Chapter 2

Biodiversity of hydrogenases

Robert Robson

With contributions from *Hermann Bothe, Klaus Fiebig, Juan Imperial, Albrecht Klein, Kornél L. Kovács, Jacques Meyer, Thomás Ruiz-Argüeso, Anita Sellstedt, Rüdiger Schulz, Paulette M. Vignais*

2.1. Introduction

The study of biodiversity is a relatively recent subject concerned not only with recording and monitoring changes in the diversity of life but also with the levels on which biology is organised and the interrelationships between them. Biodiversity is measured on three fundamental levels: ecosystems, species and genes. Questions concerning evolutionary relationships underlie all of biodiversity. This chapter considers the biodiversity of H_2 metabolism by reviewing the species in which H_2 metabolism has been studied, the importance of H_2 metabolism to those organisms and to different ecosystems, and the occurrence, function and evolution of different hydrogenases and the genes which encode them. Questions concerning the analysis of the diversity of H_2 metabolising bacteria in different environments and the potential existence of novel systems are also considered.

2.2. Species diversity

2.2.1. Overview

It is remarkable that various aspects of H_2 metabolism have been studied in some detail in over sixty species (Table 2.1). The spread of interest can be judged in relation to the universal phylogenetic tree based on 16S/18S rRNA sequence comparisons (Fig. 2.1) which shows that life (excluding viruses) divides into three domains, two of which, the Eubacteria and the Archaea, are prokaryotic, i.e. contain organisms with no nuclear membrane and the third consists of all the Eukarya, those organisms which contain a membrane-bound nucleus. Most of the organisms in which H_2 metabolism has been studied are prokaryotes and the physiological range is wide and includes both aerobes and anaerobes, autotrophs and heterotrophs, prokaryotic and eukaryoytic photosynthetic organisms, knallgas bacteria, methanogens, sulfate reducers, N_2 fixers, fermentative organisms, hyperthermophiles, parasitic protozoans, and anaerobic fungi. Particular groups of organisms have been studied because H_2 consumption is essential to their survival, e.g. the methanogens such as *Methanobacterium*, *Methanococcus* and *Methanosarcina* sp. Other organisms (e.g. *Ralstonia eutropha*, *Bradyrhizobium japonicum*, *Rhodobacter capsulatus*) can adapt to use H_2 as a sole fuel source. Particular organisms have been studied because of close interrelationships between H_2 metabolism and other biochemical and physiological

Table 2.1 Organisms in which hydrogenase metabolism, biochemistry and/or genetics have been studied

Domain	*Group*	*Order*	*Genus*	*Species*	*Abbreviation*
Archaea	Crenarchaeota		*Pyrodictium*	*brockii*	*Pbr*
	Euryarchaeota	Archaeglobales	*Archaeoglobus*	*fulgidus*	*Afu*
		Methanobacteriales	*Methanobacterium*	*marburgensis*	*Mtm*
			Methanothermus	*fervidus*	*Mfe*
		Methanococcales	*Methanococcus*	*jannaschii*	*Mja*
			Methanococcus	*thermolithotrophicus*	*Mtl*
			Methanococcus	*vanielli*	*Mva*
			Methanococcus	*voltae*	*Mvo*
		Methanomicrobiales	*Methanosarcina*	*mazei*	*Mma*
		Methanopyrales	*Methanopyrus*	*kandleri*	*Mka*
		Methanosarcinacaea	*Methanosarcina*	*barkeri*	*Mba*
		Thermococcales	*Pyrococcus*	*furiosus*	*Pfu*
			Pyrococcus	*hirokoshii*	*Phi*
			Thermococcus	*litoralis*	*Tli*
			Thermococcus	*stetteri*	*Tst*
Bacteria	Aquificales		*Aquifex*	*aeolicus*	*Aae*
			Aquifex	*pyrophilus*	*Apy*
			Calderobacterium	*hydrogenophilim*	*Chy*
			Hydrogenobacter	*thermophilus*	*Hth*
	Cyanobacteria	Chroococcales	*Synechococcus*	PCC6301	S6301
			Synechocystis	PCC6803	S6803
		Nostocales	*Anabaena*	PCC29413	A29413
			Anabaena	PCC7120	A7120
			Anabaena	*variabilis*	*Ava*
			Nostoc	PCC73102	N73102
		Oscillatoriales	*Oscillatoria*	sp.	*Osp*
		Prochlophytes	*Proclorothrix*	*hollandica*	*Pho*
	Firmicutes	Actinomycetales	*Frankia*	sp.	*Fsp*
			Streptomyces	*thermoautotrophicus*	*Sth*
		Bacillus/Clostridium	*Bacillus*	*schlegelii*	*Bsc*
			Bacillus	*tusciae*	*Btu*
			Clostridium	*acetobutylicum*	*Cac*
			Clostridium	*pasteurianum*	*Cpa*
		Corynibacterinae	*Rhodococcus*	*opacus*	*Rop*
	Proteobacteria	α	*Acetobacter*	*flavidum*	*Afl*
			Bradyrhizobium	*japonicum*	*Bja*
			Paracoccus	*denitrificans*	*Pde*
			Rhizobium	*leguminosarum*	*Rle*
			Rhodobacter	*capsulatus*	*Rca*
			Rhodobacter	*sphaeroides*	*Rsp*
		β	*Acidovorax*	*facilis*	*Afa*
			Alcaligenes	*eutrophus*	*Aeu*
			Alcaligenes	*hydrogenophilus*	*Ahy*
			Ralstonia	*eutrophus*	*Reu*
			Rhodocyclus	*gelatinosus*	*Rge*
			Thiobacillus	*plumbophilus*	*Tpl*
		δ	*Desulfomicrobium*	*baculatus*	*Dba*
			Desulfovibrio	*fructosovorans*	*Dfr*
			Desulfovibrio	*gigas*	*Dgi*
		ϵ	*Helicobacter*	*pylori*	*Hpy*
			Wolinella	*succinogenes*	*Wsu*
		γ	*Azotobacter*	*chroococcum*	*Ach*
			Azotobacter	*vinelandii*	*Avi*

Table 2.1 Continued

Domain	*Group*	*Order*	*Genus*	*Species*	*Abbreviation*
			Citrobacter	*freundii*	*Cfr*
			Chromatium	*vinosum*	*Cvi*
			Escherichia	*coli*	*Eco*
			Desulfovibrio	*vulgaris*	*Dvu*
			Pseudomonas	*carboxydovorans*	*Pca*
			Salmonella	*typhimurium*	*Sty*
			Thiocapsa	*roseopersicina*	*Tro*
	Thermotogales		*Thermotoga*	*maritima*	*Tma*
Eukarya	Ciliophora	Litostomatea	*Dasytricha*	*ruminantium*	*Dru*
		Spirotrichea	*Nyctotherus*	*ovalis*	*Nov*
	Parabasalidea	Trichomonidida	*Trichomonas*	*vaginalis*	*Tva*
	Chlorophyta	Volvocales	*Chlamydomonas*	*reinhardtii*	*Cre*
			Chlamydomonas	*moewsii*	*Cmo*
		Chlorococcales	*Scenedesmus*	*obliquus*	*Sob*
	Fungi	Chytridiomycota	*Neocallimastix*	sp. *L2*	*Nsp*

processes. N_2-fixing organisms (e.g. *Rhizobium, Anabaena, Azotobacter* and *Frankia* species) have been of special interest because H_2 is both an obligatory product and a potential inhibitor of N_2 reduction by the enzyme nitrogenase and hydrogenases may therefore enhance the efficiency of N_2 fixation (the H_2-recycling hypothesis). Some organisms have proved amenable because they are easily grown and their enzymes easily isolated. Some have good genetic systems and, once gene probes or DNA sequences become available for hydrogenases from one group of organisms, an explosion of interest in closely related species usually follows.

2.2.2. *The species coverage*

Whilst the number of organisms in which H_2 metabolism has been studied is already large, many gaps exist in our coverage. The exploration of new phylogenetic or physiological groups has often led to the discovery of new hydrogenases which in turn have advanced structural and mechanistic understanding of these enzymes and also provided potentially interesting biotechnological opportunities.

Amongst the Eubacteria (Table 2.1), the coverage is dominated particularly by studies in two major divisions, the Proteobacteria (purple photosynthetic bacteria) in which all five subdivisions (α, β, δ, ϵ, and γ) are represented, and the Cyanobacteria (of which at least three groups are represented). However, many widely divergent groups have been neglected. They include the Spirochaetes, the Cytophaga/Flexibacter/Bacteroides group, Fibrobacters, and both high and low G + C content Gram +ve bacteria in which studies of Frankia and the saccharolytic clostridia respectively almost stand alone. The green non-sulfur bacteria and the planctomyces are also neglected groups. Genome sequencing projects are filling some gaps but it is doubtful whether we will recognise truly novel systems by sequence similarity searches alone. Amongst the Archaea, studies on species in the sub-domain Euryarchaeota dominate, partly because some of these are the well-known mesophilic or moderately thermophilic methanogens. However, relatively few members of the Crenarchaeota have been studied probably because most were discovered only recently and many are extreme thermophiles which require specialised growth facilities.

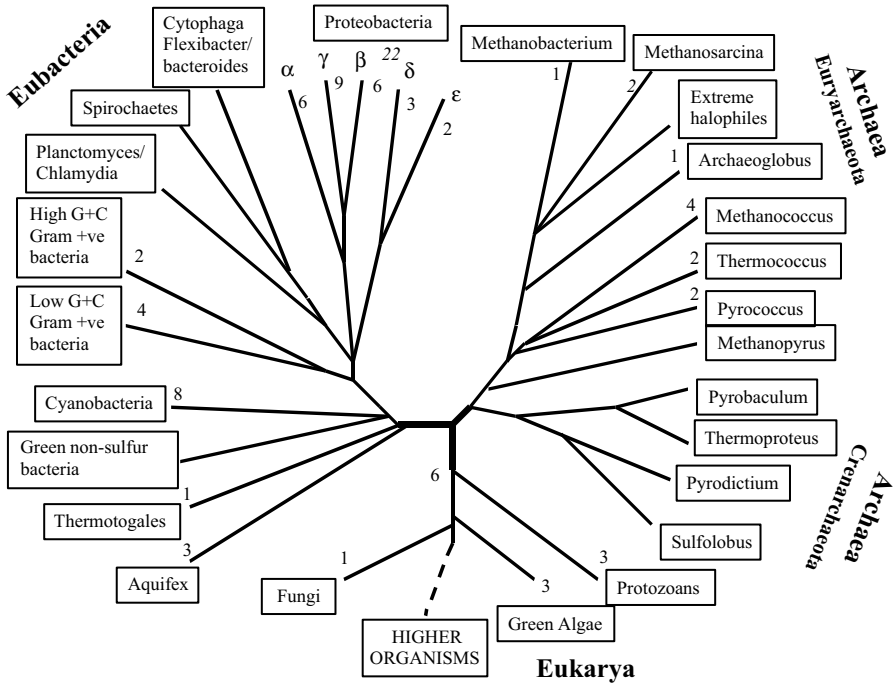


Figure 2.1 Universal phylogenetic tree. The figure shows the schematic evolutionary relationships between various organisms focusing especially on the prokaryotic eubacterial and archaeal domains. The boxes show either the various groups of Eubacteria or genera of Archaea. No details are provided for the evolutionary relationships amongst the members of the Eukarya. Numbers at the ends of the branches represent the number of species of that Group or Genus in which hydrogenases or hydrogen metabolism has been studied or in which putative hydrogenase gene sequences have been discovered through genome sequence projects. For more details see Table 2.1. It should be noted that evolutionary distances are not to scale.

Now, a growing number of complete and partial genome sequences are available for members of this sub-domain, e.g. *Sulfolobus solfataricus*, *Pyrobaculum aerophilum* and *Aeropyrum pernix*. Compared to the prokaryotes, there are relatively few studies of Eukaryotic hydrogenases despite the fact that H_2 evolution is important in some lower orders of the Eukarya including anaerobic protozoa, e.g. *Trichomonas vaginalis* and *Dasytricha ruminantium*, unicellular green algae, e.g. *Chlamydomonas* and *Scenedesmus*, and anaerobic rumen fungi, e.g. *Neocallimastix* sp.

2.3. Functional diversity of hydrogenases

2.3.1. Functions of hydrogenases: An overview

At least 13 families of hydrogenases are known. All but one are involved directly or indirectly in energy metabolism (see Table 2.2) and two main physiological functions can be discerned: either they catalyse H_2 oxidation (H_2 uptake/consumption) linked

Table 2.2 Families of hydrogenases, their occurrence and function

Number	*Hydrogenase family*	*Occurrence*	*Function*
1	Fe-only hydrogenases	Obligately anaerobic bacteria and Eukaryotes	Fermentation/ energy conservation?
2	NiFe(Se) membrane-bound or periplasmic/respiratory hydrogenases	Aerobes, facultative anaerobes, obligate anaerobes of the Proteobacteria	Energy conservation
3	NiFe-(thylakoid)uptake hydrogenases	Cyanobacteria	Energy conservation?
4	Bidirectional NAD(P)-reactive hydrogenases	Cyanobacteria	Energy conservation, Redox poising?
5	NAD(P)-reactive hydrogenases	Facultative and obligately anaerobic Eubacteria	Energy conservation
6	NAD(P)-reactive hydrogenases	Obligately anaerobic Archaea	Fermentation
7	F420-non-reactive hydrogenases	Methanogens	Energy conservation
8	F420-reactive hydrogenases	Methanogens	Energy conservation
9	NiFe-sensor hydrogenases	Chemolithotrophic/ phototrophic Proteobacteria	Hydrogen sensing components in genetic regulation of hydrogenase expression
10	NiFe-hydrogenases associated with the formate hydrogen lyases complex	Facultative and obligate anaerobes, Archaea	Fermentation
11	Ech hydrogenases	Methanogens	Methanogenesis pathway
12	Non-metal hydrogenases	Methanogens	Energy conservation
13	Soluble hydrogenase	*Anabaena cylindrica*	Unknown

to energy conserving reactions (e.g. respiration, NAD(P)H formation, methanogenesis) or they catalyse H^+ reduction (H_2 evolution) coupled to the disposal of excess reducing potential through the re-oxidation of reduced pyridine nucleotides and electron carriers. However, two other potentially related functions have emerged in recent years. One family of hydrogenases present in several autotrophic Proteobacteria (e.g. *R. eutropha*, *R. capsulatus* and *Bradyrhizobium leguminosarum*) appear to act as the H_2 sensing components of a complex genetic relay controlling the expression of other hydrogenases in these organisms (see Chapter 4). A fourth function for hydrogenases has been suggested for the so-called bidirectional hydrogenases in Cyanobacteria which may serve to poise the redox of photosynthetic and respiratory electron transport chains.

2.3.2. The multiplicity of hydrogenases

The importance of H_2 metabolism to some organisms is highlighted by their possession of more than one hydrogenase system. Four hydrogenase systems are known in *Escherichia coli* and *Methanococcus voltae*, three in *Desufovibrio vulgaris* and

R. eutropha and two in the Eukaryote, *T. vaginalis*. Several systems may be expressed simultaneously in some organisms but in others enzymes are expressed differentially depending on the growth conditions (Chapter 4). Moreover, the complement of hydrogenases may even differ between strains of a particular species as shown for sulfate reducing bacteria (Voordouw *et al.* 1990). However, some organisms contain a single system, e.g. *Azotobacter vinelandii*, and many more organisms apparently need no hydrogenase at all. No recognisable hydrogenase genes have been found in the genomes of important organisms, e.g. yeast, amongst the lower Eukaryotes, *Haemophilus influenzae* and *Mycobacterium tuberculosis* amongst the Eubacteria. All Archaea examined appear to contain hydrogenases.

2.3.3. The diversity of prosthetic groups in hydrogenases

All but two of the known families of hydrogenases are known to be metalloenzymes usually containing several metal centres including [Fe-S] clusters, NiFe centres with remarkable coordination properties, or haem groups (Chapter 6). Some also contain non-metal prosthetic groups, e.g. FAD, FMN. All the metallo hydrogenases contain Fe (Fe-only hydrogenases) but many also contain Ni (the NiFe(Se) hydrogenases) some of which, as their title suggests, contain Se in the form of selenocysteine which substitutes for one cysteinyl ligand to the Ni atom. One non-metal hydrogenase is the so-called metal-free hydrogenase involved in methanogenesis under Ni-starved conditions in *Methanobacter marburgensis* (Afting *et al.* 1998). A second potential non-metal hydrogenase has been isolated from *Anabaena cylindrica* (Ewart and Smith 1989; Ewart *et al.* 1990). It exhibits tritium exchange activity, the defining activity of members of the hydrogenase enzyme family, but its function is unknown. Generally the Fe-only enzymes are structurally related, have high turnover numbers and evolve H_2. They are found in obligate anaerobes such as saccharolytic clostridia and protozoan parasites such as *T. vaginalis* where they function in the reoxidation of reduced cofactors. However, other Fe-only enzymes function in H_2 oxidation as in the examples from the sulfate reducing bacteria such as *D. vulgaris*. The NiFe(Se) enzymes are the most diverse group and serve in energy conserving reactions in diverse organisms such as sulfate reducers, chemolithotrophs, methanogens, chemoheterotrophs and phototrophs. In addition the H_2-sensing and bidirectional hydrogenases are also NiFe enzymes. Whereas their subunit structures are quite diverse, the Ni-containing subunit appears to be conserved (Reeve and Beckler 1990). The NiFe(Se) enzymes do not appear to be especially structurally related to the Fe enzymes by primary amino acid sequence although it has emerged recently that the active site may be similar (Chapter 6). The five basic groups of hydrogenases described above can be further subdivided to give the twelve groups (Table 2.2) described in more detail below.

Group 1. The Fe-only hydrogenases

These are classically represented by the enzymes from obligate anaerobes including the sulfate reducer *D. vulgaris* (Vourdouw and Brenner 1985) and the saccharolytic *Clostridium pasteurianum* (Meyer and Gagnon 1991) in which the enzymes act in uptake or evolution modes respectively. Many examples are αβ heterodimers like the NiFe-membrane or periplasmic enzymes yet the primary sequences of the subunits of

these groups of enzymes are not similar. Nevertheless it has recently been revealed that their active sites show a great deal in common (Chapter 6). They occur in the periplasm in the case of *D. vulgaris* where they may reduce cytochrome c_3 or in the cytoplasm in the case of *C. pasteurianum*, where low-potential ferredoxins (e.g. the 2[4Fe-4S] cluster containing species) may provide the electrons *in vivo*. A novel three-subunit NADH-oxidising Fe-only hydrogenase has recently been found in *Thermotoga maritima* (Verhagen *et al.* 1999).

Group 2. The periplasmic and membrane-bound NiFe(Se) hydrogenases in the Proteobacteria

These enzymes are αβ heterodimers which are widely distributed in a variety of different organisms. These enzymes share two subunits in common, the large (α subunit) which contains the NiFe centre and the smaller (β subunit) of which usually contains three [Fe-S] clusters forming an intraprotein electron conduit, leading electrons produced at the NiFe centre to an accessory and specific electron carrier protein usually a *c* or *b*-type cytochrome. They occur either free in the periplasm (Fig. 2.2A) as in the sulfate-reducing bacteria such as *Desulfovibrio gigas* (Li *et al.* 1987) where the role of the enzyme is unclear, or most probably attached to the outer surface of the cytoplasmic membrane as in the classic examples from *B. japonicum* (Sayavedra-Soto *et al.* 1988), *E. coli* (Ballantine and Boxer 1985; Menon *et al.* 1990a,b; Menon *et al.* 1994b), *R. eutropha* (Schink and Schlegel 1979; Kortlüke *et al.* 1992), *R. capsulatus* (Leclerc *et al.* 1988), *A. vinelandii* (Seefeldt and Arp 1986; Menon *et al.* 1990a) and *Wolinella succinogenes* (Dross *et al.* 1992) (Fig. 2.2B). Based on immunological and molecular biological evidence, similar enzymes may also occur in Frankia species, the N_2-fixing symbionts of woody dicotyledenous plants such as *Alnus* and *Casuarina* (Fig. 2.3) (Sellstedt and Lindblad 1990). Though direct localisation studies have been performed for only a few cases, these enzymes are believed to be localised on the outer face of the cytoplasmic membrane because they all have signal peptides at the N-terminal sequences of the β subunit which are absent in the as isolated, mature proteins. In addition the C-terminal sequence of the β subunit is a highly hydrophobic domain missing in the soluble periplasmic *D. gigas* enzyme and which may be in part responsible for membrane attachment. The genes for the β and α subunits (*hupSL* or *hoxKG*) are usually at the proximal end of an operon (see Chapter 4) which contains a third conserved gene potentially encoding a *b*-type cytochrome as established in *W. succinogenes* where the three protein complex, as isolated, is capable of reducing quinols (Dross *et al.* 1992). The activity of these enzymes which show relatively high affinities for H_2 can be measured in membrane preparations of aerobes by following H_2-dependent O_2 consumption. Therefore they presumably act as proximal electron donors to the respiratory chain.

Group 3. The NiFe-(thylakoid) uptake hydrogenases of Cyanobacteria

These enzymes are found in filamentous Cyanobacteria, e.g. *Anabaena* 7120 (Houchins and Burris 1981), *Nostoc* sp. strain PCC73102 (Oxelfelt *et al.* 1998) and *Anabaena variabilis* (Schmitz *et al.* 1995) where they occur in the heterocysts. They may also occur in vegetative *Anabaena* cells grown under microaerobic or anaerobic non-N_2-fixing

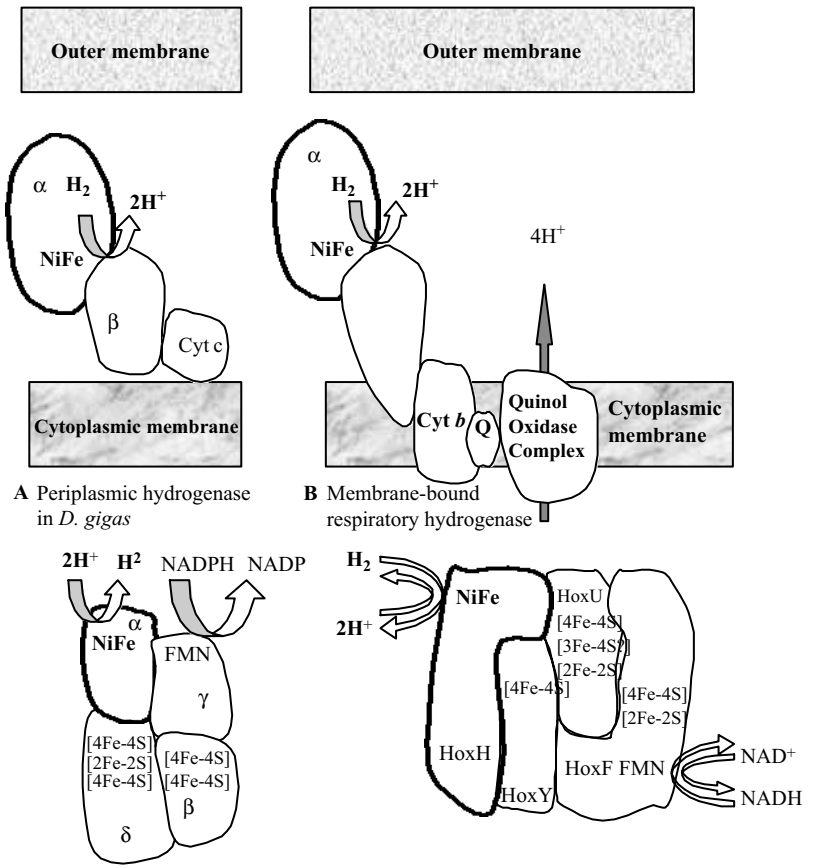
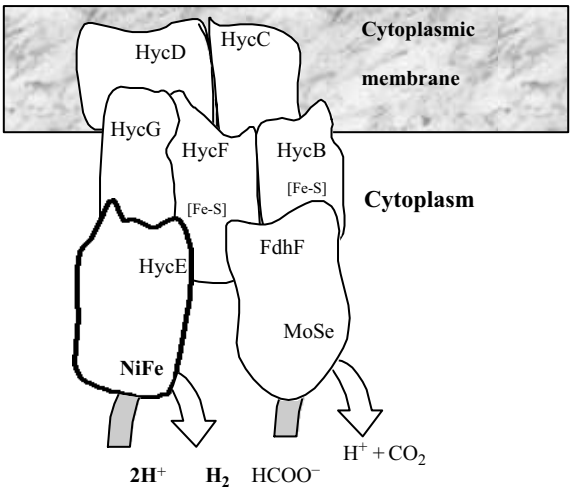
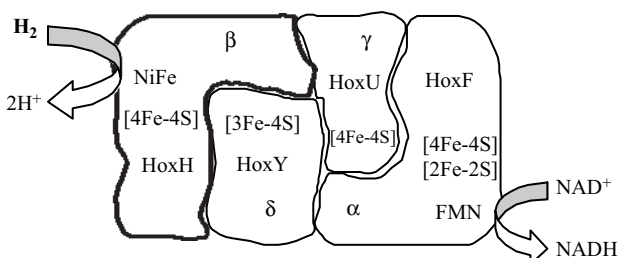


Figure 2.2 Hydrogenases: structure and function. The figure illustrates structure and functional relationships for various members of the hydrogenase enzyme family. Subunit and gene nomenclatures and composition of prosthetic groups are generally indicated. (A,B) Soluble periplasmic and membrane-bound examples of the Group 2 enzymes from Proteobacteria. (C) The Group 6 tetrameric hydrogenase from the archaean *T. litoralis*. (D) The Group 4 bidirectional hydrogenase from Cyanobacteria.

conditions. Physiological and biochemical evidence also suggests that a similar enzyme may occur in the unicellular non-N_2-fixing species *Anacystis nidulans* (*Synechococcus* sp. strain PCC6301) (Peschek 1979a,b). Although as yet these membrane-attached enzymes have not been purified to homogeneity, there is physiological evidence to support a requirement for Ni for their activity and the sequences of the *hupSL* genes from several organisms suggests that they are $\alpha\beta$ dimers, resembling the periplasmic and membrane-bound NiFe hydrogenases except that the small subunit has no transit peptide. This suggests that they are not exported to the periplasm but attached instead to the inner surface of the cytoplasmic membrane or, more likely, to the thylakoid membrane in the heterocysts. Their physiological role (Fig. 2.2B) may be to recycle H_2 produced by nitrogenase and a role in energy conservation is supported by observations which suggest that electrons first reduce cytochrome *b* or



E Formate hydrogenlyase complex from *E. coli*



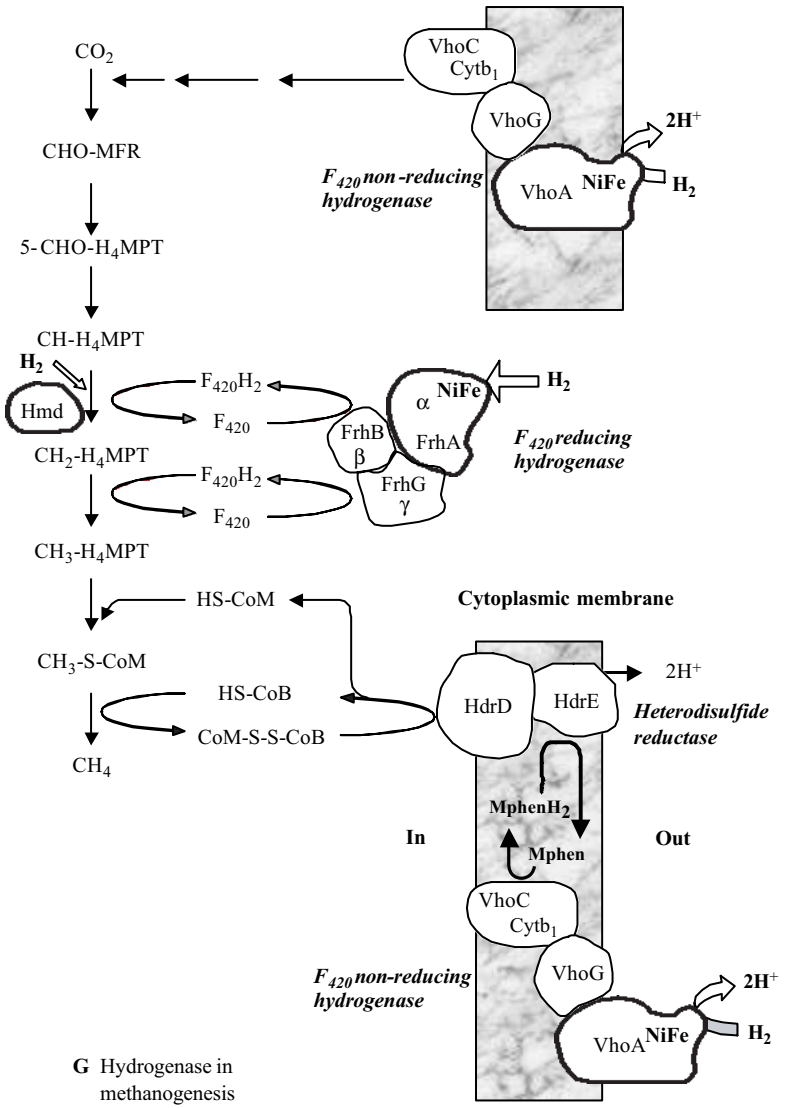
F The soluble NAD reducing hydrogenase from *A. eutrophus*

Figure 2.2 (continued) (E) The Group 10 hydrogenase associated with the formate hydrogen lyase complex from *E. coli*. (F) The Group 5 tetrameric soluble cytoplasmic hydrogenase from *R. eutropha*.

plastoquinone and are then allocated via the cytochrome *bc* complex to photosystem I or the respiratory terminal oxidase. In *Anabaena* 7120, induction of heterocyst development involves an extensive DNA excision event that affects the *hupL* gene for the NiFe subunit (Carrasco *et al*. 1995a,b), but this type of genetic rearrangement involving the *hupSL* genes is not found in all heterocyst-forming strains.

Group 4. The bidirectional-NAD(P)-reactive hydrogenases in Cyanobacteria

The so-called bidirectional or reversible hydrogenases found in some Cyanobacteria exhibit both $Na_2S_2O_4$ and methyl viologen (MV)-dependent H_2 evolution and also phenazine methosulfate or methylene-blue-dependent H_2 oxidation (uptake). They occur in filamentous N_2-fixing organisms such as *Anabaena* species (Serebrykova *et al*. 1996; Boison *et al*. 1998) and also in the unicellular non-N_2-fixing organism, e.g. *A. nidulans* (Boison *et al*. 1996) but not in *Nostoc* sp. PCC73120. These enzymes reduce $NAD(P)^+$, and sequence alignments show a distinct structural similarity between some subunits (e.g. HoxF and HoxU) of this enzyme and those of the NADH:ubiquinone oxidoreductase (Complex I) of the respiratory chain. This suggests



G Hydrogenase in methanogenesis

Figure 2.2 (continued) (G) The Group 7 (F_{420} non-reactive), Group 8 (F_{420} reactive), and Group11 (non-metal) hydrogenases from methanogens and their roles in metahanogenesis. CHO-MFR, N-formylmethanofuran; 5-CHO-H_4MPT, N^5-formyltetra dromethanopterin; CH-H_4MPT, N^5,N^{10}-methenyltrahydromethanopterin CH2-H_4MPT, N^5,N^{10}-methylenetrahydro- methanopterin; CH_3-H_4MPT N^5-methyltrahydromethanopterin; CH_3-ScoM, methyl coenzyme M; HS-CoB, Co-enzyme B or N-7-mercaptoheptanoyl-)-phospho-L-threonine (HS-HTP); Hmd is the metal-free hydrogenase.

that these enzymes are involved in energy conservation. However, the physiological role of these enzymes is unclear since these particular Cyanobacteria are incapable of autotrophic growth on H_2 and a mutant of the *hoxH* gene encoding the NiFe subunit of this hydrogenase in *A. nidulans* is unaffected in growth. In the filamentous

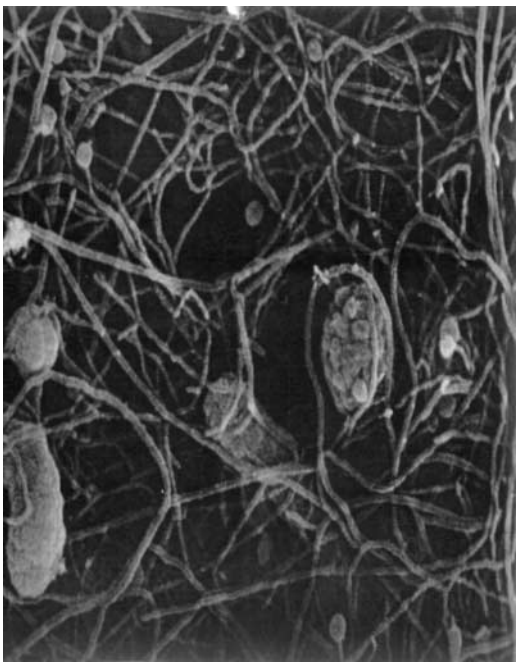


Figure 2.3 Micrograph of *Frankia* strain R43 grown in nitrogen free medium. The thread-like structures are hyphae, the small swollen spheres are vesicles and the larger structures are sporangia (Dr A. Sellstedt, University of Umeå).

N_2-fixing organisms, the enzyme is found in both the vegetative cells and the heterocysts. In *A. variabilis* and *A. nidulans* the enzyme is thought to be membrane bound. Examination of the interrelationship between photosynthesis and respiration has led to speculation that the role of the enzyme may be to control electron flow in respiration and photosynthesis (Appel and Schulz 1998, 2000) (Fig. 2.2D).

Group 5. The NAD(P)-reactive hydrogenases from bacteria

These membrane-bound or soluble multi-subunit enzymes (usually heterotetramers) are capable of reducing or oxidising nicotinamide adenine dinucleotides. They play an important part in the metabolism of knallgas bacteria which are a group of bacteria which can use H_2 as a sole energy source with O_2 as terminal electron acceptor. Also in autotrophs, the NADH formed from H_2 oxidation probably serves as the reductant for CO_2 reduction. The group includes organisms such as *R. eutropha*, *Nocardia opaca* and related organisms (Schneider and Schlegel 1976; Schneider *et al.* 1984; Hornhardt *et al.* 1990; Tran-Beckte *et al.* 1990) where H_2 can serve as sole energy source. In this well-studied system, the enzyme is soluble, reduces NAD to NADH and consists of four subunits (Fig. 2.2F) of which HoxH (β subunit) is the NiFe-containing subunit, HoxF (α subunit) contains FMN and two [Fe-S] clusters and the likely active site for NADH reduction and two other subunits, HoxY (δ subunit) and HoxU (γ subunit) both of which contain Fe-S centres which probably

are involved in electron transfer between the active sites. These enzymes are structurally related to the bidirectional hydrogenases found in Cyanobacteria.

Group 6. The NAD(P)-reactive hydrogenases from Archaea

In the hyperthermophilic Archaea, NAD(P)-reactive enzymes are involved in recycling the reduced cofactors to produce H_2 as a waste product as in the case of the NADPH oxidising hydrogenases from the hyperthermophilic Archaea, e.g. *Pyrococcus* species (Bryant and Adams 1989; Pedroni *et al*. 1995) and *Thermococcus litoralis* (Rakhely *et al*. 1999). These enzymes are also heterotetramers (Fig. 2.2C) with an apparently similar organisation of subunits and prosthetic groups to the Eubacterial examples of Group 5.

Group 7. The F_{420}-non-reducing hydrogenase of methanogens

Methanogens have an obligate requirement for H_2, which is oxidised by at least two types of hydrogenases which in turn reduce cofactors or substrates involved in the reduction of C_1 units (e.g. CO_2) to CH_4. Factor F_{420} (8-hydroxy 5-deazaflavin) is one of several unique cofactors required for methanogenesis (Fig. 2.2G). The two metal-containing hydrogenases found in methanogens can be distinguished biochemically by their ability to reduce F_{420}. The F_{420}-non-reducing hydrogenases cannot reduce F_{420} but reduce the artificial electron acceptor MV (Deppenmeier *et al*. 1992; Woo *et al*. 1993). They are thought to be important in two steps of methanogenesis: the reductive carboxylation of methanofuran to form N-formylmethanofuran, and the provision of electrons for the reductive cleavage by heterodisulfide reductase of the heterodisulfide CoM-S-S-CoB formed in the release of methane and the production of a membrane potential which is used in the synthesis of ATP (Setzke *et al*. 1994) (Fig. 2.2G). In *Methanococcus mazeii*, the F_{420} non-reducing hydrogenase and the heterodisulfide reductase may associate to form a complex known as the H_2:heterodisulfide oxidoreductase which may provide an energy-conserving proton pump (Ide *et al*. 1999). These hydrogenases are membrane bound and are of either the NiFe or NiFeSe types. Both types occur in *M. voltae* and are encoded by the *vhc* or *vhu* genes respectively (Halboth and Klein 1992). The *vhu*-encoded NiFeSe enzyme is constitutive but as the NiFe enzyme is made only when Se is limiting it appears to serve as a contingency for Se deficiency. More details of the regulation of this system appear in Chapter 4. Mostly these enzymes contain three subunits, one of which consists of the *vhc* (*vht*, *vho*)A gene which contains the NiFe(Se) centre, *vhcG* and *vhcC* which is deduced to encode a *b*-type cytochrome. However, interestingly, in *M. voltae*, VhuA is truncated at the C-terminal and lacks the cysteinyl and selenocysteinyl ligands to Ni. However, this domain is supplied by a twenty-five amino acid residue peptide encoded by a fourth gene, *vhuU*. *Methanosarcina mazei* Göl also appears to contain a pair of these enzymes encoded by the constitutively expressed *vho* genes and the *vht* genes which are not expressed when cells are grown on acetate. There is no evidence for either hydrogenase containing Se in this organism (Deppenmaier *et al*. 1992). It is interesting to note that, based on genome sequencing data, heterodisulfide reductase and hydrogenases which resemble the F_{420}-non-reactive hydrogenase also occur in the non-methanogenic hyperthermophilic, sulfate-reducing archaeon, *Archaeoglobus fulgidus*.

Group 8. The F_{420}-reducing hydrogenases of methanogens

Examples of these hydrogenases and the *frhA*, *B* and *G* genes which encode them were first characterised from *Methanothermobacter* sp. (Jacobson *et al.* 1982; Alex *et al.* 1990) and somewhat later from *Methanobacterium formicicum* (Baron and Ferry 1989). The role in methanogenesis of the F_{420} reduced by these enzymes is shown in Fig 2.2G. These are all membrane-bound NiFe(Se) enzymes which contain FAD and comprise three subunits: FrhB (β subunit), FrhG (γ subunit) and FrhA (α subunit) which bears the NiFe centre. The stoichiometry of the enzyme appears to be $\alpha_2\beta_2\gamma_1$. Most methanogens appear to possess parallel systems, one of which may be constitutive and the other expressed only during particular growth conditions, e.g. selenium deficiency, or during growth with a particular C_1 compound, e.g. with methanol but not with acetate. As for the F_{420}-non-reducing hydrogenases some methanogens contain two forms: an NiFe only form (encoded by the *frc*, *fre*, or *frh* genes) or an NiFeSe form (encoded by the *fru* genes). Both enzymes exist in *M. voltae* but the presence of Se in the growth medium suppresses expression of the NiFe-only enzyme (Halboth and Klein 1992) (see Chaper 4). *Methanosarcina barkeri* also contains a second F_{420}-reducing counterpart encoded by the *fre* genes which has a high degree of identity to the *frh*-encoded system (Fiebig and Friedrich 1989; Michel *et al.* 1995; Vaupel and Thauer 1998).

Group 9. The NiFe H_2-sensing hydrogenases

These enzymes are known in *R. eutropha* (Lenz *et al.* 1997), *B. japonicum* (Black *et al.* 1994) and *R. capsulatus* (Elsen *et al.* 1996; Vignais *et al.* 1997) which can use H_2 as an alternative sole energy source in lithotrophic situations. They have many features in common with the Group 2 enzymes in that they are heterodimers except that they are not known to couple to a cytochrome and from the gene sequences, their small subunits lack both C-terminal anchor-like and N-terminal signal domains and so presumably they are cytoplasmic rather than membrane bound or periplasmic. More details of this recently discovered group of enzymes thought to be involved in sensing H_2 can be found in Chapter 4.

Group 10. NiFe hydrogenase in the formate hydrogen lyase complex

This H_2-evolving NiFe enzyme is part of a multi-component membrane-bound enzyme system studied extensively in Enterobacteria, especially *E. coli* where it is known as hydrogenase 3 and expressed under predominantly anaerobic conditions (Sawers *et al.* 1985, 1986). The entire complex couples the reduction of protons to the oxidation of formate to form CO_2 and serves to void excess reducing potential (Fig. 2.2E). It is believed to lie on the inner face of the cytoplasmic membrane. Purification and further characterisation has been difficult because it is intrinsically unstable. However, much is known about the contiguous cluster of 8 *hyc* genes required for this hydrogenase (Bohm *et al.* 1990) of which *hycE* encodes the Ni-containing subunit. The role of a number of accessory genes in the maturation of this system will be discussed in Chapter 3. Sequence analysis of the genome of *E. coli* revealed a fourth potential NiFe hydrogenase encoded by the *hyf* gene cluster which appears to be related to hydrogenase 3 (Andrews *et al.* 1997).

Group 11. The membrane-bound Ech hydrogenases in methanogens

Relatively recently a NiFe-containing membrane-bound hydrogenase which appears to be related to the formate hydrogenase lyases has been described from *M. barkeri*. This enzyme is composed of four hydrophilic and two membrane-spanning subunits all of which show sequence identity to Complex I of respiratory chains (Künkel *et al.* 1998; Meuer *et al.* 1999). Ech does not reduce quinones but catalyses the reduction of protons to H_2 with reduced 2[4Fe-4S] ferredoxin as electron donor. This enzyme plays an important function during growth of *M. barkeri* on acetate as sole energy source in which it appears to couple to the generation of a proton motive force (Bott and Thauer 1989). The enzyme also appears to be present in *Methanothermobacter* sp. Two transcriptional units each encoding Ech-like enzymes have been identified in the genome sequence (Tersteegen and Hedderich 1999).

Group 12. The metal-free hydrogenase in methanogens

A unique metal-free hydrogenase was found first in *M. marburgensis*. It catalyses the reversible reaction of N^5,N^{10}-methenyltetrahydromethanopterin with H_2 to give N^5,N^{10}-methylenetetrahydromethanopterin and a H^+ (Fig 2.2G) (Hartmann *et al.* 1996). Several lines of evidence support the conclusion that it lacks Fe or Ni. The enzyme does not catalyse an exchange between H_2 and the protons of water nor the conversion of para H_2 to the ortho-form unless methenyl H_4MPT is present. The specific activity of the enzyme increases in cells growing under nickel-limiting conditions which suggests that it is important in methanogenesis in Ni deficient environments (Afting *et al.* 1998). The protein is encoded by the *hmd* gene which also appears in the genomes of *Methanococcus jannaschii* and *Methanopyrus kandleri*.

Group 13. The soluble hydrogenase in A. cylindrica

A. cylindrica contains two soluble hydrogenases one of 100 kDa which exhibits MV-dependent H_2 evolution and the other of 42 kDa which exhibits tritium exchange activity only. The MV-reactive enzyme is a heterodimer composed of a 50 kDa and the 42 kDa subunit form. Both proteins are required for MV reactivity (Ewart and Smith 1989). The function of this enzyme is unknown at present and the gene sequence for the tritium exchanging subunit is unrelated to any of the hydrogenases described above (Ewart *et al.* 1990).

2.4. Genetic diversity

2.4.1. Describing the genetic diversity

It is not the intention of this chapter to review the genetics of hydrogenases. However, there has been an explosion of information in this area in recent years. Table 2.3 shows that more than 27 genotypes are used to describe genes involved in the metabolism of H_2. Some are synonyms, e.g. *hox*, *hup*, *hyn*, *hya* and *hyd* have all been applied to the structural genes for the periplasmic, heterodimeric NiFe(Se) hydrogenases which form a coherent phylogenetic cluster. Unfortunately this adds to the impression of diversity for the specialist and non-specialist alike. Some of the genotypes such as *hyd*

Table 2.3 Genotypes used to describe hydrogenases, their accessory genes and related functions

Genotype	*Functions (known or presumed)*	*Organisms*
anf	Alternative nitrogenase (Fe-only)	Avi, A7120, Rca, Rru
eha	Energy converting hydrogenase a	Mta
ehb	Energy converting hydrogenase b	Mta
fhl	Formate hydrogen lyase	Eco
frc	F_{420}-reducing hydrogenases	Mvo
fre	F_{420}-reducing hydrogenases	Mba
frh	F_{420}-reducing hydrogenases	Mba, Mtm, Mvo, Mja
fru	F_{420}-reducing hydrogenases	Mvo
hdr	Heterodisulfide reductase	Mka
hev	Hydrogen evolving hydrogenase	Eco
hox	Membrane bound/Periplasmic/ soluble NADH-reducing hydrogenases (structural and accessory genes)	Aeu, Ava, Avi, Pca, Pho, S63801, S6803, Rop
hmd	Metal-free hydrogenase	Mtl
hup	Membrane-bound/Periplasmic hydrogenases (structural and accessory genes)	A7120, Ach, Ahy, Ava, Avi, Bja, N73102, Phy, Rle, Rca, Rsp, Rge, Tro
hya	Membrane-bound hydrogenase 1	Eco, Cfr, Hpy
hyb	Membrane-bound hydrogenase 2	Eco
hyc	Membrane-bound hydrogenase 3 associated with formate-hydrogen lyase	Eco
hyd	Various hydrogenases	Cac, Dfr, Dgi, Dvu, Hpy, Pfu, Tma, Tva, Wsu
hyf	Postulated hydrogenase 4 related to formate hydrogen lyase	Eco
hyp	Pleiotropic hydrogenase maturation genes	
hyn	Periplasmic (NiFe)	Dvu
hys	Periplasmic (NiFeSe)	Dvu
mbhl	Methylene blue reducing	Aae
mvh	Methyl viologen reducing	Mta, Mfe
nif	Nitrogen fixation (Mo system)	A7120, Ach, Avi, Bja, Dvu, Rle, Rca, Mvo, Mta, N73102, Tro
vhc	F_{420} non-reducing	Mvo
vho	F_{420} non-reducing	Mma
vht	F_{420} non-reducing	Afu, Mma,
vhu	Methyl viologen reducing	Mja
vnf	Alternative nitrogenase (V system)	A7120, Ach, Avi

Organisms: A7120, *Anabaena* 7120, Aae, *Aquifex aeolicus*; Afl, *Acetobacter flavidum*; Aeu, *Alcaligenes eutropha*; Ahy, *Alcaligenes hydrogenophilus*; Ach, *Azotobacter chroococcum*; Afu, *Archaeoglobus fulgidus*; Ava, *Anabaena variabilis*; Avi, *Azotobacter vinelandii*; Bja, *Bradyrhizobium japonicum*; Cfr, *Citrobacter freundii*; Cac, *Clostridium acetobutylicum*; Cpa, *Clostridium pasteurianum*; Dba, *Desulfomicrobium baculatus*; Dfr, *Desulfovibrio fructosovorans*; Dgi, *Desulfovibrio gigas*; Dvu, *Desulfovibrio vulgaris*; Eco, *Escherichia coli*; Hpy, *Helicobacter pylori*; Mba, *Methanosarcina barkeri*; Mfe, *Methanothermus fervidus*; Mja, *Methanococcus jannaschii*; Mka, *Methanopyrus kandleri*; Mma, *Methanosarcina mazei*; Mta, *Methanobacterium thermoautotrophicum*; Mtl, *Methanococcus thermolithotrophicus*; Mvo, *Methanococcus voltae*; N73102, *Nostoc* 73102; Pho, *Proclorothrix hollandica*; Pca, *Pseudomonas carboxydovorans*; Phy, *Pseudomonas hydrogenovora*; Pfu, *Pyrococcus furiosus*; Phi, *Pyrococcus horikoshii*; Reu, *Ralstonia eutrophus*; Rle, *Rhizobium leguminosarum*; Rca, *Rhodobacter capsulatus*; Rsp, *Rhodobacter sphaeroides*; Rop, *Rhodococcus opacus*; Rge, *Rhodocyclus gelatinosus*; S6301, *Synechococcus* 6301; S6803, *Synechocystis* 6803; Tma, *Thermotoga maritima*; Tro, *Thiocapsa roseopersicina*; Tva, *Trichomonas vaginalis*; Wsu, *Wolinella succinogenes*.

are confusing and are often employed as a generic term for various groups of hydrogenases. Rationalisation would appear to be required but the adoption of a simpler nomenclatures would add further confusion and tend to obscure earlier literature. A similar situation appears to be developing in the nomenclature for various hydrogenases in the Archaea and especially in the methanogens.

One of the fascinating insights to have emerged from the study of hydrogenase genetics is that many genes, in addition to regulatory and structural genes, are required for the functioning of the various hydrogenases. These accessory genes are required for the uptake of metals such as Ni, or the synthesis and insertion of metal clusters and their ligands into the nascent enzymes, the targeting of the assembled enzymes into cell compartments such as the periplasm in Eubacteria, the chloroplast of green algae, or the hydrogenosome of certain anaerobic fungi and protozoans. In addition, many enzymes need accessory proteins to become physiologically relevant, e.g. to couple to their cognate redox chains, energy conserving complexes or fermentation pathways (see Chapter 3). What we have learned from this of biotechnological importance is that the structural genes cannot be transferred from one organism to another without the appropriate accessory genes.

In the prokaryotes, structural, regulatory and accessory genes are usually organised into polycistronic, transcriptional units (operons) (Chapter 4). In Eukaryotes all genes appear to be monocistronic and there appears to be little grouping of related genes. The significance of this is that the location of accessory genes is much more difficult in the Eukaryotes and as yet though hydrogenase structural genes have been characterised, no accessory genes have been described though they must surely be present. Different operon structures occur in different organisms (see Section 5.5 and Chapter 4) and these reflect functional diversification or different regulatory strategies.

2.5. The evolution of hydrogenases

2.5.1. H_2 metabolism and the origins of life

The ability to catalyse the evolution or oxidation of H_2 may have been exploited by the earliest life forms as H_2 would have been present in the early prebiotic environments. The origins of the proton-dependent chemiosmotic mechanism for ATP synthesis may also reflect the formation of proton gradients created by hydrogenases on either side of the cytoplasmic membrane. In addition, it has been speculated that the coupling of H_2 and S metabolisms was also of fundamental importance in the origin of life. These two processes seem intimately coupled in the bifunctional sulfhydrogenase found in *Pyrococcus furiosus* (a combination of subunits for hydrogenase and sulfite reductase) which can dispose of excess reductant either by the reduction of protons to H_2 or S^0 to H_2S (Ma *et al*. 1993; Pedroni *et al*. 1995).

2.5.2. Revealing hydrogenase evolution by protein comparison

Insights into the evolution of H_2 metabolism can be gained by the comparison of different hydrogenases at the primary sequence level. The choice of which proteins to compare is influenced by the number of sequences available, their occurrence and the presumption of functional analogy. This approach was thoroughly explored by Wu and

Mandrand-Berthelot (1993). Since that time many more sequences and several new hydrogenases have emerged. Also, not only can the hydrogenase structural genes be compared but also the accessory genes, e.g. those involved in hydrogenase maturation (see Chapter 3). Other information (e.g. the comparative arrangement of hydrogenase gene clusters in different organisms) can be incorporated into the analysis.

2.5.3. How phylogenetic trees are built

The construction of phylogenetic trees relies on the computer-aided alignment of amino acid sequences and the calculation of the degree of similarity between the sequences. This results in a similarity matrix from which trees of relatedness can be constructed. Entire gene/protein sequences may be used but some degree of editing of the sequences is performed to maximise the alignment. Alternatively, it is possible to compare only those segments of molecules which are most easily aligned, i.e. they contain few or no gaps and show high sequence identity and may also represent important conserved domains within a protein. The statistical significance of such trees is usually assessed by determining how many times the tree or elements of the tree are likely to have been produced at random. The significance of such trees increases when a large number of truly homologous proteins or genes are available for comparison as for the 16S/18S ribosomal RNA subunit sequences mentioned earlier. Two basic types of trees can be constructed. Rooted trees compute or assume a common ancestor. However the rooting of trees becomes highly problematic where sequences are quite divergent, and where functionality is not analogous or unknown. This applies in most cases when considering both the hydrogenase structural and accessory gene products. Alternatively, the construction of unrooted trees avoids many such assumptions.

2.5.4. Relationships between the basic groups of hydrogenases

The first conclusion to emerge from attempts to align the sequences of the H_2-activating subunits of the metal-containing hydrogenases (Group 11 and 12 hydrogenases are not considered here) is that the Fe-only hydrogenases, the NiFe(Se) hydrogenases, and the hydrogenase component of the formate hydrogen lyases form three distinct families. It is tempting to speculate that all hydrogenases diverged from a single ancestral enzyme. However, there is little evidence to distinguish between convergent or divergent evolutionary pathways. This picture may change if novel intermediate enzymes are discovered. It is possible that certain domains, e.g. [Fe-S] cluster binding domains within the proteins, have been derived from common ancestral proteins which at one time may have been encoded by smaller genes. In this respect it is interesting to note that two of the metal-ligands forming residues of the NiFeSe centre in the F_{420}-reducing hydrogenase from *M. voltae* are borne on a separately encoded peptide. So it is possible that the Ni subunits of the NiFe(Se) enzyme family are the product of one or more gene fusion events.

Relationships in the NiFe(Se) enzyme family

Within the NiFe(Se) hydrogenase family, the unrooted tree (Fig. 2.4) clearly reveals several major lineages. As might be expected, the enzyme groups discussed above all emerge as distinct clades which reflect the major prokaryotic groups and the enzyme

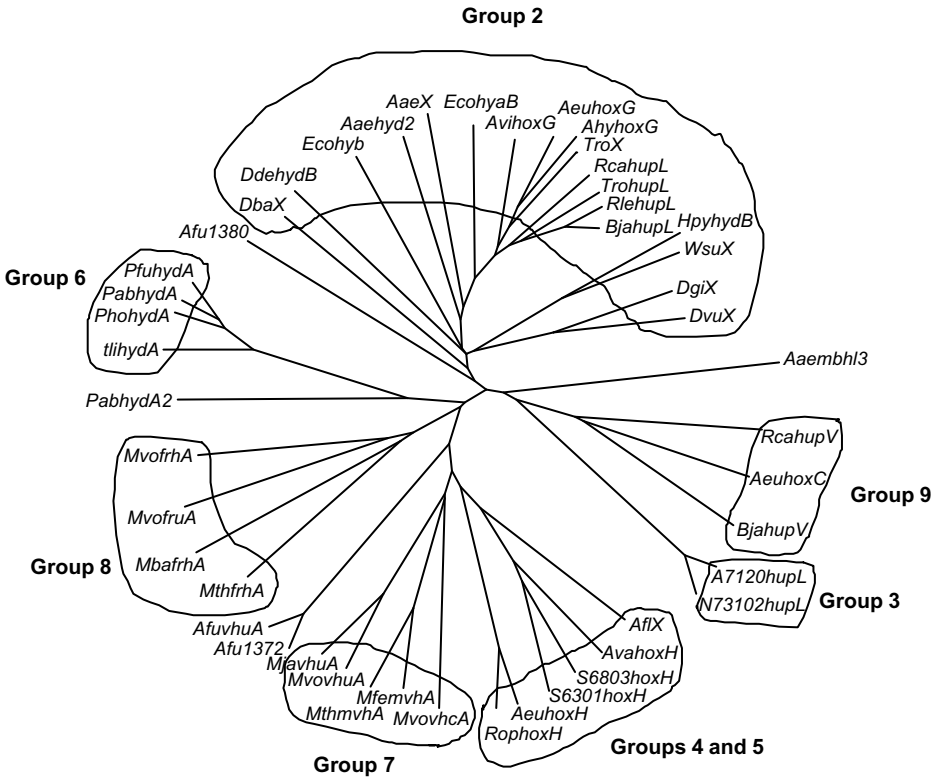


Figure 2.4 Evolutionary relationships between various members of the NiFe(Se) hydrogenase families based on Ni-binding subunits. Amino acid sequences for Ni-bearing (large) subunits of various members of the NiFe(Se) family were aligned by CLUSTAL W (Thompson *et al.* 1994). Trees were drawn using the program Treeview (Page 1996). Numbers refer to the various hydrogenase family groups shown in Table 2.2. Gene products are indicated as follows. The first three letters indicate the species (see Table 2.1) with the exception of the Cyanobacteria where the Genus and the species number is indicated. The following four letters indicate the gene product derived from the gene name where defined.

functionality within that group. It is difficult to root such trees with any degree of certainty and so comment on the very early origins of this enzyme family. For example, it is not possible to deduce that they emerged originally in an ancestor of the hyperthermophilic Archaea which are proposed to be the closest descendents of the universal common ancestor based on 16S rRNA data. However, the arrangement within each clade broadly accords with the phylogenetic relationships between the constituent organisms. This is clearly illustrated in the lineage of the Group 2 periplasmic and membrane-bound NiFe(Se) hydrogenases where the overall clustering follows the likely evolutionary origin of the α, β, δ, γ and ϵ groups of the Proteobacteria in which all these enzymes occur. Enzymes from the α, β and γ groups of the Proteobacteria (represented by *R. capsulatus*, *R. eutropha* and *E. coli* hydrogenase 1 respectively) form one cluster (Group 2a) whilst functionally comparable enzymes from the δ and

ϵ groups of the Proteobacteria (Group 2b represented by *D. gigas* and *W. succinogenes* respectively) form two other divergent clusters. This large group is easily expanded to include hydrogenase 2 (encoded by the *hyb* genes) from *E. coli*. But it is interesting to note that hydrogenases 1 and 2 of *E. coli* (encoded by the *hya* and *hyb* genes) clearly did not arise from a gene duplication event in a recent ancestor of the Enterobacteraceae. Probably these enzymes are not functionally interchangeable. But if hydrogenase 2 from *E. coli* is to be included in this group, then two hydrogenase sequences from *Desulfitobacter dehalogens* and *Aquifex aeolicus* must also be included. Yet more distantly rooted is the important NiFeSe enzyme from *Desulfomicrobium baculatum*, which does not group with the other enzymes from the δ group of the Proteobacteria and which appears as one of the most ancient of this functionally related group.

Another especially interesting feature of the tree concerns the position of the Group 3 thylakoid uptake NiFe hydrogenases from *Anabaena* and *Nostoc*, which are presumed to be functionally analogous to the Group 2 enzymes in H_2 recycling from nitrogenase in the N_2-fixing Proteobacteria (e.g. members *Azotobacteraceae* and *Rhizobiaceaea*). However, they cluster most closely with the cytoplasmic H_2-sensing hydrogenases (Group 9) found in the Proteobacteria. One common feature of these two enzyme types is that neither has a transit peptides and so they must be assumed to be cytoplasmic, or, at least in the Cyanobacteria, to be bound to thylakoid membranes. These data suggest that the sensor enzymes and the thylakoid hydrogenases may have evolved from a common ancestor.

As might be expected, the heterotetrameric NAD(P)-reactive enzymes in the Eubacteria and the Archaea form two distinct clusters within a broad lineage, which more distantly includes the F_{420}-reducing and F_{420}-non-reducing hydrogenases from methanogens. Amongst the bacteria, the cyanobacterial bidirectional Group 4 enzymes are clearly distinct from the Group 5 enzymes of the knallgas bacteria. The *Acetobacter flavidum* hydrogenase is also related to this group.

The F_{420}-reducing hydrogenases found in methanogens form a lineage (Group 7) which is quite distinct from the F_{420}-non-reducing hydrogenases (Group 6) yet both enzymes appear to have emerged from a common ancestor. Included in this group are the deduced hydrogenase sequences from *M. jannaschii*, which lies close to the universal ancestor. Also, it appears that the NiFe (*frhA*) and NiFeSe (*fruA*) enzymes which both occur in *M. voltae*, though quite closely related evolved, after a gene duplication event which occurred after the divergence of the major groups of methanogens. It is interesting that the relatively ancient *M. jannaschii* appears from the genome sequence data to contain both NiFe and NiFeSe enzymes.

The H_2-evolving multi-subunit NAD(P)-reactive hydrogenases from the Thermococcales group of the hyperthermophilic Crenarchaeota, including *Pyrococcus* sp. and *T. litoralis*, also form a distinct lineage (Group 6). Total genome sequence data suggests that *Pyrococcus abyssi* contains an isoenzyme which may have arisen as a result of a very early gene duplication in an ancestor of this group.

Despite the multiplicity of hydrogenases which occur in many organisms, there are relatively few examples of very recent gene duplications within the NiFe(Se) enzyme group. The situation is different amongst the Fe hydrogenases (see later). Hydrogenases 1 and 2 of *E. coli* emerged before the *Enterobacteriaceaea* appeared as a distinct lineage. Also, uptake hydrogenases 1 and 2 of *Thiocapsa roseopersicina* clearly fall

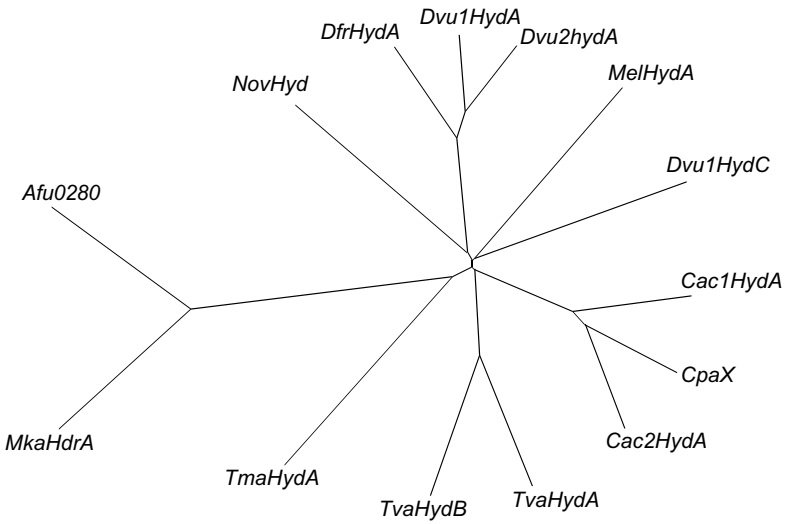


Figure 2.5 Evolutionary relationships amongst the Fe-only hydrogenases. Amino acid sequences of large subunits of Fe hydrogenases from Archaea, Eubacteria and Eukaryotes were used to construct the tree, as described for Fig. 2.3.

into two different lineages originating after the divergence of the δ, ϵ and α, β, γ clades but prior to the divergence of the latter three groups. The origin of the multiplicity of hydrogenases in the methanogens is also particularly interesting especially since it raises the question of the origin of the Se-containing enzymes. Although relatively few sequences are available at present, Se hydrogenase are found only in the Archaea within the methanococci which are the oldest of the methanogen lineages. In *M. voltae* the Se-containing enzyme appears to expressed unless Se is absent whereupon the NiFe only enzymes are expressed. *M. jannaschii* which is one of the most ancient of the methanococci also apparently contains pairs of NiFe and NiFeSe enzymes. Despite the sparse evidence, it appears as if the Se enzymes of methanogens are ancient.

Relationships amongst the Fe-only hydrogenases

Only the Fe-only hydrogenases are found in all three domains (Eubacteria, Archaea, and Eukaryotes). Therefore the evolutionary history of these enzymes is particularly interesting. They may have been present in the universal ancestor and radiated throughout all three domains or alternatively they may have evolved in a prokaryotic lineage and transferred to the Eukaryotes as a result of prokaryotic endosymbiotic events such as led to the origin of mitochondria and chloroplasts. In this respect it is particularly interesting to note that the hydrogenosome, an organelle-like structure in the protozoan *Nyctotherus ovalis*, contains its own genome much as the chloroplast and the mitochondria do. The tree for the Fe hydrogenases (Fig 2.5) shows that though the *N. ovalis* hydrogenase is deeply rooted it falls in the same lineage as the Fe hydrogenases from the sulfate-reducing γ Proteobacteria bacteria but does not group with the hydrogenases from the other protozoan *T. vaginalis*. Fe-hydrogenase isozymes are

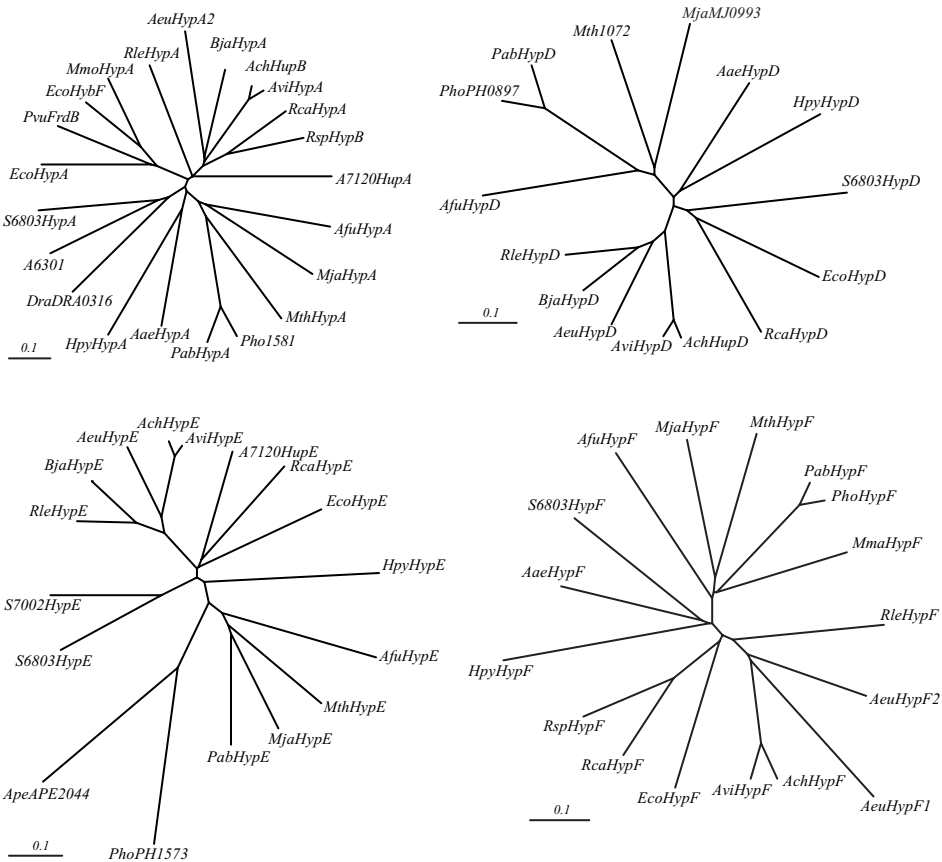


Figure 2.6 Evolutionary relationships between various *hyp* gene products. Trees were constructed for each of the *hypA*, *hypD*, *hypE*, and *hypF* encoded proteins as described in the legend for Fig. 2.3. The HypA tree includes the FrdB protein from the fumarate reductase operon of *Proteus vulgaris*.

found in the Eukaryote *T. vaginalis* and the Eubacterium *Clostridium acetobutylicum*. These isozymes appear to have arisen from gene duplications but not particularly recently in evolutionary terms. Studies of hydrogenase(s) in protozoans such as *Giardia* would be interesting especially from an evolutionary point of view since it possesses a prokaryotic 16S-type rather than eukaryotic 18S-type rRNA and may be one of the least evolved Eukaryotes.

2.5.5. *Relationships amongst accessory* (hyp) *genes*

Hydrogenase accessory genes can also be used to study the evolutionary relationships between various hydrogenase systems. This is particularly the case with the *hypA*, *B*, *C*, *D*, *E* and *F* genes most of which were originally identified as being required for the maturation of the several NiFe hydrogenases in *E. coli* (Bohm *et al.* 1990; Lutz *et al.* 1991).

Now, they are known to occur in Eubacteria and Archaea though at present none have been found in any Eukaryote. However the *hyp* genes are apparently absent in the genome of *T. maritima* which contains only an Fe-only hydrogenase. Hence the *hyp* genes may only provide information about the history of the NiFe(Se) hydrogenases. Comparison of the trees for each of the related genes should prove informative. If all the genes have followed the same evolutionary histories, then the trees should be consistent. However, at present relatively few *hyp* gene sequences are available. Those derived from genome sequencing projects can be included, but with caution, since we can only assume analogous functions. For example, from genomic sequence data we can recognise putative *hyp* genes in methanogens and hyperthermophiles but as yet there remains no direct demonstration that these genes are required for any of the hydrogenases in these organisms.

Hyp genes in the Eubacteria

The trees derived for the HypA, D, E and F proteins (Fig. 2.6) clearly divide the Archaea and eubacterial examples. Close inspection of the trees reveals interesting features. For example, in the Proteobacteria, for which the greatest number of sequences are available, the trees are remarkably consistent for each of the Hyp proteins examined. This is evident in the groupings of all four proteins in *Ralstonia, Azotobacter, R. leguminosarum, R. capsulatus* and *Rhodobacter sphaeroides*. But these trees do not mirror the evolutionary pathway deduced from 16S rRNA sequences (see Fig. 2.1). It is interesting to note that in all these organisms, the structural, regulatory and accessory genes are all tightly clustered in the genomes, and moreover in *R. eutropha*, and *B. leguminosarum*, the hydrogenase genes are even plasmid borne. Therefore it is likely that such hydrogenase gene clusters have undergone lateral gene transfer events, especially between the Proteobacteria. The positions of the *E. coli* sequences are interesting because these might be expected to group consistently with other members of the γ subdivision (e.g. the Azotobacters). However, the *E. coli* Hyp proteins are consistently more deeply rooted and more closely match the relative positions of the hydrogenase 2 (*hyb*) genes rather than the hydrogenase 1 (*hya*) genes which group more closely with other γ group members. It is also interesting to note that the *hyp* genes in *E. coli* unlike those in other Proteobacteria cluster with the *hyc* (formate hydrogenase lyase) genes rather than the *hya, hyb*, or *hyf* genes. In *E. coli*, *hypF* is separate from the main *hyp* gene cluster (Maier *et al.* 1996) but nevertheless HypF has an evolutionary history which also closely matches those of hydrogenase 2.

Another interesting observation that arises from the Hyp protein trees is the inconsistency in the apparent evolutionary history of the cyanobacterial proteins despite the fact that the Nostocales (*Anabaena* 7120) and the Chroococcales (*Synechococcus* and *Synechocystis*) are not greatly divergent according to 16S rRNA sequences. However, although there are only a few sequences available as yet for this group, a clear difference has already emerged between the positions of the HypAs and HypEs. In *Anabaena* they group with the Proteobacteria which more closely accords with the locations of the Group 2 uptake hydrogenase genes but the Hyp proteins from the Chroococcales are more deeply rooted. It appears as if the structural genes and accessory genes in the Cyanobacteria do not have parallel evolutionary histories.

Relationships between hyp genes in the Archaea

Unlike the situation with the Eubacteria, the archaeal Hyp protein trees are surprisingly inconsistent with one another. This may be because the relatively few sequences available derive from widely divergent organisms.

Diversity in the organisation of the Hyp gene clusters

There is a remarkable variation in the organisation of *hyp* genes in different organisms. In the relatively ancient organism *M. jannaschii*, the *hypA*, *B*, *C*, *D*, *E* and *F* genes are all present but they are scattered around the genome in a remarkable way and moreover none is linked to any of the hydrogenase structural gene clusters. In *M. marburgensis* the situation is similar with the exception that *hypA* and *B* are contiguous and no *hypF* has been found. It is argued that polycistronic operons allow functionally related genes to be controlled coordinately with economic use of promoter regions and regulatory proteins. However in the methanogens, expression of the *hyp* genes is probably constitutive and there seems to have been no selective advantage to be gained from the clustering of the genes. Alternatively, no particular selective disadvantage has arisen from their scattering throughout the genome. This is in striking contrast to the situation in *A. fulgidus*, their obligately sulfate reducing non-methanogenic relative, where the *hyp* genes are contiguous and sandwiched between sets of putative hydrogenase and H_2:heterodisulfide oxido reductase genes.

One can envisage how the evolutionary histories of structural and accessory genes could be clearly distinct and how the ancestry of the accessory genes may better reflect the early evolutionary origins of the NiF(Se) hydrogenase systems. In the cases of gene clusters introduced into any organism by lateral transfer events, the pre-existence of a functional set of *hyp* genes in that organism capable of supporting the newly arrived structural genes would most likely lead to the loss of the incoming set whereas the structural genes would be retained if of selective value. It is therefore interesting to note that in the remarkable assembly of megaplasmid-borne hydrogenase genes in *R. eutropha*, duplications of the *hypA*, *B* and *F* genes occur (Wolf *et al.* 1998). These genes are even partially interchangeable. Inspection of the HypF tree shows that the gene duplication was not recent but possibly occurred prior to the separation of the γ and β groups of the Proteobacteria. Therefore it appears most likely that *R. eutropha* is in the process of losing one set of *hyp* genes rather than gaining a second set by gene duplication.

2.6. Extending the horizons

One goal of biodiversity is the development of tools with which to survey and catalogue hitherto unknown genes and species. Biologists have accumulated knowledge about an astonishing 1.5 million species but an even more surprising statistic is that only 5,000 of these (~3 per cent) are prokaryotes despite the relative antiquity of these organisms. Though different species concepts are used by different groups of biologists, nevertheless, it appears that microbiologists may only have isolated and cultivated 1 per cent or less of all the existing prokaryotes. Clearly this poses an enormous challenge but offers a potential wealth of novel systems. Modern technologies

have provided new tools with which to explore this 'great unknown'. The use of the polymerase chain reaction (PCR) has enabled amplification and sequencing of genes from DNA extracted directly from environmental samples. The existence of many potential novel organisms in different environments has been predicted on the basis of the amplification and sequencing of their 16S rRNA genes which are so highly conserved that universal primers can be designed. Although hydrogenase genes are not so well conserved, it should still be possible to design sets of primers for the amplification of different hydrogenase genes from environmental samples. Such an approach has already been used with environmental samples and bioreactors to survey the diversity and expression of a 440 bp fragment of genes potentially encoding large (α) subunits of [NiFe] hydrogenases most similar to those in *Desulfovibrio* (Wawer *et al*. 1997). A considerable diversity of genes and therefore species was observed in the environmental samples but less in bioreactor samples. Also, cDNA amplified by RT-PCR of RNA obtained from the samples was used to show differential expression of the NiFe-hydrogenase genes under denitrifying and methanogenic growth conditions.

The increasing number of genome sequencing projects being undertaken means that more examples of hydrogenase genes will be uncovered. In some cases the roles of these genes may be reasonably deduced but in other cases matches to the existing databases may be ambiguous. Robust phylogenetic trees (e.g. Fig. 2.4) can help to resolve and identify potentially interesting systems. The example of *A. fulgidus*, one of the first members of the Archaea to be entirely sequenced, is illustrative. The sequence reveals a large contiguous cluster of seventeen genes encoding two sets of putative F_{420}-non-reducing hydrogenases, a putative heterodisulfide reductase and a complete cluster of *hyp* genes. However, the phylogenetic tree (Fig. 2.4) shows that in *Archaeoglobus*, the putative Ni-binding subunits AF1372 (VhuA) and AF1380 (VhtA) have roots which are quite distinct from their methanogenic counterparts with which they share closest sequence identity and which they should be most closely related on the basis of 16S rRNA analysis. Whereas the roles of these genes in methanogens is linked to methanogenesis, their role in *Archaeoglobus* is unclear since it is an obligately sulfate-reducing bacterium. The contiguity of the genes in *Archaeoglobus* also suggests that they may only be expressed under certain growth conditions unlike the constitutive expression likely in their methanogen counterparts. In this case the database searches identify convincing matches but neither the physiology of the organism nor the detailed location of these genes in the overall phylogenetic tree could convince one of the actual functions of the hydrogenases in this organism which clearly remain to be established.