Cell. Mol. Life Sci. 64 (2007) 1063 – 1084
1420-682X/07/091063-22
DOI 10.1007/s00018-007-6477-4
© Birkhäuser Verlag, Basel, 2007

Cellular and Molecular Life Sciences

Review

[FeFe] hydrogenases and their evolution: a genomic perspective

J. Meyer

Chimie et Biologie des Métaux, UMR 5249 CEA-CNRS-UJF, IRTSV, CEA-Grenoble, 38054 Grenoble (France), Fax: +33-4-3878-5487, e-mail: jacques.meyer@cea.fr

Received 7 November 2006; received after revision 21 December 2006; accepted 1 February 2007
Online First 13 March 2007

Abstract. Most hydrogenases (H_2ases), the enzymes that produce or oxidize dihydrogen, possess dimetallic active sites and belong to either one of two phylogenetically distinct classes, the [NiFe] and the [FeFe] H_2 ases. These families of H_2ases share a number of similarities regarding active site structure and reaction mechanism, as a result of convergent evolution. They are otherwise alien to each other, in particular with respect to protein sequence and structure, maturation mechanisms, and distribution among the realms of life. One of the interesting features of [FeFe] H_2ases is their occurrence in anaerobic bacteria, anaerobic protists, and mitochondriate eukaryotes. They thus have the potential to report on important evolutionary events, including transitions from the prokaryote to the eukaryote lifestyle. Genome sequences yield a variety of [FeFe] H_2ase sequences that have been implemented to shed light on the evolution of these proteins and their host organisms.

Keywords. Hydrogenase, iron-sulfur, evolution, bacteria, protists, eukaryotes, bioenergetics.

Introduction

Hydrogenases (H_2ases) are redox metalloenzymes that catalyze the reversible reaction: $H_2 \leftrightarrow 2H^+ + 2e^-$. They are found in a wide range of microorganisms, allowing some of them to use H_2 as a source of reducing power, and others to use protons as oxidants for the disposal of excess reducing equivalents [1, 2]. H_2ases belong to either one of two broad classes, which are defined by the metal composition of their diatomic active sites [1]: two iron atoms in the [FeFe] H_2ases [3] and one iron and one nickel atom in the [NiFe] H_2ases [4]. H_2-forming methylenetetrahydromethanopterin dehydrogenases (Hmd) [5] may be considered as a third class of H_2ase. However, they differ fundamentally from the [FeFe] and [NiFe] H_2 ases: they have a mononuclear Fe active site, and catalyze a reaction differing from the one indicated above [5]. In addition, due to their extremely narrow distribution (they occur in but a few methanogenic archaea) they are of marginal phylogenetic relevance. They will therefore not be further discussed here.

Since their discovery over 70 years ago [6], H_2ases have been investigated in considerable detail. Starting in the 1960s and 1970s, they were purified and characterized from a variety of organisms [7, 8]. From the mid 1980s, cloning and sequencing of increasing numbers (now many hundreds) of H_2ase genes have allowed classification of these enzymes and rationalization of some of their properties ([1] and references therein). From the early 1990s, genes and proteins involved in the maturation of [NiFe] H_2ases have been identified and characterized [9], which paved the way for molecular engineering [10]. In that same decade, the first crystal structures of H_2ases were elucidated, yielding long-expected three-dimensional models [11, 12], and allowing interpretation of biochemical and spectroscopic properties in unprecedented detail. Knowledge of the active site structural frameworks also triggered a renaissance of branches

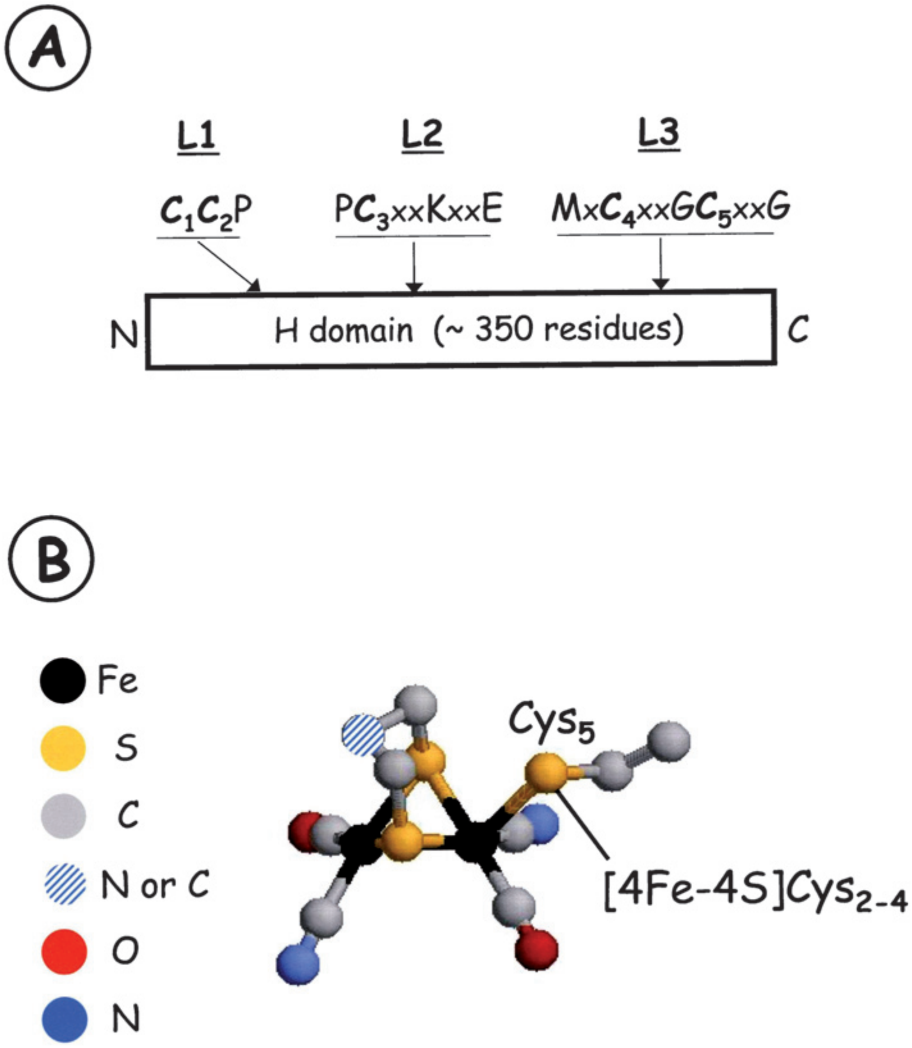

Figure 1. (A) Schematic representation of the H domain. The L1-L3 sequence signatures are simplified versions of previously published ones [1]. Cys 2–5 are ligands of the [4Fe-4S] cluster, Cys5 bridges the [FeFe] and [4Fe-4S] clusters. Cys 1 is believed to act as an acid/base near the active site [3]. (B) The [FeFe] active site with the dithiolate bridging ligand, CO and CN ligands, and Cys5 bridging one Fe to the [4Fe-4S] cluster [3]. The active site shown here is from *D. desulfuricans*; in *C. pasteurianum* a CO molecule bridges the two iron atoms in the as-isolated enzyme [3].

of inorganic chemistry concerned with the elaboration of H_2ase active site models [13, 14].

Altogether, these developments have produced a body of data that has largely outgrown the scope of a single review or even volume. The present review is restricted to recent developments in the genetics and biochemistry of [FeFe] H_2ases, with particular emphasis on insights derived from genome sequences.

[FeFe] and [NiFe] H_2ases

While [FeFe] H_2ases are the main concern of this article, [NiFe] H_2ases are referred to for comparison. The main similarities and differences between these two classes of enzymes are summarized here.

The active sites of [FeFe] and [NiFe] H_2ases share a dimetallic framework and the unique presence of CO and CN as iron ligands [3, 4, 13]. These similarities appear to be requirements of their common catalytic activity [13, 14]. Otherwise, the H_2ase active sites differ not only by their metal composition but also by their attachment to the protein moiety: the [NiFe] site is bound to the polypeptide chain through the thiolate functions of four cysteine residues (or three cysteines and one selenocysteine in some cases [2, 4]), while the [FeFe] site has only one protein ligand, a cysteine which bridges it to a [4Fe-4S] cluster (Fig. 1).

[FeFe] and [NiFe] H_2ases are highly modular proteins consisting of a conserved structural core indispensable for stability and activity, complemented by variable numbers of accessory domains or subunits. The structural cores differ considerably between the two classes of H_2ases. In [FeFe] H_2ases it consists of a ca. 40-kDa (350 amino acids) domain, which has been named ”H domain” [1, 3] for its accommodating the “H cluster” [3, 7] (Fig. 1). The latter includes the binuclear [FeFe] site and the [4Fe-4S] cluster that is connected to it. The main characteristics and sequence signatures of the H domain have been highlighted by crystal structures [12, 15] and multiple sequence alignments [1]. The most conserved sequence patterns encompass the cysteine ligands of the H cluster (Fig. 1). In the [NiFe] enzymes, the structural core consists of a ca. 60-kDa subunit (large subunit) containing the [NiFe] active site and a ca. 30-kDa subunit (small subunit) containing one to three [4Fe-4S] or [3Fe-4S] clusters [4]. The core polypeptides of the two classes of H_2ase do not share any similarity, in either sequence [1] or three-dimensional fold [3, 4].

The accessory domains or subunits contain redox centers, most of which are iron-sulfur (Fe-S) clusters, and are committed to carrying electrons towards and away from the active site. Most accessory domains are shared between the two classes of H_2ases, as well as with other redox proteins, *e.g.*, ferredoxins or complex I (NADH-ubiquinone-oxidoreductase) [1, 16, 17].

In addition to the structural genes, both classes of H_2ases require genes whose products catalyze complex sets of maturation reactions. The maturation systems implemented by the two classes of H_2ases are biochemically distinct, and phylogenetically unrelated [9].

[FeFe] and [NiFe] H_2ases also differ largely by their distribution among organisms [1]. Altogether, there is compelling evidence from several independent lines that [FeFe] and [NiFe] H_2ases are phylogenetically unrelated catalysts of the same reaction, and thus constitute a clear case of convergent evolution.

Structural diversity of [FeFe] H_2ases

Unlike the [NiFe] H_2ases, which are at least dimeric (see above, and [1]), many [FeFe] H_2ases are monomeric. The smallest [FeFe] H_2ases (M, ~45 kDa), which are found in green algae [18–20], may be regarded as minimal forms of these enzymes, as they consist of barely more than the catalytic domain (M1 structural framework, see Fig. 2). The existence of a much smaller (16-kDa) [FeFe] H_2ase in the facultative anaerobe *Enterobacter cloacae* has been proposed [27]. The purported sequence is an assemblage of segments of the H domain, which together amount to only one half of that domain, and lack the important "CCP" (L1) pattern. In fact, the DNA fragment encoding the putative H_2ase has been cloned by PCR and not demonstrated to encompass a complete gene. Furthermore, there is no evidence that the polypeptide of interest contains an active site, and its reported H_2ase activity is at best marginal [27]. The data thus fail to be fully convincing, and for the time being the smallest H_2ases remain the green algal ones [18–20]. The sizes of other monomeric [FeFe] H_2ases cover a wide range, due to the presence, in addition to the H domain, of one (M ~50-kDa, M2 framework) [21], three (M ~65-kDa, M3 framework) [12, 22], and up to five (M ~120-kDa, M5 framework) [25] accessory domains containing Fe-S clusters (Fig. 2).

[FeFe] H_2ases also exist as trimers, as in *Thermotoga maritima* [24], or tetramers, as in *Desulfovibrio fructosovorans* [28] and *Thermoanaerobacter tengcongensis* [23]. In these cases, the catalytic subunit consists of the H domain and three [23, 28] or four [24] Fe-S-containing accessory domains. The additional subunits, which also accommodate Fe-S clusters, are homologous to two peripheral subunits of NADH-ubiquinone oxidoreductase [1, 17]. The latter similarity is in keeping with the fact that these H_2ases are NAD(P)(H) dependent. Trimeric and tetrameric H_2ases are closer kin than suggested by their quaternary structure: indeed, one of the accessory subunits of the trimeric enzyme [24] approximately matches fusions of two of the accessory subunits of the tetrameric ones [1, 23].

All [FeFe] H_2ases listed above are soluble and cytoplasmic, notwithstanding indications to the contrary which are most often erroneous sequence annotations. Periplasmic [FeFe] H_2ases nevertheless do exist, mostly in sulfate-reducing bacteria. These enzymes are dimers [29] derived from monomeric H_2 ases by splitting of the 3' terminus of the gene and insertion of a periplasmic-targeting signal sequence [15].

Membrane-bound [FeFe] H_2ases are a rarity. One occurrence has been inferred from the sequence of a formate dehydrogenase–hydrogenase cluster of genes in *Eubacterium acidaminophilum* [30]. The putative H_2ase operon is composed of four genes, three of which encode a soluble trimeric [FeFe] H_2ase, and the last one a transmembrane protein. However, the catalytic activity of this putative H_2ase remains to be proven, and may be questioned on the grounds that a cysteine ligand of the H cluster (C2 in the L1 pattern, Fig. 1) is replaced by a serine [30] (see below for further discussion of this issue).

Yeast and higher eukaryotes have been found to contain proteins (Narf or Nar1p) homologous to [FeFe] H_2ases [31]. Similar proteins are encoded in all higher eukaryote genomes, and their sequences are indicative of an M2 structural framework. However, the cysteine ligands of one of the accessory clusters are missing, and Cys1 is absent from the active-site sequence pattern L1. The Nar1p protein from *S. cerevisiae* has been characterized in detail and shown to be an Fe-S protein [32] containing two 4Fe-4S clusters [33]. It is involved in the maturation of cytosolic and nuclear, but not mitochondrial, Fe-S proteins [32, 33]. It localizes to the cytosol [32], while the human protein had previously been shown to be nuclear [31]. The latter discrepancy may arise from the fact that the human protein (Narf) is not the true ortholog of the yeast one (Nar1p) [32]. Indeed, more than one copy of these genes is present in some

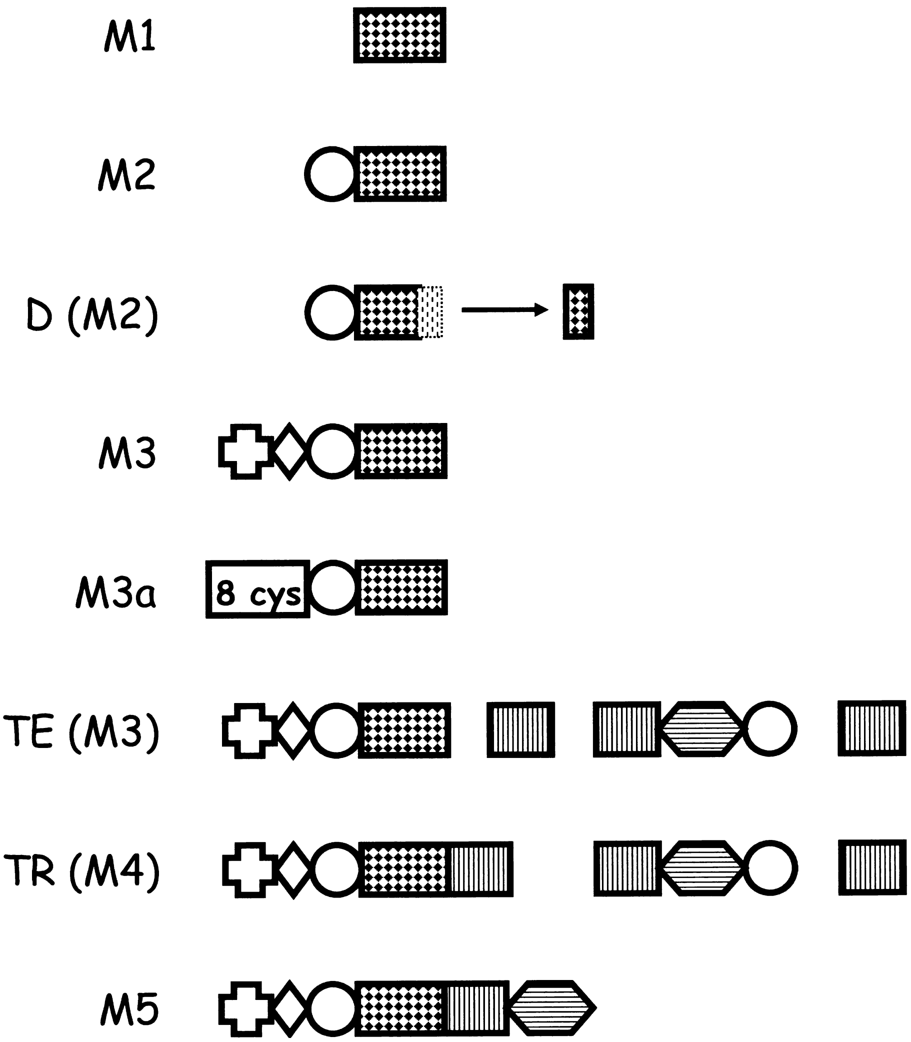

Figure 2. Modular structures of [FeFe] H_2ases. "M" is for monomeric, the numbers indicate increasing size. "D": dimeric, "TR": trimeric, "TE": tetrameric. M1-type H_2ases are unique to green algae [18–20]. Enzymes typical of other frameworks are found in *M. elsdenii* (M2 [21]), *D. desulfuricans* {D(M2) [15]}, *C. pasteurianum* {M3 [12, 22]}, *T. tengcongensis* {TE(M3) [23]}, *T. maritima* {TE(M3) [24]}, and *N. ovalis* (M5 [25]). Domains are not to scale, and subunits are not in the order of the encoding genes on the operons. In principle, HydA subunits are those containing the H domain (checkered rectangle), as indicated in Table 1; however, there are many inconsistencies in the literature. Proposals for an updated nomenclature have been put forward (see Fig. 11 in [1]) but are not repeated here. Codes for the domains are as in [1]: cross, [2Fe-2S] plant-type ferredoxin-like [12]; lozenge, [4Fe-4S] cluster with one His and three Cys ligands [12]; circle, 2[4Fe-4S] ferredoxin-like [12, 15]; square with vertical stripes, similar to thioredoxin-like [2Fe-2S] Fd [26] and NuoE [17]; hexagon with horizontal stripes, similar to NuoF [17]. The square inscribed with "8 cys" is a sequence with eight conserved cysteines, unique to M3a H_2ases (Table 1).

eukaryotic genomes (see below). Biochemical data, as well as the sequences of Nar1p-like proteins [32–34], suggest that these proteins are unable to metabolize H_2. This assumption is further borne out by the absence of maturase genes in the genomes of yeast and higher eukaryotes (see below).

Maturation of [FeFe] H_2ases

With respect to maturation, our understanding of [FeFe] H_2ases has long been considerably weaker than that of [NiFe] H_2ases. At a time when all genes committed to the assembly of [NiFe] H_2ases had been isolated, their products characterized, and most of their functions unraveled [35], nothing was known about their putative counterparts of [FeFe] H_2ases. The main problem with the latter was that, unlike [NiFe] H_2ases, the known structural genes were apparently not located in operons, and in association with putative maturase-encoding genes. The observation that homologs of [NiFe] maturation genes were absent in genomes encoding [FeFe] H_2ases only, led to the prediction that the two classes of H_2ases should have distinct chaperone systems [1]. Also, from the fewer anticipated requirements for [FeFe] H_2ase maturation (only one metal, only one subunit in most cases, no C-terminal processing), it was predicted that a smaller number of genes might be involved [1].

The long-expected breakthrough was achieved through genetic analysis of H_2ase-deficient mutants of the green alga *Chlamydomonas reinhardtii* [36]. Two genes, *hydEF* and *hydG* (the letters A to D being reserved for structural genes [1]), were identified and predicted to encode Fe-S proteins of the radical *S*-adenosyl-L-methionine (SAM) family [36, 37]. The function of *hydEF* and *hydG* was demonstrated by showing that an active [FeFe] H_2ase is obtained when they are coexpressed with the *C. reinhardtii* structural gene *hydA* in *E. coli*, which is otherwise unable to synthesize active [FeFe] H_2ase [36]. *HydEF* was designated with two letters because it is split into two genes, *hydE* and *hydF,* in most other organisms. The sequences of the HydE, HydF and HydG proteins are highly conserved [36, 38], and accordingly the maturation proteins from clostridia are proficient in assembling algal H_2ases [39].

Characterization of the expression products in *E. coli* of the maturation genes from *T. maritima* revealed that HydE is a radical-SAM protein containing two [4Fe-4S] clusters [40], HydF is a GTPase containing one [4Fe-4S] cluster [41], and HydG is a radical-SAM protein containing one [4Fe-4S] and one undefined Fe-S cluster [40]. The functional relevance of these Fe-S clusters has subsequently been assessed by site-directed mutagenesis of their predicted cysteine ligands [42]. The results indicate that both HydE and HydG contain an essential SAM-binding [4Fe-4S] cluster involving three cysteines in the N-terminal region. Additional essential Fe-S clusters are bound to cysteine residues in the C-terminal regions of HydF and HydG [9, 42].

The HydE, HydF, HydG sequences have been implemented to identify homologous proteins in the hydrogenosome of *Trichomonas vaginalis*, *i.e.*, the subcellular compartment where H_2ase itself is found [38]. In *C. reinhardtii* the [FeFe] H_2ases are chloroplastic [20], thus HydEF and HydG might be expected to occur in the same organelle. Indeed, both proteins possess N-terminal extensions [36], but there is no evidence that these sequences are chloroplast-targeting, and the localization of the *C. reinhardtii* maturation proteins has so far remained undetermined.

The maturation machineries of [NiFe] and [FeFe] H_2 ases have very recently been reviewed in detail and compared [9]. At present, the main issue concerns the actual reactions catalyzed by HydE, HydF, and HydG and the mechanisms of [FeFe] H_2ase maturation. The minimally required reactions include synthesis of the CO and CN ligands, assembly and incorporation of the Fe/CO/CN units, and synthesis of the propanedithiolate or di(thiomethyl)amine [34] ligand of the [FeFe] active site. Hypotheses regarding these questions have been put forward [36, 40, 43] and discussed in detail [9]. While the combination of HydE, HydF, and HydG appears to be competent for the assembly of [FeFe] active sites in most organisms, in cannot yet be excluded that additional maturation genes may be required in some cases [9]. Regarding the accessory Fe-S clusters, previous work suggests that their assembly is independent of the biosynthesis of the active site [9, 44], and most probably carried out by the Isc or Suf systems [45].

Identification of [FeFe] H_2ases in sequence databases

In view of the considerable body of data available on all varieties of [FeFe] H_2ases, one would expect the identification of genes encoding these enzymes to be straightforward. Nevertheless, inspection of sequence databases reveals some confusion, and the more so in genome sequences. Many [FeFe] H_2ase-encoding genes are erroneously annotated, notwithstanding the generalized use of powerful search tools such as BLAST [46]. The main source of the problems is the implementation of non-selective bait sequences. Indeed, the modular structures of [FeFe] H_2ases include domains, *e.g.*, ferredoxin- or complex I-like, that are not unique to H_2ases (Fig. 2). Thus, searches implementing full-length H_2ase sequences as baits will retrieve H_2ases, but also large numbers of unrelated sequences: hence the frequent annotations of H_2ase-encoding genes as "ferredoxin" or "NADH-ubiquinone oxidoreductase" (Fig. 3). With the exponential development of massive sequencing, existing errors are transferred to new sequences by automated annotation, and thereby amplified beyond control. To achieve correct and thorough identification of H_2 ase sequences, it is mandatory to bait searches with sequence segments unique to [FeFe] H_2ases, *i.e.*, H domains. For the [NiFe] H_2ases likewise, error-free annotations require the implementation of specific baits such as the NiFe-containing large subunit [1].

Distribution of [FeFe] *versus* [NiFe] H_2ases

In the most recent comprehensive compilation of H_2 ases [1], the following trends were reported. Firstly, [NiFe] H_2ases are more numerous and have a wider distribution among prokaryotes than [FeFe] H_2ases. Secondly, the two classes of H_2ases have qualitatively different distributions: [NiFe] H_2ases occur only in Archaea and Bacteria, while [FeFe] H_2ases occur only in Bacteria and Eukarya. In the intervening years, the volume of sequence data has been growing considerably, through acquisition of complete genome sequences [47], as well as through environmental (metagenomic) sequencing [48]. These data open a much wider field for comparing the distributions of the two classes of H_2ases.

The main sets of environmental sequences consist of data from sea surface waters [49], deep-sea whale falls, and mine drainage [50], which together amount to well over a billion basepairs of non-redundant sequences. While the microbial population diversity of these three biotopes varies considerably, together they include a large number of microbial species, probably in the thousands [49]. Nevertheless, few hits were recorded in BLAST searches implementing [NiFe] and [FeFe] H_2ase sequences as baits, 30 and 10 (with E values better than 10^{-10}), respectively. Metagenomes

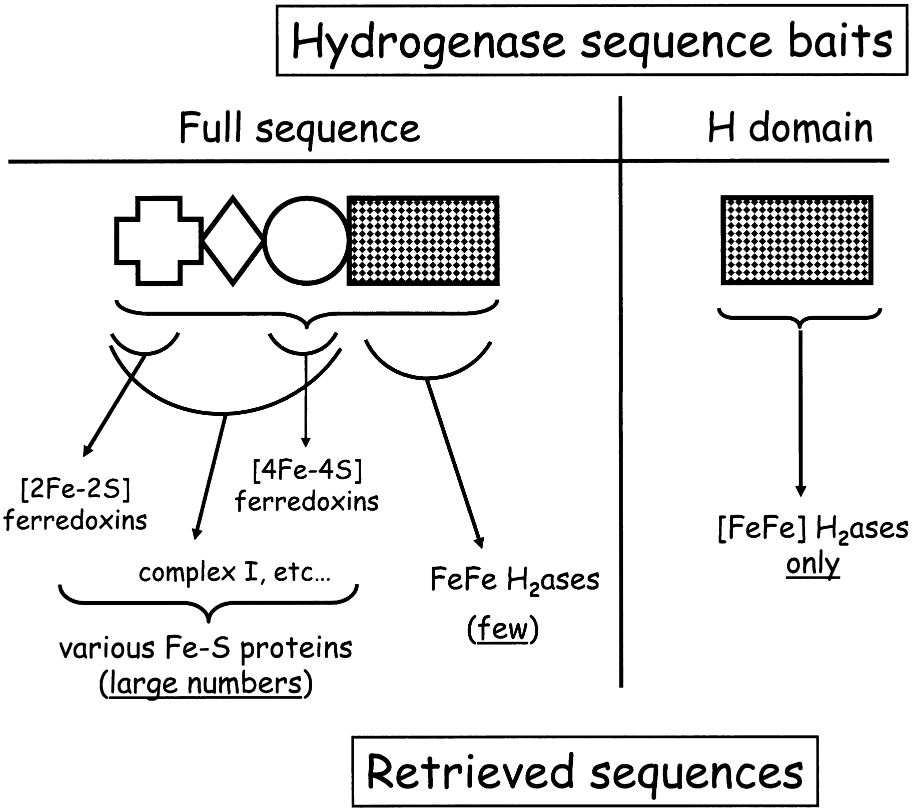

Figure 3. Nonspecific searches and mistaken gene annotations of H_2ases and other Fe-S proteins. Bait sequences are shown in the upper part of the figure, retrieved ones in the lower part. Due to the modular structure of [FeFe] H_2ases, searches baited with full-length sequences (left) yield large numbers of diverse Fe-S proteins among which [FeFe] H_2ases are only a minority. In contrast, searches baited with specific sequences (H domain, to the right) yield only H_2ase sequences. Conversely, searches baited with ferredoxins and other Fe-S proteins may retrieve H_2ases as well. Misidentifications go both ways.

thus suggest that, notwithstanding the importance of H_2 metabolism in the biosphere [1], relatively few microbial species are endowed with the ability to take up or evolve this gas. Environmental sequences also confirm that [NiFe] H_2ases are more numerous than the [FeFe] ones.

An even more formidable mass of data is stored in the non-redundant GenBank (ca. 100 billion basepairs). The numbers of hits (E values better than 10^{-30}) were 208 for the [NiFe] H_2ases and 131 for the [FeFe] H_2 ases. The latter included 55 sequences from higher eukaryotes, which are homologous to [FeFe] H_2ases but lack their enzyme activity (see above, and [32]). Upon exclusion of these, the ratio of [NiFe] to [FeFe] H_2ases is about 3:1. Analysis of the GenBank data also confirmed that [NiFe] H_2ases do not occur in Eukarya, and that Archaea are devoid of [FeFe] H_2ases.

While less massive than the preceding ones, genome sequence databases have the potential to provide clear-cut answers regarding the presence or absence of H_2ase-encoding genes in a given organism, and thus provide reliable information concerning the distribution of H_2ases. Fully sequenced microbial genomes (Table 1) were first scanned by BLAST searches, which showed that putative [NiFe] H_2ase- and [FeFe] H_2ase-encoding genes occur in 87 and 25 genomes, respectively, out of a total of 371 genomes. Twelve genomes, mostly from sulfate-reducing bacteria, apparently contain genes from both classes of H_2ases (see below). Among the protist genome sequences (22 altogether, finished and unfinished), 10 contain [FeFe] H_2ase-encoding genes, while none encode [NiFe] H_2 ases. This rapid overview of genome sequences confirms that H_2ases are present in only a minority of microorganisms, that the [NiFe] H_2ases outnumber the [FeFe] ones by a factor of three, and that the former are restricted to Bacteria and Archaea, while the latter only occur in Bacteria and Eukarya. The next section presents a more detailed analysis of sequenced genomes with respect to genes encoding [FeFe]-H_2ases and their maturases.

[FeFe] H_2ase sequences in genomes

Fully sequenced bacterial genomes containing putative [FeFe] H_2ases are listed in Table 1. Redundancies have been minimized by taking no more than one strain from each species.

Structural genes

Some genomes (*Carboxydothermus hydrogenoformans* [51] and *Geobacter* [52], not included in Table 1) may at first sight appear to encode [FeFe] H_2ases. However, these sequences are in fact sigma-54-dependent transcriptional regulators having an

[FeFe] H_2ase-like sensory module. While these regulators have been hypothesized to activate gene expression in *C. hydrogenoformans* in response to hydrogen availability [51], they are most unlikely to be active H_2ases as they lack essential features of the canonical H domain [52]. In addition, the *Geobacter* genome lacks [FeFe] H_2ase maturation genes altogether. The genomes of *C. hydrogenoformans* and *Geobacter* encode several [NiFe] H_2ases that are probably in charge of the H_2 metabolism [51, 52].

Two *hydA* genes listed in Table 1 (TTE0887 and Tm1421) have not been included in the alignments because they miss important parts of the H domain, and are thus unlikely to encode active enzymes. One of the four *Desulfitobacterium hafniense* sequences (DSY0936), has not been included either: upon its M3-like framework it sports a 500-residue insertion (flavoprotein-like) following the N-terminal [2Fe-2S] domain, and a 100-residue C-terminal extension. A number of other sequences have been included, but are starred and deemed hypothetical because they possess mutations of the important cysteines 1 and 2 of signature L1 (Fig. 1 and Table 1). Cysteine 2 is a ligand of the [4Fe-4S] component of the H cluster, and therefore expected to be important for structure and activity (Fig. 1). In addition, preliminary experiments indicate that mutations of C1 or C2 to serine severely impair H_2ase activity (P. King, private communication). It is therefore intriguing that H_2ases carrying the C2S mutation have been reported to be active [23, 30]. Confirmation of the presence of these mutations and further activity measurements on other proteins known to have similar mutations would be most welcome.

The analysis of several hundred (371, including 28 Archaea) microbial genome sequences has confirmed that [FeFe] H_2ases do not occur in Archaea and are present in only a small number of Bacteria (25). A majority of the latter belongs to the classes Clostridia (10) or δ-proteobacteria (6) (Table 1).

Genome sequences have apparently not revealed novel structural frameworks of [FeFe] H_2ases: most sequences encode enzyme types previously characterized by biochemical methods or classical gene cloning techniques (Table 1). It should nevertheless be kept in mind that enzyme subunits might be overlooked in genomes when the encoding genes are not adjacent or co-cistronic. Also, some interesting variations of the M2 and M3 frameworks have been uncovered (see Fig. 2 and Table 1). Also, the hypothetical DSY0936 sequence from *D. hafniense* would encode a very unusual H_2ase.

While genome sequences of green alga and protists have not been finalized and annotated, several of them are complete enough to allow gene inventories. The organisms listed in Table 1, with the noteworthy exception of the diatom *Thalassiosira pseudonana,* were previously known to contain [FeFe] H_2ase-encoding genes [53]. The *T. vaginalis* genome is remarkable in that it reveals the presence of six or seven such genes, whereas ”only” three or four had been identified by classical cloning techniques [38, 53, 54]. While *T. vaginalis* is an extreme case, it is worth noting that most genomes containing [FeFe] H_2ases genes possess at least two of them.

As indicated above, all higher eukaryotes appear to contain [FeFe] H_2ase-like proteins. Table 1 lists only four of them, which have been implemented in sequence alignments and dendrograms (see below). Interestingly, vertebrate genomes are unique in their containing two [FeFe] H_2ase-like genes, whereas other mitochondriate eukaryotes contain a single copy (see below).

Maturation genes

While most [FeFe] H_2ase-containing genomes accommodate multiple copies of the structural genes, the presence of a single set of maturation genes appears to be the rule. Exceptions are *Desulfotalea psychrophila* (two copies of *hydE*, *hydF* and *hydG*), *Pelobacter carbinolicus* (three copies of *hydE*), *Syntrophobacter fumaroxidans* (two copies of *hydE* and *hydG*), and *T. vaginalis* (triplicates and duplicates of *hydE* and *hydG*, respectively) (Table 1). Rather surprisingly, several [FeFe] H_2ase-encoding genomes appear to contain no maturation genes at all. As the latter have been shown to be mandatory for the production of active enzyme, at least in bacteria [9, 36, 42], it is most likely that the relevant bacteria do not contain active [FeFe] H_2ases. *D. hafniense*, despite containing four *hydA* genes, is devoid of *hydG*, while the candidate *hydE* and *hydF* have too low similarities to their respective families to encode maturases. *D. ethenogenes* contains no maturation genes at all, and likewise the spirochete *T. denticola* and the α-proteobacteria *R. palustris* and *R. rubrum.* The latter two bacteria are known to contain [NiFe] H_2ases, which are in charge of their H_2 metabolism [1]. Altogether, the absence of *hydE*, *hydF* and *hydG* genes in several genomes thus reduces the number of bacterial classes likely to contain active [FeFe] H_2ases.

Among the sequenced eukaryotic genomes, only those of *C. reinhardtii* and *T. vaginalis* appear to encode [FeFe] H_2ase maturases. The absence of such genes was not surprising in higher eukaryotes [38], which do not contain active H_2ases. It was more unexpected, however, in anaerobic protists that are known to produce dihydrogen, such as *E. histolytica* and *G. lamblia* [55]. It should nevertheless be kept in mind that H_2ase activity in these two organisms is at

Table 1. [FeFe] H_2ase sequences from fully sequenced bacterial genomes, and from selected eukaryotic genomes.

Species	Website^{a}	Class	HydA^{b}	Abbrev^{c}	Type^{d}	HydE^{b}	HydF^{b}	HydG^{b}	NiFe^{e}
Bacteroides fragilis NCTC 9343	1	bacteroidetes	BF3662 (-18)	1	M3a^{f}	BF3663 (-59)	BF3665 (-83)	BF3664 (-125)	no
Bacteroides thetaiotaomicron VPI-5482	1	bacteroidetes	BT0124 (-59)	2a	TR (M3)^{g}				no
			BT1834 (-15)	2b	M3a^{f}	BT1835 (-67)	BT1837 (-74)	BT1836 (-125)	
Clostridium acetobutylicum ATCC 824	1	clostridia	CAC0028 (-49)	3a	M3	CAC1631 (-102)	CAC1651 (-94)	CAC1356 (-118)	yes^{h}
			CAC3230 (-17)	3b	M2a^{i}				
Clostridium botulinum ATCC 3502	2	clostridia	CBO1847 (-64)	4a	TR (M3)^{g}	CBO3582 (-76)	CBO3584 (-73)	CBO3583 (-110)	no
			CBO3302 (-24)	4b	M3a^{f}				
			CBO3550 (-21)	4c	M2a^{i}				
Clostridium difficile strain 630	2	clostridia	CD3407 (-61)	5a	TR (M3)^{g}	CD2155 (-64)	CD2154 (-82)	CD2156 (-113)	no
			CD0893 (-23)	5b	M3a^{f}				
			CD0894 (-13)	5c	M3a^{f}				
Clostridium novyi-NT	1	clostridia	NT01CX_1699 (-63)	6a	M2	NT01CX_1484 (-86)	NT01CX_1486 (-78)	NT01CX_1485 (-122)	no
			NT01CX_2259 (-55)	6b	M3				
			NT01CX_0464 (-26)	6c	M3a^{f}				
			NT01CX_1006 (-22)	6d	M2a^{i}				
Clostridium perfringens strain 13	1	clostridia	CPE2346 (-58)	7a	M3	CPE2460 (-99)	CPE0108 (-68)	CPE0110 (-129)	no
			CPE2575 (-24)	7b	M3a^{f}				
			CPE0276 (-19)	7c	M2a^{i}				
Clostridium tetani E88	1	clostridia	CTC02417 (-31)	8a	M3a^{f}	CTC01327 (-76)	CTC01325 (-84)	CTC01326 (-132)	no
			CTC00184 (-20)	8b	M2a^{i}				
Desulfitobacterium hafniense Y51	1	clostridia	DSY4712 (-53)	9a	M2	DSY4958 (-29)	DSY5055 (-11)	none	yes
			DSY0803 (-20)	9c	M2a^{i}				
			DSY4326 (-15)	9b	M3a^{f}				
			DSY0936 (-52)	*^{j}					
Moorella thermoacetica ATCC 39073	1	clostridia	Moth_1717 (-61)	10a*^{j}	M3	Moth_1735 (-98)	Moth_1277 (-96)	Moth_1734 (-129)	no
			Moth_1883 (-45)	10b	M2^{k}				
Thermoanaerobacter tengcongensis MB4	1	clostridia	TTE0894 (-64)	11*^{j,l}	TE(M3)	TTE1073 (-95)	TTE0814 (-98)	TTE1568 (-149)	yes^{l}
			TTE0887 (-12)	*^{j}					
Syntrophomonas wolfei	1	clostridia	Swol_2436 (-63)	12a*	M3				no
			Swol_1017 (-61)	12b*	TR (M3)	Swol_1020 (-68)	Swol_0318 (-92)	Swol_1022 (-154)	
			Swol_1925 (-52)	12c	D (M2)				
Dehalococcoides ethenogenes 195	1	dehalococcoidetes	DET0147 (-59)	13	HYM (M3)^{g}	none	none	none	yes
Symbiobacterium thermophilum IAM 14863		actinobacteria	STH3209 (-59)	14b*^{j}	TR (M3)	STH353 (-89)	STH352 (-86)	STH354 (-126)	no
			STH3293 (-58)	14a	M2^{k}				
Thermotoga maritima MSB8	1	thermotogae	Tm1426 (0)	15a	TR (M4)^{g}	Tm1269 (0)	Tm0445 (0)	Tm1267 (0)	no
			Tm0201 (-65)	15b	M3				
			Tm1421 (-12)	*^{j}					
Treponema denticola ATCC 35405	1	spirochaetes	TDE1593 (-50)	16a	M3	none	none	none	no
			TDE1277 (-21)	16b	M3a^{f}				

Table 1 *(Continued)*

Species	Website[a]	Class	HydA[b]	Abbrev[c]	Type[d]	HydE[b]	HydF[b]	HydG[b]	NiFe[e]
Rhodopseudomonas palustris CGA009	1	α-proteobacteria	RPA0134 (-48)	17	M3	none	none	none	yes
Rhodospirillum rubrum ATCC 11170	1	α-proteobacteria	Rru_A0310 (-25)	18*[j]	M2[k]	none	none	none	yes
Shewanella oneidensis MR-1	1	γ-proteobacteria	SO_3920 (-43)	19	D (M2)[g]	SO_3925 (-56)	SO_3926 (-86)	SO_3923 (-127)	yes
Desulfotalea psychrophila LSv54	1	δ-proteobacteria	DP2379 (-18)	20b	M3a[f]	DP2376 (-60)	DP2378 (-79)	DP2377 (-166)	yes
			DP0479 (-15)	20a	M2[k]	DP0475 (-56)	DP0474 (-24)	DP0476 (-129)	
Desulfovibrio desulfuricans G20	1	δ-proteobacteria	Dde_0475 (-58)	21a	M2[k]				yes
			Dde_2281 (-45)	21c	D (M2)[g]	Dde_2277 (-62)	Dde_2276 (-94)	Dde_2278 (-142)	
			Dde_0081 (-35)	21d	D (M2)[g]				
			Dde_0725 (-14)	21b	M3a[f]				
Desulfovibrio vulgaris Hildenborough	1	δ-proteobacteria	DVU1769 (-39)	22a	D (M2)[g]	DVU1767 (-69)	DVU1768 (-90)	DVU1765 (-127)	yes
			DVU1771 (-51)	22b	M3				
Pelobacter carbinolicus DSM 2380	1	δ-proteobacteria	Pcar_1605 (-62)	23a	TE (M3)[g]	Pcar_1607 (-71)			no
			Pcar_1633 (-59)	23b	TE (M3)[g]	Pcar_1632 (-75)	Pcar_1629 (-94)	Pcar_1631 (-153)	
			Pcar_2502 (-48)	23c*[j]	M3	Pcar_1722 (-72)			
Syntrophobacter fumaroxidans MPOB	1	δ-proteobacteria	Sfum_0844 (-58)	24a*[j]	TR (M3)	Sfum_0841 (-87)		Sfum_0842 (-138)	yes
			Sfum_0848 (-45)	24b	D (M2)	Sfum_1844 (-68)	Sfum_1843 (-93)	Sfum_1841 (-138)	
Syntrophus aciditrophicus SB	1	δ-proteobacteria	SYN_01370 (-60)	25*[j]	M3	SYN_02134 (-78)	SYN_01372 (-87)	SYN_02133 (-130)	yes
Chlamydomonas reinhardtii (U)[m]	3	chlorophyceae	AAL23573 (-51)	CR2	M1	AAS92601 (-60)	AAS92601 (-34)	AAS92602 (-119)	no
			AAL23572 (-43)	CR1	M1				
Trichomonas vaginalis strain (U)[m]	4	parabasalia	92802.m00036 (-54)	TV1	M3	93793.m00242 (-48)	87122.m00057 (-44)	88811.m00081 (-120)	no
			95912.m00238 (-54)	TV2	M3[n]	91566.m00125 (-46)		96484.m00069 (-102)	
			81599.m00076 (-54)	TV3	M3	81202.m00098 (-45)			
			95725.m00262 (-52)	TV4	M3[n]				
			97007.m00150 (-47)	TV5*[j]	M5a[n,o]				
			86485.m00476 (-45)	TV6	M2[p]				
			91127.m00212 (-41)	TV7	M2[p]				
Entamoeba histolytica HM-1:IMSS (U)[m]	5	archamoebae	XP_656685 (-36)	EH1	M2[p]	none	none	none	no
			XP_652137 (-36)	EH2	M2[p]				
			XP_652839 (-34)	EH3	M3a[f]				
Giardia lamblia ATCC 50803 (U)[m]	6	diplomonadida	EAA39802 (-32)	GL	M2[p]	none	none	none	no
Thalassiosira pseudonana CCMP1335 (U)[m]	7	bacillariophyceae	none[q] (-44)	TP1	M4	none	none	none	no
			none[q] (-15)	TP2	M2b[r]				
Saccharomyces cerevisiae	1	saccharomycetes	AAT93011 (-13)	SC	M2b[r]	none	none	none	no
Oryza sativa	1	liliopsida	XP_469746 (-)	OS	M2b[r]	none	none	none	no
Drosophila melanogaster	1	insecta	AAL48960 (-25)	DM	M2b[r]	none	none	none	no

Table 1 *(Continued)*

Species	Website[a]	Class	HydA[b]	Abbrev[c]	Type[d]	HydE[b]	HydF[b]	HydG[b]	NiFe[e]
Homo sapiens	1	mammalia	NP_071938 (-29)	HS1	M2b[r]	none	none	none	no
			NP_036468 (nd)	HS2	M2b[r]	none	none	none	no

[a] 1: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=Genome; 2: http://www.sanger.ac.uk/Projects/Microbes/; 3: (http://genome.jgi-psf.org/Chlre3/Chlre3.home.html; 4: http://www.tigr.org/tdb/e2k1/tvg/; 5: http://www.tigr.org/tdb/e2k1/eha1/; 6: http://gmod.mbl.edu/perl/site/giardia; 7: http://genome.jgi-psf.org/thaps1/thaps1.home.html

[b] Accession numbers of the protein sequences. In parenthesis, expected (E) values (exponents of 10) from BLAST [46] searches. E values for *T. maritima* sequences are zero because these sequences have been used as baits in searches.

[c] Abbreviations used in Figure 5. A number has been attributed to each bacterial species, followed by a lower case letter to differentiate paralogs. Two capital letters (initials of genus and species) have been used for eukaryotes, followed by numbers to differentiate paralogs.

[d] Modular structural framework (see Fig. 1).

[e] Presence of [NiFe] H_2ase encoding genes in genome.

[f] In M3a the two N-ter accessory domains of M3 are replaced by a 140-residue sequence assuming an unknown structure, unique to [FeFe] H_2ases, and containing eight Cys residues distributed in three runs: CC, Cx2C, Cx2Cx4Cx3C. M3a sequences are shorter (ca. 500 residues) than M3 ones because of their smaller H domain.

[g] TR: trimeric; in parenthesis, modular structure of catalytic subunit; likewise, HYM: membrane-bound, TE: tetrameric and D: dimeric. D-type H_2ases are periplasmic.

[h] Encoded on the Sol plasmid [87].

[i] In M2a the first run of four cysteines (Cx2Cx2Cx3C) in the accessory domain is replaced by Cx3Cx8Cx3C, and the L1 ligand pattern is TSCCCPMW.

[j] Asterisks mark sequences that for various reasons have to be considered as putative H_2ases. Three of them have not been given abbreviated names and have been omitted from the alignments for reasons given in the text. The other ones have single point mutations on one of the L1 pattern cysteines (C1 or C2, see Fig. 1) and have been kept in the alignments. The mutations are indicated hereafter in one letter amino acid code: 10a*, 11*, 14b*, 23c*, 24a* and 25* (C2S); 18* (C1S); TV5* (C1W).

[k] Differs from the canonical M2 framework (e.g. *M. elsdenii* H_2ase [21]) by a 20-residue insertion between the accessory cluster and the H domain. 14a and 20a may be oligomeric, associated with a formate dehydrogenase.

[l] This H_2ase has been purified and characterized [23]: it contains an H cluster which is an efficient catalyst, even though intramolecular electron transfer appears to be limiting the reaction. Confirmation of the presence of the C2S mutation in H_2ase 11* would indicate that C2 is dispensable, at least in this case (see text). The [NiFe] H_2ase from *T. tengcongensis* has been studied in the same article [23].

[m] (U): unfinished.

[n] The ligands of the second accessory cluster (His-ligated [4Fe-4S]) are missing.

[o] Length similar to that of the M5 framework, but the C-ter section (ca 500 residues) is homologous to NAD(P)H oxidoreductases. In M5 in contrast, the C-ter section is related to NuoE+NuoF (Fig. 2).

[p] Differs from the canonical M2 framework (e.g., *M. elsdenii* H_2ase [21]) by a 20-residue insertion between the two Cx2Cx2Cx3C runs of the accessory domain.

[q] Protein sequences were not annotated in genome, and were generated by translation of nucleotide sequences.

[r] In M2b the ligands of one of the [4Fe-4S] clusters of the accessory domain are missing. C1 in the L1 ligand pattern is mutated to A,S or V (see [34]). These sequences have not been included among the putative ones (starred) because they are known to have other functions than H_2ase activity.

least one order of magnitude lower than in *T. vaginalis* [55]. A low level of poorly assisted H_2ase maturation, by mechanisms that remain to be elucidated, might account for this puzzling observation.

In stark contrast with the structural genes of [FeFe] H_2 ases, which are rather variable across and even within species, the sequences of the maturation genes are highly conserved [38], as witnessed by the very low E values for each of the three maturases (Table 1). HydE and HydG are homologous to biotin synthase and ThiH, respectively, but their sequences are clearly separated from those of the latter enzymes in dendrograms derived from sequence alignments [38].

While the genes encoding the [NiFe] H_2ases and their maturases are most often clustered within operons (*Aquifex aeolicus* being a significant exception [1]), [FeFe] H_2ases are an example to the contrary [9]. Data relevant to the latter have recently been discussed [9], but a more extensive compilation is attempted here based on the comprehensive genomic data gathered in Table 1. A variety of situations are encountered, as illustrated in Figure 4: all genes may be isolated (*C. acetobutylicum*), some (*M. thermoacetica*, *T. maritima*) or all (*C. tetani*) of the maturation genes may be grouped within an operon separately from the structural genes, or together with them (*B. thetaiotaomicron*, *D. vulgaris*). In addition, the succession of genes within operons is highly variable. A few of these situations have previously been discussed, and attempts have been made to correlate the various kinds of gene association with protein localization or metabolic regulations [9]. Inspection of a larger sample reveals a considerable diversity, even within classes of bacteria, *e.g.*, the clostridia, where H_2ases are cytosolic and have presumably similar functions (Fig. 4). Nevertheless, clostridia share at least one feature: none of the structural genes are associated with any of the maturase-encoding ones. The reverse is true in the case of the proteobacteria, where structural genes are associated with at least one or two maturase-encoding genes, regardless of the localization (cytosol or periplasm) of the H_2ases (Fig. 4). Whether these features have any functional relevance, or are merely imprints of H_2ase evolution, remains unclear at this stage.

Comparisons of [FeFe] H_2ase sequences

[FeFe] H_2ase sequences have been available for several years in numbers allowing informative sequence alignments. Initially implemented for the identification of functionally significant features [22], alignments were soon used to retrieve phylogenetic information [1, 53, 56]. The number of [FeFe] H_2 ase sequences from genomes (Table 1) is roughly similar to the number of those previously obtained through classical cloning [1]. Preliminary alignments were carried out with all available sequences (not shown). However, genomic and non-genomic sequences are mostly redundant, as they are retrieved from the same species, or at least closely related ones. Therefore, dendrograms including all sequences do not afford significantly improved phylogenetic insights. On the contrary, they are less readable due to overcrowding, particularly in some "bushy" and less robust regions of the tree (see below). Thus, comparisons were performed with the sequences of Table 1, and only a few additional significant ones were included (Fig. 5).

Alignments were run using ClustalW [57], and trees were built with the SplitsTree [58] software. Alignments of whole-length sequences were checked for consistency (mainly verifying that all ligand patterns of the H cluster and accessory clusters were well aligned), and either used as such, or implemented to edit the sequences and keep only the H domain (from M211 to D534 in the *C. pasteurianum* sequence [22]) common to all of them. The H domains were then realigned to construct a new dendrogram, according to a previously used rationale [1]. The latter procedure was aimed at simplifying the alignments and ruling out possible interferences from the widely differing lengths and domain compositions of the various sequences. In fact, dendrograms depend little on whether they are derived from full-length or H domain sequence alignments (not shown), as previously noted [1]. The latter observation indicates that the H domains and their cognate accessory domains have for a large part coevolved, which in turn suggests that the acquisition or loss of accessory domains has assumed a similar pace as the overall evolution of [FeFe] H_2ases.

The dendrogram derived from the H domain alignments displays several distinct regions (Fig. 5). The largest cluster of sequences to the left is rather heterogeneous and is the least robust part of the tree: its internal organization is very sensitive to the addition or removal of sequences, and the bootstrap values are low (not shown). However, it is stable as a substructure distinct from other parts of the overall tree, and its main features are the following: it includes sequences assuming each of the main structural frameworks (M1–M5, D, TE, TR, HYM, see Fig. 2), as well as sequences from nearly all phyla. The only missing ones are the higher eukaryotes and the protists having cytosolic H_2ases. Next to this large subtree, the H_2ase-like sequences from higher eukaryotes form a distinct and robust subset. Alignments of larger numbers of such sequences produce trees consistent with known eukaryotic phylogeny (see below). Next to the higher eukaryote branch, another

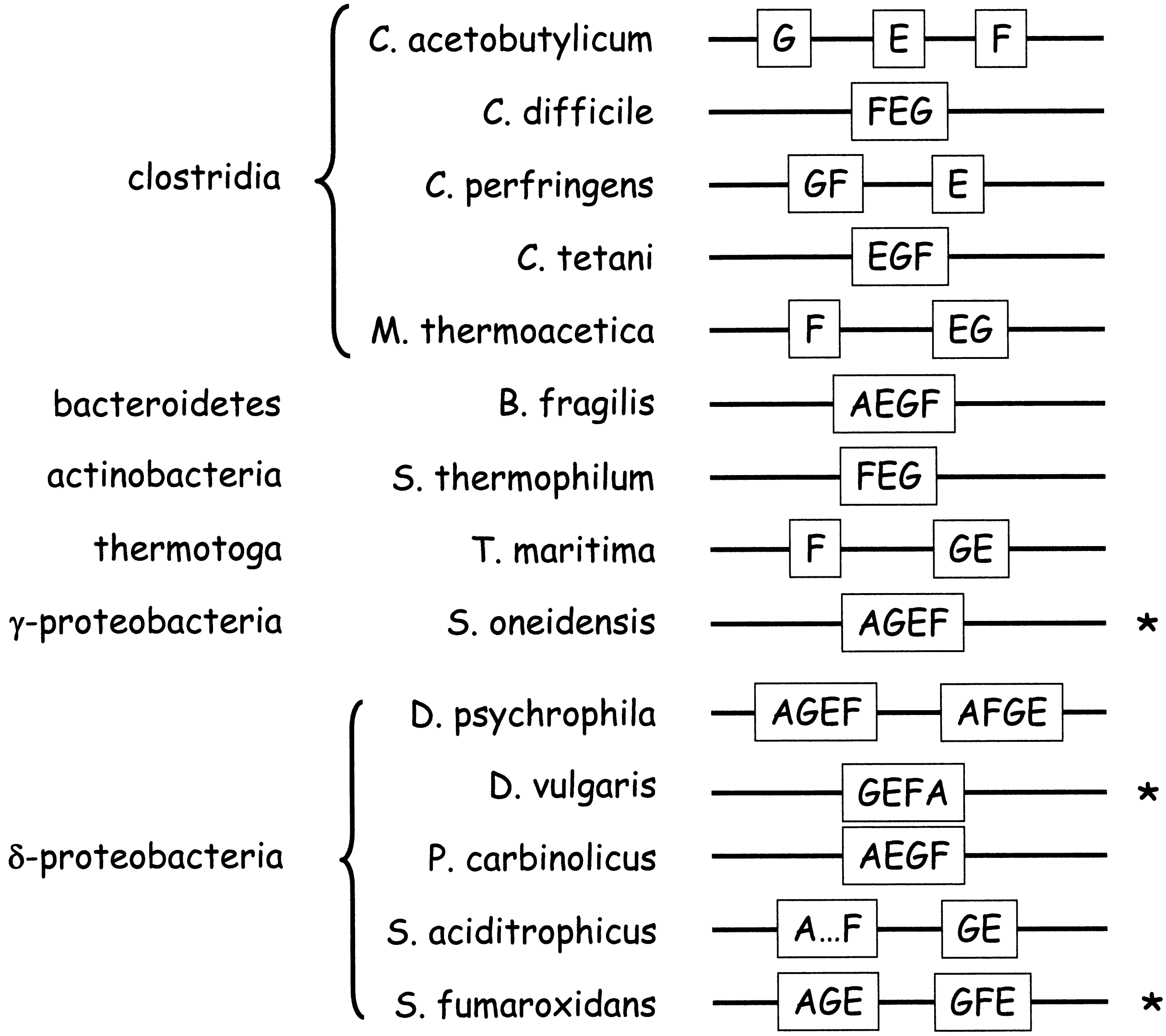

Figure 4. Distribution of *hyd* genes in sequenced bacterial genomes. *Hyd* genes inscribed in the same box are vicinal or at least presumed to belong to the same operon. These operons may include other (possibly non *hyd*) genes not indicated here. Transcription is from left to right in boxes containing more than one gene. *HydA* genes are indicated only when they are grouped with at least one maturase encoding gene (*e.g.*, not in the case of clostridia). Some genomes (taken from Table 1) displaying redundant *hyd* patterns have been included, in order to have at least one member of each class of bacteria. In the case of *S. aciditrophicus*, the association of *hydA* and *hydF* is uncertain. Stars indicate organisms having periplasmic H_2ases, otherwise they are cytosolic. Scales are arbitrary.

branch includes the sequences of cytosolic H_2ases from anaerobic protists. The only exception is EH3, which occurs in a subtree to the right, and has most likely been laterally transferred from a bacterium [59]. The rightmost part of the tree is composed of two or three subsets of sequences. The most distinct of these (top right) consists of five very similar sequences, all from clostridia, which share a variation (M2a) of the M2 structural framework. In M2a the first run of four cysteines ($Cx_2Cx_2Cx_3C$) in the accessory domain is replaced by ($Cx_3Cx_8Cx_3C$), and the L1 pattern is TSCCCPMW (Table 1). The two branches below the M2a branch are more heterogeneous, but all sequences therein possess the so-called M3a structural framework. In M3a, the two N-terminal accessory domains of M3 are replaced by a sequence unique to these proteins, and which contains eight conserved cysteines distributed in three runs: CC, Cx_2C, and $Cx_2Cx_4Cx_3C$ (Table 1). Many of the M3a sequences are from clostridia, but other bacterial classes are represented (Table 1), as well as *E. histolytica* with the anomalous EH3 sequence (see above). Altogether, with the exception of the large ramified subtree to the left, all branches are rather homogeneous, with respect to structural framework as well as host organisms.

Inspection of the dendrogram in Fig. 5 reveals the following general trends. With respect to structural

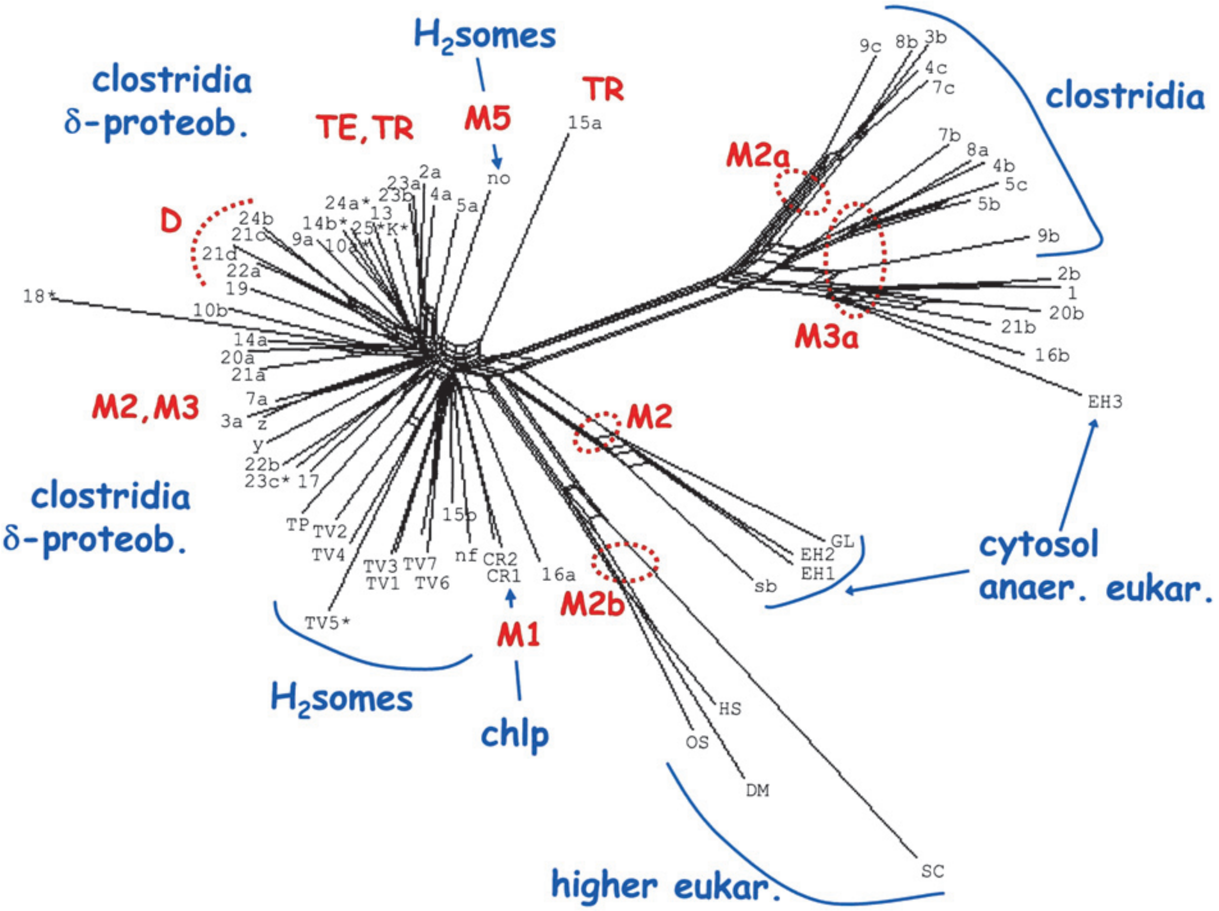

Figure 5. Dendrogram derived from sequence alignments of H domains of [FeFe] H_2ases. Full sequence alignments [57] were used to edit sequences and keep only the H domains. Alignments of the latter alone were used [58] to construct the dendrogram. Distances were calculated with the maximum likelihood algorithm. The sequences are those listed in Table 1, with the following ones added: *M. elsdenii* (y, AAF22114), *C. pasteurianum* (z, AAA23248), *Spironucleus barkhanus* (sb, AAG31038), *N. ovalis* (no, AAU14235), and *Neocallimastix frontalis* (nf, AAK60409). In parenthesis: label in tree, GenBank accession. H_2ase structural frameworks are marked in red, phyla in blue. Abbreviations: H_2somes, hydrogenosomes; chlp, chloroplasts.

frameworks, M2 is by far the most widespread, as it occurs in nearly all main branches. Otherwise, sequences assuming similar structural frameworks are often clustered together (M3a, D, TE and TR). With respect to organisms, clostridia and to a lesser degree δ-proteobacteria, are the most widespread. A few categories of organisms are apparently monophyletic with respect to H_2ases (*e.g.*, green algae), but for most of the other ones the presence of several distinct paralogs appears to be the rule (hydrogenosomes, clostridia, proteobacteria).

Emergence and evolution of [FeFe] H_2ases in Bacteria

The confirmation that all known Archaea are devoid of [FeFe] H_2ases suggests that these enzymes have emerged later than the separation between the archaeal and bacterial domains. However, as these enzymes predominantly occur in clostridia and δ-proteobacteria, which are the deepest branches within the firmicutes and proteobacteria, respectively [60], an early appearance within the domain Bacteria is most likely. More precise information relevant to the earliest hosts of [FeFe] H_2ases cannot be derived from the cladograms (Fig. 5), because the tree is unrooted and parts of it are not very robust. The likelihood that the M2 structural framework is the ancestral one (see below and [1]) is of little help in that respect, because of its ubiquity.

Subsequent evolution of bacterial [FeFe] H_2ases is also difficult to trace because of the scanty distribution of [FeFe] H_2ases, and the presence of several paralogs in most genomes (Table 1), particularly clostridia and δ-proteobacteria. The rare occurrence of [FeFe] H_2 ases in bacterial classes other than those of their presumptive origin is clearly visible in Table 1, where α-proteobacteria are a case in point: while the genomes of 44 species of that class have been fully sequenced, only two of them (*R. rubrum* and *R.*

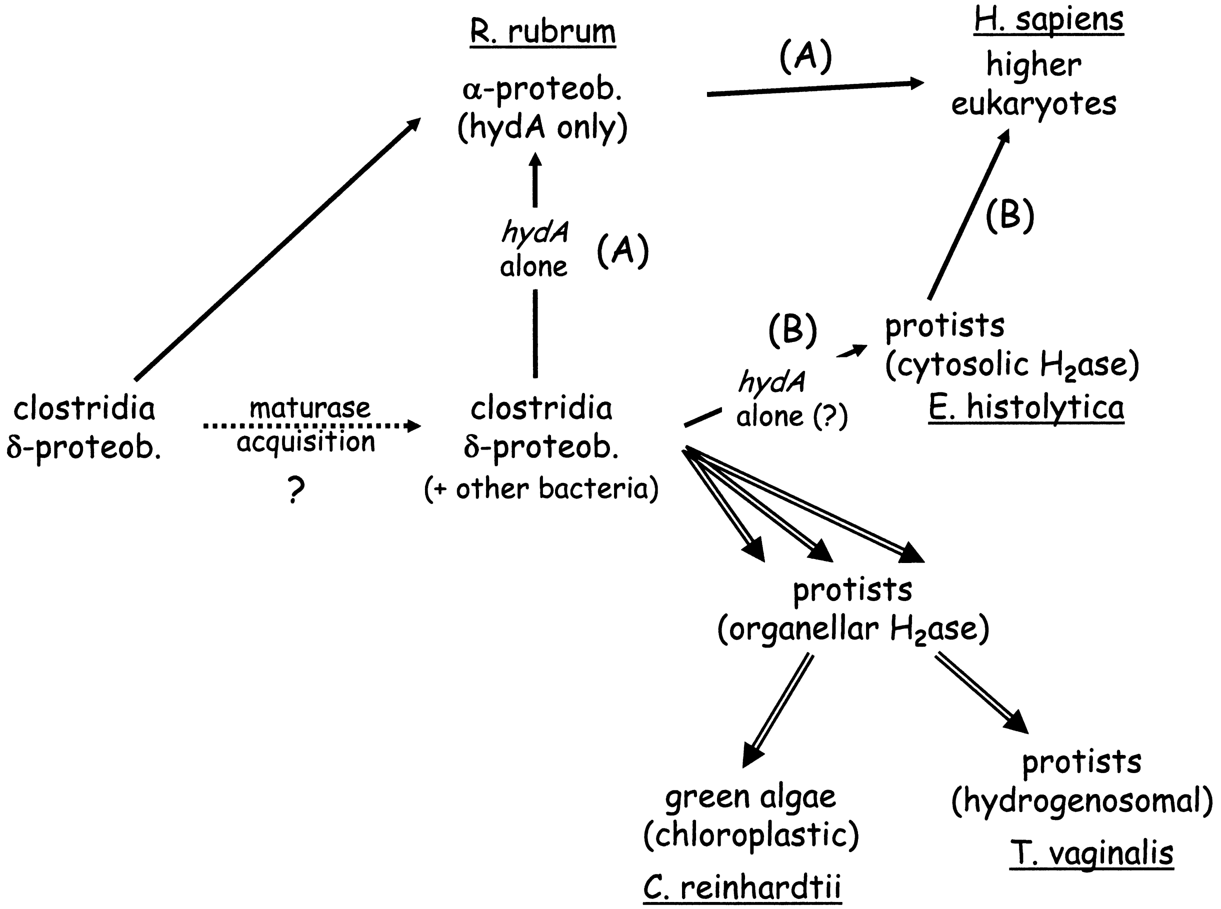

Figure 6. Scheme outlining possible pathways of [FeFe] H_2ase evolution. (*A*) and (*B*) are two alternative pathways for the transfer of [FeFe] H_2ase-encoding genes from bacteria to higher eukaryotes. Single-line arrows indicate transfer of structural genes only, double-line arrows indicate transfer of structural and maturation genes. The means of H_2ase transfer, either simple gene transfer or symbiosis, are not detailed. A few significant species of extant organisms are indicated (underlined).

palustris) contain [FeFe] H_2ase-encoding genes. Furthermore, such genes are present in only two out of five *R. palustris* strains. The absence of maturation genes in α-proteobacteria suggests either lateral gene transfer of the structural gene alone, or transfer of the latter before the emergence of maturases (see Fig. 6). The answer depends in part on whether [FeFe] H_2ases have appeared together with their maturases, or earlier, an issue that is discussed below.

Evolution of [FeFe] H_2ases in Eukarya

Protists. Most of the active eukaryotic [FeFe] H_2ases are found in the hydrogenosomes or cytosols of anaerobic protists [53]. The structural diversity of H_2 ases in these organisms nearly matches that of bacterial H_2ases, but for their quaternary structure: all known protist H_2ases are monomeric (Table 1). It is most likely that bacterial H_2ases have made their way into the eukaryotic domain first through anaerobic protists. Hydrogenosomal and cytosolic H_2ase sequences occur in distinct regions of the dendrogram (Fig. 5), which may suggest that cytosolic and hydrogenosomal H_2ases have entered protists via two distinct events. This hypothesis is bolstered by differences with respect to maturases: presently available evidence indicates that maturases are present in hydrogenosomes (*T. vaginalis*) [38], but not in hydrogenosome-less protists (*G. lamblia* and *E. histolytica*, Table 1). Confirmation of this trend with a wider range of organisms would indeed be supportive of two distinct transfers: possibly a symbiosis with a clostridium or δ-proteobacterium in the case of the hydrogenosomal H_2ases, and merely gene transfer in the case of the cytosolic ones (Fig. 6). Transfer of the structural gene alone is not unlikely in view of the separation of the structural and maturation genes in many genomes, particularly the clostridial ones (Fig. 4). Hydrogenosomes have been proposed to be polyphyletic [53, 56, 61] based on the observation that the relevant sequences are not clustered together in dendrograms, but are nested in distinct places among bacterial sequences [53, 56]. Indeed the sequences of *Neocallimastix frontalis* and *T. vaginalis* on one hand, and *N. ovalis* on the other hand, are well separated in the leftmost part of tree shown here (Fig. 5). However, considering that this region is not very robust, inferences regarding the phylogeny of hydrogenosomes should be drawn with caution.

Green algae and diatoms. Only a few species of green algae contain [FeFe] H_2ases, but they have been of considerable significance in the study of these enzymes ([9, 62] and references therein). The origin of green algal H_2ases is enigmatic in at least two ways. The first is that they are localized in an organelle, the chloroplast, derived from a cyanobacterial symbiont, while no [FeFe] H_2ases have yet been found in cyanobacteria [63]. The second is that algal H_2ases are uniquely assuming the M1 structural framework [18–20, 62]. The similarity of the H_2ases (Fig. 5), and maturases [38] from *T. vaginalis* and *C. reinhardtii*, suggests a common origin for the hydrogenosomal and green algal H_2ases, most likely through a non-photosynthetic symbiosis (Fig. 6) distinct from the one that gave rise to the chloroplast. Green algal H_2ases would subsequently have been targeted to the chloroplast and acquired their M1 structure, possibly as an adaptation to their interaction with the photosynthetic electron transfer chain [62].

The presence of an [FeFe] H_2ase in the diatom *T. pseudonana* (Table 1) appears to be a rarity among these organisms. The case of this diatom illustrates that organisms from very diverse eukaryotic phyla [64] contain genes encoding potentially active [FeFe] H_2 ases. However, while the gene sequence displays all the idiosyncrasies of a genuine [FeFe] H_2ase, maturases are not encoded in the genome of *T. pseudonana.* Regardless of whether *T. pseudonana* contains an active H_2ase, this organism is endowed with the remarkable property of containing both an [FeFe] H_2ase gene and a higher eukaryote [FeFe] H_2ase-like gene (Table 1). In this respect it may be regarded as related to both the anaerobic and the mitochondriate eukaryotes, which contain only either one of these two forms of [FeFe] H_2ases.

Another unusual case is that of the apicomplexan *Cryptosporidium parvum*, an intestinal parasite that in many respects resembles anaerobic protists [65]. This organism has been shown to contain an [FeFe] H_2ase-like gene typical of higher eukaryotes [65]. However, it does not, as revealed by its genome sequence [66], contain any protistal-type [FeFe] H_2ase.

Mitochondriate eukaryotes. The [FeFe] H_2ase-like proteins from mitochondriate eukaryotes are a highly conserved group of proteins, both with respect to structural framework (M2, Fig. 1) and sequence similarity. Specific features of these sequences and their structural bearings have been discussed elsewhere [34]. [FeFe] H_2ase-like sequences from higher eukaryotes form a distinct group branching off a single point in the general tree (Fig. 5), and alignments of these sequences produce cladograms that parallel the phylogeny of the relevant organisms (Fig. 7, see below). [FeFe] H_2ase-like proteins are thus most likely monophyletic, but the gateway through which they entered mitochondriate eukaryotes remains to be identified. One possibility (pathway B in Fig. 6) is suggested by their proximity in the tree (Fig. 5) to the sequences of cytosolic H_2ases from protists, as well as by the absence of maturases in both phyla (Table 1). Thus, only transfer of the structural gene from a protist would have been required. Another pathway (A in Fig. 6) may be inferred from the observation that [FeFe] H_2ases without maturases occur in a few α-proteobacteria (Table 1), and from the widely held opinion that these bacteria have been involved in the symbiosis that gave rise to mitochondria [67, 68]. Schemes A and B differ by the H_2ase donor (eukaryotic in A, bacterial in B), but loss of maturases would be required in neither of them (Fig. 6). The initial absence of maturases in higher eukaryotes may have been advantageous in preventing H_2ases from assuming their nominal function and thus possibly facilitating their acquisition of a novel one. The latter has most likely occurred early in the evolution of higher eukaryotes, because H_2ase-like proteins appear to have frozen in an idiosyncratic structural framework that has subsequently undergone limited evolution.

The presence of two genes encoding H_2ase-like proteins in the human genome, whereas only one is present in the yeast genome, has been noted previously [32]. Further inspection of sequenced genomes from mitochondriate eukaryotes revealed that the case of the human genome (two H_2ase-like genes) is common to all known craniates, while a single gene is found in all organisms predating the chordates, including the urchin (http://www.hgsc.bcm.tmc.edu/projects/seaurchin/) and the tunicates (http://genome.jgi-psf.org/Cioin2/Cioin2.home.html;). The dendrogram derived from alignments of eukaryotic H_2ase-like sequences (Fig. 7) is consistent with the known evolution of these organisms (see also Fig. 5). The two sub-trees corresponding to the vertebrate paralogs display similar patterns, and are well separated from each other. This may be taken as an indication that the presence of two genes in the craniates is due to a duplication that has occurred early in vertebrate evolution, and certainly long before the separation of the tetrapod and teleost branches. It would now certainly be interesting to

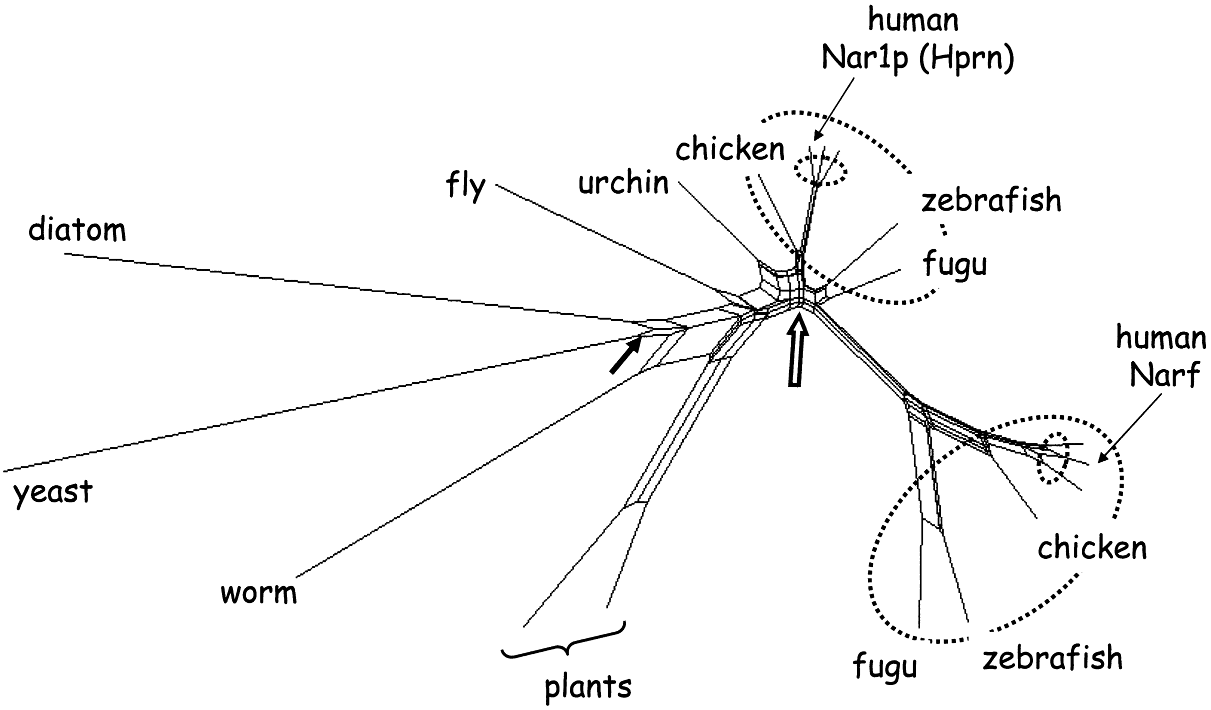

Figure 7. Dendrogram derived from alignments of H_2ase-like sequences from mitochondriate eukaryotes. Procedures were as in Fig. 5. The large ellipses highlight the two sub-trees of vertebrate sequences that arose from a duplication early in the vertebrate lineage. The smaller ellipses surround the mammalian sequences: human (*H. sapiens*, see Table 1), mouse (*M. musculus*, NP_080514 and NP_080548) and dog (*C. familiaris*, XP_547207 and XP_537935). Other sequences are from chicken (*G. gallus*, XP_414836 and XP_415606), zebrafish (*D. rerio*, XP_707447 and NP_001002342), fly (*D. melanogaster*, NP_001036428), worm (*C. elegans*, NP_498092), yeast (*S. cerevisiae*, NP_014159), rice (*O. sativa*, XP_469746), *Arabidopsis* (*A. thaliana*, NP_567496). The sequences from fugu (*F. rubripes*; http://www.fugu-sg.org/), urchin (*S. purpuratus*; http://www.hgsc.bcm.tmc.edu/projects/seaurchin/), and diatom (*T. pseudonana*; http://genome.jgi-psf.org/thaps1/thaps1.home.html) are not yet indexed in protein databanks. They were retrieved from non-annotated genome sequences using TBLASTN [46] and translated. The root of the dendrogram, as determined with bacterial H_2ase sequences, is located in the region indicated by the solid arrow. The open arrow points to the duplication of the H_2ase-like gene early in the emergence of the vertebrate lineage.

learn whether the H_2ase-like gene duplication had already occurred in the cephalochordate *Amphioxus* [69], the closest well-known relative of the craniates. Regarding the function of the gene products, recent data suggest that in craniates the two paralogs may be assuming distinct functions [32, 70].

Possible paths of [FeFe] H_2ase transfer into, and propagation among, eukaryotes are summarized in Figure 6. The scheme is based mainly on similarities among H_2ases and on the distribution of maturases. Notwithstanding some uncertainties, the conclusion that [FeFe] H_2ases have been transferred at least two or three times into the domain eukaryotes is difficult to escape. The main features of the scheme outlined in Figure 6 are difficult to reconcile with the proposal that mitochondria and hydrogenosomes are all derived from a common ancestor, surmised to be an α-proteobacterium [67, 68], and most likely a purple non-sulfur bacterium [67]. A previously underscored weakness of that model is the presence in hydrogenosomes of typically anaerobic enzymes, in particular [FeFe] H_2ases with their maturases and pyruvate-ferredoxin oxidoreductases (PFO), which are all missing in extant mitochondria and α-proteobacteria [61, 67, 68, 71, 72]. The existence of another hydrogenosomal enzyme highly similar to counterparts in firmicutes, which has recently been assigned to lateral gene transfer [73], may alternatively be taken as yet another argument against a common origin of mitochondria and hydrogenosomes. These observations, as well as the distribution pattern of [FeFe] H_2ases among eukaryotes, would be in better agreement with the emergence of hydrogenosomes and mitochondria through more than one symbiosis. A new concept has been elaborated recently, which is concerned mainly with motility, but also bears on eukaryotic evolution at large: it calls upon successive symbiotic events, where an ancestral α-proteobacterium would only be involved as a secondary symbiont [74]. Serial symbioses might likewise provide a rationale for separate origins of mitochondria and hydrogenosomes. While such evolutionary schemes would appear to be demanding in terms of symbiotic events, repeated observations of contemporarily occurring secondary symbioses [75,

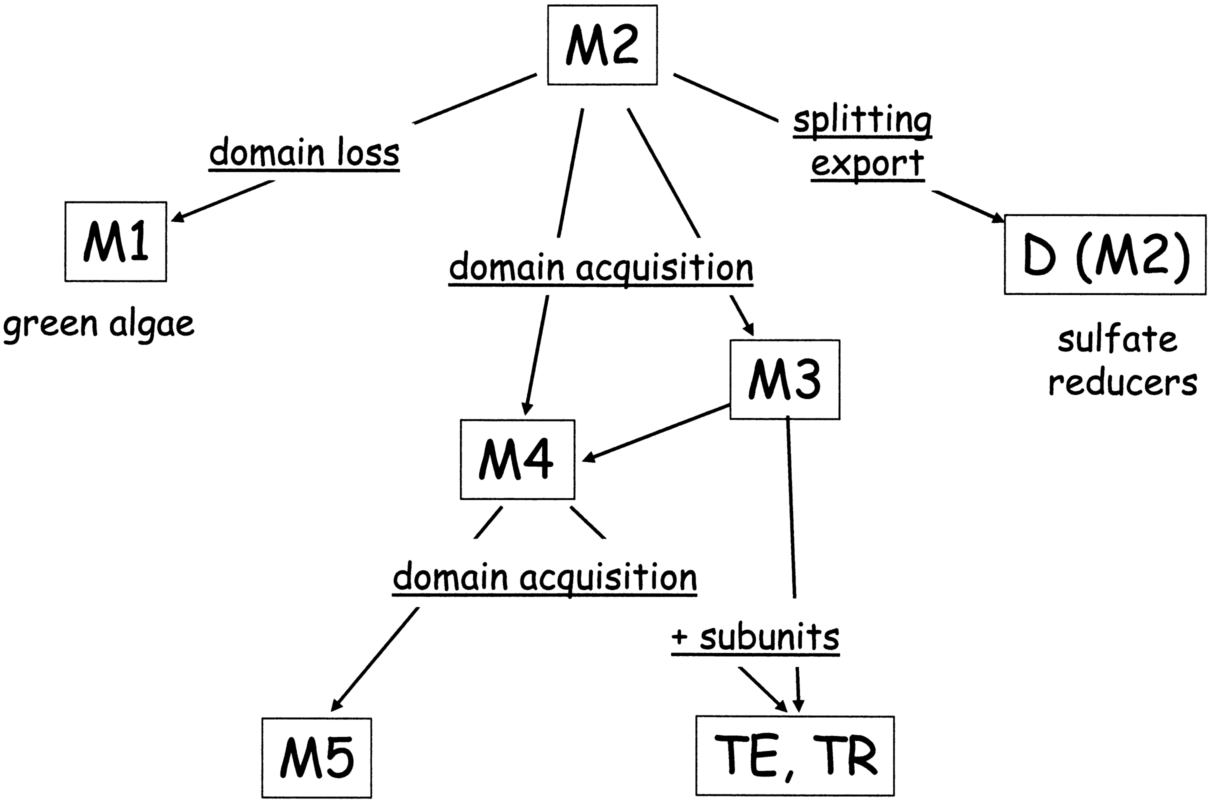

Figure 8. Modular structure of [FeFe] H_2ases: a tentative evolutionary scheme. Codes for structural frameworks are as in Figs 1 and 5.

76] suggest that events of the like might have been widespread in early stages of eukaryotic evolution.

Evolution of [FeFe] H_2ase structures

While the H domain accommodating the [FeFe] active site is by necessity common to all [FeFe] H_2ases, these enzymes have undergone extensive variation through addition or loss of accessory domains. It should nevertheless not be inferred from this that the H domains have evolved less than the accessory ones. In fact, the dendrogram framework (Fig. 5) depends very little upon whether the accessory domains are or are not included in the alignments. This indicates that accessory and catalytic domains have evolved at similar paces for long periods of time and that domain reshuffling has not been extensive in the recent history of these enzymes.

In a previous review [1], the M2 framework (Fig. 1) had been proposed to be the most likely ancestral form. Amendments of that model have subsequently been put forward, including the suggestion that the ancestral form might instead be larger than M2, and that the main evolutionary trend might have involved loss rather than acquisition of accessory modules [56]. However, the genomic data compiled here confirm and extend the overwhelming predominance of the M2 structure in all realms of life and in most individual organisms: it has even been found in all clostridial genomes (Table 1), while earlier biochemical [7] and genetic [22] evidence indicated M3 to be the paradigm of clostridial H_2ase structure. Likewise, M2-type H_2 ases are the most widespread among δ-proteobacteria as well as anaerobic protists. Among the latter, the M3 type is less frequently encountered (Table 1), and larger H_2ases (M4 or M5) are exceptional [25]. As to the higher eukaryotic H_2ase-like proteins, they uniformly assume the M2 framework.

Collectively, these observations bolster the previously proposed phylogenic scheme [1] according to which M2-type H_2ases appeared first, and have remained dominant in most evolutionary lines. M2 was most likely the primordial structural framework, which gave rise to larger H_2ases by serial accretion of N- or C-terminal domains. In contrast, the smaller M1 H_2 ases, which are unique to green algae [18–20], would have been derived by loss of the N-terminal 2[4Fe-4S] domain (Fig. 8). It should nevertheless be kept in mind that even within restricted groups of organisms (*e.g.*, clostridia and sulfate reducers, see Table 1) H_2ases display highly diverse modular and quaternary structures. The modular structure of an [FeFe] H_2ase should therefore not be used indiscriminately as an indicator of its phylogeny or host organism.

The case of the lost clostridial H_2-oxidizing [FeFe] H_2 ase

The normal physiological function of clostridial H_2 ases is the disposal of excess reducing equivalents through proton reduction. The best known of these enzymes is from *C. pasteurianum* ([7] and references therein). It has been thoroughly studied by biochemical and spectroscopic methods [7], the encoding gene has been sequenced [22], and a three-dimensional model has been put forward [12]. While this enzyme is equally active in both directions, another [FeFe] H_2ase (dubbed "II" to distinguish it from the better known "I" [7, 77]) has been isolated from the same organism and its metal sites have also been analyzed in detail by several biophysical methods [7]. Quite unlike H_2ase I, H_2ase II is almost irreversible: it is considerably more active as an H_2 oxidant, and is synthesized in late growth stages of batch cultures [7, 77]. H_2ase II is smaller (53–55 kDa vs. 63 kDa) than H_2ase I, and, while H_2ase I and II were found to have basically similar H clusters, they possess different numbers of accessory Fe-S clusters, namely four in H_2ase I, and two in H_2ase II [7]. In addition, one of the two latter ones was reported to have an unusually high redox potential (– 180 mV vs. – 400 mV) [7]. Efforts aimed at cloning the H_2ase II encoding gene were unsuccessful [22] and this H_2ase has not been further investigated since. The catalytic idiosyncrasies of H_2 ase II, which might now be rationalized using available structural models [12, 15], should renew interest in this protein and motivate efforts towards its re-isolation. Its putative physiological role, *e.g.*, energy recovery at the end of the growth phase or under starvation, suggests that it may be present is other species of clostridia.

Indeed, inspection of the cladogram of Figure 5 reveals a cluster of closely similar clostridial sequences that would have the expected size (ca. 500 amino acids) and number (two) of accessory clusters. These sequences are a slightly altered version (M2a) of the M2 structural framework, where the spacing of the first tetrad of cysteine residues is $Cx_3Cx_8Cx_3C$. This and other features of the accessory domain might be the cause of the anomalous redox potential of one of the [4Fe-4S] clusters. Indeed, the cysteine ligand pattern is somewhat similar to that observed in some bacterial 2[4Fe-4S] ferredoxins where the two clusters have very different redox potentials [78]. Other idiosyncrasies of M2a sequences concern the H domain, in particular TSCCCPMW as an L1 ligand pattern. The possibility that clostridial M2a genes might encode H_2-oxidizing [FeFe] H_2ases would certainly be worth exploring using the recently elaborated expression methods [36, 39, 42]. It may even be recommendable to implement the latter methods to sample other subgroups of [FeFe] H_2ases with the expectation of unveiling catalytic peculiarities associated with specific features of the sequences and structures.

Evolution of maturases

While extant [FeFe] H_2ases need maturases for the synthesis of their active site (keeping aside the intriguing cases of *G. lamblia* and *E. histolytica*, see above), earlier versions of these enzymes may have dispensed of this requirement. In archaean anaerobic biotopes rich in iron ions, sulfide, and possibly CO and CN, one could envision a suite of reactions involving self assembly [79, 80] of the [4Fe-4S] part of the H cluster, and subsequent grafting upon it of the [FeFe] subcluster, through reactions analogous to those implemented in contemporary chemistry [13, 14]. In such an event, the assistance of chaperones for the assembly of Fe-S clusters [45] and [FeFe] active site [9] would only have become mandatory in later eons, under conditions less favorable to self assembly. The emergence of maturation proteins later than H_2ases would indeed be in keeping with the more conserved sequences of the former (Table 1). However, establishing a chronology for the origin of H_2ases and their maturases, and identifying the organisms within which they emerged, is certainly not an easy task with the data at hand (Fig. 5, see also cladograms in [38]). Another standing question is the exceptional variety of *hyd* gene distribution patterns in genomes (Fig. 4). Whether this has any phylogenetic or functional meaning remains to be established.

[FeFe] *versus* [NiFe] H_2ases: which came first?

[FeFe] H_2ases have so far been unreported in Archaea. In contrast, [NiFe] H_2ases occur in both Archaea and Bacteria, and have a wider distribution in Bacteria than the [FeFe] enzymes [1]. This may suggest that [NiFe] H_2ases appeared earlier than the [FeFe] ones, an inference that is otherwise supported by indications of a close association of iron and nickel in prebiotic chemistry and primitive biochemistry [81].

On the other hand, the nearly universal association of the structural and maturation genes of the [NiFe] H_2 ases within operons suggests that these H_2ases might have been associated with their chaperone system since their very origin. This may be taken as an indication that they appeared at a time when spontaneous assembly of the active site was no longer feasible, *i.e.*, later than the [FeFe] H_2ases. In that case, the wider distribution of the [NiFe] H_2ases might be the outcome, not of a greater ancestry, but of more facile inter-specific transfer, thanks to the grouping of all wanted genes on a single operon.

Conclusions and prospects

While genome sequences have significantly increased the body of sequence data relevant to [FeFe] H_2ases, neither the diversity of known structural frameworks nor the spectrum of host organisms have been spectacularly enhanced. As to the various metagenome sequencing projects, they have yielded only very few [FeFe] H_2ase sequences ([48–50], see also http://www.ncbi.nlm.nih.gov/blast/). More focused approaches, such as the retrieval of H_2ase sequences from anaerobic bioreactor sludges by PCR amplification with specific primers, have brought forth relatively larger numbers of sequences, but all expectedly very similar to known clostridial ones [82]. Novelties might still be expected from the exploration of some rich anaerobic niches, *e.g.*, vertebrate guts [83]. However, the latter biome has so far yielded only four [FeFe] H_2ase sequences encompassing complete H domains (our unpublished data), and all are very similar to known clostridial sequences. These observations, added to the fact that Archaea appear to be devoid of [FeFe] H_2ases, suggest that the biodiversity of these enzymes in prokaryotes has been rounded up for the most part.

Eukaryotic anaerobes, on the contrary, appear to be largely undersampled [56]. [FeFe] H_2ase sequences drawn from but a few species (Table 1) occur in several distinct branches of the cladogram (Fig. 5), even though only a tiny fraction of the gut commensals of hundreds of insect species has been explored, and probably less so for free-living organisms [56]. It may therefore be anticipated that the known diversity of [FeFe] H_2ases will greatly benefit from the investigation of eukaryotic anaerobes. In addition, data from a wider spectrum of organisms would shed more light on the phylogeny of cytosolic and hydrogenosomal H_2 ases (Table 1 and Fig. 6), and thereby provide new insights into the early stages of eukaryote evolution.

The most significant breakthrough in recent years has been the identification and partial characterization of proteins in charge of the maturation of [FeFe] H_2ases [9]. Elucidation of this complex process is expected to unveil previously unknown biochemical reactions. Also, analysis of the sequences, distribution, and biochemistry of the maturases should afford valuable information relevant to the evolution of [FeFe] H_2 ases. From a practical point of view, implementation of the maturases is allowing heterologous expression of [FeFe] H_2ase genes, and has opened the way towards the long-awaited molecular engineering of these enzymes [36, 39, 42]. These new approaches will also be instrumental for the functional exploration of natural variants of [FeFe] H_2ases. Targets for such investigations have already been identified, namely the M2a-type proteins (Table 1, Fig. 5), which, as discussed above, may be suspected to be H_2-uptake H_2 ases, and the M3a subfamily (Fig. 5). The N-terminal domain of the latter exhibits idiosyncrasies that might eventually reveal novelties with respect to Fe-S folds as well as H_2ase function. Another sequence of interest is the very unusual DSY0936 putative gene from *D. hafniense* (Table 1), of which the protein product would have a unique set of accessory domains.

The diversity of domain frameworks found in [FeFe] H_2ases (Fig. 1) is probably greater than in most redox enzymes, including the counterpart [NiFe] H_2ases. It has long been known [1, 8, 16, 17] that many of these domains contain Fe-S clusters and are shared between numerous proteins (ferredoxins, [NiFe] H_2ases, complex I, *etc.*), thus allowing transfer of structural information among these redox enzymes. These relationships are spectacularly illustrated in the recently elucidated structure of a bacterial complex I [17]: some subunits of the latter display folds similar to those previously uncovered in ferredoxins [26], [FeFe]- [12] or [NiFe]-H_2ases [11]. Conversely, other subunits of complex I display novel structures that are known from sequence data to have counterparts in both classes of H_2ases [1, 8, 16, 17]. In fact, structural models are now available, directly or indirectly, for all known [FeFe] H_2ase domains displayed in Fig. 2. One may also mention in passing that complex I includes a counterpart of the structural core of [NiFe] H_2ases, even though the NiFe site has been replaced by a quinone site, but no counterpart of the structural core (H domain) of [FeFe] H_2ases [17].

Early investigations (reviewed in [7, 8]) of H_2ases revealed stability and catalytic properties, which in the main still hold today. For instance, most [FeFe] H_2ases are irreversibly inactivated by oxygen, while [NiFe] H_2 ases can be reactivated or are merely inhibited. Also, [FeFe] H_2ases are generally known to be reversible catalysts, while [NiFe] H_2ases catalyze preferentially H_2 uptake. In view of the projected use of H_2ases (or models thereof) as catalysts for the production of H_2 or its oxidation in fuel cells, the stability of these enzymes [84] and the optimization of their directionality [85] are turning into issues or economic relevance. Rapid progress along these lines of research should be favored by the recently achieved possibility of engineering [FeFe] H_2ases, as well as by the analysis of natural variants brought forth by genome sequences.

Research on H_2ases has occasionally, and notably in the 1970 s and 2000 s, benefited from enhanced support aimed at developing the use of whole cells [86], isolated H_2ases [85], or biomimetic organometallic complexes [13, 14], as catalysts for the production of hydrogen from renewable energy sources and its implementation as an energy currency. A number of studies have thus been jumpstarted or bolstered through funding aimed at tackling energy or environmental issues [2]. It should nevertheless be recalled that the investigation of H_2ases has been carried on continuously and unflaggingly for more than three decades, and that many decisive breakthroughs in the field (in biochemistry, genetics, and structure determination) have been driven by mere curiosity.

Summary

1. The absence of any phylogenetic link between [FeFe] and [NiFe] H_2ases is now compellingly supported by evidence from sequences, structures, and maturation systems. These enzymes nevertheless catalyze the same reaction by similar chemical means, and thus present a striking example of convergent evolution.
2. [FeFe] and [NiFe] H_2ases have very different distributions among organisms: while the [NiFe] enzymes occur only in prokaryotes (Bacteria and Archaea), [FeFe] H_2ases are present in Bacteria and Eukaryotes, but not in Archaea. Also, [NiFe] H_2ases outnumber [FeFe] H_2ases by at least a factor of three, with respect to host organisms as well as overall gene numbers. The question of whether the [FeFe]- or [NiFe]-H_2ases appeared first in the course of biological evolution remains unanswered at this stage.
3. [FeFe] H_2ases are remarkable by their highly diverse modular structure. Inventories and sequence comparisons reveal that the most frequently occurring structural framework is a combination of the active site domain (H domain) and a 2[4Fe-4S] ferredoxin-like domain. This structural framework is probably the primordial one, and most others have been derived from it by addition of modules or subunits. Green algal H_2ases, which consist of a mere H domain, are the only known exception and probably arose through loss of the ferredoxin-like domain.
4. [FeFe] H_2ases are absent in Archaea but widespread among clostridia and δ-proteobacteria, which are held to be among the most ancient lineages in bacterial evolution. The origin of these H_2ases may therefore be assigned to early stages of bacterial evolution, after the split between the archaeal and bacterial lineages. Even though the available data are not conclusive, there are indications suggesting that [FeFe] H_2ases may have appeared earlier than their maturases.
5. [FeFe] H_2ases have most likely entered the eukaryotic domain of life through anaerobic protists. In the latter organisms, H_2ases are either cytosolic or hydrogenosomal. Sequence comparisons strongly suggest that cytosolic and hydrogenosomal H_2ases have arisen through at least two distinct events, be they symbioses or gene transfers. In contrast, the question of whether hydrogenosomal H_2ases are mono- or polyphyletic has not received a clear answer. As to green algal H_2 ases, they are closely related to some hydrogenosomal H_2ases, which may be suggestive of a common origin, possibly through a non-photosynthetic symbiosis.
6. Mitochondriate eukaryotes contain [FeFe] H_2ase-like proteins that appear to be derived from protist cytosolic H_2ases. They have lost their nominal function and at least some of them are involved in the assembly of Fe-S clusters. They are present as single copies in all genomes, except in vertebrates, which contain two of them. The latter results most probably from a gene duplication early in the evolution of the vertebrate lineage.

Acknowledgements. The author wishes to thank Paul King for communicating and discussing unpublished data, as well as providing thoughtful comments on the manuscript.

References

1 Vignais, P. M., Billoud, B. and Meyer, J. (2001) Classification and phylogeny of hydrogenases. FEMS Microbiol. Rev. 54, 455 – 501.
2 Cammack, R., Frey, M. and Robson, R. (2001) Hydrogen as a Fuel. Learning from Nature. Taylor & Francis, London.
3 Nicolet, Y., Lemon, B. J., Fontecilla-Camps, J. C. and Peters, J. W. (2000) A novel FeS cluster in Fe-only hydrogenases. Trends Biochem. Sci. 25, 138 – 143.
4 Frey, M., Fontecilla-Camps, J. C. and Volbeda, A. (2001) Nickel-iron hydrogenases. In: Handbook of Metalloproteins, pp. 880 – 896, Messerschmidt, A., Huber, R., Poulos, T. and Wieghardt, K. (eds.), John Wiley & Sons, Chichester.
5 Pilak, O., Mamat, B., Vogt, S., Hagemeier, C. H., Thauer, R. K., Shima, S., Vonrhein, C., Warkentin, E. and Ermler, U. (2006) The crystal structure of the apoenzyme of the iron–sulphur cluster-free hydrogenase. J. Mol. Biol. 358, 798 – 809.
6 Stephenson, M. and Stickland, L. H. (1931) Hydrogenase: a bacterial enzyme activating molecular hydrogen: The properties of the enzyme. Biochem. J. 25, 205 – 214.
7 Adams, M. W. W. (1990) The structure and mechanism of iron-hydrogenases. Biochim. Biophys. Acta 1020, 115 – 145.
8 Albracht, S. P. J. (1994) Nickel hydrogenases: in search of the active site. Biochim. Biophys. Acta 1188, 167 – 204.
9 Böck, A., King, P. W., Blokesch, M. and Posewitz, M. C. (2006) Maturation of Hydrogenases. Adv. Microb. Physiol. 51, 1 – 72.
10 Friedrich, B., Buhrke, T., Burgdorf, T. and Lenz, O. (2005) A hydrogen-sensing multiprotein complex controls aerobic hydrogen metabolism in *Ralstonia eutropha.* Biochem. Soc. Trans. 33, 97 – 101.
11 Volbeda, A., Charon, M. H., Piras, C., Hatchikian, E. C., Frey, M. and Fontecilla-Camps, J. C. (1995) Crystal structure of the nickel-iron hydrogenase from *Desulfovibrio gigas.* Nature 373, 580 – 587.
12 Peters, J. W., Lanzilotta, W. N., Lemon, B. J. and Seefeldt, L. C. (1998) X-ray crystal structure of the Fe-only hydrogenase (CpI) from *Clostridium pasteurianum* to 1. 8 Angstrom resolution. Science 282, 1853 – 1858.
13 Evans, D. J. and Pickett, C. J. (2003) Chemistry and the hydrogenases. Chem. Soc. Rev. 32, 268 – 275.
14 Tye, J. W., Darensbourg, M. Y. and Hall, M. B. (2006) *De novo* design of synthetic di-iron(I) complexes as structural models of

the reduced form of iron-iron hydrogenase. Inorg. Chem. 45, 1552 – 1559.

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

16 Hedderich, R. and Forzi, L. (2005) Energy-converting [NiFe] hydrogenases: more than just H_2 activation. J. Mol. Microbiol. Biotechnol. 10, 92 – 104.

17 Sazanov, L. A. and Hinchliffe, P. (2006) Structure of the hydrophilic domain of respiratory complex I from *Thermus thermophilus.* Science 311, 1430 – 1436.

18 Florin, L., Tsokoglou, A. and Happe, T. (2001) A novel type of iron hydrogenase in the green alga *Scenedesmus obliquus* is linked to the photosynthetic electron transport chain. J. Biol. Chem. 276, 6125 – 6132.

19 Winkler, M., Heil, B. and Happe, T. (2002) Isolation and molecular characterization of the [Fe]-hydrogenase from the unicellular green alga *Chlorella fusca.* Biochim. Biophys. Acta 1576, 330 – 334.

20 Forestier, M., King, P., Zhang, L., Posewitz, M., Schwarzer, S., Happe, T., Ghirardi, M. L. and Seibert, M. (2003) Expression of two [Fe]-hydrogenases in *Chlamydomonas reinhardtii* under anaerobic conditions. Eur. J. Biochem. 270, 2750 – 2758.

21 Atta, M. and Meyer, J. (2000) Characterization of the gene encoding the [Fe]-hydrogenase from *Megasphaera elsdenii.* Biochim. Biophys. Acta 1476, 368 – 371.

22 Meyer, J. and Gagnon, J. (1991) Primary structure of hydrogenase I from *Clostridium pasteurianum.* Biochemistry 30, 9697 – 9704.

23 Soboh, B., Linder, D. and Hedderich, R. (2004) A multisubunit membrane-bound [NiFe] hydrogenase and an NADH-dependent Fe-only hydrogenase in the fermenting bacterium *Thermoanaerobacter tengcongensis.* Microbiology 150, 2451 – 2463.

24 Verhagen, M. F., O'Rourke, T. and Adams, M. W. (1999) The hyperthermophilic bacterium, *Thermotoga maritima*, contains an unusually complex iron-hydrogenase: amino acid sequence analyses versus biochemical characterization. Biochim. Biophys. Acta 1412, 212 – 229.

25 Akhmanova, A., Voncken, F., van Alen, T., van Hoek, A., Boxma, B., Vogels, G., Veenhuis, M. and Hackstein, J. H. P. (1998) A hydrogenosome with a genome. Nature 396, 527 – 528.

26 Yeh, A. P., Chatelet, C., Soltis, S. M., Kuhn, P., Meyer, J. and Rees DC. (2000) Structure of a thioredoxin-like [2Fe-2S] ferredoxin from *Aquifex aeolicus.* J. Mol. Biol. 300, 587 – 595.

27 Mishra, J., Khurana, S., Kumar, N., Ghosh, A. K. and Das, D. (2004) Molecular cloning, characterization, and overexpression of a novel [Fe]-hydrogenase isolated from a high rate of hydrogen producing *Enterobacter cloacae* IIT-BT 08. Biochem. Biophys. Res. Commun. 324, 679 – 685.

28 Malki, S., Saimmaime, I., De Luca, G., Rousset, M., Dermoun, Z. and Belaich, J.-P. (1995) Characterization of an operon encoding an NADP-reducing hydrogenase in *Desulfovibrio fructosovorans.* J. Bacteriol. 177, 2628 – 2636.

29 Voordouw, G. and Brenner, S. (1985) Nucleotide sequence of the gene encoding the hydrogenase from *Desulfovibrio vulgaris* (Hildenborough). Eur. J. Biochem. 148, 515 – 520.

30 Graentzdoerffer, A., Rauh, D., Pich, A. and Andreesen, J. R. (2003) Molecular and biochemical characterization of two tungsten- and selenium-containing formate dehydrogenases from *Eubacterium acidaminophilum* that are associated with components of an iron-only hydrogenase. Arch. Microbiol. 179, 116 – 130.

31 Barton, R. M. and Worman, H. J. (1999) Prenylated prelamin A interacts with Narf, a novel nuclear protein. J. Biol. Chem. 274, 30008 – 30018.

32 Balk, J., Pierik, A. J., Aguilar Netz, D. J., Muhlenhoff, U. and Lill, R. (2004) The hydrogenase-like Nar1p is essential for maturation of cytosolic and nuclear iron-sulphur proteins. EMBO J. 23, 2105 – 2115.

33 Balk, J., Pierik, A. J., Aguilar Netz, D. J., Muhlenhoff, U. and Lill, R. (2005) Nar1p, a conserved eukaryotic protein with similarity to Fe-only hydrogenases, functions in cytosolic iron-sulphur protein biogenesis. Biochem. Soc. Trans. 33, 86 – 89.

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

35 Blokesch, M., Paschos, A., Theodoratou, E., Bauer, A., Hube, M., Huth, S. and Böck, A. (2002) Metal insertion into NiFe-hydrogenases. Biochem. Soc. Trans. 30, 674 – 680.

36 Posewitz, M. C., King, P. W., Smolinski, S. L., Zhang, L., Seibert, M. and Ghirardi, M. L. (2004) Discovery of two novel radical S-adenosylmethionine proteins required for the assembly of an active [Fe] hydrogenase. J. Biol. Chem. 279, 25711 – 25720.

37 Nicolet, Y. and Drennan, C. L. (2004) AdoMet radical proteins – from structure to evolution – alignment of divergent protein sequences reveals strong secondary structure element conservation. Nucleic Acids Res. 32, 4015 – 4025.

38 Pütz, S., Dolezal, P., Gelius-Dietrich, G., Bohacova, L., Tachezy, J. and Henze, K. (2006) Fe-hydrogenase maturases in the hydrogenosomes of *Trichomonas vaginalis.* Eukaryot. Cell 5, 579 – 586.

39 Girbal, L., von Abendroth, G., Winkler, M., Benton, P. M., Meynial-Salles, I., Croux, C., Peters, J. W., Happe, T. and Soucaille, P. (2005) Homologous and heterologous overexpression in *Clostridium acetobutylicum* and characterization of purified clostridial and algal Fe-only hydrogenases with high specific activities. Appl. Environ. Microbiol. 71, 2777 – 2781.

40 Rubach, J. K., Brazzolotto, X., Gaillard, J. and Fontecave, M. (2005) Biochemical characterization of the HydE and HydG iron-only hydrogenase maturation enzymes from *Thermotoga maritima.* FEBS Lett. 579, 5055 – 5060.

41 Brazzolotto, X., Rubach, J. K., Gaillard, J., Gambarelli, S., Atta, M. and Fontecave, M. (2006) The [Fe-Fe]-hydrogenase maturation protein HydF from *Thermotoga maritima* is a GTPase with an iron-sulfur cluster. J. Biol. Chem. 281, 769 – 774.

42 King, P. W., Posewitz, M. C., Ghirardi, M. L. and Seibert, M. (2006) Functional studies of [FeFe] hydrogenase maturation in an *Escherichia coli* biosynthetic system. J. Bacteriol. 188, 2163 – 2172.

43 Peters, J. W., Szilagyi, R. K., Naumov, A. and Douglas, T. (2006) A radical solution for the biosynthesis of the H-cluster of hydrogenase. FEBS Lett. 580, 363 – 367.

44 Voordouw, G., Hagen, W. R., Kruse-Wolters, K. M., van Berkel-Arts, A. and Veeger, C. (1987) Purification and characterization of *Desulfovibrio vulgaris* (Hildenborough) hydrogenase expressed in *Escherichia coli.* Eur. J. Biochem. 162, 31 – 36.

45 Johnson, D. C., Dean, D. R., Smith, A. D. and Johnson, M. K. (2005) Structure, function, and formation of biological iron-sulfur clusters. Annu. Rev. Biochem. 74, 247 – 281.

46 Altschul, S. F., Gish, W., Miller, W., Myers, E. W. and Lipman, D. J. (1990) Basic local alignment search tool. J. Mol. Biol. 215, 403 – 410.

47 Fraser-Liggett, C. M. (2005) Insights on biology and evolution from microbial genome sequencing. Genome Res. 15, 1603 – 1610.

48 Tringe, S. G. and Rubin, E. M. (2005) Metagenomics: DNA sequencing of environmental samples. Nat. Rev. Genet. 6, 805 – 814.

49 Venter, J. C., Remington, K., Heidelberg, J. F., Halpern, A. L., Rusch, D., Eisen, J. A., Wu, D., Paulsen, I., Nelson, K. E., Nelson, W., Fouts, D. E., Levy, S., Knap, A. H., Lomas, M. W., Nealson, K., White, O., Peterson, J., Hoffman, J., Parsons, R., Baden-Tillson, H., Pfannkoch, C., Rogers, Y. H. and Smith, H. O. (2004) Environmental genome shotgun sequencing of the Sargasso Sea. Science 304, 66 – 74.

50 Tringe, S. G., von Mering, C., Kobayashi, A., Salamov, A. A., Chen, K., Chang, H. W., Podar, M., Short, J. M., Mathur, E. J., Detter, J. C., Bork, P., Hugenholtz, P. and Rubin, E. M. (2005)

Comparative metagenomics of microbial communities. Science 308, 554 – 557.

51 Wu, M., Ren, Q., Durkin, A. S., Daugherty, S. C., Brinkac, L. M., Dodson, R. J., Madupu, R., Sullivan, S. A., Kolonay, J. F., Haft, D. H., Nelson, W. C., Tallon, L. J., Jones, K. M., Ulrich, L. E., Gonzalez, J. M., Zhulin, I. B., Robb, F. T. and Eisen, J. A. (2005) Life in hot carbon monoxide: the complete genome sequence of *Carboxydothermus hydrogenoformans* Z-2901. PLoS Genet. 1, e65.

52 Coppi, M. V. (2005) The hydrogenases of *Geobacter sulfurreducens:* a comparative genomic perspective. Microbiology 151, 1239 – 1254.

53 Horner, D. S., Heil, B., Happe, T. and Embley, T. M. (2002) Iron hydrogenases – ancient enzymes in modern eukaryotes. Trends Biochem. Sci. 27, 148 – 153.

54 Bui, E. T. and Johnson, P. J. (1996) Identification and characterization of [Fe]-hydrogenases in the hydrogenosome of *Trichomonas vaginalis.* Mol. Biochem. Parasitol. 76, 305 – 310.

55 Lloyd, D., Ralphs, J. R. and Harris, J. C. (2002) *Giardia intestinalis*, a eukaryote without hydrogenosomes, produces hydrogen. Microbiology 148, 727 – 733.

56 Hackstein, J. H. P. (2005) Eukaryotic Fe-hydrogenases – old eukaryotic heritage or adaptive acquisitions? Biochem. Soc. Trans. 33, 47 – 50.

57 Higgins, D., Thompson, J., Gibson, T., Thompson, J. D., Higgins, D. G. and Gibson, T. J. (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res. 22, 4673 – 4680.

58 Huson, D. H. and Bryant, D. (2006) Application of phylogenetic networks in evolutionary studies. Mol. Biol. Evol. 23, 254 – 267.

59 Nixon, J. E., Field, J., McArthur, A. G., Sogin, M. L., Yarlett, N., Loftus, B. J. and Samuelson, J. (2003) Iron-dependent hydrogenases of *Entamoeba histolytica* and *Giardia lamblia:* activity of the recombinant entamoebic enzyme and evidence for lateral gene transfer. Biol. Bull. 204, 1 – 9.

60 Ciccarelli, F. D., Doerks, T., von Mering, C., Creevey, C. J., Snel, B. and Bork, P. (2006)Toward automatic reconstruction of a highly resolved tree of life. Science 311, 1283 – 1287.

61 van der Giezen, M., Tovar, J. and Clark, C. G. (2005) Mitochondrion-derived organelles in protists and fungi. Int. Rev. Cytol. 244, 175 – 225.

62 Melis, A., Seibert, M. and Happe, T. (2004) Genomics of green algal hydrogen research. Photosynth. Res. 82, 277 – 288.

63 Ludwig, M., Schulz-Friedrich, R. and Appel, J. (2006) Occurrence of hydrogenases in cyanobacteria and anoxygenic photosynthetic bacteria: implications for the phylogenetic origin of cyanobacterial and algal hydrogenases. J. Mol. Evol. 63, 758 – 768.

64 Keeling, P. J., Burger, G., Durnford, D. G., Lang, B. F., Lee, R. W., Pearlman, R. E., Roger, A. J. and Gray, M. W. (2005) The tree of eukaryotes. Trends Ecol. Evol. 20, 670 – 676.

65 Stejskal, F., Slapeta, J., Ctrnacta, V. and Keithly, J. S. (2003) A Narf-like gene from Cryptosporidium parvum resembles homologues observed in aerobic protists and higher eukaryotes. FEMS Microbiol. Lett. 229, 91 – 96.

66 Abrahamsen, M. S., Templeton, T. J., Enomoto, S., Abrahante, J. E., Zhu, G., Lancto, C. A., Deng, M., Liu, C., Widmer, G., Tzipori, S., Buck, G. A., Xu, P., Bankier, A. T., Dear, P. H., Konfortov, B. A., Spriggs, H. F., Iyer, L., Anantharaman, V., Aravind, L. and Kapur, V. (2004) Complete genome sequence of the apicomplexan, *Cryptosporidium parvum.* Science 304, 441 – 445.

67 Cavalier-Smith, T (2006) Origin of mitochondria by intracellular enslavement of a photosynthetic purple bacterium. Proc. R. Soc. B 273, 1943 – 1952.

68 Martin, W. (2005) The missing link between hydrogenosomes and mitochondria. Trends Microbiol. 13, 457 – 459.

69 Panopoulou, G., Hennig, S., Groth, D., Krause, A., Poustka, A. J., Herwig, R., Vingron, M. and Lehrach, H. (2003) New evidence for genome-wide duplications at the origin of vertebrates using an amphioxus gene set and completed animal genomes. Genome Res. 13, 1056 – 1066.

70 Huang, J., Song, D., Flores, A., Zhao, Q., Mooney, S. M., Shaw, L. M. and Lee, F. S. (2007) IOP1, a novel hydrogenase-like protein that modulates hypoxia-inducible factor-1α activity. Biochem. J. 401, 341 – 352.

71 Dyall, S. D. and Johnson, P. J. (2000) Origins of hydrogenosomes and mitochondria: evolution and organelle biogenesis. Curr. Opin. Microbiol. 3, 404 – 411.

72 Embley, T. M. (2006) Multiple secondary origins of the anaerobic lifestyle in eukaryotes. Phil. Trans. R. Soc. B 361, 1055 – 1067.

73 Mukherjee, M., Brown, M. T., McArthur, A. G. and Johnson, P. J. (2006) Proteins of the glycine decarboxylase complex in the hydrogenosome of *Trichomonas vaginalis.* Eukaryot. Cell 5, 2062 – 2071.

74 Margulis, L., Chapman, M., Guerrero, R. and Hall, J. (2006) The last eukaryotic common ancestor (LECA): Acquisition of cytoskeletal motility from aerotolerant spirochetes in the Proterozoic Eon. Proc. Natl. Acad. Sci. USA 103, 13080 – 13085.

75 Okamoto, N. and Inouye, I. (2005) A secondary symbiosis in progress? Science 310, 287.

76 Andersson, S. G. E. (2006) The bacterial world gets smaller. Science 314, 259 – 250.

77 Chen, J.-S. and Blanchard, D. K. (1978) Isolation and properties of a unidirectional H2-oxidizing hydrogenase from the strictly anaerobic N2-fixing bacterium *Clostridium pasteurianum* W5. Biochem. Biophys. Res. Commun. 84, 1144 – 1150.

78 Giastas, P., Pinotsis, N., Efthymiou, G., Wilmanns, M., Kyritsis, P., Moulis, J.-M. and Mavridis, I. M. (2006) The structure of the 2[4Fe-4S] ferredoxin from *Pseudomonas aeruginosa* at 1. 32-Å resolution: comparison with other high-resolution structures of ferredoxins and contributing structural features to reduction potential values. J. Biol. Inorg. Chem. 11, 445 – 458.

79 Malkin, R. and Rabinowitz, J. C. (1966) The reconstitution of clostridial ferredoxin. Biochem. Biophys. Res. Commun. 22, 822 – 827.

80 Venkateswara Rao, P. and Holm, R. H. (2004) Synthetic analogues of the active sites of iron-sulfur proteins. Chem. Rev. 104, 527 – 559.

81 Russell, M. J. and Martin, W. (2004) The rocky roots of the acetyl-CoA pathway. Trends Biochem. Sci. 29, 358 – 363.

82 Fang, H. H. P., Zhang, T. and Li, C. (2006) Characterization of Fe-hydrogenase genes diversity and hydrogen-producing population in an acidophilic sludge. J. Biotechnol. 126, 357 – 364.

83 Gill, S. R., Pop, M., Deboy, R. T., Eckburg, P. B., Turnbaugh, P. J., Samuel, B. S., Gordon, J. I., Relman, D. A., Fraser-Liggett, C. M. and Nelson, K. E. (2006) Metagenomic analysis of the human distal gut microbiome. Science 312, 1355 – 1359.

84 Vincent, K. A., Parkin, A., Lenz, O., Albracht, S. P., Fontecilla-Camps, J. C., Cammack, R., Friedrich, B. and Armstrong, F. A. (2005) Electrochemical definitions of O_2 sensitivity and oxidative inactivation in hydrogenases. J. Am. Chem. Soc. 127, 18179 – 18189.

85 Vincent, K. A., Cracknell, J. A., Parkin, A. and Armstrong, F. A. (2005) Hydrogen cycling by enzymes: electrocatalysis and implications for future energy technology. Dalton Trans. 3397 – 3403.

86 Prince, R. C. and Kheshgi, H. S. (2005) The photobiological production of hydrogen: potential efficiency and effectiveness as a renewable fuel. Crit. Rev. Microbiol. 31, 19 – 31.

87 Nolling, J., Breton, G., Omelchenko, M. V., Makarova, K. S., Zeng, Q., Gibson, R., Lee, H. M., Dubois, J., Qiu, D., Hitti, J., Wolf, Y. I., Tatusov, R. L., Sabathe, F., Doucette-Stamm, L., Soucaille, P., Daly, M. J., Bennett, G. N., Koonin, E. V. and Smith, D. R. (2001) Genome sequence and comparative analysis of the solvent-producing bacterium *Clostridium acetobutylicum.* J. Bacteriol. 183, 4823 – 4838.