The Redox Protein Construction Kit; Pre-LUCA Evolution of Energy Conserving Enzymes

Frauke Baymann, Evelyne Lebrun, Myriam Brugna, Barbara Schoepp-Cothenet, Marie-Thérèse Giudici-Orticoni, Wolfgang Nitschke

THE ROYAL SOCIETY

Published online 16 December 2002

The redox protein construction kit: pre-last universal common ancestor evolution of energy-conserving enzymes

Frauke Baymann, Evelyne Lebrun, Myriam Brugna†, Barbara Schoepp-Cothenet, Marie-Thérèse Giudici-Orticoni and Wolfgang Nitschke*

Laboratoire de Bioénergétique et Ingénierie des Protéines (CNRS UPR 9036), Institut de Biologie Structurale et Microbiologie, 31 chemin Joseph-Aiguier, 13402 Marseille Cedex 20, France

Genome analyses and the resolution of three-dimensional structures have provided evidence in recent years for hitherto unexpected family relationships between redox proteins of very diverse enzymes involved in bioenergetic electron transport. Many of these enzymes appear in fact to be constructed from only a limited set of building blocks. Phylogenetic analysis of selected units from this 'redox enzyme construction kit' indicates an origin for several prominent bioenergetic enzymes that is very early, lying before the divergence of Bacteria and Archaea. Possible scenarios for the early evolution of selected complexes are proposed based on the obtained tree topologies.

Keywords: evolution of bioenergetic enzymes; phylogeny; last universal common ancestor; redox proteins

1. THE LIMITS OF INTERSPECIES COMPARISONS IN THE STUDY OF THE EVOLUTION OF BIOENERGETICS

Up to the early 1980s, efforts to understand the origin and evolution of energy-conserving mechanisms were mainly based on theoretical considerations and geochemical evidence. The recognition of the prokaryotic roots of chloroplasts and mitochondria as well as the diversity of prokaryotic bioenergetics introduced the comparison of energy-conserving systems in as wide as possible a sample of organisms as a new approach to study evolution. This approach progressively unravelled the evolutionary history of photosynthesis and aerobic respiration backwards from plastids and mitochondria towards their earlier and more primitive precursors (see Castresana *et al.* 1995; Blankenship & Hartman 1998; Martin & Müller 1998; Xiong *et al.* 2000; Baymann *et al.* 2001; Pereira *et al.* 2001).

Molecular phylogeny and the compilation of characteristic biomarkers have shown that all organisms living on Earth fall into one of three distinct domains (or kingdoms): Bacteria, Archaea and Eukarya. Although there is some controversy on this topic, most phylogenetic analyses consider Archaea and Eukarya as sister kingdoms, thereby placing the root of the tree, i.e. the divergence from the LUCA, in between the Archaea and Bacteria sub-trees. Molecular dating and geochemical evidence suggest that the divergence of the Archaea and the Bacteria occurred between 3.5×10^9 and 4×10^9 years ago (Schopf 1993; Martin 1996; Feng *et al.* 1997). This would leave the comparatively short time of at most a few hundred million years from the time Earth became habitable (at *ca.* 3.8×10^9 years ago) to the presence of fully differentiated prokaryotes.

During the past decade, as a result of sampling in extreme environments, many hitherto unknown 'exotic' (with respect to habitat, aspect and/or metabolism) microbial species have been discovered. The availability of these species provided a new impetus to the study of the molecular evolution of bioenergetics. When a few selected species situated at strategic positions on the tree were studied in detail and their genomes sequenced, the part of the bioenergetic community interested in evolution expected to see the origin and evolution of energy-conserving mechanisms unfold before their eyes. The main questions were the number and nature of bioenergetic mechanisms operating in LUCA and their evolution into the multitude of energy-conserving electron transfer chains we find in extant prokaryotes. However, the picture that emerged from examining the bioenergetic capabilities of these species was sobering (Castresana & Moreira 1999; Castresana 2001; Forst & Schulten 2001). With the possible exception of photosynthesis and methanogenesis, which have so far only been found in Bacteria and Archaea, respectively, all major energy-conserving electron transport chains are operative in both prokaryotic kingdoms (Castresana & Moreira 1999). Phylogenetic analyses of constituent enzymes in most cases argue against lateral gene transfer of these mechanisms between

* Author for correspondence (nitschke@ibsm.cnrs-mrs.fr).
† Present address: Department of Cell and Organism Biology, Division of Microbiology, Lund University, Sölvegatan 35, SE-223 62 Lund, Sweden.

One contribution of 21 to a Discussion Meeting Issue 'Chloroplasts and mitochondria: functional genomics and evolution'.

kingdoms—a notable exception being sulphate reduction where evidence for inter-kingdom lateral transfer has been reported (Klein *et al.* 2001).

LUCA thus indeed seems to have been a bioenergetic 'species for all seasons'. This puzzling fact subsequently became comprehensible by the recognition that LUCA most probably does not represent a well-defined species but rather a gene pool formed by organisms with a low or non-existent species barrier rapidly and efficiently exchanging genetic information within the whole pool (Martin 1999; Woese 2000). The finding that many early-branching phyla contain a large fraction (20–30%) of alien DNA (Nelson *et al.* 1999; Olendzenski *et al.* 2000; Ruepp *et al.* 2000) suggests a high frequency of lateral gene transfer for these organisms and indicates that the divergence into Bacteria and Archaea should be regarded more appropriately as a progressive build-up of species barriers. Consequently, in the following the term 'LUCA' will be used to refer to the time period when the species barriers became less and less permeable rather than to a single diversification event.

Interspecies comparisons thus only allow analysis of the vertical and horizontal evolution (i.e. by inheritance and by lateral gene transfer, respectively) of individual constituent enzymes of the respective bioenergetic chains (Castresana *et al.* 1995; Xiong *et al.* 1998, 2000; Schütz *et al.* 2000; Klein *et al.* 2001; Pereira *et al.* 2001; Vignais *et al.* 2001; Lemos *et al.* 2002). Almost the full diversity of distinct bioenergetic mechanisms, however, has apparently evolved during the 'gene pool' infancy of life on earth and thus cannot be addressed by interspecies comparisons.

Does this mean that the relevant information is definitively buried in the primordial gene pool and cannot be retrieved? In the following, we show that the combination of genomic data, bioinformatic tools and, most importantly, the recent resolution of three-dimensional (3D) structures of many enzymes involved in bioenergetic mechanisms have helped to unravel previously unexpected family relationships between components of almost all key enzymes of bioenergetics. In certain cases this fact, in turn, allows the deduction of the genealogy of the family, providing hints for evolutionary events even before LUCA.

2. CONSTITUENTS OF REDOX ENZYMES FOR WHICH COMMON ANCESTRIES ARE ESTABLISHED

A few fundamental redox protein units, such as the [2Fe-2S] ('plant-type'), the 2[4Fe-4S] ('bacterial-type') ferredoxins, the thioredoxins or the type I monohaem cytochromes, occur as separate units or as domains in a vast number of enzymes. These classes of redox proteins will not be treated here for two reasons: (i) their occurrence extends far beyond the realm of bioenergetics; and (ii) they are 'small', i.e. contain very few informative sites which significantly complicates the task of constructing reliable phylogenies. The present analysis will therefore restrict itself to redox proteins that (so far) have been found predominantly in bioenergetic enzymes, and which have clearly recognizable representatives (either by 3D-structure or clear-cut sequence homology) as part of at least two distinct types of enzyme and for which sufficiently robust phylogenetic trees can be obtained. The following proteins will be considered.

(i) The molybdopterin unit (figure 1*a*) present in the dissimilatory and assimilatory nitrate reductases, in DMSO reductases, in formate dehydrogenase (Fdh-H, -O and -N; Jormakka *et al.* 2002), in formate hydrogenlyase (Andrews *et al.* 1997) and in arsenite oxidase (Ellis *et al.* 2001). The essential redox cofactors in this unit are the molybdopterin–guanidine dinucleotide and the cubane [Fe-S]-cluster, which is a [4Fe-4S]-centre in all enzymes, except arsenite oxidase where it occurs as a [3Fe-4S]-cluster. Several DMSO reductases entirely lack the [Fe-S] cluster.

(ii) The 'unusual' 4×[4Fe-4S]-protein (figure 1*b*). Several redox subunits in electron transfer enzymes have been described which contain the typical cluster binding motif of 2[4Fe-4S] (bacterial-type) ferredoxins in a tandem repeat structure. The initial interpretation therefore was that this protein corresponded with a fusion of two ferredoxin domains. A mutagenesis/spectroscopy study, however, strongly indicated a ligation pattern significantly different from such an arrangement (Guigliarelli *et al.* 1996; Blasco *et al.* 2001). The recent 3D-structure of formate dehydrogenase-N (Jormakka *et al.* 2002) confirmed the results of the mutagenesis work. Rather than being composed of N- and C-terminal 2[4Fe-4S]-ferredoxin domains, the protein consists of a 2[4Fe-4S] domain formed by the N- and the C-terminal quarters (not taking the C-terminal membrane anchor into account) and a second similar domain formed by the second and third quarters of the sequence. In addition to dissimilatory nitrate reductase and formate dehydrogenase (Fdh-N), representatives of this atypical four-cluster protein are present in group 4 [NiFe] hydrogenases (Vignais *et al.* 2001), formate hydrogenlyase (Andrews *et al.* 1997), the Hmc-complex involved in sulphate reduction (Keon & Voordouw 1996) and the recently described heterodisulphide reductase-related enzyme (Mander *et al.* 2002) from the sulphate reducing Archaeon *Archaeoglobus fulgidus*.

(iii) The [NiFe] catalytic domain (figure 1*c*) from all four groups of [NiFe] hydrogenases (Vignais *et al.* 2001; Brugna-Guiral *et al.* 2002) and from formate hydrogenlyase-2 (Andrews *et al.* 1997). This protein has been shown to be closely related to a central subunit of complex I-type enzymes, i.e. the so-called nuoD (or 49 kDa) protein (Friedrich & Scheide 2000).

(iv) Similar to the [NiFe] catalytic subunit, the 'proximal cluster domain' (figure 1*d*) from all four groups of [NiFe] hydrogenases and from formate hydrogenlyase-2 has an obvious common ancestry with the nuoB (or PSST) protein from complex I.

(v) A membrane-integral dihaem *b*-type cytochrome, in the following called 'cyt b^{I}' (figure 1*e*) is common to group I hydrogenases and formate dehydrogenase-N (Jormakka *et al.* 2002).

(vi) A second membrane-integral dihaem *b*-type cytochrome (cyt b^{II}) with a probably somewhat different

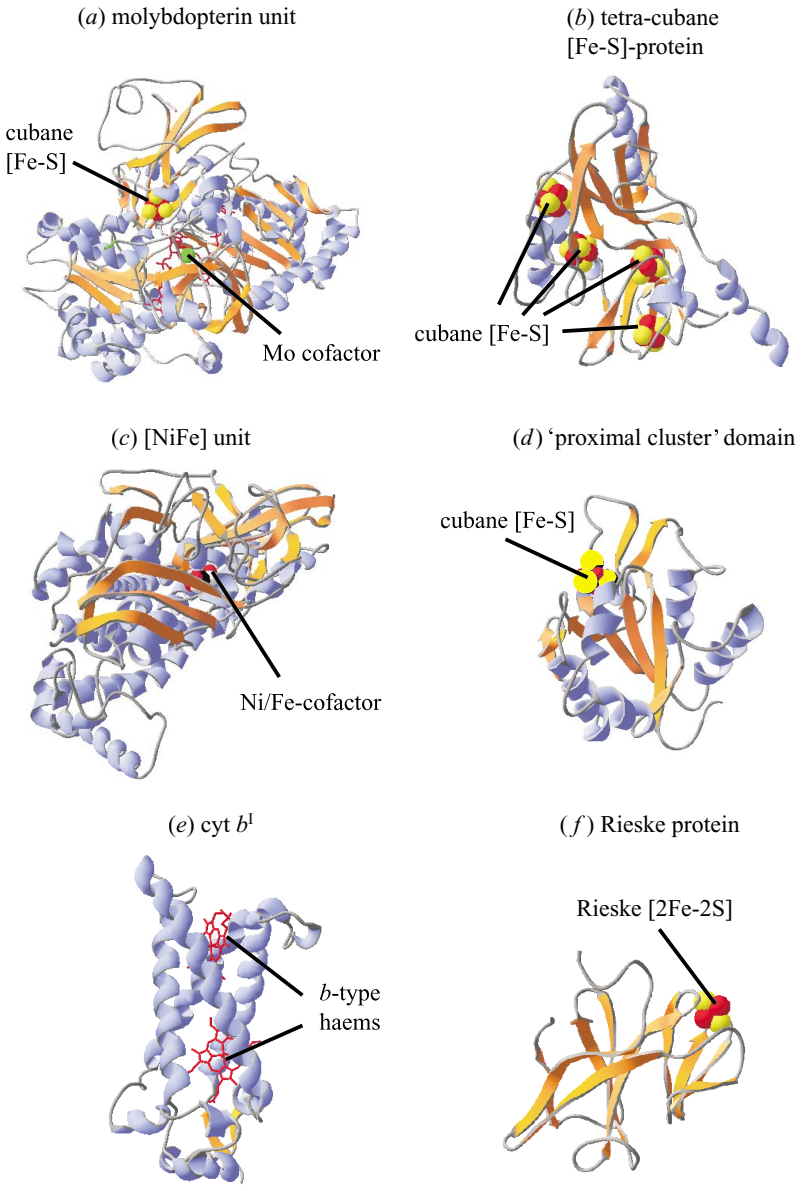

Figure 1. Cartoon representations of the six out of eight redox proteins shown in figure 2 for which 3D structures are available. The molybdopterin unit (*a*) is represented by the arsenite oxidase catalytic subunit (pdb-entry 1G8J), and the atypical tetra-cubane [Fe-S] protein (*b*) by the respective formate dehydrogenase-N subunit (1KQF). The [NiFe] unit (*c*) and the proximal cluster domain (*d*) correspond with the proteins from [NiFe] hydrogenases (1H2R), cyt b^{I} (*e*) to that of formate dehydrogenase-N (1KQF). The Rieske protein (*f*) is from arsenite oxidase (1G8J). Heteroatoms are shown in Corey, Pauling and Kultun-colour coding, the pterin–guanidine nucleotide in (*a*) is marked in red, α-helical and β-sheet structural elements are indicated in cyan and orange, respectively.

arrangement of haem groups is part of dissimilatory nitrate reductase (Blasco *et al.* 2001), a variant of group I hydrogenases (Rakhely *et al.* 1998), the Hmc-complex of sulphate reduction (Keon & Voordouw 1996) and of the heterodisulphide reductase related enzyme in *A. fulgidus* (Mander *et al.* 2002). No 3D-structure for any representative of the cyt b^{II} family is available so far. It is tempting to assume a common evolutionary origin for cyt b^{I} and cyt b^{II}. Sequence analyses, however, do not reveal any significant homology and the fact that the axial ligands to the two *b*-type haems are located in differing transmembrane helices for cyt b^{I} and cyt b^{II} has been put forward as an argument against evolutionary links between these haem proteins (Jormakka *et al.* 2002). The currently available 3D-structures of membrane integral di-haem *b*-type cytochromes show that the haem groups are invariably situated in a cavity formed by a four-helix bundle of transmembrane helices. In all cases, the connectivity pattern and the arrangement of the four-helix bundles are conserved. Although four-helix bundles may lend themselves to harbouring the two haem groups (see Shifman *et al.* 2000), it still appears intriguing to us that the general layout of the bundle turned out to be similar for all systems for which a structure has been solved so far.

(vii) The Rieske protein (figure 1*f*). The first Rieske protein was discovered almost 40 years ago in Hans Beinert's laboratory as a subunit of mitochondrial

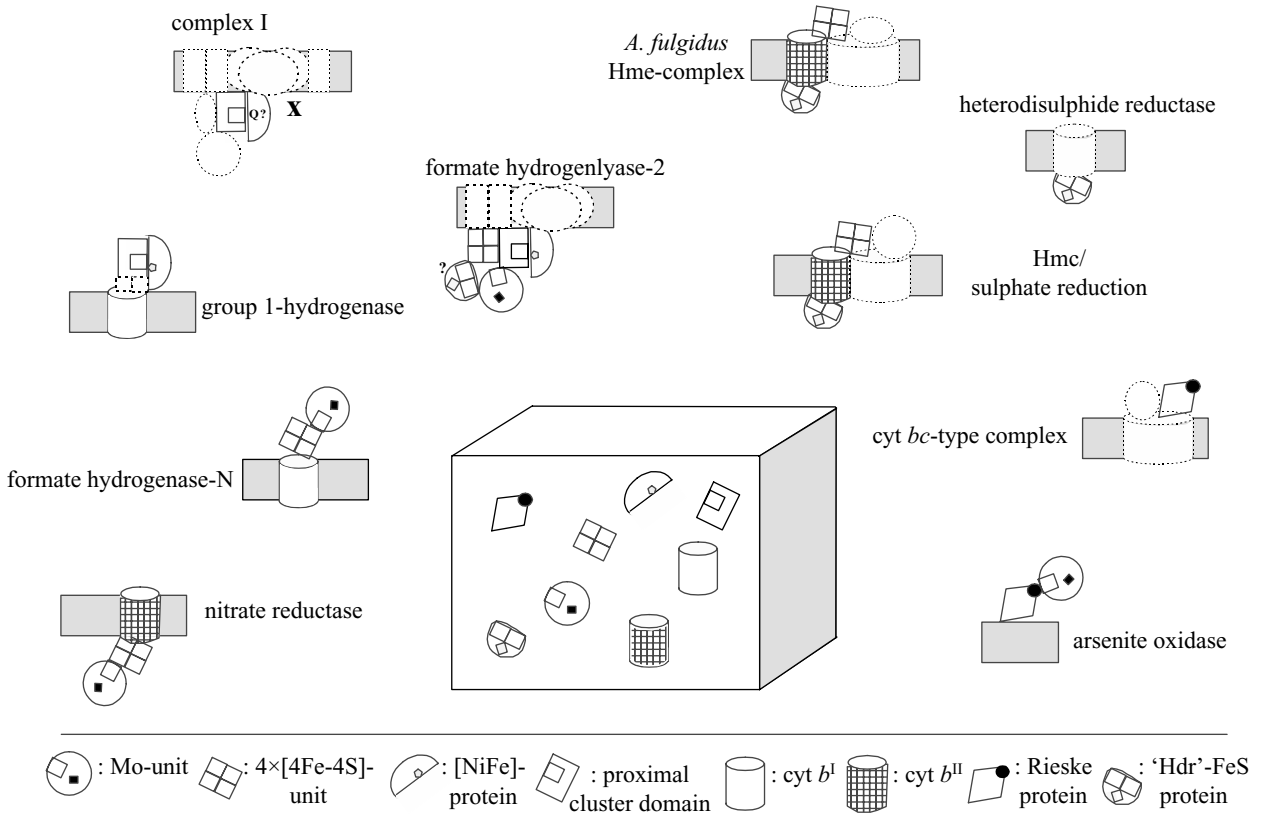

Figure 2. Schematic representation of the elements of the redox protein construction kit (inside the box) and enzymes put together mainly by using elements from the kit. In the bottom of the figure, the correspondence between the graphic symbols and the redox units referred to in the text is defined.

complex III. These enzymes, now generally referred to as cytochrome *bc*-type complexes (or, more appropriately, as Rieske/cyt*b* complexes), are present in a vast variety of bioenergetic electron transport chains of both Bacteria and Archaea (Schütz *et al.* 2000). So-called Rieske-type proteins have furthermore been found as part of several bacterial aromatic ring cleaving dioxygenases (for a review, see Schmidt & Shaw 2001). A third enzyme, arsenite oxidase, has been shown to contain a Rieske protein subunit (Anderson *et al.* 1992; Ellis *et al.* 2001).

(viii) The heterodisulphide reductase-related [Fe-S]-protein. The catalytic subunit of heterodisulphide reductase from the methanogenic archaeon *Methanosarcina barkeri* has been shown to contain a novel metal (probably iron) centre in addition to two cubane [4Fe-4S] clusters (Künkel *et al.* 1997). Homologous proteins are part of the Hmc complex involved in sulphate reduction and of the recently described Hme-complex from *A. fulgidus* (Mander *et al.* 2002). A subunit of formate hydrogenlyases (Andrews *et al.* 1997) may also be related to this family.

3. THE REDOX ENZYME CONSTRUCTION KIT

Figure 2 schematically shows that a large part of the currently known principal bioenergetic enzymes of Bacteria and Archaea are put together using various combinations of these eight basic units. The depicted ensemble of complexes comprises enzymes involved in mechanisms as diverse as nitrate respiration, sulphate reduction or hydrogen metabolism as well as 'ubiquitous' energy conserving systems such as complex I or the cytochrome *bc*-type complexes. Figure 2 is far from exhaustive because: (i) enzymes for which the authors feel incompetent have deliberately been omitted; and (ii) new members are being discovered at a rapid pace (see for example the *A. fulgidus* Hme complex or arsenite oxidase). Potentially Fe-only hydrogenases and their related enzyme subunits might also enter the scheme. For the time being, the respective family relationships appear less clear-cut than those for the depicted units and Fe-only hydrogenases have therefore not been considered in this figure.

Units from prominent bioenergetic complexes that remain unaccounted for are the core subunits of members from the cytochrome oxidase/nitric oxide reductase family, the photosynthetic reaction centres, cytochrome *f* from the b_6f complex and the membrane-integral di-haem *b*-type cytochromes from fumarate reductase/succinate dehydrogenase and the cytochrome *bc* complexes. The iron–sulphur and catalytic subunits of the fumarate reductase/succinate dehydrogenase family have numerous homologues outside the field of bioenergetics. The phylogenetic analyses for these subunits appear too complicated at present but may be integrated in this approach in the future.

4. MANY (MOST?) OF THESE ENERGY-CONSERVING ENZYMES APPEAR TO BE PRE-LUCA CONSTRUCTS

Many bioenergetic enzymes are put together by picking out a few convenient redox proteins from a very limited

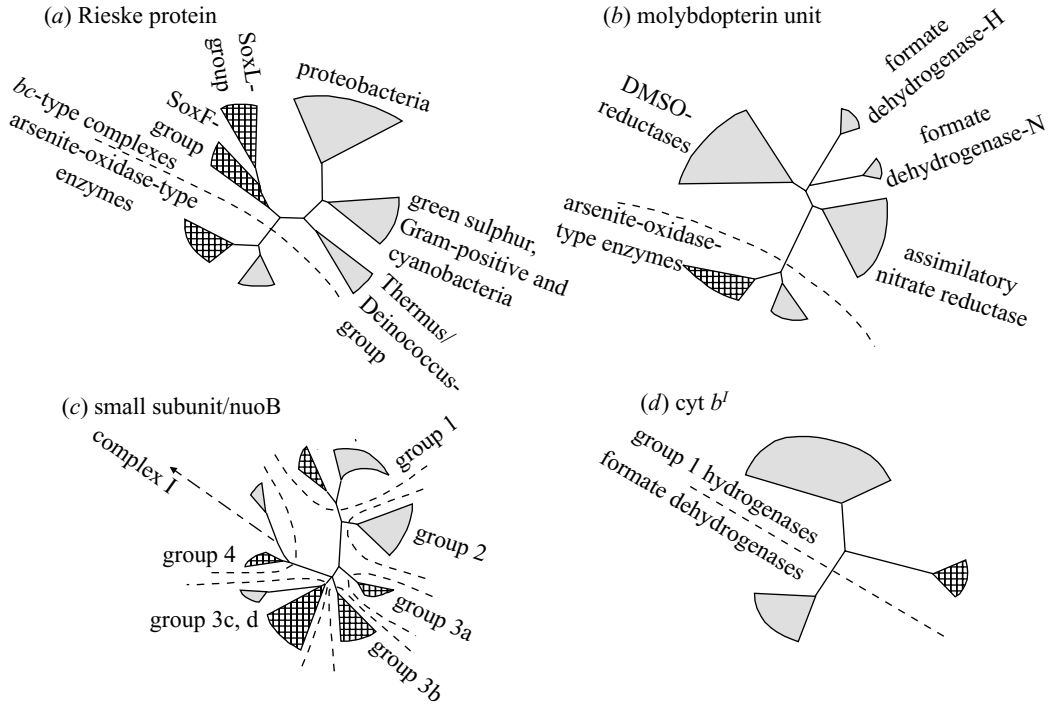

Figure 3. Schematic representation of phylogenetic trees obtained for (*a*) the Rieske protein (see figure 1*f*); (*b*) the molybdopterin catalytic subunit (see figure 1*a*); (*c*) the [NiFe] hydrogenase small/NuoB-subunit (see figure 1*d*); and (*d*) membrane-integral dihaem cyt b^{I} (see figure 1*e*). Bacterial radiations are shown in grey shade and archaeal ones are marked by hatches. Phylogenetic trees were constructed based on the neighbour-joining algorithm of Saitou and Nei using CLUSTALX (Thompson *et al.* 1997). The multiple alignment used for tree-building in the case of the Rieske proteins is based on structural comparisons in addition to Clustal-alignment (E. Lebrun and W. Nitschke, unpublished data).

basic set and rearranging them according to catalytic requirements. This raises the question whether this process is based on lateral gene transfer, i.e. is potentially still going on today, or whether these enzymes were constructed once and for all very early in life's history. A means to address this question empirically is to calculate phylogenetic trees for each of the recognized building blocks and inspect the respective tree topologies for the sequence of diversification events. We have done this for a few selected redox proteins.

5. THE RIESKE PROTEIN

Recently, we have detected arsenite-oxidase-type operons in several species of Bacteria and Archaea and have studied the phylogenetic relationship of its constituent proteins to related homologues (Lebrun *et al.* 2003). The Rieske protein (figure 1*f*) from these arsenite oxidase-type enzymes turned out to be much more closely related to the cytochrome *bc* Rieskes than the dioxygenase proteins. This fact, together with the presence of a 3D-structure for the *Alcaligenes* arsenite oxidase (Ellis *et al.* 2001), allows the construction of reliable multiple alignments between Rieske proteins from arsenite oxidase and cytochrome *bc* complexes (Lebrun *et al.* 2003), a task that proved significantly more difficult when the dioxygenase Rieske-type proteins were taken into account. For the sake of tree robustness, the present analysis therefore only considers the *bc*- and the arsenite oxidase Rieske proteins. Figure 3*a* shows a schematic phylogenetic tree constructed from the full sequences of Rieske proteins from cytochrome *bc* complexes and from arsenite-oxidase-type enzymes. The tree topology shows the *bc*- and the arsenite oxidase Rieske proteins on well-separated sub-trees. Each of these two sub-trees features the divergence in an Archaea and Bacteria branch (bacterial radiations are shown in grey shades, whereas the Archaea are marked by hatches in figure 3). This topology strongly indicates that both an arsenite oxidase and a *bc*-complex already existed in the common ancestor and that each of these two entities subsequently evolved predominantly by vertical evolution.

6. THE MOLYBDOPTERIN UNIT

The phylogram of figure 3*b* analyses the relationship of the molybdopterin subunits of DMSO reductases, formate dehydrogenases, nitrate reductases and arsenite oxidases (figure 1*a*). The obtained tree again suggests the existence of an arsenite-oxidase-type enzyme before the divergence of the prokaryotic domains (a detailed version of the tree will be published elsewhere; Lebrun *et al.* 2003). Owing to the lack of *bona fide* Archaea representatives of the other families, no conclusions can be drawn so far about the time of diversification into these enzyme families. The tree topology, however, does indicate a pre-LUCA branching into arsenite-oxidase-type enzymes on one side and the ancestor of the remaining enzymes on the other side, rather than a diversification of a bacterial arsenite oxidase representative into one or all of the other enzyme families.

7. THE [NiFe] HYDROGENASES AND COMPLEX I

Figure 3*c* shows the global topology of a phylogenetic tree of the small ([Fe-S]-) subunit of the [NiFe] hydrogenases (figure 1*d*). The tree shown in figure 3*c* corresponds closely with that reported earlier by Vignais *et al.* (2001). The topology of the phylogram based on the large (catalytic) subunit (figure 1*c*) closely resembles that of the small subunit.

The dominant feature of the tree is its splitting into well-separated sub-trees (corresponding with enzymes termed group 1–4 [NiFe] hydrogenases). For group 2–4 hydrogenases, the number of species studied is still rather limited although the ongoing genome sequencing projects provide fertile soil for the future proliferation of these sub-trees. The most complete sub-tree is that concerning group 1 hydrogenases and it again shows the splitting into Archaea and Bacteria kingdoms. The whole tree therefore suggests the coexistence of at least two (groups 1 and 4) and probably more distinct [NiFe] hydrogenases already in the common ancestor. Interestingly, the related complex I subunits do not form a separate sub-tree but seem to emerge from among the group 4 hydrogenases as already noticed previously (Friedrich & Scheide 2000). The significance of this fact will be discussed below.

8. CYTOCHROME b^{I}

The phylogram constructed from available cyt b^{I} sequences (figure 3*d*) indicates that indeed the entire group I hydrogenase enzyme (i.e. including catalytic, [Fe-S]- and cytochrome subunits) is a pre-LUCA construct. As already mentioned for its companion subunit, the molybdopterin protein, the phylogeny of cyt b^{I} from formate dehydrogenase-N (figure 1*e*) so far lacks clear-cut representatives from Archaea.

The examples discussed above admittedly only represent a fraction of the bioenergetic enzymes shown in figure 2. However, they stand out by the many representatives for which sequence information is available and therefore are the ones most amenable to this kind of analysis currently. In most cases discussed above, the most parsimonious interpretation of the trees suggests that the build-up of these enzymes occurred before the diversification of kingdoms, i.e. within the pre-LUCA gene pool era. However, other bioenergetic enzymes may well have been distributed in the post-LUCA era by lateral gene transfer. Key enzymes of the sulphate reducing pathways may be an example for such processes (Klein *et al.* 2001). A more detailed examination of the cases not addressed here will help to assay the extent to which the conclusions arrived at above may be generalized.

The finding of a pre-LUCA origin for the bioenergetic enzymes discussed above is particularly intriguing in the light of a novel hypothesis on the origin of life recently put forward by Russell and co-workers (Russell & Hall 1997; Russell *et al.* 2003) and extended by Martin and Russell (Martin & Russell 2003). According to this scheme, the earliest 'cellular' structures may have been cavities in colloidal FeS precipitates of ocean-floor hydrothermal systems. Many of the basic metabolic processes (see Martin & Russell 2003) would have evolved in these structures and it was not before the replacement of the inorganic walls by lipids that free-living cells came into being. In this model, two independent lineages would have invented fundamentally different lipids and biosynthetic pathways thereof, eventually giving birth to Archaea and Bacteria. One of the main elements of this scheme, i.e. the development of fully-fledged metabolic systems already within the mineral-confined, quasi-cellular structures, therefore predicts that many bioenergetic enzyme systems should have existed in this early phase before the Archaea/Bacteria diversification. The results shown above are thus perfectly in line with this crucial premise in Russell and Martin's model.

9. A GLIMPSE OF THE EVOLUTION OF BIOENERGETIC ENZYMES DURING THE GENE-POOL INFANCY OF LIFE

As discussed above, the splitting into Archaea and Bacteria sub-trees in the phylograms of figure 3 strongly indicates pre-LUCA origins for several of these enzymes. If one accepts this scenario then other topological parameters of the trees can be exploited for drawing conclusions on details of evolutionary pathways of these enzymes further back in time, i.e. beyond the LUCA boundary. These conclusions will become increasingly reliable as more and more sequences are taken into account for building the respective trees. The currently available data, however, already allow a few schemes to be proposed.

10. [NiFe] HYDROGENASES AND COMPLEX I

As noted previously (Vignais *et al.* 2001), the impressive diversity of [NiFe] hydrogenases already in the common ancestor witnesses the importance of hydrogen metabolism for life striving under environmental conditions of the early earth. The pre-LUCA gene pool may thus already have learnt to use H_2 as a powerful electron donor as well as protons as ubiquitous electron acceptors. The evolutionary advantage of recruiting the protons of water as a substrate for electron transfer reactions is only paralleled by the (almost certainly post-LUCA) invention of water oxidation in oxygenic photosynthesis. The topology of the phylogenetic trees of the [NiFe] hydrogenases (including complex I subunits) strongly suggests that complex I only developed after the diversification of [NiFe] hydrogenases. Recent data on subunit composition of formate hydrogenlyase-2 (Andrews *et al.* 1997) have been interpreted to suggest a very close evolutionary relationship between this enzyme and complex I, a scenario that is well-supported by the phylogenetic trees (Vignais *et al.* 2001; Brugna-Guiral *et al.* 2002). The currently available data thus indicate that a formate hydrogenlyase-related enzyme may have been converted into complex I-type enzymes by mere alteration of substrate specificities. Future analyses including subunits from the Archaea complex I-type enzymes (employing substrates different from those of their bacterial homologues) may allow the deduction that the complex I family is monophyletic or is due to a repeated divergence from an enzyme such as formate hydrogenlyase.

11. ARSENITE OXIDASE

Arsenite oxidase was discovered about a decade ago in the proteobacterium *Alcaligenes faecalis* where it is expressed when arsenite is present in the growth medium; it was therefore considered to play the part of a detoxification enzyme (Anderson *et al.* 1992). In addition to the Rieske protein (figure 1*f*), it contains the molybdopterin subunit (figure 1*a*) and the electrons resulting from the oxidation of arsenite by the molybdenum cofactor are passed on via the Rieske protein to soluble cytochrome c_2. The reducing equivalents thus ultimately end up on a terminal oxidase contributing to the overall energy coupling reactions of the parent organism. A specific search of recently sequenced genomes for arsenite-oxidase-related genes yielded arsenite-oxidase-type operons in at least three further cases, i.e. one Bacteria (*Chloroflexus aurantiacus*) and two Archaea (*Sulfolobus tokodaii* and *Aeropyrum pernix*) species (Lebrun *et al.* 2003). As shown in figure 3*a*,*b*, both subunits of these arsenite-oxidase-related enzymes cluster together on phylogenetic trees of the constituent redox units. As mentioned above, the deep branching into Archaea and Bacteria arsenite-oxidase-related enzymes strongly indicates the presence of this enzyme in pre-LUCA cells. Ecological studies of deeply branching hyperthermophilic species have demonstrated that they are able to use arsenite as electron donor for their bioenergetic chains when H_2 and reduced sulphur compounds become limiting (Jackson *et al.* 2001). Non-negligible concentrations of arsenite are indeed present in the outflow from deep-sea hydrothermal vents and in volcanic hot pools. All these facts taken together raise the possibility that electron donation from arsenite towards a terminal oxidase may be a very ancient bioenergetic mechanism. Has arsenite oxidase possibly even given birth to cytochrome *bc*-type complexes by associating with a membrane integral di-haem *b*-type cytochrome? For the time being, these questions may seem almost philosophical. It is noteworthy, however, that additional 'orphan' Rieske protein genes are found in whole genome sequences (such as in *Aquifex aeolicus*, *Deinococcus radiodurans*, *Pyrobaculum aerophilum*). Based on operon structure and primary sequences, these Rieske proteins are associated neither with a *bc*-complex nor with an arsenite oxidase (nor with a dioxygenase). A better understanding of these Rieske proteins may ultimately allow a rooting of the cyt *bc*/arsenite oxidase phylogeny and thus allow us to draw conclusions on their evolutionary origins.

12. CYTOCHROME *bc*-TYPE COMPLEXES

In contrast to the case of the [NiFe] hydrogenases, there was no indication of a pre-LUCA diversity of cytochrome *bc* complexes until recently. Genome sequencing projects of Archaea in particular, however, have provided evidence for the presence of at least two different *bc*-complexes coexisting in several species. In Bacteria, only one case of such a 'twin-*bc* organism', *Acidithiobacillus ferrooxidans*, has been detected so far (Brasseur *et al.* 2002). In *Acidithiobacillus*, phylogenetic analysis clearly shows that the diversification event is recent, having occurred within the *Thiobacillus* lineage (E. Lebrun and W. Nitschke, unpublished data). This diversification event has probably been driven by the advantage of optimizing one of the two complexes for reverse electron transfer when *Acidithiobacillus* grows exclusively by oxidizing the weak electron donor Fe^{2+} (Brasseur *et al.* 2002). In Archaea, however, a deep bifurcation between the two *bc*-type (or more precisely Rieske/cyt*b*) complexes is observed in the phylogenetic trees based on the Rieske protein (figure 3*a*), leaving room for speculation on the presence of multiple Rieske/cyt*b* complexes in the common ancestor. The possibility of rooting phylogenetic trees of the *bc*-type complex by the arsenite oxidase Rieske protein suggests that such a pre-LUCA diversity should be discarded, because in these trees the duplication event occurs within the Archaea kingdom. If ever the pre-LUCA gene pool should have contained several diversified *bc*-type complexes, then only one of them apparently has succeeded to make it into the post-LUCA era.

13. OUTLOOK

As can be seen from figure 2, the number of enzymes potentially informative for the above approach largely exceeds the few cases discussed in the preceding sections. It can therefore be expected in the near future that more complete data on bioenergetic mechanisms in Archaea, and in particular in methanogens, will allow an assessment of the validity of this approach and, we hope, a substantial extension of the kind of arguments discussed in this work.

The authors thank Michi Schütz (Regensburg, Germany), Mike Russell (Glasgow, UK), Bill Martin (Düsseldorf, Germany) and John M. Olson (Amherst, Massachusetts, USA) for many stimulating discussions. We also gratefully acknowledge the availability of gene sequences from the unfinished genome of *Chloroflexus aurantiacus* (these data have been provided freely by the US DOE Joint Genome Institute for use in this publication only). F.B. was supported by a FEBS long-term fellowship.

REFERENCES

Anderson, G. L., Williams, J. & Hille, R. 1992 The purification and characterization of arsenite oxidase from *Alcaligenes faecalis*, a molybdenum-containing hydroxylase. *J. Biol. Chem.* **267**, 23 674–23 682.

Andrews, S. C., Berks, B. C., McClay, J., Ambler, A., Quail, M. A., Golby, P. & Guest, J. R. 1997 A 12-cistron *Escherichia coli* operon (hyf) encoding a putative proton-translocating formate hydrogenlyase system. *Microbiology* **143**, 3633–3647.

Baymann, F., Brugna, M., Mühlenhoff, U. & Nitschke, W. 2001 Daddy, where did (PS)I come from? *Biochim. Biophys. Acta* **1507**, 291–320.

Blankenship, R. E. & Hartman, H. 1998 The origin and evolution of oxygenic photosynthesis. *Trends Biochem. Sci.* **23**, 94–97.

Blasco, F., Guigliarelli, B., Magalon, A., Asso, M., Giordano, G. & Rothery, R. A. 2001 The coordination and function of the redox centres of the membrane-bound nitrate reductases. *Cell. Mol. Life Sci.* **58**, 179–193.

Brasseur, G., Bruscella, P., Bonnefoy, V. & Lemesle-Meunier, D. 2002 The bc_1 complex of the iron-grown acidophilic chemolithotrophic bacterium *Acidithiobacillus ferrooxidans* functions in the reverse but not in the forward direction. Is there a second bc_1 complex? *Biochim. Biophys. Acta* **555**, 37–43.

Brugna-Guiral, M., Tron, P., Nitschke, W., Stetter, K.-O., Burlat, B., Guigliarelli, B., Bruschi, M. & Guidici-Orticoni,

M.-T. 2002 Biochemical, biophysical and phylogenetic characterization of hyperthermostable [NiFe] hydrogenases from *Aquifex aeolicus*. *Extremophiles*. (In the press.)

Castresana, J. 2001 Comparative genomics and bioenergetics. *Biochim. Biophys. Acta* **1506**, 147–162.

Castresana, J. & Moreira, D. 1999 Respiratory chains in the last common ancestor of living organisms. *J. Mol. Evol.* **49**, 453–460.

Castresana, J., Lübben, M. & Saraste, M. 1995 New archaebacterial genes coding for redox proteins: implications for the evolution of aerobic mechanisms. *J. Mol. Biol.* **250**, 202–210.

Ellis, P. J., Conrads, T., Hille, R. & Kuhn, P. 2001 Crystal structure of the 100 kDa arsenite oxidase from *Alcaligenes faecalis* in two crystal forms at 1.64 Å and 2.03 Å. *Structure* **9**, 125–132.

Feng, D.-F., Cho, G. & Doolittle, R. F. 1997 Determining divergence times with a protein clock: update and reevaluation. *Proc. Natl Acad. Sci. USA* **94**, 13 028–13 033.

Forst, C. V. & Schulten, K. 2001 Phylogenetic analysis of metabolic pathways. *J. Mol. Evol.* **52**, 471–489.

Friedrich, T. & Scheide, D. 2000 The respiratory complex I of bacteria, archaea and eukarya and its module common with membrane-bound multisubunit hydrogenases. *FEBS Lett.* **479**, 1–5.

Guigliarelli, B., Magalon, A., Asso, M., Bertrand, P., Frixon, C., Giordano, G. & Blasco, F. 1996 Complete coordination of the four Fe-S centers of the beta subunit from *Escherichia coli* nitrate reductase. Physiological, biochemical and EPR characterization of site-directed mutants lacking the highest or lowest potential [4Fe-4S] clusters. *Biochemistry* **35**, 4828–4836.

Jackson, C. R., Langner, H. W., Doahoe-Christiansen, J., Inskeep, W. P. & McDermott, T. R. 2001 Molecular analysis of microbial community structure in an arsenite-oxidizing acidic thermal spring. *Environ. Microbiol.* **3**, 532–542.

Jormakka, M., Toernroth, S., Byrne, B. & Iwata, S. 2002 Molecular basis of proton motive force generation: structure of formate dehydrogenase-N. *Science* **295**, 1863–1868.

Keon, R. & Voordouw, G. 1996 Identification of the HmcF and topology of the HmcB subunit of the Hmc complex of *Desulfovibrio vulgaris*. *Anaerobe* **2**, 231–238.

Klein, M., Friedrich, M., Roger, A. J., Hugenholtz, P., Fishbain, S., Abicht, H., Blackall, L. L., Stahl, D. A. & Wagner, M. 2001 Multiple lateral transfers of dissimilatory sulfite reductase genes between major lineages of sulfate-reducing prokaryotes. *J. Bacteriol.* **183**, 6028–6035.

Künkel, A., Vaupel, M., Heim, S., Thauer, R. K. & Hedderich, R. 1997 Heterodisulfide reductase from methanol-grown cells of *Methanosarcina barkeri* is not a flavoenzyme. *Eur. J. Biochem.* **244**, 226–234.

Lebrun, E., Brugna, M., Baymann, F., Muller, D., Lièvremont, D., Lett, M. C. & Nitschke, W. 2003 Arsenite oxidase, an ancient bioenergetic enzyme? *Mol. Biol. Evol.* (Submitted.)

Lemos, R. S., Fernandes, A. S., Pereira, M. M., Gomes, C. M. & Teixeira, M. 2002 Quinol:fumarate oxidoreductases and succinate:quinone oxidoreductases: phylogenetic relationships, metal centres and membrane attachment. *Biochim. Biophys. Acta* **1553**, 158–170.

Mander, G. J., Duin, E. C., Linder, D., Stetter, K.-O. & Hedderich, R. 2002 Purification and characterisation of a membrane-bound enzyme complex from the sulfate-reducing archaeon *Archaeoglobus fulgidus* related to heterodisulfide reductase from methanogenic archaea. *Eur. J. Biochem.* **269**, 1895–1904.

Martin, W. 1996 Is something wrong with the tree of life? *BioEssays* **18**, 523–527.

Martin, W. 1999 Mosaic bacterial chromosomes: a challenge en route to a tree of genomes. *BioEssays* **21**, 99–104.

Martin, W. & Müller, M. 1998 The hydrogen hypothesis for the first eukaryote. *Nature* **392**, 37–41.

Martin, W. & Russell, M. J. 2003 On the origin of cells: a hypothesis for the evolutionary transitions from prebiotic geochemistry to membrane-bounded chemoautotrophic prokaryotes, and from prokaryotes to nucleated cells. *Phil. Trans. R. Soc. Lond.* B **358**. (In this issue.) (DOI 10.1098/rstb.2002.1183.)

Nelson, K. E. (and 27 others) 1999 Evidence for lateral gene transfer between Archaea and Bacteria from genome sequence of *Thermotoga maritima*. *Nature* **399**, 323–329.

Olendzenski, L., Liu, L., Zhaxybayeva, O., Murphey, R., Shin, D. G. & Gogarten, J. P. 2000 Horizontal transfer of archaeal genes into the Deinococcaceae: detection by molecular and computer-based approaches. *J. Mol. Evol.* **51**, 587–599.

Pereira, M. M., Santana, C. & Teixeira, M. 2001 A novel scenario for the evolution of haem-copper oxygen reductases. *Biochim. Biophys. Acta* **1505**, 185–208.

Rakhely, G., Colbeau, A., Garin, J., Vignais, P. M. & Kovacs, K. L. 1998 Unusual organization of the genes coding for HydSL, the stable [NiFe]hydrogenase in the photosynthetic bacterium *Thiocapsa roseopersicina* BBS. *J. Bacteriol.* **180**, 1460–1465.

Ruepp, A., Graml, W., Santos-Martinez, M.-L., Koretke, K. K., Volker, C., Mewes, H. W., Frishman, D., Stocker, S., Lupas, A. & Baumeister, W. 2000 The genome sequence of the thermoacidophilic scavenger *Thermoplasma acidophilum*. *Nature* **407**, 508–513.

Russell, M. J. & Hall, A. J. 1997 The emergence of life from iron monosulphide bubbles at a submarine hydrothermal redox and pH front. *J. Geol. Soc. Lond.* **154**, 377–403.

Russell, M. J., Hall, A. J. & Mellersh, A. R. 2003 On the dissipation of thermal and chemical energies on the early Earth: the onsets of hydrothermal convection, chemiosmosis, genetically regulated metabolism and oxygenic photosynthesis. In *Natural and laboratory-simulated thermal geochemical processes* (ed. R. Ikan). (In the press.) Dordrecht, The Netherlands: Kluwer.

Schmidt, C. & Shaw, L. 2001 A comprehensive phylogenetic analysis of Rieske and Rieske-type iron-sulphur proteins. *J. Bioenerg. Biomembr.* **33**, 9–26.

Schopf, J. W. 1993 Microfossils of the early Archaean apex chart: new evidence for the antiquity of life. *Science* **260**, 640–646.

Schütz, M. (and 11 others) 2000 Early Evolution of cytochrome *bc* complexes. *J. Mol. Biol.* **300**, 663–675.

Shifman, J. M., Gibnoy, B. R., Sharp, R. E. & Dutton, P. L. 2000 Heme redox potential control in de novo designed four-alpha-helix-bundle proteins. *Biochemistry* **39**, 14 813–14 821.

Thompson, J. D., Gibson, T. J., Plewniak, F., Jeanmougin, F. & Higgins, D. G. 1997 The ClustalX Windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res.* **24**, 4876–4882.

Vignais, P. M., Billoud, B. & Meyer, J. 2001 Classification and phylogeny of hydrogenases. *FEMS Microbiol. Rev.* **25**, 455–501.

Woese, C. R. 2000 Interpreting the universal phylogenetic tree. *Proc. Natl Acad. Sci. USA* **97**, 8392–8396.

Xiong, L., Inoue, K. & Bauer, C. E. 1998 Tracking molecular evolution of photosynthesis by characterization of a major photosynthesis gene cluster from *Heliobacillus mobilis*. *Proc. Natl Acad. Sci. USA* **95**, 14 851–14 856.

Xiong, L., Fischer, W. M., Inoue, K., Nakahara, M. & Bauer, C. E. 2000 Molecular evidence for the early evolution of photosynthesis. *Science* **289**, 1724–1730.

GLOSSARY

DMSO: dimethylsulphoxide

LUCA: last universal common ancestor