local environment of E589 happens to be similar to the aqueous medium as suggested by the pK_a values in these final structures of unprocessed HupL (data not shown): 3.0, 2.9, 3.5, and 3.8 for trajectories 1–4, respectively, values very close to the standard pK_a^0 of glutamate.

2.3. Molecular dynamics simulations reveal global structural changes provoked by the removal of the C-terminal tail

MD 50-ns simulations evidence the high flexibility of the CTT in unprocessed forms of hydrogenase LSUs (Figs. 3, S3 and S4). For instance, whereas the complete *Rlv* HupL protein (residues 1-596) shows a relatively low mobility with RMSD values ~2.5 Å that stabilize after the first 10 ns (Fig. S3A), the CTT exhibits a large mobility with RMSD values > 6 Å just after the first 5 ns (Figs. 3A and S3B). Even though this flexible segment spans only 15 amino acids in a protein with nearly 600 residues, its effect is noticed in the complete structure. RMSD values averaged over the simulation time are the following: 2.30 Å (std. dev. 0.24 Å) for processed HupL composed of residues 1-581, 2.53 Å (std. dev. 0.28 Å) for the unprocessed form composed of residues 1-596, and 6.97 Å (std. dev. 1.04 Å) for CTT residues 582-596, although this last average is not very significant in a highly flexible segment as the large standard deviation illustrates. MD results support thus the disordered nature of CTT segment presented above on the grounds of initial, static model structures and sequence-based predictions. In the absence of the CTT, the processed form of HupL shows an overall decreased mobility with respect to the unprocessed form (Figs. 3B and S3C).

This MD analysis was also applied to the unprocessed forms of *H. marinus* HoxH and *E. coli* Hyd1 LSUs (Fig. S4). While large mobility of the CTT is also found in HoxH, Hyd1 shows a different behaviour with much smaller RMSD values of the CTT segment than those in HupL or HoxH (Fig. S4). These MD results thus suggest a much less disordered tail in the *E. coli* LSU. As noted above with regard to disorder predictions (see Section 2.1), Hyd1 was the unique case among the four hydrogenases studied for which IUPred, one of the three predictor methods employed, failed to predict disorder for the CTT sequence (Table 1).

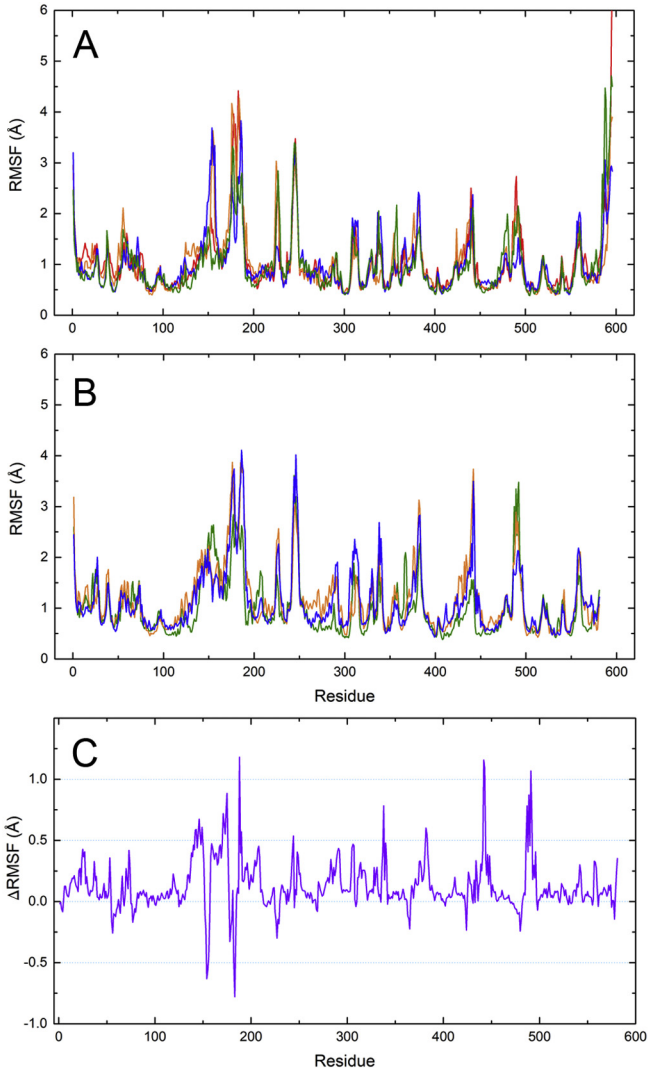
Structure fluctuations measured by RMSFs computed from four and three trajectories in unprocessed and processed HupL respectively, consistently show several regions with large backbone fluctuation (Fig. 4A and B). The following residues have RMSF values > 2.5 Å: 176-187, 244-247, and 587-596 (these in the CTT) in unprocessed HupL, and 174-188, 243-248, 382-383, 442, and 489-492 in processed HupL. Most of these residues also show large RMSF differences between unprocessed and processed MD results (Fig. 4C) and, in all cases, they are located in loop regions of disparate length (Fig. 4D).

The abundance of amino acids in coil-type segments is a feature of hydrogenase architecture that provides structural freedom to several segments (Figs. 1A and 4D). While ~40% of total residues are in alpha helices and ~20% are in beta strands, the remaining ~40% are in coil segments (Table 3). These figures refer to (i) static structures and (ii) processed forms of hydrogenase LSUs as this is the form purified from bacteria to obtain crystal structures. Our model initial structures of *Rhizobium* HupL show that, discounting the 15 residues in the disordered CTT, the alpha/beta/coil proportion is not very distinct from that found in other bacteria (Table 3). Even though the same prediction method is applied to obtain model structures for two sequences that share the first 581 residues and only differ in that the unprocessed form has 15 additional amino acids, significant differences are noticed. Processed HupL has 19 residues in secondary structure (10 alpha and 9 beta) which are in coil segments in unprocessed HupL ("initial" data in Table 3). However, dynamic effects after 50-ns MD simulations affect similarly to both forms ("final" data in Table 3) as the number of residues that change secondary structure from alpha/beta to coil is 45 in unprocessed and 39 in processed. With this, the final structure of unprocessed HupL shows the largest number of residues (307) in coil segments and the smallest number of residues in beta strands in Table 3.

2.4. Molecular dynamics data identify local effects on the active site region provoked by the removal of the C-terminal tail

The comparison of average RMSDs of non-hydrogen atoms in both backbone and side chains of the four cysteines and the metallic cluster in the active site between unprocessed and processed HupL is illuminating (Fig. 5). To analyze these local effects, it must be stressed that MD simulations were performed starting at initial geometries in which the preformed $Fe(CN)_2CO$ cofactor and the Ni^{2+} cation were positioned as explained above (Section 2.1) and in Methods Sections 4.5 and 4.7. The starting position of Ni might not reflect the real one in unprocessed form, since it is not known where HypA delivers the cation. However, this analysis allows us to evaluate how stable is Ni at the position it occupies at the end of the maturation process. This way, no bonds between protein atoms and cofactor atoms or nickel were initially defined. The underlying intention was to explore local effects on the structural region where the active site will be configured after CTT removal predicted by MD on the basis of the purely interatomic interactions calculated in the simulations. Final geometries in the immediate vicinity of the active site region after simulations are depicted in Fig. 6.

Lower RMSD values are found for cysteines, iron cofactor and nickel in processed HupL as compared with unprocessed HupL although only C78 among the four cysteines shows a markedly different behaviour (Fig. 5) as this cysteine is predicted to bind iron atom in both unprocessed and processed forms (Fig. 6). C575 and C578 are predicted to orient towards nickel but without forming bonds in unprocessed HupL (Fig. 6A). As they bind nickel in processed HupL (Fig. 6B) these results suggest that Ni is not in its final position when the protein is in its



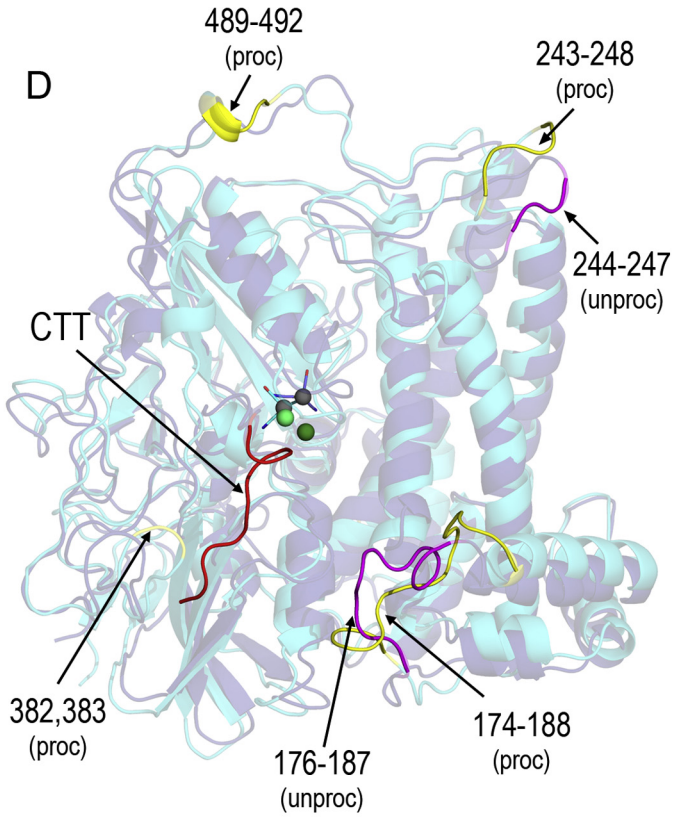


Fig. 4. RMSF of α carbons in 50-ns MD simulations. (A). Results of four trajectories for the complete structure (residues 1-596) of unprocessed HupL. (B) Results of three trajectories for the complete structure (residues 1-581) of processed HupL. Same scale is used in plots A and B for comparison purposes. (C) RMSF difference (processed – unprocessed) computed with average RMSF values for the trajectories in (A) and (B). (D). Superimposed backbones of representative final structures after 50-ns simulations of unprocessed (deep blue) and processed (cyan) HupL. $Fe(CN)_2CO$ cofactors are shown as sticks with Fe in deep grey and Ni shown as spheres in deep green (unprocessed) and light green (processed). Residues showing RMSF > 2.5 Å are labeled and their backbone colored in magenta (unprocessed) and yellow (processed).

Table 3
Total number of residues and number of residues in secondary structure elements alpha helix, beta strand, and coil, in some hydrogenase LSUs.

Bacterium	Structure (PDB id.)	Total	Alpha	Beta	Coil
Desulfovibrio vulgaris	Crystal (1WUL)	534	223	105	206
D. vulgaris strain Miyazaki F	Crystal (1UBK)	534	222	102	210
D. desulfuricans	Crystal (1E3D)	537	222	96	219
Escherichia coli Hyd-1	Crystal (3UQY)	587	226	114	247
E. coli	Crystal (5A4M)	581	227	108	246
Hydrogenovibrio marinus	Crystal (3AYX)	595	234	114	247
Allochromatium vinosum	Crystal (3MYR)	559	235	88	236
Ralstonia eutropha	Crystal (3RGW)	602	238	112	252
Rhizobium leguminosarum	Unprocessed, initial[a]	596	230	104	262
R. leguminosarum	Unprocessed, final[b]	596	203	86	307
R. leguminosarum	Processed, initial[a]	581	240	113	228
R. leguminosarum	Processed, final[b]	581	201	113	267

[a] I-TASSER model structure of HupL used as initial geometry in MD simulations.

[b] Final structures of HupL after MD 50-ns simulations.

precursor form. C75 is predicted to place at a somewhat intermediate location between iron and nickel atoms in all cases (Fig. 6A and B). As a consequence, MD predicts rather similar mobility for $Fe(CN)_2CO$ in both unprocessed and processed forms whereas nickel exhibits rather different RMSD variations in the two forms (Fig. 5). These final geometries include R508 at a close distance of cofactor atoms (Fig. 6A and B) in both unprocessed and processed forms. As noted above, this arginine was found to have an anomalously high pK_a indicating that its protonated charged state is strongly favoured.

Of special interest was the finding that in one of the MD simulations on unprocessed HupL, the final geometry showed nickel considerably moved outwards from cysteines in the binding site and linked by two nearby glutamates E27 and E589 (Fig. 6C). Given that E589 belongs to the disordered CTT and that MD results show an extreme flexibility for this segment, one might not expect to find a particular spatial location of the tail frequently reproduced. Distances between nickel and those two glutamates along MD simulations in both unprocessed and processed LSUs provide some clues (Fig. 7). To avoid fluctuations due to flapping motion of carboxylate oxygens, these distances to nickel were computed with the Cδ atom instead of oxygen atoms of glutamates

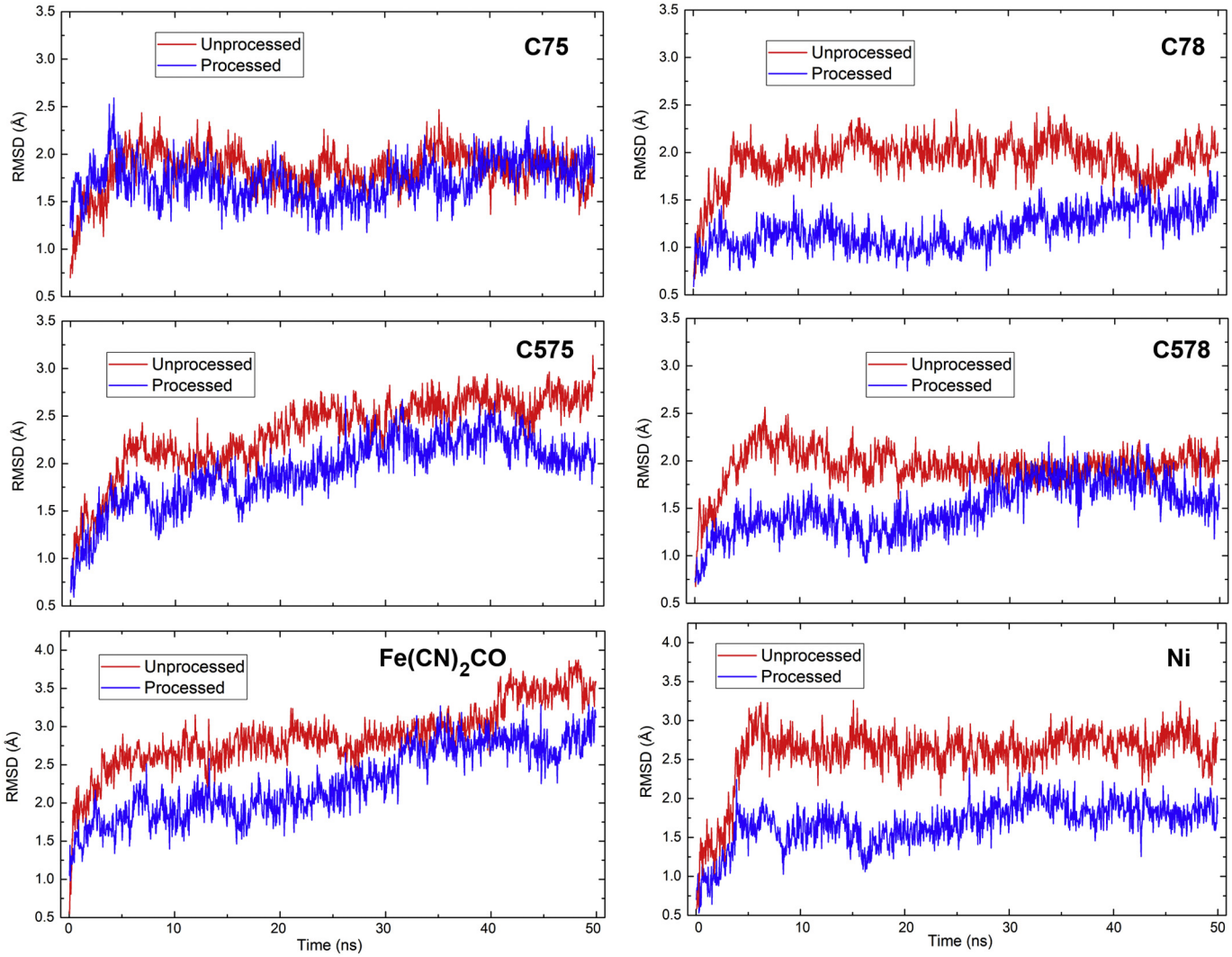


Fig. 5. Average RMSD values of non-hydrogen atoms (backbone and side chain) of the four cysteines in the active site, $Fe(CN)_2CO$ cofactor, and Ni atom in 50-ns MD simulations. Plots compare averages of four trajectories in unprocessed HupL with averages of three trajectories in processed HupL except for Ni atom in which the average of unprocessed trajectories does not include trajectory #1.

(detailed analyses of the MD output shows that Ni-glutamate links are established with the Oε1 atom of both E27 and E589: data not shown). In the case of this particular trajectory (hereafter dubbed "number 1"), after the first 3 ns, E27 locates at a much closer distance to Ni than in any other trajectory of either unprocessed or processed HupL (compare plots in Fig. 7A). At ~9 ns, this Ni-E27·Cδ distance stabilizes at 2.75 Å (std. dev. 0.086 Å). In processed HupL, this distance is so large that it precludes direct nickel···E27 interaction (Fig. 7A). The equivalent Ni-E589·Cδ distance for this glutamate in the highly flexible CTT takes longer to stabilize but at ~35 ns remains at a similar value 2.77 Å (std. dev. 0.062 Å, Fig. 7B). This result led us to conjecture that the disordered nature of the CTT might provoke that some molecules in a given population of HupL could have these two glutamates linking Ni^{2+} (at least for a time lapse) prior to its final incorporation into the active site. Since it is known that nickel is inserted by an accessory metal-binding protein (see, for example refs. [7, 8, 47]), and also that nickel must be recognized by the protease to carry out proteolytic processing [20], this conjecture should imply that E27 and E589 could play a role in enhancing or facilitating such an insertion/recognition.

Upon removal of the CTT, a local change occurs in the vicinity of the site harboring the metallic cluster. It has been traditionally assumed that cleavage of the CTT provokes a conformational change in hydrogenase LSU that buries the active site. Available structures of already processed LSUs show in fact buried sites but that change has not been directly observed. Our MD results provide a new insight into this point. Although protein architecture is preserved, the structures found after 50-ns simulations in unprocessed and processed forms of HupL show changes in many of their coil regions (Fig. 4D). Since there is a large number of coil residues (Table 3) that move with respect to their adjacent secondary structure alpha or beta elements (Fig. 4D), they must have an effect on the exposure of many parts of the structure. The local space in the neighborhood of active site is particularly prone to this effect due to the close proximity of the CTT, and MD results suggest significant differences in the local topography of the surface in this region (Fig. 8). In the presence of the CTT, this local surface defines a pocket which leaves an open access to the site (Fig. 8A) while in the absence of the CTT, this surface is more closed and leaves no access to solvent (Fig. 8B). This might arise not from a marked conformational change but merely from a local backbone variation in coil segments that otherwise had considerable freedom to orient in space.

2.5. HupL E27Q and E589Q mutants are impaired in Ni-dependent processing and hydrogenase activity

The results from Propka and MD analyses shown above indicate a potentially relevant role for glutamate residues E27 and E589 in HupL

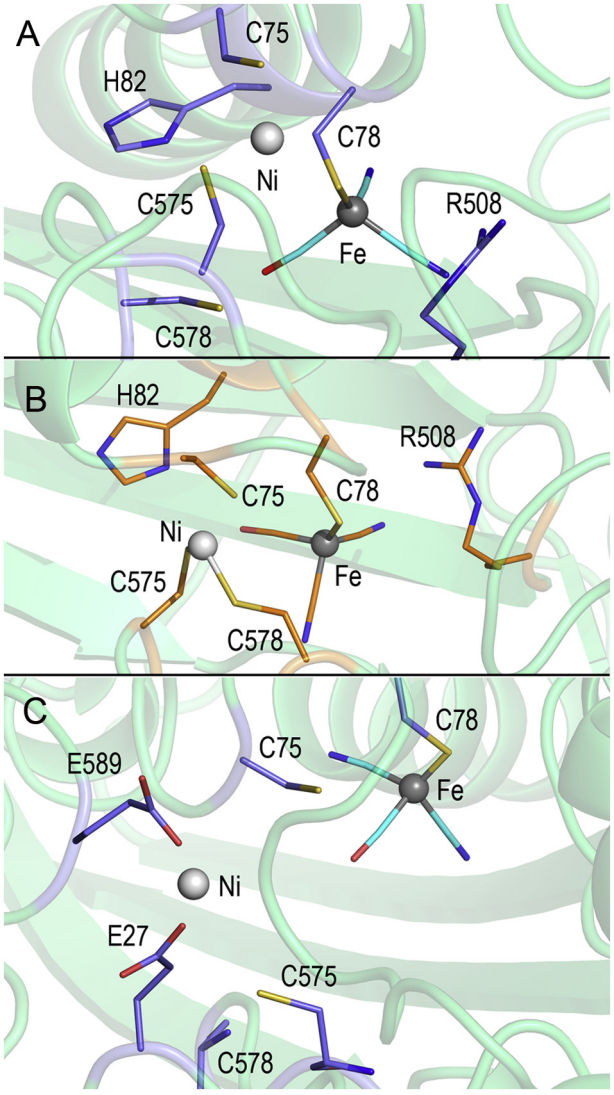


Fig. 6. Final geometries in the site defined by residues within 3.5 Å from both $Fe(CN)_2CO$ cofactor and Ni atom found after 50-ns MD simulations. (A) Representative geometry in three out of four simulations for unprocessed HupL. (B) Representative geometry in the three simulations for processed HupL. (C) Geometry found in one simulation for unprocessed HupL.

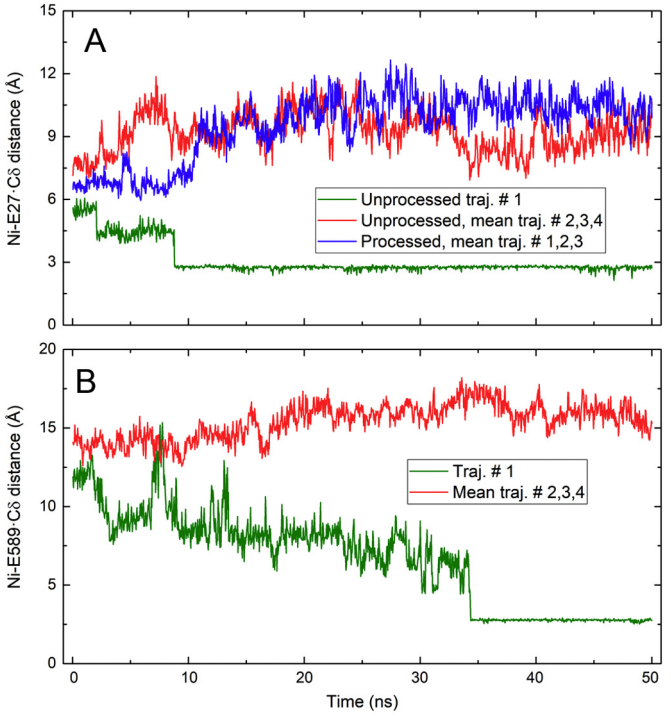


Fig. 7. Distances between Ni atom and carboxylate Cδ atom of Glu27 and Glu589 along 50-ns MD simulations of HupL. (A) Ni-Glu27 distance in trajectory #1 and mean values of three other trajectories (#2,3,4) for unprocessed HupL together with mean values of this distance in three trajectories (#1,2,3) for processed HupL. (B) Similar data for Ni-Glu589 distance only in unprocessed HupL (this CTT amino acid is absent in processed HupL).

protein processing/activity. In order to obtain experimental evidence for this possibility, we constructed mutant versions of the protein in which these residues were changed into glutamines, a non-ionizable but structurally similar residue. Starting from plasmid pPM505, harboring the *hupL*$_{ST}$ gene (encoding HupL with an N-terminal StrepTag fusion, and considered here as wild-type version), site-directed mutants E27Q and E589Q, and double mutant E27Q/E589Q were generated, resulting in plasmids pPM505.E27Q, pPM505.E589Q, and pPM505.E27Q/E589Q, respectively. These were introduced into the *hupL*-deficient strain *Rlv* UPM1155 (pALPF2), and bacterial cultures of the corresponding transconjugant strains were induced for hydrogenase activity under microaerobic conditions in a standard medium (YMB), and in the same medium supplemented with Ni^{2+} (5 μM). In these experiments, the recipient *hupL*-deleted strain showed no detectable hydrogenase activity and was used as negative control (Table 4). The level of general induction of the entire *hup/hyp* cluster was comparable in all strains, as shown for the similar levels of HypB protein in the extracts (Fig. 9, bottom panel). The presence of plasmid pMP505 in the *hupL*-deficient strain was associated with high levels of hydrogenase activity in cultures microaerobically induced in YMB medium, and this level was increased by ca. 30% when the induction medium was supplemented with additional Ni^{2+} (Table 4). Crude extracts from the same microaerobic cultures were used for immunodetection experiments with anti-HupL antisera, using a strain deficient in the endoprotease HupD as control of unprocessed protein. These experiments showed that the presence of plasmid pPM505 was associated with two bands, corresponding to the unprocessed and processed forms of LSU (Fig. 9). The intensity of the faster-migrating, processed form of HupL$_{ST}$ was significantly increased when additional nickel was present in the medium, consistently with the data on hydrogenase activity and also with the nickel-dependent limitation of *Rlv* hydrogenase processing and activity previously observed [34].

We then examined hydrogenase activity and processing associated with the mutation E589Q, affecting the conserved glutamate located in the CTT of hydrogenase LSU. Microaerobic cultures carrying pPM505.E589Q induced levels of hydrogenase activity that were ca. 40% lower than those from cells carrying the wild-type enzyme (Table 4). More interestingly, this level of activity was not increased upon nickel addition; in fact, a reduction in the level of activity of mutant E589Q was observed under these conditions. Western blot immunodetection of HupL$_{ST}$ in the extracts from the strain expressing the E589Q variant induced in YMB revealed the presence of bands corresponding to both processed and unprocessed forms of the hydrogenase LSU. Both bands were less intense than in the case of the wild type (Fig. 9). Consistently with the observed data on hydrogenase activity, the level of processed band in E589Q mutant did not increase, but rather decreased, in response to nickel addition to the medium and higher levels of the unprocessed form of HupL were observed. Our interpretation of this difference with the wild-type strain is that E589Q mutation affects the way through which nickel incorporates into the LSU protein prior to protein processing. In addition to that, the observed lower level of unprocessed HupL in the E589Q strain suggests that this mutation might also cause some reduction on the stability of

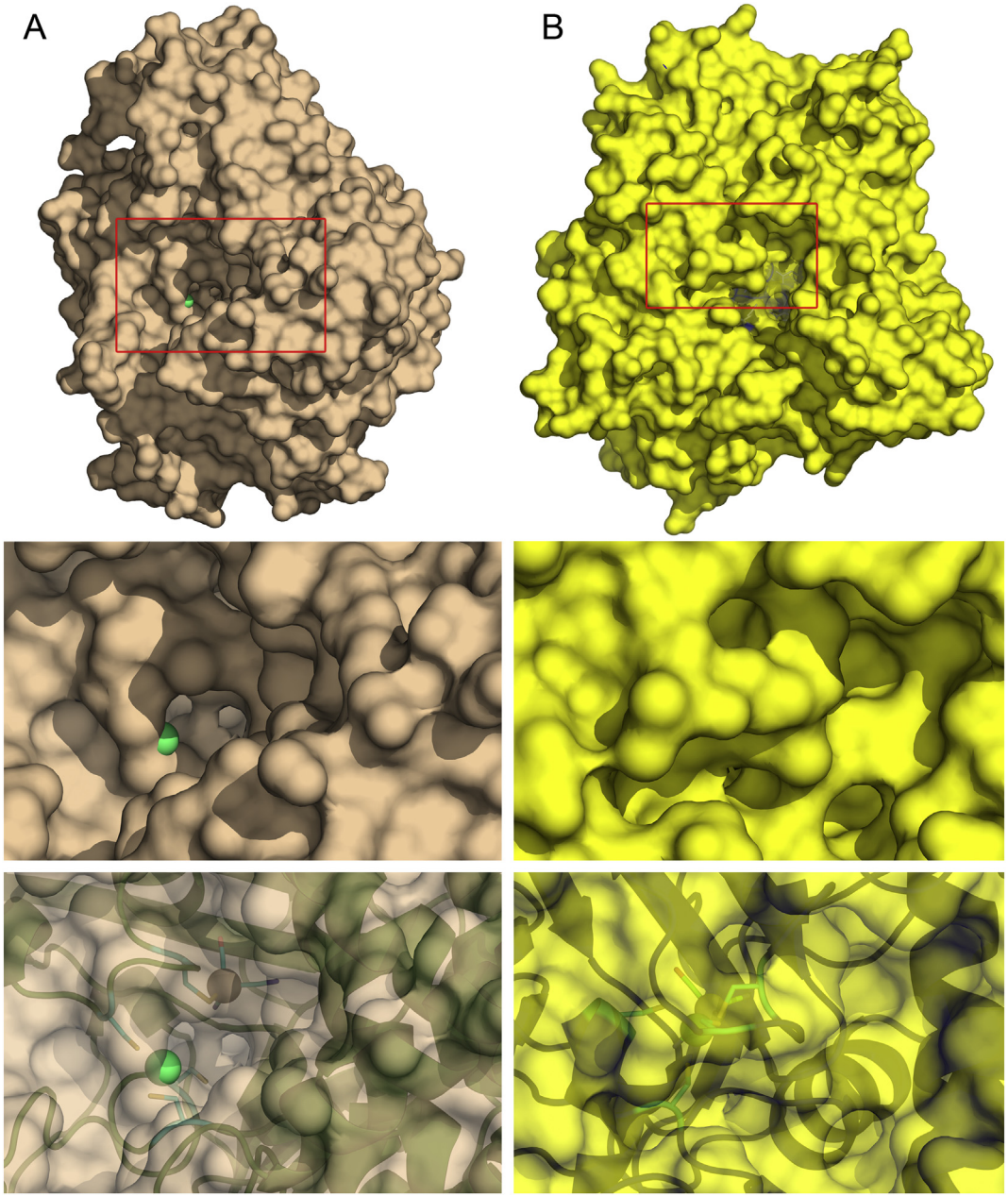


Fig. 8. Protein surfaces of representative final structures after 50-ns simulations of HupL. (A) Unprocessed form. Surface region marked with a red box corresponds to the binding site with $Fe(CN)_2CO$ cofactor and Ni atom (green sphere) that is zoomed below as opaque and partially translucent surfaces. (B) Same for processed HupL.

the unprocessed enzyme.

Analysis of $HupL_{ST}$.E27Q variant showed a stronger phenotype than in the case of E589Q. Microaerobic cultures of E27Q mutant showed very low levels of hydrogenase activity (ca. 6% of the wild type), that decreased to undetectable levels in the presence of additional nickel in the induction medium (Table 4). Immunoblot analyses of extracts from this mutant revealed the presence of both unprocessed and processed forms of $HupL_{ST}$. The band corresponding to the processed form was less intense than that of the wild type (Fig. 9). Furthermore, the addition of nickel again resulted in a decrease on the amount of processed hydrogenase LSU. These results indicate that residue E27 might have a role in both the processing of the hydrogenase LSU and the activity of the mature enzyme. The double mutant E27Q/E589Q showed a phenotype even stronger than mutant E27Q, with almost undetectable levels of hydrogenase activity and low levels of LSU processing when induced in plain YMB medium. These levels became even lower when extra nickel was added to the medium (Table 4 and Fig. 9).

2.6. Structural models of the complex between unprocessed HupL and endoprotease HupD suggest interactions between disordered segments from both proteins

The structure of the endoprotease HupD from *R. leguminosarum* was obtained by homology modelling with SwissModel [48] that selected as template the crystal structure at 2.20 Å resolution (PBD id: 1CFZ) of the HybD endopeptidase from *E. coli* [49]. Recently, the crystal structure at 1.82 Å resolution of the HybD protease from *Thermococcus kodakarensis* (PBD id: 5IJA) has permitted a better understanding of the substrate recognition by providing atomic details on the metal-binding site [50]. In addition, that report revealed the existence of three loops L4, L5, and L6 (defined in terms of identified secondary structure elements) with high B-factor values, implying that their mobility should be relevant for their biological activity. The authors proposed that the flexibility of

Table 4
Effect of substitutions on *R. leguminosarum* HupL glutamate residues E27 and E589 on hydrogenase activity.

Complementing plasmid	Genotype	Hydrogenase activity[a]	
		YMB	YMB + $NiCl_2$
None	Δ*hupL*	< 100	< 100
pPM505	Δ*hupL*/*hupL*$_{ST}$	4500 ± 285	5929 ± 524
pPM505.E27Q	Δ*hupL*/*hupL*$_{ST}$.E27Q	290 ± 146	< 100
pPM505.E589Q	Δ*hupL*/*hupL*$_{ST}$.E589Q	2673 ± 378	2247 ± 269
pPM505.E27Q/E589Q	Δ*hupL*/*hupL*$_{ST}$.E27Q/E589Q	183 ± 83	< 100

The indicated plasmids were used for complementation of the Δ*hupL R. leguminosarum* strain UPM1155(pALPF2) strain. Hydrogenase activities are expressed in nmoles H_2 taken up h^{-1} (mg protein)$^{-1}$ with oxygen as electron acceptor. Values are the average of three independent assays ± S.E. Bacteria were microaerobically induced in YMB medium or in the same medium supplemented with 5 μM Ni_2Cl.

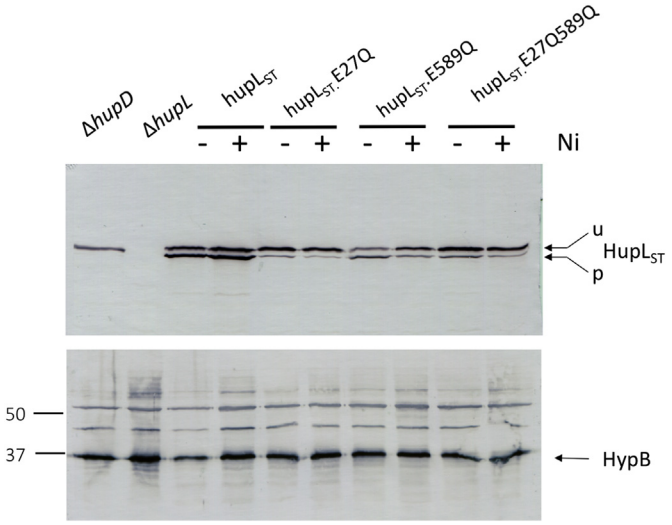


Fig. 9. Effect of mutations E27Q and E589Q on *R. leguminosarum* HupL processing. The pictures show the immunological detection of hydrogenase LSU HupL (top panel) and HypB (bottom panel) in crude extracts from *R. leguminosarum* strains carrying the indicated mutations in HupL$_{ST}$. Proteins were resolved in 9% (HupL) or 12% (HypB) acrylamide SDS-PAGE gels. Numbers on the left margin of the panel indicate the position of the molecular weight standards, in kDa. Marks on the right margin indicate the location of the two forms of HupL$_{ST}$ protein: unprocessed HupL (u, 66 kDa), and processed HupL (p, 65 kDa). Cells were induced microaerobically either in YMB standard medium (−) or in the same medium supplemented with 5 μM $NiCl_2$ (+).

these loops could facilitate their appropriate conformation for interaction with the CTT of immature hydrogenase [50].

The model structure of *R. leguminosarum* HupD superimposes well with this crystal structure of *T. kodakarensis* HybD not used for modelling (Fig. 10A). The structural alignment shows that loops L4, L5, and L6 (defined in HupD with the same criteria as in HybD [50]) reasonably match in both proteins. The crystal structure of HybD from *T. kodakarensis* has no metal, but that from *E. coli* used for homology modelling includes a zinc cation which was incorporated by SwissModel to the HupD model. Replacing this cation with nickel shows that the metal is coordinated in HupD by close E28, D74, and H105 residues (Fig. 10B). A 50-ns MD simulation of this model structure showed a very low mobility of HupD (Fig. S5A). RMSD computed with backbone atoms and averaged over simulation time is 2.17 Å for the complete protein and 1.98 Å excluding loops L4-L6. These loops exhibit a different behaviour. While L5 has an averaged RMSD = 1.65 Å, even lower than that of the whole protein, L4 increases its motion at about half the

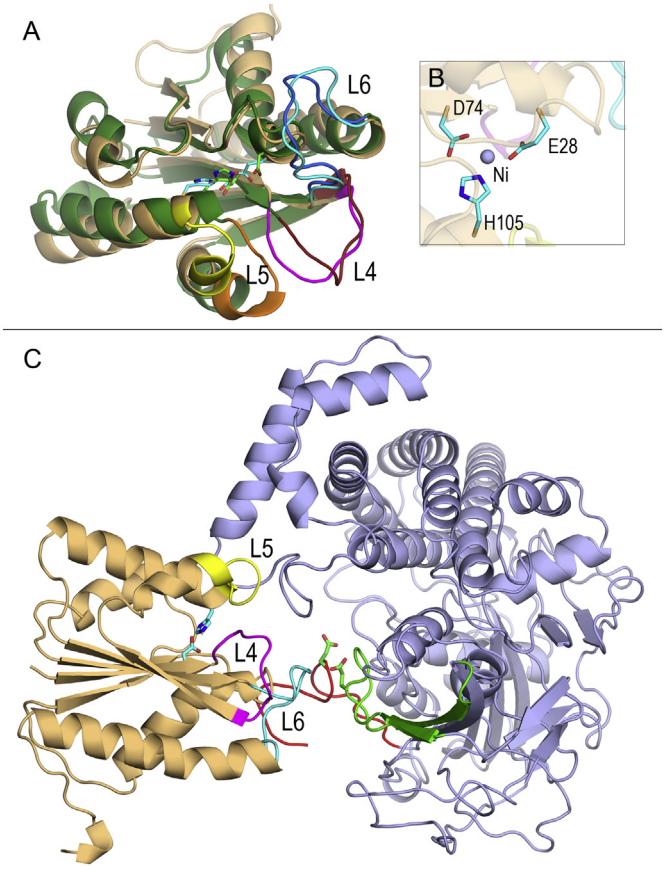


Fig. 10. (A) Superposition of model structure of endoprotease HupD from *R. leguminosarum* (light orange) and crystal structure of endoprotease HybD from *T. kodakarensis* (deep green). Metal-binding residues are shown as sticks with carbons in cyan (HupD) and green (HybD). Flexible loops L4 (residues 76-83 in HupD and 60-68 in HybD), L5 (98-104 in HupD and 81-87 in HybD), and L6 (133-144 in HupD and 119-129 in HybD) are colored magenta, yellow, and cyan, respectively, in HupD and deep red, orange, and marine in HybD. (B) Close view of metal-binding residues in HupD with a nickel ion at the position in which the homology modelling prediction locates a zinc ion. (C) Best cluster solution of protein-protein docking obtained with HADDOCK for the HupD (light orange)-unprocessed HupL LSU (light blue) complex. L4, L5, and L6 loops and metal-binding residues in HupD are colored as in (A). N-terminal and C-terminal regions of HupL proposed to participate in the interaction with the endoprotease are colored green and red, respectively.

simulation time (averaged RMSD = 2.68 Å), and L6 shows a large motion from the very beginning (Fig. S5A) reflected into a high averaged RMSD = 4.14 Å. If one considers the much shorter extent of L5 with respect to L4 and L6 together with the fact that it connects two close helices (features found in both the HupD model and the crystal structure of HybD: Fig. 10A), one could expect a lower mobility for L5 as compared to that of L4 and L6 as these RMSD results illustrate. Metal-binding residues and the nickel ion show a steady low mobility: RMSD computed with all non-hydrogen atoms averaged over simulation time are (Å): 1.40 for E28, 0.85 for D74, 1.03 for H105, and 1.17 for Ni (Fig. S5B).

Possible effects arising from the flexible CTT of unprocessed HupL in the complex formed with HupD were addressed by preparing model static structures of the complex with exploratory protein-protein docking calculations. Blind docking modelling with ClusPro [51] predicted a consistent orientation of the endoprotease with respect to HupL LSU. The 10 best models (most populated clusters) agree in locating HupD in front of the active site region in which the CTT is also present (Fig. S6). This predicted "proper" orientation of HupD for posterior CTT cleavage is interesting if one considers that blind docking implies no a

priori spatial constraints. 5 out of these 10 best models also predict L4-L6 loops oriented towards the CTT/active site (Fig. S6A) although the other 5 models situate the loops at different orientations (Fig. S6B). The best 3 models have 127, 91, and 60 members in the solution clusters (the 4th has 50 and the 5th only 37 members) and they agree in positioning the loops near the CTT (1st and 2nd best models are colored cyan and yellow, respectively, in Fig. S6A).

Using the very best ClusPro model as reference, the geometry of the HupL-HupD complex was refined with HADDOCK [52] by setting in input E28 and H105 of HupD and E27 and E589 of HupL as residues in the protein-protein interface. HADDOCK clustered 103 structures in 8 clusters which represents 51.5% of the water-refined models generated. The three clusters with scores (−113, −112, and −108) clearly better than the remaining (−98 for the fourth cluster) locate HupD loops L4-L6 in the proximity of the CTT of HupL (Fig. S6C). Taking finally the number 1 best structure of the best HADDOCK cluster as initial exploratory static structure of the HupL-HupD complex (Fig. 10C), it is observed that the more mobile loops L4 and L6 of HupD are close to both N- and C-terminal segments of HupL. Since the distance between Ni-binding centers in both proteins is too large in this initial structure, it seems obvious that local changes must occur in all the flexible segments of both HupD and HupL to facilitate appropriate conformation for interaction, as it has been proposed before [50].

3. Discussion

Maturation of the precursor form of NiFe hydrogenase LSU by cofactor insertion and removal of its C-terminal peptide is a characteristic trait of the biosynthetic pathway of these enzymes. It has been suggested that the presence of the CTT is required for maintaining a conformation suitable for binding of LSU precursor with the Hyp machinery and NiFe cofactor insertion [13]. Other studies had indicated that CTT itself does not contain specific binding sites of the LSU to recruit the Hyp machinery nor recognition sequences for the endopeptidase [15,53]. Our finding that the CTT is an intrinsically disordered region opens new possibilities for understanding hydrogenase assembly and functionality. There is currently significant evidence that IDRs have many specific functions that complement functions of ordered proteins (for recent reviews see [54–56]). An IDR can participate in formation of specific but weak complexes so that the combination of high specificity with low affinity is particularly suited to regulatory interactions in which turning a signal on and off is central [56]. Of particular interest is the suggestion that IDRs should enhance interaction with proteases [57] as cleavage sites have often been predicted to be intrinsically disordered [58,59]. In 1994, long before the current intense activity on protein disorder research, it was suggested that to interact with a protease, a protein segment must adopt an extended conformation with at least 12 residues with propensity for local unfolding [60]. In 2010, a survey of experimental structures of proteases confirmed that in almost all cases peptides in protease active sites adopt extended conformations [61]; and quite recently Kwon et al. have proposed that disorder in both the N- and C-terminal segments of hydrogenase LSUs has to play a key role in hydrogenase interactions with auxiliary metalloproteins [62]. According to this proposal, unstructured very flexible regions able to modify conformation should be involved in the association with the protein that inserts nickel in the active site as well as in recognition by the protease that cleaves the CTT in final stages of hydrogenase maturation [62]. It should be highlighted that a central feature of intrinsic disorder in a protein is the ability to modify conformation depending on its interaction partner rather than the very absence of structure [54–56].

Our results on disorder predicted for the CTT of LSUs and the evidence suggesting that IDRs are sensitive substrates for proteases led us to propose that for being removed at hydrogenase maturation, the CTT needs to be intrinsically disordered. This should be a recognition requirement for the endopeptidase to cleave the H-V peptide link between the last residue in the DPCLACSTH consensus motif and the first one in the 15-residue VMSxD/EGxExxxVxVR tail. Since the endopeptidase recognizes the consensus motif only when the nickel atom has been already inserted [20], the presence of the metal is the very first requirement likely needed to anchor the endopeptidase which once placed on site, cleaves the backbone at the H-V link. For this, one could conjecture that a second requirement should be a marked difference in the local electrostatic environment of residues at the cleavage site as it happens in the TH-VM link with polar threonine and histidine just preceding non-polar valine and methionine. And a third requirement should be a large mobility of the backbone chain starting just at this H-V link, which is the same as saying that the peptide segment to be cleaved must be disordered. We stress that our proposals support the suggestions by Kwon et al. made on the basis of a crystal structure of the complex formed with the unprocessed LSU from *T. kodakarensis* and the Ni-inserting protein in this bacterium [62].

An additional effect of the presence of the CTT regards the environment of nickel atom. Our MD simulations indicate that it is different in unprocessed vs. processed forms of HupL. Binding of nickel to cysteines C575 and C578 is predicted only in the processed form whereas MD data suggest that nickel is not stable in the vicinity of the active site in the unprocessed forms, with the potential to bind E27 and E589 in a stable structure. As previously reported [34,35] and also shown here, the presence of high levels of nickel lead to increases on hydrogenase activity and processing of *R. leguminosarum* hydrogenase. The observed decreases on the levels of hydrogenase activity and processing in mutants E27Q and E589Q in response to nickel supplementation suggest that both residues might participate on an adequate, productive nickel incorporation leading to efficient processing by protease. Based on the parallel phenotype regarding Ni-dependent processing associated to E27Q and E589Q mutations, and also on the fact that results from MD simulations do not lead to Ni binding by the cysteine residues known to participate in such binding in the processed form of the LSU (Fig. 6), we propose that the contribution of N-terminal E27 and the C-terminal E589 residues might allow the unprocessed LSU to capture nickel ions and to present them to protease with higher efficiency. These residues could be the first nickel insertion site after delivery of this metal by HypA, thus keeping the ion in an exposed position that might be more accessible for interaction with endoprotease [7]. In addition to this role, our experimental data also indicate the high relevance of E27 for catalysis (Table 4). Analysis of a similar mutant of *Desulfovibrio fructosovorans* LSU (E25Q) showed that this mutant lacks H_2-uptake activity but retains para/ortho H_2 conversion activity, suggesting that the mutation leads to impaired proton transfer but is compatible with the interaction of H_2 with the Ni atom in the active site [63].

The relevance of residues at both N- and C-*termini* of hydrogenase LSU for maintaining a conformation accessible for the interaction with components of the maturation machinery was shown in previous work with *E. coli* HycE [64]. Furthermore, the recent work by Kwon et al. on *T. kodakarensis* shows that HypA in fact interacts with both N- and C-terminal regions of the LSU [62]. In the absence of the CTT, the LSU has no longer a largely flexible, mobile segment in the vicinity of the active site and hence, the local conformation is less susceptible to changes that would leave the active site exposed at times. Our MD results suggest that this is the major effect in removing the CTT in hydrogenase LSUs. In addition to this role in nickel incorporation, our experimental data also indicate a major involvement of E27 in catalysis, given the almost complete suppression of hydrogen uptake activity in the E27Q mutant. This mutation is likely blocking proton transfer in the vicinity of the active site, but still allows the interaction of H_2 with the active site, as deduced from the high levels of para/ortho H_2 conversion activity reported in a similar mutant in *D. fructosovorans* hydrogenase [63].

Computational prediction of pK_a values of ionizable side chains from model structures of unprocessed and processed forms of *Rlv* hydrogenase LSU also provides information about differences that arose

from the presence or absence of the CTT. Of particular interest are the pK_a's predicted for three residues in the vicinity of the active site: E27, E589, and R508. The former has $pK_a \sim 7$ in unprocessed and ~9 in processed HupL, so it is apparent that the spatial proximity of the CTT (also close to the active site region) must be the major cause of that difference. E27 was proposed years ago to play a central role in activation and anaerobic inactivation processes of [NiFe] hydrogenases [65]. The strong phenotype of the $HupL_{ST}$.E27Q mutant described here confirms the relevant role of this residue in the case of *R. leguminosarum*. Experimental evidence suggested that E25 in *D. fructosovorans* hydrogenase (E27 in *R. leguminosarum*) could form hydrogen bond with a water molecule coming out of the active site during activation. The carboxylic group of this residue would be protonated upon taking a proton from the water molecule which would facilitate the entrance of an OH^- bridge between iron and nickel to the active site as it was suggested in 2004 [63,65]. The recent crystal structure at ultra-high resolution of *D. vulgaris* hydrogenase has permitted the detection of hydrogen atoms and confirmed that suggestion [18]. Time-resolved photolysis established this glutamate as proton acceptor during the transition between Ni_a-C and Ni_a-S states in *Pyrococcus furiosus* hydrogenase [66], in consistency with studies showing that proton transfer in this system is mediated by a residue with $pK_a > 7$ [67]. Our pK_a results for E27 in HupL are consistent with the existence of a glutamate not expected to be negatively charged and spatially close to the active site. While protonation should be the favoured state of E27 in proton transfer associated to the enzymatic activity, our MD simulations open the possibility that E27 could also participate in nickel-binding before hydrogenase maturation (and thus before proton transfer activity). Since the partner glutamate in that process, E589, with an estimated $pK_a \sim 10$ is also predicted not to be charged and is part of the highly flexible CTT, the intrinsic disordered nature of this segment could lead to local conformations in some LSU molecules at some time lapse that would facilitate association with E27 to bind nickel. Interaction with Ni-loaded HypA might increase the possibility of such interaction. Once LSU processing has occurred, interactions of E27 with residues of the SSU need to be taken into consideration to fully understand the actual position of this residue, as it has been shown that mutations in a threonine residue of the SSU affect the interaction of the *D. fructosovorans* E27-equivalent with the environment of Ni atom at the active site [68].

Regarding R508, a $pK_a \sim 19$ found for it in the mature enzyme implies that at pH 7 the concentration of its protonated charged state is about 12 orders of magnitude greater than the concentration of its deprotonated neutral state. Furthermore, as suggested by our MD simulations, R508 could play a prominent role in the interactions around the active site of relevance in possible proton transfer pathways. While it has been claimed that the standard pK_a^0 of arginine in proteins should be revised to 13.8 instead of the value 12.5 conventionally used in Biochemistry [69], R508 (or equivalent numbering in other bacteria) is a particularly strong base in hydrogenases. As recently summarized [12], this highly conserved arginine has been proposed to participate in the H_2 activation step of NiFe hydrogenases through an equilibrium between a state defined by the protonated base in the close proximity of the cofactor with a hydride anion bridging Fe and Ni atoms and a state defined by the deprotonated base oxidizing H_2 which would form with the hydride and the released proton ([70]: see Fig. 3 in this reference). However, this proposal poses a deprotonated state of R508 which, according to our results in the model structures of HupL forms, could only be feasible if the local environment or Arg-508 is strongly modified so as to drastically change its pK_a. This issue could be addressed only after analysing the system formed by the cofactor site plus R508 in the mature, active form of hydrogenase through quantum calculations in order to study electron effects in the suggested equilibrium.

The results presented here suggest that the CTT, in addition to the known role in allowing interaction with Hyp machinery and avoiding premature interaction with hydrogenase small subunit [13,15], might participate as an intracellular chaperone in a double way: i) by facilitating nickel insertion, and ii) by keeping critical residues around active site in an inactive configuration. Once removed by endoprotease, the predicted local structural changes in critical residues close to the heterobimetallic cofactor would lead to proper insertion of nickel and triggering of the catalytic function of the active site. Further confirmations using in vitro hydrogenase maturation systems and analyses of protein-protein interactions will be required to advance on the understanding of the mechanisms underlying the late steps for the synthesis of this fascinating metalloenzyme.

4. Materials and methods

4.1. Bacterial strains, plasmids, and growth conditions

Strains and plasmids used in this study are described in Table S2. *R. leguminosarum* strains were routinely grown at 28 °C in yeast mannitol broth (YMB), tryptone-yeast extract (TY), or *Rhizobium* minimal (Rm) media [13,15,33]. *E. coli* DH5α [71] was used for standard cloning procedures, and *E. coli* S17.1 [72] was used for conjugative plasmid transfer between *E. coli* and *R. leguminosarum*. Antibiotic concentrations used were as follows ($\mu g \cdot ml^{-1}$): kanamycin, 50; tetracycline, 5 (for *R. leguminosarum*) or 10 (for *E. coli*).

4.2. DNA manipulation techniques and mutant construction

DNA manipulations, including purification, restriction, ligation, agarose gel electrophoresis, PCR amplification, and transformation into *E. coli* cells were carried out by standard methods [73]. In order to express $hupL_{ST}$ gene in microaerobic cultures of *R. leguminosarum* strains, plasmid pPM505 was constructed. To this end, $hupL_{ST}$ gene was PCR amplified using plasmid pALPF1 as template and primers NSTHupL5 and HupL3X (Table S3). The amplified fragment was cloned as an *Nde*I-*Xba*I fragment in pPM1350 plasmid, a pBBRMCS derivative containing the microaerobically-inducible P*fixN* promoter [31]. $hupL_{ST}$ protein variants altered in codons for residues E27 and E589 were generated by site-directed mutagenesis of the $hupL_{ST}$ gene on plasmid pPM505. To this aim, PCR amplification using complementary oligonucleotides with adequate centered nucleotide substitutions (Table S3) was performed, followed by *Dpn*I digestion, thus generating plasmids pMP505.E27Q and pMP505.E589Q. The double mutant E27QE589Q was constructed by repeating the site-directed mutagenesis process on plasmid pMP505.E27Q with oligonucleotides corresponding to the other mutation. Substitution of target nucleotides as well as absence of additional mutations was verified by sequencing.

4.3. Hydrogenase activity assays

Hydrogenase activity in free living microaerobic cell cultures was measured by an amperometric method using a Clark-type hydrogen sensor (Unisense©, Denmark) with oxygen as electron acceptor [30]. Hydrogenase activity in vegetative cells was induced in bacterial cultures grown under microaerobiosis conditions (1% O_2) as previously described [74]. In these assays *Rlv* cultures were grown in YMB or in YMB supplemented with 5 μM $NiCl_2$. This concentration was selected as it was 1/10 of the minimal concentration showing toxicity (data not shown). Protein content of cell cultures was determined by the bicinchoninic acid method [75] after alkaline digestion of cells at 90 °C in NaOH for 10 min, with bovine serum albumin as the standard.

4.4. Western immunoblot analysis

Immunological detection of proteins was carried out by Western blot assays as described previously [34] using antisera raised against *R. leguminosarum* HupL (1:400 dilution) or HypB (1:2000 dilution). Immunoblot analyses were performed with 60 μg (total protein) of crude

extracts, for HupL, or 10 μg for HypB detection.

4.5. Structure modelling

Model structures were constructed for both unprocessed and processed forms of *R. leguminosarum* HupL hydrogenase LSU with I-TASSER [38,39]. This hybrid predictor method was also used to model complete structures of unprocessed LSUs that include the CTT segment for the following [NiFe] hydrogenases: *H. marinus* HoxH, *R. eutropha* HoxG, and *E. coli* Hyd1. The NiFe cofactor was added as explained in Results Section 2.1 and then the local geometries in a neighborhood of 8 Å around $Fe(CN)_2CO$ and Ni atoms were optimized in water with VMD 1.9.2 [76] in AutoIMD mode using the same force field specifications given in Section 4.7 below. Model structure of *R. leguminosarum* endoprotease HupD was obtained by homology modelling with Swiss-Model [48]. Disorder prediction from sequence was achieved by employing the predictors DisEMBL [77], PONDR [78], and IUPred [79,80]. Model structures of the HupL-HupD complex were initially prepared at protein-protein blind docking calculations with ClusPro [51]. The best model (most populated cluster in ClusPro results) was then refined with HADDOCK [52]. Molecular graphics and structure analyses were performed with PyMOL 2.0.3 [81] and Chimera 1.12 [82].

4.6. Calculations of pKa and protonation states

pKa values of ionizable side chains were computed with Propka 3.1 [45,46]. This is an updated version of the Propka predictor that incorporates a new algorithm for modelling non-covalently coupled residues which can influence the titration of each other because of their close spatial proximity [46]. We used Propka 3.1 implemented in the pdb2pqr server [83,84] which adds hydrogens as needed by each protonation state and optimizes local conformations to fix possible steric clashes.

4.7. Molecular dynamics calculations

50-ns MD simulations were performed for unprocessed forms of the following hydrogenases: *R. leguminosarum* HupL, *H. marinus* HoxH, and *E. coli* Hyd1. Four simulations were run for HupL, two for HoxH, and two for Hyd1 starting in all cases at the I-TASSER modelled structures. Three 50-ns simulations were run for processed HupL. These calculations used the CHARMM 3.6 force field for proteins [85,86] and were performed with the high performance computing Linux-Power-MPI version of NAMD 2.12 [87] in the Magerit supercomputer of the Technical University of Madrid. CHARMM parameters and atomic charges of the $Fe(CN)_2CO$ cofactor were obtained from quantum calculations for this molecular moiety as explained before [88,89]. Our topological analysis of the quantum electron density calculated for Fe $(CN)_2CO$ revealed that Fe-C links have a considerable covalent character whereas its interaction with sulfide groups of cysteines is essentially ionic with a small participation of covalent bonding between Fe and S atoms [89]. Parameters for Ni^{+2} were adapted from the values of Cd^{2+} and Zn^{2+} included in the general CHARMM 3.6 set for ions and the values of Fe^{2+} obtained in our parameterization of $Fe(CN)_2CO$. This way, no specific force field parameters were used for metal-protein interactions that were thus treated with the conventional non-bonded terms in the CHARMM force field energy [86]. Negatively charged, deprotonated states were set for thiolate groups of the four cysteines in the cofactor-binding site during all the MD simulations. Proteins were immersed in 3D periodic solvation boxes with 15 Å spacing, water molecules were added according to the TIP3P model [90], and Na^+ and Cl^- ions added to counter total charges of proteins while providing 0.150 M salt concentration. This procedure resulted in systems composed of about 80,000 atoms. The particle-mesh Ewald summation method was used for long-range electrostatics [91] and a 10 Å cutoff was set for short-range non-bonded interactions. Initial geometries were optimized at 5000 conjugate-gradient minimization steps, water was equilibrated at 298 K and 1 atm for 100 ps at 2 fs time steps, and 50-ns simulations at 2 fs time steps (25 million steps per calculation) were run in the NPT ensemble at 298 K and 1 atm. Langevin dynamics for T control and Nosé-Hoover Langevin piston method for P control were used. With the total number of atoms in the molecular systems studied and even with the considerable computing power of the Magerit supercomputer, each trajectory took NAMD running between 45 and 50 h in 512 processors. NAMD output was stored every 20,000 steps rendering trajectories composed of 1250 frames which were processed and analyzed with VMD 1.9.2 [76] and Carma [92,93]. Root mean square deviation (RMSD) were computed for backbone atoms in proteins and for cofactor atoms with respect to initial protein structures. Root mean square fluctuations (RMSF) were computed for alpha carbons along the whole simulations. Both RMSD and RMSF values were obtained with VMD 1.9.2.

Transparency document

The Transparency document associated with this article can be found, in online version.

Acknowledgments

All molecular dynamics calculations were performed on the Magerit supercomputer of Universidad Politécnica de Madrid. The authors thankfully acknowledge the computer resources, technical expertise, and assistance provided by the Supercomputing and Visualization Center of Madrid (CeSViMa). We also thank Dr. H. Manyani for the construction of pMP505 plasmid. We dedicate this paper to Professor Tomás Ruiz-Argüeso on his 75th birthday, as a recognition to his outstanding and inspiring work on hydrogenases from endosymbiotic bacteria.

Conflict of interest

The authors declare that they have no conflict of interest with the contents of this article.

Author's contribution

Marta Albareda: Investigation, Writing-Reviewing and Editing, Visualization. Luis Fernandez-Pacios: Investigation, Software, Writing-original draft preparation, Visualization, Supervision. Jose Palacios: Conceptualization, Writing-Reviewing and Editing, Supervision, Funding acquisition.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.bbabio.2019.01.001. This work was funded by Spain's MINECO grants BIO2013-4043 (to J.P.) and BIO2017-84548 (to L.F.P.), and UPM grant PINV-18-ADPEFJ-9-4LIMUX (to M.A.).

References

[1] W. Lubitz, H. Ogata, O. Rudiger, E. Reijerse, Hydrogenases, Chem. Rev. 114 (2014) 4081–4148.
[2] J.W. Peters, G.J. Schut, E.S. Boyd, D.W. Mulder, E.M. Shepard, J.B. Broderick, P.W. King, M.W. Adams, [FeFe]- and [NiFe]-hydrogenase diversity, mechanism, and maturation, Biochim. Biophys. Acta 1853 (2015) 1350–1369.
[3] R.P. Happe, W. Roseboom, A.J. Pierik, S.P. Albracht, K.A. Bagley, Biological activation of hydrogen, Nature 385 (1997) 126.
[4] A. Volbeda, E. Garcin, C. Piras, A.L. De Lacey, V.M. Fernandez, C. Hatchikian, M. Frey, J.C. Fontecilla-Camps, Structure of the [NiFe] hydrogenase active site: evidence for biologically uncommon Fe ligands, J. Am. Chem. Soc. 118 (1996) 12989–12996.

[5] A. Böck, P.W. King, M. Blokesch, M.C. Posewitz, Maturation of hydrogenases, Adv. Microb. Physiol. 51 (2006) 1–71.
[6] M.J. Lacasse, D.B. Zamble, [NiFe]-Hydrogenase maturation, Biochemistry 55 (2016) 1689–1701.
[7] M.J. Lacasse, C.D. Douglas, D.B. Zamble, Mechanism of selective nickel transfer from HypB to HypA, *Escherichia coli* [NiFe]-hydrogenase accessory proteins, Biochemistry 55 (2016) 6821–6831.
[8] S. Watanabe, T. Kawashima, Y. Nishitani, T. Kanai, T. Wada, K. Inaba, H. Atomi, T. Imanaka, K. Miki, Structural basis of a Ni acquisition cycle for [NiFe] hydrogenase by Ni-metallochaperone HypA and its enhancer, Proc. Natl. Acad. Sci. U. S. A. 112 (2015) 7701–7706.
[9] R. Rossmann, M. Sauter, F. Lottspeich, A. Böck, Maturation of the large subunit (HYCE) of *Escherichia coli* hydrogenase 3 requires nickel incorporation followed by C-terminal processing at Arg537, Eur. J. Biochem. 220 (1994) 377–384.
[10] P.A. Ash, R. Hidalgo, K.A. Vincent, Proton transfer in the catalytic cycle of [NiFe] hydrogenases: insight from vibrational spectroscopy, ACS Catal. 7 (2017) 2471–2485.
[11] E. Szőri-Dorogházi, G. Maróti, M. Szőri, A. Nyilasi, G. Rákhely, K.L. Kovács, Analyses of the large subunit histidine-rich motif expose an alternative proton transfer pathway in [NiFe] hydrogenases, PLoS ONE 7 (2012) e34666.
[12] H. Tai, Y. Higuchi, S. Hirota, Comprehensive reaction mechanisms at and near the Ni-Fe active sites of [NiFe] hydrogenases, Dalton Trans. 47 (2018) 4408–4423.
[13] M. Senger, S.T. Stripp, B. Soboh, Proteolytic cleavage orchestrates cofactor insertion and protein assembly in [NiFe]-hydrogenase biosynthesis, J. Biol. Chem. 292 (2017) 11670–11681.
[14] E. Theodoratou, A. Paschos, S. Mintz-Weber, A. Böck, Analysis of the cleavage site specificity of the endopeptidase involved in the maturation of the large subunit of hydrogenase 3 from *Escherichia coli*, Arch. Microbiol. 173 (2000) 110–116.
[15] C. Thomas, E. Muhr, R.G. Sawers, Coordination of synthesis and assembly of a modular membrane-associated [NiFe]-hydrogenase is determined by cleavage of the C-terminal peptide, J. Bacteriol. 197 (2015) 2989–2998.
[16] S.E. Beaton, R.M. Evans, A.J. Finney, C.M. Lamont, F.A. Armstrong, F. Sargent, S.B. Carr, The structure of hydrogenase-2 from *Escherichia coli*: implications for H_2-driven proton pumping, Biochem. J. 475 (2018) 1353–1370.
[17] Y. Higuchi, T. Yagi, N. Yasuoka, Unusual ligand structure in Ni-Fe active centre and an additional Mg site in hydrogenase revealed by high resolution X-ray structure analysis, Structure 5 (1997) 1671–1680.
[18] H. Ogata, K. Nishikawa, W. Lubitz, Hydrogens detected by subatomic resolution protein crystallography in a [NiFe] hydrogenase, Nature 520 (2015) 571–574.
[19] T. Maier, A. Böck, Generation of active [NiFe] hydrogenase in vitro from a nickel-free precursor form, Biochemistry 35 (1996) 10089–10093.
[20] E. Theodoratou, A. Paschos, A. Magalon, E. Fritsche, R. Huber, A. Böck, Nickel serves as a substrate recognition motif for the endopeptidase involved in hydrogenase maturation, Eur. J. Biochem. 267 (2000) 1995–1999.
[21] T. Palmer, B.C. Berks, The twin-arginine translocation (Tat) protein export pathway, Nat. Rev. Microbiol. 10 (2012) 483–496.
[22] J. Kalms, A. Schmidt, S. Frielingsdorf, T. Utesch, G. Gotthard, D. von Stetten, P. van der Linden, A. Royant, M.A. Mroginski, P. Carpentier, O. Lenz, P. Scheerer, Tracking the route of molecular oxygen in O_2-tolerant membrane-bound [NiFe] hydrogenase, Proc. Natl. Acad. Sci. U. S. A. 115 (2018) E2229–E2237.
[23] F. Oteri, M. Baaden, E. Lojou, S. Sacquin-Mora, Multiscale simulations give insight into the hydrogen in and out pathways of [NiFe]-hydrogenases from *Aquifex aeolicus* and *Desulfovibrio fructosovorans*, J. Phys. Chem. B 118 (2014) 13800–13811.
[24] Y. Rippers, T. Utesch, P. Hildebrandt, I. Zebger, M.A. Mroginski, Insights into the structure of the active site of the O_2-tolerant membrane bound [NiFe] hydrogenase of *Ralstonia eutropha* H16 by molecular modelling, Phys. Chem. Chem. Phys. 13 (2011) 16146–16149.
[25] D.M.A. Smith, Y. Xiong, T.P. Straatsma, K.M. Rosso, T.C. Squier, Force-field development and molecular dynamics of [NiFe] hydrogenase, J. Chem. Theory Comput. 8 (2012) 2103–2114.
[26] I. Sumner, G.A. Voth, Proton transport pathways in [NiFe]-hydrogenase, J. Phys. Chem. B 116 (2012) 2917–2926.
[27] A. Leyva, J.M. Palacios, T. Mozo, T. Ruiz-Argüeso, Cloning and characterization of hydrogen uptake genes from *Rhizobium leguminosarum*, J. Bacteriol. 169 (1987) 4929–4934.
[28] C. Sánchez-Cañizares, B. Jorrín, D. Durán, S. Nadendla, M. Albareda, L. Rubio-Sanz, M. Lanza, M. González-Guerrero, I.R. Prieto, B. Brito, G.M. Giglio, L. Rey, T. Ruiz-Argüeso, M.J. Palacios, J. Imperial, Genomic diversity in the endosymbiotic bacterium *Rhizobium leguminosarum*, Genes 9 (2018).
[29] J.M. Palacios, H. Manyani, M. Martínez, A.C. Ureta, B. Brito, E. Báscones, L. Rey, J. Imperial, T. Ruiz-Argüeso, Genetics and biotechnology of the H2-uptake [NiFe] hydrogenase from *Rhizobium leguminosarum* bv. viciae, a legume endosymbiotic bacterium, Biochem. Soc. Trans. 33 (2005) 94–96.
[30] T. Ruiz-Argüeso, F.J. Hanus, H.J. Evans, Hydrogen production and uptake by pea nodules as affected by strains of *Rhizobium leguminosarum*, Arch. Microbiol. 116 (1978) 113–118.
[31] H. Manyani, L. Rey, J.M. Palacios, J. Imperial, T. Ruiz-Argüeso, Gene products of the *hupGHIJ* operon are involved in maturation of the iron-sulfur subunit of the [NiFe] hydrogenase from *Rhizobium leguminosarum* bv. viciae, J. Bacteriol. 187 (2005) 7018–7026.
[32] B. Brito, M. Martínez, D. Fernández, L. Rey, E. Cabrera, J.M. Palacios, J. Imperial, T. Ruiz-Argüeso, Hydrogenase genes from *Rhizobium leguminosarum* bv. viciae are controlled by the nitrogen fixation regulatory protein NifA, Proc. Natl. Acad. Sci. U. S. A. 94 (1997) 6019–6024.
[33] B. Brito, J.M. Palacios, J. Imperial, T. Ruiz-Argüeso, Engineering the *Rhizobium leguminosarum* bv. viciae hydrogenase system for expression in free-living microaerobic cells and increased symbiotic hydrogenase activity, Appl. Environ. Microbiol. 68 (2002) 2461–2467.
[34] B. Brito, J.M. Palacios, E. Hidalgo, J. Imperial, T. Ruiz-Argüeso, Nickel availability to pea (*Pisum sativum* L.) plants limits hydrogenase activity of *Rhizobium leguminosarum* bv. viciae bacteroids by affecting the processing of the hydrogenase structural subunits, J. Bacteriol. 176 (1994) 5297–5303.
[35] A.C. Ureta, J. Imperial, T. Ruiz-Argueso, J.M. Palacios, *Rhizobium leguminosarum* biovar viciae symbiotic hydrogenase activity and processing are limited by the level of nickel in agricultural soils, Appl. Environ. Microbiol. 71 (2005) 7603–7606.
[36] M. Albareda, A. Rodrigue, B. Brito, T. Ruiz-Argueso, J. Imperial, M.A. Mandrand-Berthelot, J. Palacios, *Rhizobium leguminosarum* HupE is a highly-specific diffusion facilitator for nickel uptake, Metallomics 7 (2015) 691–701.
[37] F. Sargent, The model [NiFe]-hydrogenases of *Escherichia coli*, in: R.K. Poole (Ed.), Adv. Microb. Physiol, Academic Press, 2016, pp. 433–507.
[38] A. Roy, A. Kucukural, Y. Zhang, I-TASSER: a unified platform for automated protein structure and function prediction, Nat. Protoc. 5 (2010) 725–738.
[39] J. Yang, R. Yan, A. Roy, D. Xu, J. Poisson, Y. Zhang, The I-TASSER Suite: protein structure and function prediction, Nat. Methods 12 (2015) 7–8.
[40] Y. Shomura, K.-S. Yoon, H. Nishihara, Y. Higuchi, Structural basis for a [4Fe-3S] cluster in the oxygen-tolerant membrane-bound [NiFe]-hydrogenase, Nature 479 (2011) 253–256.
[41] J. Fritsch, P. Scheerer, S. Frielingsdorf, S. Kroschinsky, B. Friedrich, O. Lenz, C.M.T. Spahn, The crystal structure of an oxygen-tolerant hydrogenase uncovers a novel iron-sulphur centre, Nature 479 (2011) 249–253.
[42] A. Volbeda, P. Amara, C. Darnault, J.-M. Mouesca, A. Parkin, M.M. Roessler, F.A. Armstrong, J.C. Fontecilla-Camps, X-ray crystallographic and computational studies of the O2-tolerant [NiFe]-hydrogenase 1 from *Escherichia coli*, Proc. Natl. Acad. Sci. U. S. A. 109 (2012) 5305–5310.
[43] H. Ogata, P. Kellers, W. Lubitz, The crystal structure of the [NiFe] hydrogenase from the photosynthetic bacterium *Allochromatium vinosum*: characterization of the oxidized enzyme (Ni-A State), J. Mol. Biol. 402 (2010) 428–444.
[44] H. Ogata, S. Hirota, A. Nakahara, H. Komori, N. Shibata, T. Kato, K. Kano, Y. Higuchi, Activation process of [NiFe] hydrogenase elucidated by high-resolution X-ray analyses: conversion of the ready to the unready state, Structure 13 (2005) 1635–1642.
[45] M.H.M. Olsson, C.R. Søndergaard, M. Rostkowski, J.H. Jensen, PROPKA3: consistent treatment of internal and surface residues in empirical pKa predictions, J. Chem. Theory Comput. 7 (2011) 525–537.
[46] C.R. Søndergaard, M.H.M. Olsson, M. Rostkowski, J.H. Jensen, Improved treatment of ligands and coupling effects in empirical calculation and rationalization of pKa values, J. Chem. Theory Comput. 7 (2011) 2284–2295.
[47] F.C. Blum, H.Q. Hu, S.L. Servetas, S.L. Benoit, R.J. Maier, M.J. Maroney, D.S. Merrell, Structure-function analyses of metal-binding sites of HypA reveal residues important for hydrogenase maturation in *Helicobacter pylori*, PLoS ONE 12 (2017) e0183260.
[48] S. Bienert, A. Waterhouse, T.A. de Beer, G. Tauriello, G. Studer, L. Bordoli, T. Schwede, The SWISS-MODEL repository-new features and functionality, Nucleic Acids Res. 45 (2017) D313–D319.
[49] E. Fritsche, A. Paschos, H.G. Beisel, A. Bock, R. Huber, Crystal structure of the hydrogenase maturating endopeptidase HYBD from *Escherichia coli*, J. Mol. Biol. 288 (1999) 989–998.
[50] S. Kwon, Y. Nishitani, S. Watanabe, Y. Hirao, T. Imanaka, T. Kanai, H. Atomi, K. Miki, Crystal structure of a [NiFe] hydrogenase maturation protease HybD from *Thermococcus kodakarensis* KOD1, Proteins 84 (2016) 1321–1327.
[51] D. Kozakov, D.R. Hall, B. Xia, K.A. Porter, D. Padhorny, C. Yueh, D. Beglov, S. Vajda, The ClusPro web server for protein-protein docking, Nat. Protoc. 12 (2017) 255–278.
[52] G.C.P. van Zundert, J.P.G.L.M. Rodrigues, M. Trellet, C. Schmitz, P.L. Kastritis, E. Karaca, A.S.J. Melquiond, M. van Dijk, S.J. de Vries, A.M.J.J. Bonvin, The HADDOCK2.2 Web server: user-friendly integrative modeling of biomolecular complexes, J. Mol. Biol. 428 (2016) 720–725.
[53] M. Albareda, G. Buchanan, F. Sargent, Identification of a stable complex between a [NiFe]-hydrogenase catalytic subunit and its maturation protease, FEBS Lett. 591 (2017) 338–347.
[54] M. Fuxreiter, Á. Tóth-Petróczy, D.A. Kraut, A.T. Matouschek, R.Y.H. Lim, B. Xue, L. Kurgan, V.N. Uversky, Disordered proteinaceous machines, Chem. Rev. 114 (2014) 6806–6843.
[55] F. Meng, V.N. Uversky, L. Kurgan, Comprehensive review of methods for prediction of intrinsic disorder and its molecular functions, Cell. Mol. Life Sci. 74 (2017) 3069–3090.
[56] V.N. Uversky, A decade and a half of protein intrinsic disorder: biology still waits for physics, Protein Sci. 22 (2013) 693–724.
[57] B. Eldridge, R.N. Cooley, R. Odegrip, D.P. McGregor, K.J. FitzGerald, C.G. Ullman, An in vitro selection strategy for conferring protease resistance to ligand binding peptides, Protein Eng. Des. Sel. 22 (2009) 691–698.
[58] A. Fontana, P. Polverino de Laureto, B. Spolaore, E. Frare, M. Zambonin, Detecting disordered regions in proteins by limited proteolysis, in: V.N. Uversky, S. Longhi (Eds.), Assessing Structures and Conformations of Intrinsically Disordered Proteins, John Wiley & Sons, Wiley, NY, 2010.
[59] M.J. Suskiewicz, J.L. Sussman, I. Silman, Y. Shaul, Context-dependent resistance to proteolysis of intrinsically disordered proteins, Protein Sci. 20 (2011) 1285–1297.
[60] S.J. Hubbard, F. Eisenmenger, J.M. Thornton, Modeling studies of the change in conformation required for cleavage of limited proteolytic sites, Protein Sci. 3 (1994) 757–768.
[61] P.K. Madala, J.D.A. Tyndall, T. Nall, D.P. Fairlie, Update 1 of: proteases universally recognize beta strands in their active sites, Chem. Rev. 110 (2010) PR1–PR31.

[62] S. Kwon, S. Watanabe, Y. Nishitani, T. Kawashima, T. Kanai, H. Atomi, K. Miki, Crystal structures of a [NiFe] hydrogenase large subunit HyhL in an immature state in complex with a Ni chaperone HypA, Proc. Natl. Acad. Sci. U. S. A. 115 (2018) 7045–7050.
[63] S. Dementin, B. Burlat, A.L. De Lacey, A. Pardo, G. Adryanczyk-Perrier, B. Guigliarelli, V.M. Fernandez, M. Rousset, A glutamate is the essential proton transfer gate during the catalytic cycle of the [NiFe] hydrogenase, J. Biol. Chem. 279 (2004) 10508–10513.
[64] A. Magalon, A. Bock, Analysis of the HypC-hycE complex, a key intermediate in the assembly of the metal center of the *Escherichia coli* hydrogenase 3, J. Biol. Chem. 275 (2000) 21114–21120.
[65] A.L. De Lacey, A. Pardo, V.M. Fernández, S. Dementin, G. Adryanczyk-Perrier, E.C. Hatchikian, M. Rousset, FTIR spectroelectrochemical study of the activation and inactivation processes of [NiFe] hydrogenases: effects of solvent isotope replacement and site-directed mutagenesis, J. Biol. Inorg. Chem. 9 (2004) 636–642.
[66] B.L. Greene, G.E. Vansuch, C.-H. Wu, M.W.W. Adams, R.B. Dyer, Glutamate gated proton-coupled electron transfer activity of a [NiFe]-hydrogenase, J. Am. Chem. Soc. 138 (2016) 13013–13021.
[67] B.L. Greene, C.-H. Wu, G.E. Vansuch, M.W.W. Adams, R.B. Dyer, Proton inventory and dynamics in the Nia-S to Nia-C transition of a [NiFe] hydrogenase, Biochemistry 55 (2016) 1813–1825.
[68] A. Abou-Hamdan, P. Ceccaldi, H. Lebrette, O. Gutierrez-Sanz, P. Richaud, L. Cournac, B. Guigliarelli, A.L. De Lacey, C. Leger, A. Volbeda, B. Burlat, S. Dementin, A threonine stabilizes the NiC and NiR catalytic intermediates of [NiFe]-hydrogenase, J. Biol. Chem. 290 (2015) 8550–8558.
[69] C.A. Fitch, G. Platzer, M. Okon, B. Garcia-Moreno, E.L. McIntosh, Arginine: its pKa value revisited, Protein Sci. 24 (2015) 752–761.
[70] S.B. Carr, R.M. Evans, E.J. Brooke, S.A.M. Wehlin, E. Nomerotskaia, F. Sargent, F.A. Armstrong, S.E.V. Phillips, Hydrogen activation by [NiFe]-hydrogenases, Biochem. Soc. Trans. 44 (2016) 863–868.
[71] D. Hanahan, Studies on transformation of *Escherichia coli* with plasmids, J. Mol. Biol. 166 (1983) 557–580.
[72] R. Simon, U. Priefer, A. Pühler, Vector plasmids for in-vivo and in-vitro manipulations of Gram-negative bacteria, in: A. Pühler (Ed.), Molecular Genetics of the Bacteria-plant Interactions, Springer-Verlag, Berlin, 1983, pp. 98–106.
[73] J. Sambrook, D.W. Russell, Molecular Cloning: A Laboratory Manual, 3rd ed., Cold Spring Harbor, N.Y., 2002.
[74] M. Albareda, H. Manyani, J. Imperial, B. Brito, T. Ruiz-Argueso, A. Bock, J.M. Palacios, Dual role of HupF in the biosynthesis of [NiFe] hydrogenase in *Rhizobium leguminosarum*, BMC Microbiol. 12 (2012) 256.
[75] P.K. Smith, R.I. Krohn, G.T. Hermanson, A.K. Mallia, F.H. Gartner, M.D. Provenzano, E.K. Fujimoto, N.M. Goeke, B.J. Olson, D.C. Klenk, Measurement of protein using bicinchoninic acid, Anal. Biochem. 150 (1985) 76–85.
[76] W. Humphrey, A. Dalke, K. Schulten, VMD: visual molecular dynamics, J. Mol. Graph. 14 (1996) 33–38.
[77] R. Linding, L.J. Jensen, F. Diella, P. Bork, T.J. Gibson, R.B. Russell, Protein disorder prediction: implications for structural proteomics, Structure 11 (2003) 1453–1459.
[78] B. Xue, R.L. Dunbrack, R.W. Williams, A.K. Dunker, V.N. Uversky, PONDR-FIT: a meta-predictor of intrinsically disordered amino acids, Biochim. Biophys. Acta 1804 (2010) 996–1010.
[79] Z. Dosztányi, V. Csizmok, P. Tompa, I. Simon, IUPred: web server for the prediction of intrinsically unstructured regions of proteins based on estimated energy content, Bioinformatics 21 (2005) 3433–3434.
[80] Z. Dosztányi, V. Csizmók, P. Tompa, I. Simon, The pairwise energy content estimated from amino acid composition discriminates between folded and intrinsically unstructured proteins, J. Mol. Biol. 347 (2005) 827–839.
[81] PyMol, The PyMOL Molecular Graphics System, Schrödinger, LLC, New York, NY, USA, 2017.
[82] E.F. Pettersen, T.D. Goddard, C.C. Huang, G.S. Couch, D.M. Greenblatt, E.C. Meng, T.E. Ferrin, UCSF Chimera—a visualization system for exploratory research and analysis, J. Comput. Chem. 25 (2004) 1605–1612.
[83] T.J. Dolinsky, P. Czodrowski, H. Li, J.E. Nielsen, J.H. Jensen, G. Klebe, N.A. Baker, PDB2PQR: expanding and upgrading automated preparation of biomolecular structures for molecular simulations, Nucleic Acids Res. 35 (2007) W522–W525.
[84] T.J. Dolinsky, J.E. Nielsen, J.A. McCammon, N.A. Baker, PDB2PQR: an automated pipeline for the setup of Poisson–Boltzmann electrostatics calculations, Nucleic Acids Res. 32 (2004) W665–W667.
[85] R.B. Best, X. Zhu, J. Shim, P.E.M. Lopes, J. Mittal, M. Feig, A.D. MacKerell, Optimization of the additive CHARMM all-atom protein force field targeting improved sampling of the backbone ϕ, ψ and side-chain $\chi 1$ and $\chi 2$ dihedral angles, J. Chem. Theory Comput. 8 (2012) 3257–3273.
[86] A.D. McKerell, D. Bashford, M. Bellott, R.L. Dunbrack, J.D. Evanseck, M.J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph-McCarthy, L. Kuchnir, K. Kuczera, F.T.K. Lau, C. Mattos, S. Michnick, T. Ngo, D.T. Nguyen, B. Prodhom, W.E. Reiher, B. Roux, M. Schlenkrich, J.C. Smith, R. Stote, J. Straub, M. Watanabe, J. Wiórkiewicz-Kuczera, D. Yin, M. Karplus, All-atom empirical potential for molecular modeling and dynamics studies of proteins, J. Phys. Chem. B 102 (1998) 3586–3616.
[87] J.C. Phillips, R. Braun, W. Wang, J. Gumbart, E. Tajkhorshid, E. Villa, C. Chipot, R.D. Skeel, L. Kalé, K. Schulten, Scalable molecular dynamics with NAMD, J. Comput. Chem. 26 (2005) 1781–1802.
[88] M. Albareda, L.F. Pacios, H. Manyani, L. Rey, B. Brito, J. Imperial, T. Ruiz-Argueso, J.M. Palacios, Maturation of *Rhizobium leguminosarum* hydrogenase in the presence of oxygen requires the interaction of the chaperone HypC and the scaffolding protein HupK, J. Biol. Chem. 289 (2014) 21217–21229.
[89] M. Albareda, J.-M. Palacios, J. Imperial, L.F. Pacios, Computational study of the Fe(CN)2CO cofactor and its binding to HypC protein, J. Phys. Chem. B 117 (2013) 13523–13533.
[90] W.L. Jorgensen, J. Chandrasekhar, J.D. Madura, R.W. Impey, M.L. Klein, Comparison of simple potential functions for simulating liquid water, J. Chem. Phys. 79 (1983) 926–935.
[91] T. Darden, D. York, L. Pedersen, Particle mesh Ewald: an N·log(N) method for Ewald sums in large systems, J. Chem. Phys. 98 (1993) 10089–10092.
[92] N.M. Glykos, Software news and updates carma: a molecular dynamics analysis program, J. Comput. Chem. 27 (2006) 1765–1768.
[93] P.I. Koukos, M. Glykos Nicholas, Grcarma: a fully automated task-oriented interface for the analysis of molecular dynamics trajectories, J. Comput. Chem. 34 (2013) 2310–2312.