Journal of Bioenergetics and Biomembranes, Vol. 33, No. 1, 2001

A Comprehensive Phylogenetic Analysis of Rieske and Rieske-Type Iron-Sulfur Proteins

Christian L. Schmidt[1,3] and Lee Shaw[2]

Received July 25, 2000; accepted October 14, 2000

The Rieske iron-sulfur center consists of a [2Fe–2S] cluster liganded to a protein via two histidine and two cysteine residues present in conserved sequences called Rieske motifs. Two protein families possessing Rieske centers have been defined. The Rieske proteins occur as subunits in the cytochrome bc_1 and cytochrome $b_6 f$ complexes of prokaryotes and eukaryotes or form components of archaeal electron transport systems. The Rieske-type proteins encompass a group of bacterial oxygenases and ferredoxins. Recent studies have uncovered several new proteins containing Rieske centers, including archaeal Rieske proteins, bacterial oxygenases, bacterial ferredoxins, and, intriguingly, eukaryotic Rieske oxygenases. Since all these proteins contain a Rieske motif, they probably form a superfamily with one common ancestor. Phylogenetic analyses have, however, been generally limited to similar sequences, providing little information about relationships within the whole group of these proteins. The aim of this work is, therefore, to construct a dendrogram including representatives from all Rieske and Rieske-type protein classes in order to gain insight into their evolutionary relationships and to further define the phylogenetic niches occupied by the recently discovered proteins mentioned above.

KEY WORDS: Rieske; iron–sulfur; phylogenetics; evolution; cytochrome bc_1; cytochrome $b_6 f$; oxygenase; ferredoxin; archaea; electron transfer.

INTRODUCTION

There now exists a substantial body of evidence to suggest that iron and sulfur played a decisive role in the chemical processes leading to the appearance of terrestrial life (Wächtershäuser, 1992; Blöchl *et al.*, 1992; Russell *et al.*, 1994). It has been suggested that together these elements not only provided the necessary chemical driving force for these events, but also served as catalysts for many of the reactions. Perhaps, not unexpectedly, iron and sulfur are frequently encountered as integral components in the metabolism, in particular, the energy metabolism of present-day organisms. One of the key pieces of evidence for the catalytic function of iron and sulfur in early and possibly prebiotic reactions is the existence of enzymes possessing essential iron-sulfur centers, consisting of a cluster of iron and inorganic sulfide ions bound to a polypeptide chain via specific amino acid residues. Iron-sulfur enzymes are ubiquitous throughout the living world, fulfilling a variety of functions in general metabolism and energy conservation (Palmer, 1975; Cammack, 1992; Beinert *et al.*, 1997). In these proteins, the iron-sulfur centers frequently play a role in redox reactions, in which the iron ions alternate between Fe^{III} and Fe^{II} states, although their involvement in regulatory processes and in the chemical transformation of substrates is also common (Rouault and Klausner, 1996; Cammack, 1992).

The functional diversity of iron–sulfur proteins is reflected in the variety of cluster structures encountered, as well as in the strategies employed to ligand the iron–sulfur cluster to the polypeptide chain. The rubredoxins, which contain one iron atom liganded by four cysteine side chains, represent the simplest form of a sulfur-liganded iron cofactor (Palmer, 1975). However, the majority of true iron–sulfur clusters occur in the form of [2Fe–2S],

[1] Institut für Biochemie der Medizinischen Universität Lübeck, Ratzeburger Allee 160, 23538 Lübeck, Germany.
[2] Biochemisches Institut, Universität Kiel, Olshausenstr. 40, 24098 Kiel, Germany. e-mail: Ishaw@biochem.uni-kiel.de
[3] To whom correspondence should be addressed; e-mail: schmidt@biochem.muluebeck.de

0145-479X/01/0200-0009$19.50/0 © 2001 Plenum Publishing Corporation

[4Fe–4S], and [3Fe–4S] structures and are bound via their Fe ions to the sulfur atoms of cysteine side chains of the respective proteins. The same is also true for a number of more complex iron–sulfur centers, such as the cofactor clusters of nitrogenase (Chan *et al.*, 1993). However, the so-called Rieske iron–sulfur center is a notable exception to this general rule. This iron–sulfur cluster is of the [2Fe–2S] type, one of the Fe ions being liganded to two cysteine sulfur atoms and the other Fe ion being bound to the ε-imidazole nitrogens of two histidine residues.

The Rieske iron–sulfur center was first discovered in the mitochondrial cytochrome bc_1 complex by EPR spectroscopy (Rieske *et al.*, 1964). In its reduced form, the suspected [2Fe–2S] cluster exhibited a similar EPR spectrum to the purely cysteine-bound reduced [2Fe–2S] clusters, although the average *g* value of the Rieske cluster was shifted, indicative of a different mode of liganding to the polypeptide chain. Further research revealed that iron–sulfur clusters with similar spectroscopic properties occur in cytochrome *bc* complexes of bacterial respiratory chains (Trumpower and Gennis, 1994) and in the cytochrome $b_6 f$ complex of the photosynthetic electron transport chain of plants and photosynthetic bacteria (Riedel *et al.*, 1991). Subsequent investigations also revealed the presence of Rieske centers in the respiratory chains of the archaea (Schäfer *et al.*, 1996). Since the Rieske center was discovered in the respiratory and photosynthetic electron transport chains, the respective components of these systems are commonly referred to as Rieske proteins.

Interestingly, Rieske centers also occur in the electron transport chains associated with a number of bacterial hydroxylases and oxygenases (Batie *et al.*, 1991; Mason and Cammack, 1992; Butler and Mason, 1997). These proteins are frequently termed Rieske-type proteins. Bacterial Rieske ferredoxins, which deliver electrons to oxygenases possessing a di-iron center of the type found in methane monooxygenase have also been reported (Pikus *et al.*, 1996; Small and Ensign, 1997). Although the existence of bacterial oxygenases containing a Rieske center has long been recognized, recent research has revealed several Rieske oxygenases of eukaryotic origin. In contrast to the bacterial enzymes, which are generally involved in the degradation of exogenous hydrophobic, usually aromatic compounds, the eukaryotic oxygenases appear to act on a number of unrelated endogenous substrates, for example, cytidine monophosphate–*N*-acetylneuraminic acid hydroxylase (Schlenzka *et al.*, 1996), choline monooxygenase (Rathinasabapathi *et al.*, 1997), and two proteins of unknown function postulated to be Rieske oxygenases on the basis of cDNA sequences (Caliebe *et al.*, 1997; Gray *et al.*, 1997).

Sequence analyses of the various Rieske- and Rieske-type proteins revealed that, in general, all possess a common homologous region with the sequence -C-X-H-X_{15-17}-C-X-X-H- (Neidle *et al.*, 1991; Mason and Cammack, 1992; Castresana *et al.*, 1995; Carrell *et al.*, 1997). However, exceptions to this general rule can be found in certain Rieske proteins of archaeal origin in which the two cysteine–histidine boxes are separated by a longer stretch of amino acids (Henninger *et al.*, 1999). In functionally related proteins, more extensive sequence homology is observed both within the above motif and in other regions. The results of spectroscopic studies (Gurbiel *et al.*, 1991; Kuila *et al.*, 1992; Riedel *et al.*, 1995; Gurbiel *et al.*, 1996) and *in vitro* mutagenesis experiments (Davidson *et al.*, 1992) suggested that the conserved cysteine and histidine residues in this sequence motif were responsible for liganding the Rieske iron–sulfur center. Final confirmation of this hypothesis has come from recent X-ray crystallographic studies on the Rieske proteins from the mitochondrial cytochrome bc_1 complex (Iwata *et al.*, 1996), the cytochrome $b_6 f$ complex of chloroplasts (Carrell *et al.*, 1997), as well as the oxygenase component of naphthalene dioxygenase from *Pseudomonas* sp. NCIB 9816-4 (Kauppi *et al.*, 1998).

Although the proteins with Rieske [2Fe–2S] centers exhibit quite different catalytic functions, their possession of a common sequence motif suggests that they are in some way related, presumably by divergent evolution from a common ancestor. Limited sequence alignment studies have yielded information on the phylogenetic relationships between proteins of a particular type, *e.g.*, of the bacterial Rieske oxygenases (Asturias *et al.*, 1995; Nakatsu *et al.*, 1995), the Rieske ferredoxins (Nakatsu *et al.*,1995), the Rieske proteins of the cytochrome *bc* complexes (Castresana *et al.*, 1995), and the archaeal Rieske proteins (Schäfer *et al.*, 1996; Henninger *et al.*, 1999). In this paper, more extensive alignments, encompassing all the Rieske cluster-containing protein types have been performed in order to gain insight into the origins and the mode of evolution of these proteins. In view of the difficulties associated with alignments of a large number of distantly related sequences, our approach has involved the separate alignment of the Rieske and Rieske-type proteins followed by a global alignment of the prealigned sequence groups.

The results of this extensive study not only confirm the subdivision of proteins containing Rieske [2Fe–2S] centers into the Rieske and Rieske-type proteins, they also

reveal some new and unexpected phylogenetic relationships within these protein groups. On the basis of these results, we suggest a course for the evolution of the various redox proteins possessing a Rieske center.

MATERIALS AND METHODS

Sequences Employed in this Study

The sequences of the following Rieske and Rieske-type proteins used in the analysis reported here were retrieved from GenBank via PubMed (http://www.ncbi.nlm.nih.gov/PubMed/). The abbreviated name of the protein is followed by the accession number in parenthesis and, where appropriate, the full designation and origin of the respective protein.

Mitochondrial Rieske Proteins from the Cytochrome bc_1 Complex

RATrat (P20788) rat (*Rattus rattus*) (Nishikimi *et al.*, 1989); BOStau (P13272) ox (*Bos taurus*) (Brandt *et al.*, 1993); SACcer (P08067) *Saccharomyces cerevisiae* (Beckmann *et al.*, 1987); NEUcra (P07056) *Neurospora crassa* (Harnisch *et al.*, 1985); SOLtub (P37841) potato (*Solanum tuberosum*) (Emmermann *et al.*, 1994); NICtab (P49729) tobacco (*Nicotiania tabacum*) (Huang *et al.*, 1991).

Chloroplast Rieske Proteins from the Cytochrome $b_6 f$ Complex

VOLcar (AAD55565) *Volvox carteri* (Meissner *et al.*, 1998); CHLrei (P49728) *Chlamydomonas reinhardtii* (de Vitry, 1994); NICtab (CAA45705) tobacco (*Nicotiania tabacum*) (Palomares *et al.*, 1991); SPIole (P08980) spinach (*Spinacia oleracea*) (Steppuhn *et al.*, 1987); PISsat (P26291) pea (*Pisum sativum*) (Salter *et al.*, 1992).

Bacterial Rieske Proteins

RHOrub (P23136) *Rhodospirillum rubrum* (Majewski and Trebst, 1990); PARden (P05417) *Paracoccus denitrificans* (Kurowski and Ludwig, 1987); RHOcap (P08500) *Rhodobacter capsulatus* (Davidson and Daldal, 1987a); RHOsph (CAA27194) *Rhodopseudomonas sphaeroides* (Davidson and Daldal, 1987b); CHRvin (O31214) *Allochromatium vinosum* (Chen *et al.*, 1997); HELpyl (AAD07047) *Helicobacter pylori* (Alm *et al.*, 1999); AQUaeo (AAC06423) *Aquifex aeolicus* (Deckert *et al.*, 1998); *Nostoc* (P14698) *Nostoc* PCC7906 (Kallas *et al.*, 1988); SYNcys (P26290) *Synechocystis* PCC 6803; SYNcoc (P26292) *Synechococcus* PCC7002 (Mayes and Barber, 1991); PHOlam (CAA70823) *Phormidium laminosum* (Wagner *et al.*, 1996); ANAsp. (P70758) *Anabaena* PCC7120 (Ramaswarmy *et al.*, 1996); ANAvar (CAB72244) *Anabena variabilis* (Arnold, 2000); CHLlim (Q46136) *Chlorobium limnicola* (Schütz *et al.*, 1994); BACsub (P46911) *Bacillus subtilis* (Yu *et al.*, 1995); HELmob (AAC84018) *Heliobacillus mobilis* (Xiong *et al.*, 1998); THEthe (AAB91482) *Thermus thermophilus* (Gatti *et al.*, 1997); STRliv (AAD04932) *Streptomyces lividans* (Parro and Mellado, 1998); MYCtub (Q10387) *Mycobacterium tuberculosis* (Cole *et al.*, 1998).

Archaeal Rieske Proteins

PYRaer (AAF02198) *Pyrobaculum aerophilum* (Henninger *et al.*, 1999); SULaci soxF (S56156) *Sulfolobus acidocaldarius* soxF (Schmidt *et al.*, 1996); SULaci soxL (CAA65777) *Sulfolobus acidocaldarius* soxL (Schmidt *et al.*, 1996); AERper soxL (BAA80725) *Aeropyrum pernix* (soxL homologous protein) (Kawarabayasi *et al.*, 1999). Two archaeal Rieske protein sequences were retrieved from the homepage of the *Sulfolobus solfataricus* P2 project: One displaying a high degree of similarity to the soxL protein of *S. acidocaldarius:* SULsol soxL (http://niji.imb.nrc.ca/sulfolobus/sh02c0748/Sequences/AA/c48_008.aa) as well as a second sequence, SULsol RFeS (http://niji.imb.nrc.ca/sulfolobus/sh19h1230/Sequences/AA/c30_022.aa).

Rieske Ferredoxins from the following Oxygenase Systems

PSEput TODB (J04996.1) toluene dioxygenase from *Pseudomonas putida* (Zylstra and Gibson, 1989); PSEput BENZC (P08086) benzene-1,2-dioxygenase from *Pseudomonas putida* (Irie *et al.*, 1987); PSEput BEDB (Q07947) benzene-1,2-dioxygenase from *Pseudomonas putida* (strain ML2) (Tan *et al.*, 1993); PSEsp. bphF (P37332) biphenyl dioxygenase from *Pseudomonas* strain LB400 (Erickson and Mondello, 1992); BURsp. dntAb (U62430.1) 2,4-dinitrotoluene dioxygenase from

Burkholderia sp. (Suen *et al.*, 1996); PSEput NDOA (JN0643) naphthalene dioxygenase from *Pseudomonas putida* (Kurkela *et al.*, 1988); RHIleg mocE (AF076240) ferredoxin component of the rhizopine degradation pathway of *Rhizobium leguminosarum* (Bahar *et al.*, 1998); PSEsp. phnR (AF061802) unknown function in *Pseudomonas* DJ77 (Kim and Park, 1998); PSEmen tmoC (M65106) toluene-4-monooxygenase from *Pseudomonas mendocina* (Yen *et al.*, 1991); XANsp. xamoC (AJ012090) alkene monooxygenase from *Xanthobacter* strain Py2 (Zhou *et al.*, 1999).

Prokaryotic Rieske Oxygenases

BURcep ophA2 (AF095748) phthalate dioxygenase from *Burkholderia cepacia* (Chang and Zylstra, 1998); ALCsp. CBAA (Q44256) 3-chlorobenzoate-3,4-dioxygenase from *Alcaligenes* sp. (Nakatsu *et al.*, 1995); COMtes tsaM (U32622) toluene sulfonate methylmonooxygenase from *Comamonas testosteroni* (Junker *et al.*, 1997); PSEput oxoO (Y12655.1) 2-oxo-1,2-dihydroquinoline 8-monooxygenase from *Pseudomonas putida* (Rosche *et al.*, 1997); PSEput TODC1 (J04996.1) toluene dioxygenase (α-subunit) from *Pseudomonas putida* (Zylstra and Gibson, 1989); PSEput BNZA (P08084) benzene 1,2-dioxygenase (α subunit) from *Pseudomonas putida* (Irie *et al.*, 1987); PSEput BEDC1 (L04642/L04643) benzene dioxygenase (α subunit) from *Pseudomonas putida* (strain ML2) (Tan *et al.*, 1993); PSEsp. bphA (P37333) biphenyl dioxygenase (α subunit) from Pseudomonas strain LB400 (Erickson and Mondello, 1992); SPHsp. dxnA1 (AJ223220.1) dioxin dioxygenase (α subunit) from *Sphingomonas* sp. (Armengaud *et al.*, 1998); PSEput NDOB (M23914) naphthalene dioxygenase (α-subunit) from *Pseudomonas putida* (strain NCIB9816) (Kurkela *et al.*, 1988); nahAc (JN0644) naphthalene dioxygenase (α subunit) from *Pseudomonas putida* (strain G7) (Simon *et al.*, 1993); PSEabi DitA1 (AF119621) 7-oxodehydroabietic acid dioxygenase (large subunit) from the diterpenoid-degrading bacterium *Pseudomonas abietaniphila* BKME-9 (Martin and Mohn, 1999); METmet msmA (AF091716) methanesulfonic acid monooxygenase (α subunit) from *Methylosulfonomonas methylovora* (de Marco *et al.*, 1999); ACIsp. benA (AF009224) benzoate-1,2-dioxygenase from *Acinetobacter* sp. (strain ADP1) (Neidle *et al.*, 1991); PSEput xylX (M64747) benzoate-1,2-dioxygenase from *Pseudomonas putida* (TOL plasmid) (Harayama *et al.*, 1991); BURcep CbdC (X79076) 2-halobenzoate-1,2-dioxygenase from *Burkholderia cepacia* (Haak *et al.*, 1995).

Eukaryotic Rieske Oxygenases

SPIole Cholmon (U85780) choline monooxygenase from spinach (*Spinacia oleracea*) (Rathinasabapathi *et al.*, 1997); ARAtha LIS1 (U77347) the lethal leaf spot 1 gene from *Arabidopsis thalia* (Gray *et al.*, 1997); ZEAmai LIS1 (U77436) the lethal leaf spot 1 gene from maize (*Zea mays*) (Gray *et al.*, 1997); PISsat Tic55 (AJ000520) 52 kDa protein involved in protein translocation across the inner membrane of chloroplasts from pea (*Pisum sativum*) (Caliebe *et al.*, 1997); PANtro CMPNAc (AF074481) cytidine monophosphate-*N*-acetylneuraminate (CMP-Neu5Ac) hydroxylase from chimpanzee (*Pan troglodytes*) (Chou *et al.*, 1998); MUSmus CMPNAc (A57469) CMP-Neu5Ac hydroxylase from mouse (*Mus musculus*) (Kawano *et al.*, 1994).

Sequence Alignments and Calculation of Phylogenetic Trees

The program ClustalX (version 1.64b) (Thompson *et al.*, 1997) was used to calculate the alignments and the phylogenetic trees. Initially, the Rieske and the Rieske-type proteins were aligned separately. Four criteria were used to evaluate the quality of the alignments:

1. The alignment of residues known to be homologous such as the histidine and cysteine ligands of the iron–sulfur cluster and the residues of the Pro-loop (Iwata *et al.*, 1996) of the Rieske proteins.
2. The above condition should be met by introducing a minimal number of gaps.
3. The phylogenetic trees calculated from the alignments should reflect the widely accepted monophyletic origin of chloroplasts and mitochondria, which was also previously demonstrated to apply to the Rieske proteins of the cytochrome bc_1 and $b_6 f$ complexes of these organelles (Henninger *et al.*, 1999; Castresana *et al.*, 1995).
4. While fulfilling the first three conditions, the resulting trees should be as robust as possible as indicated by the bootstrap values.

All gaps were reset before the beginning of the individual alignments of the two protein groups.

Pairwise Alignment Parameters

The identity matrix was used. All other parameters remained as preset by the program.

Multiple Alignment Parameters

The threshold for delaying divergent sequences was set to 10%. Two slightly different parameter sets were found to produce optimal alignments for the Rieske and Rieske-type proteins: The BLOSUM series was used for both sets of proteins. The "gap separation distance" was set to 16 for the Rieske-type proteins. All other parameters remained as preset by the program. In the case of the Rieske proteins, slightly better results were obtained when the "gap opening penalty" was increased to 15 and the "gap separation distance" was set to 12. All other parameters remained as preset by the program. Positions containing gaps were excluded from the calculation of the phylogenetic trees.

The combined alignment and the phylogenetic tree of the Rieske and the Rieske-type proteins was calculated from the individual alignments of both protein groups. Both alignments were loaded and a complete alignment was calculated without resetting the gaps. Initially, the identity matrix was used for the pairwise as well as for the multiple alignments. The "gap opening penalty" for the multiple alignments was set to 15 and the threshold for delaying divergent sequences was set to 10%. All other parameters remained as preset by the program. "Gap only positions" were removed after calculating the alignment. The central part of this preliminary alignment was refined by calculating a second alignment using the identity matrix for the pairwise comparison and PAM series for the multiple alignments. All other settings remained as before. Subsequently, a number of sequences, which could be represented by closely related proteins, were deleted in order to simplify the combined phylogenetic tree of the Rieske and Rieske-type proteins (see Fig. 2). Positions containing gaps were not excluded from the calculation of this dendrogram, since the position and extent of sequence insertions or deletions also contains phylogenetically relevant information.

The trees were drawn using the programs "NJplot" by Perrière and Gouy (1996) and "unrooted" (http://pbil.univ-lyon1.fr/software/unrooted.html) and the layout modified using a standard graphics program.

RESULTS AND DISCUSSION

A multiple sequence alignment of the various proteins possessing a Rieske iron–sulfur center revealed that only a stretch of about 70 amino acids, encompassing the [2Fe–2S]-cluster binding domain, exhibited any overall sequence homology. This section of the alignment is presented in Fig. 1. For simplicity, representative examples of each protein type, rather than all the proteins included in the comprehensive alignment (see section Materials and Methods), are shown. The only globally conserved amino acids were the cysteine and histidine residues in the Cys-X-His and Cys-X-X-His boxes whose side chains are responsible for coordinating the iron atoms of the iron–sulfur cluster (shaded residues in Fig. 1). However, two regions consisting of 3 to 4 hydrophobic residues were found situated three residues N-terminally from the Cys-X-His box (labeled H1) and a variable distance C-terminally from the Cys-X-X-His box (labeled H2) of all the sequences investigated. In addition, a single conserved hydrophobic residue situated two amino acids N-terminally from the Cys-X-X-His box was also observed.

Although no obvious further global homologies within the the Rieske-type proteins are visible, a number of homologies between the Rieske proteins could be identified within the Rieske motif region. For example, a conserved Gly–Cys peptide is located two residues C-terminally of the Cys-X-His box, thus expanding its consensus sequence to Cys-X-His-X-Gly-Cys. The Cys-X-X-His box also harbors a strictly conserved Cys, which has been shown in structural studies to form a disulfide bond with the additional Cys in the Cys-X-His box (indicated in Fig. 1) (Iwata *et al.*, 1996; Carrell *et al.*, 1997). Furthermore, two residues C-terminal to the Cys-X-X-His box is a Ser-X-Tyr motif, which is only conserved in proteins from organisms utilizing high potential quinones, such as plasto-, ubi- or calderiella quinone. In organisms utilizing menaquinone, one or both of these latter residues is frequently replaced by a nonhydroxylated amino acid residue (Fig. 1, indicated by arrows). Finally, the residues of the so-called Pro-loop, *i.e.*, Gly-Pro-Ala-Pro, located 11 to 27 residues C-terminally from the Cys-X-X-His box, are also reasonably conserved in the Rieske proteins.

A dendrogram constructed on the basis of this alignment is depicted in Fig. 2. The dendrogram shows that the two fundamental groups of proteins containing Rieske centers, *i.e.*, the Rieske and the Rieske-type proteins can be readily identified. The phylogenetic divisions within each group are discussed in detail below.

Rieske Proteins

In the phylogenetic tree constructed from the alignment of all proteins (Fig. 2), three major groups of Rieske proteins can be identified; (1) those belonging to the prokaryotic and eukaryotic cytochrome bc_1 complexes, (2) the cytochrome $b_6 f$ complexes from cyanobacteria

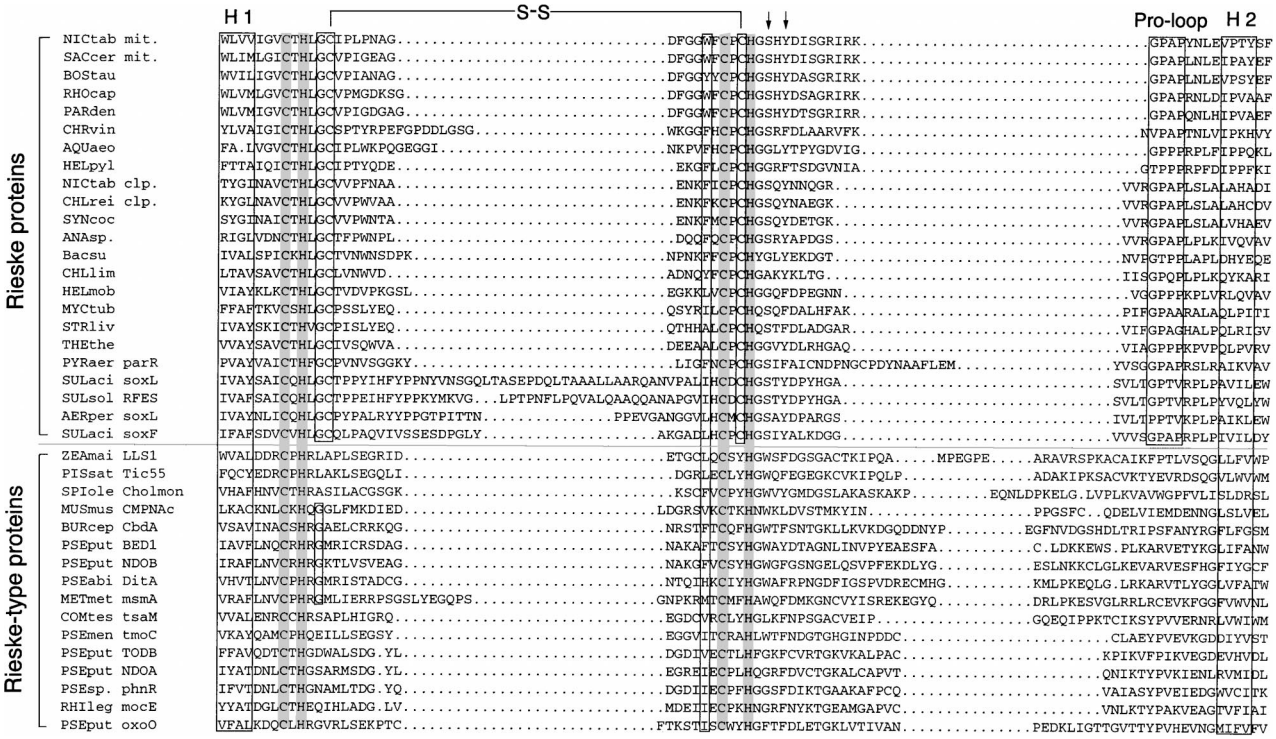

Fig. 1. Alignment of the Rieske sequence motifs and adjacent primary structure from representative members of the various Rieske and Rieske-type proteins. The Cys-X-His and Cys-X-X-His boxes are shaded in gray and the internal disulfide bridge is shown as -S-S-. The Ser-X-Tyr motif and variants thereof are indicated with arrows. Further sequences, such as the Pro-loop, the H1 and H2 motifs, and the hydrophobic amino acids two residues N-terminal from the Cys-X-X-His motif are boxed in.

and chloroplasts, as well as the sequences from *Bacillus* and *Chlorobium*, and (3) a heterogeneous group containing the archaeal Rieske proteins as well as proteins from bacterial species, such as *Thermus thermophilus*, *Streptomyces lividans*, and *Mycobacterium tuberculosis.* This threefold division is reflected well in a detailed alignment performed solely with the Rieske protein sequences (Fig. 3). However, in this latter alignment, the third group of sequences appears to be better resolved. Moreover, with the exception of the segregation of the sequences from *Thermus* and *Pyrobaculum*, which will be discussed below, the tree shown in Fig. 3 is compatible with previously published dendrograms (Henninger *et al.*, 1999; Schmidt *et al.*, 1996). Any apparent differences between these and the present dendrograms result from the fact that the former trees were unrooted, although in dendrograms, such as the ones published by Schmidt *et al.* (1996) and Castresana *et al.*, (1995), it is tempting to place the root at the lowest branching point. In contrast, the sequences of the Rieske-type proteins serve as the outgroup for the Rieske proteins in the dendrograms presented here (Figs. 2 and 3). The phylogenetic data used as the basis for Fig. 3 could be equally well represented by a dendrogram showing the *Sulfolobus* and *Aeropyrum* sequences as the lowest branch (not shown), comparable to previously published trees (Castresana *et al.*, 1995; Schmidt *et al.*, 1996). The representation shown in Fig. 3 was chosen since it reflects the rooting of this tree inferred in Fig. 2.

Rieske Proteins of Respiratory Cytochrome bc_1 Complexes

With the exception of the *Aquifex* Rieske protein, all sequences arranged within this group are from proteobacteria, or the closely related mitochondria (Olsen *et al.*, 1994). The cytochrome *bc* complexes of these organisms consist of a minimal structure of three redox active subunits, (1) the Rieske protein, (2) cytochrome *b*, and (3) cytochrome c_1. The genes encoding the bacterial complexes are typically arranged within an operon in the order shown here. Whereas the genes encoding the subunits of the eukaryotic complexes are scattered throughout the mitochondrial and the nuclear genomes, the mitochondrial Rieske proteins are generally encoded

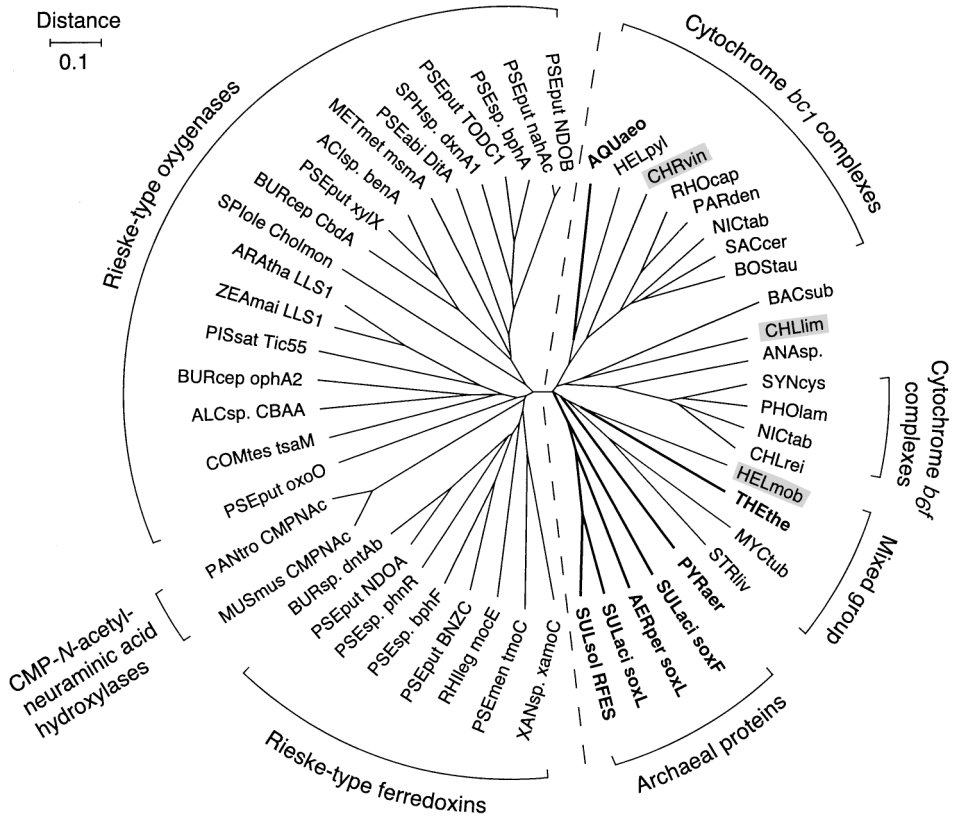

Fig. 2. Dendrogram constructed on the basis of the alignment of all Rieske and Rieske-type sequences listed in the section Materials and Methods. Hyperthermophilic species and their lineages are shown in bold type. Obligate anaerobic species are indicated by shading in gray.

by nuclear genes. Despite the modifications that were undoubtedly necessary to adapt these genes and the encoded proteins to the requirements of a eukaryotic transcription, translation, and protein translocation machinery, the mitochondrial Rieske proteins have retained a remarkable similarity to their proteobacteria-like ancestor. The branching pattern of this region of the dendrogram reflects the relatively close phylogenetic relationship between the mitochondria and the α subdivision of the proteobacteria as inferred from comparisons of rRNA sequences (Olsen *et al.*, 1994). The Rieske proteins of *Allochromatium* and *Helicobacter,* which belong to the γ- and ε-subdivisions of the proteobacteria, respectively, are clearly more distantly related to the α-proteobacterial and mitochondrial analogs. The positioning of the *Aquifex* sequence within this group is puzzling, since it contrasts with the phylogeny inferred from the rRNA sequences, according to which this genus forms the lowest branch at the base of the bacterial domain. However, this observation can be explained by lateral gene transfer as discussed by Schütz *et al.* (2000).

Rieske Proteins of the Photosynthetic Cytochrome b_6 f and Related Complexes

The chloroplast Rieske proteins of eukaryotic origin (pea, spinach, tobacco, *Volvox*, and *Chlamydomonas*) form a branch separate from the equivalent proteins of cyanobacteria, but apparently share a common ancestor, as predicted by the endosymbiosis hypothesis. This is also well reflected in the trees calculated on the basis of the rRNA sequences (Olsen *et al.*, 1994). Although the positioning of the sequence from *Anabena* sp. (PCC7120) appears to contradict to this tendency, closer inspection of the gene locus reveals that it is not part of a *petAC* operon as is the case for the other cyanobacterial *petC* genes found in the data base (data not shown). In contrast, the recently published *petC* gene from *Anabena variabilis* is part of a typical *petAC* operon and is accordingly positioned within the cluster formed by the other cyanobacterial sequences (Fig. 3). The unexpectedly low similarity of *Anabena* sp. (PCC7120) to the proteins from the closely related cyanobacteria, as well as the deviating genomic

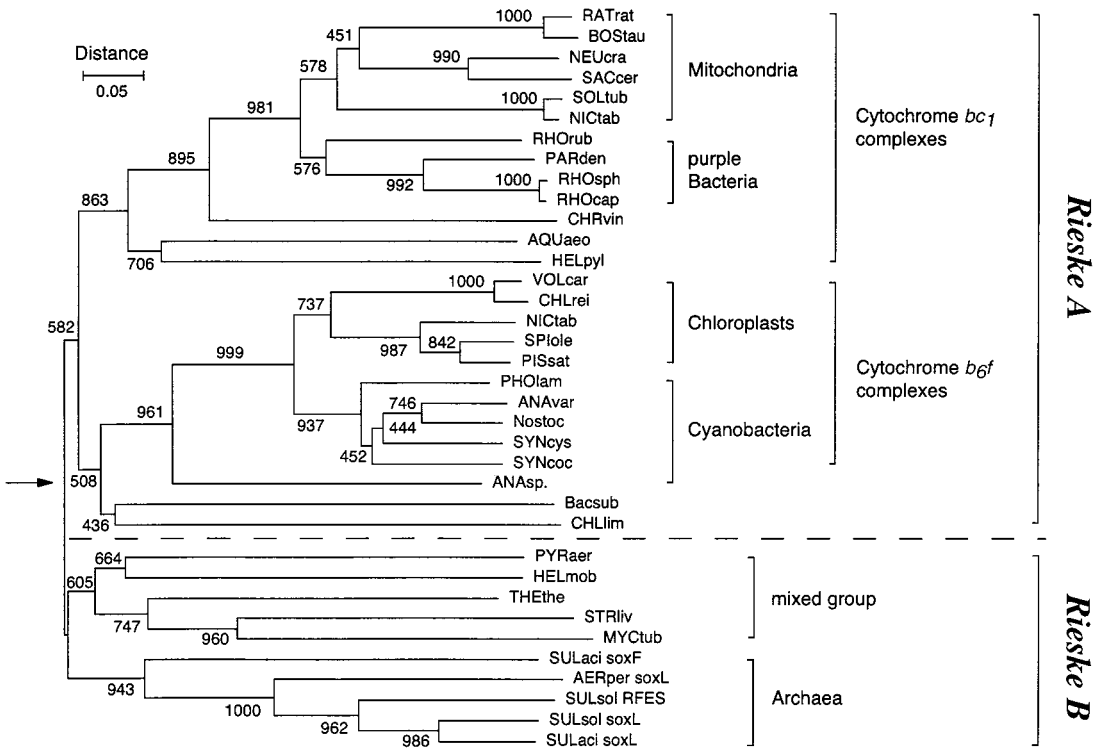

Fig. 3. Dendrogram constructed solely from the alignment of the Rieske protein sequences given in the section Materials and Methods. The bootstrap values (out of 1000 trials) are indicated in the figure. The arrow indicates the root of the tree as inferred from Fig. 2.

context of this sequence, raise questions as to whether the gene is actually expressed in *Anabena* or even if the protein is part of the cytochrome $b_6 f$ complex. The possibility that the gene was acquired from an organism more distantly related to the cyanobacteria should also be considered.

The sequences from *Bacillus*, belonging to the Gram-positive bacteria (Olsen *et al.*, 1994), and *Chlorobium*, a member of the Cytophaga/Flexibacter/Bacteroides group (Olsen *et al.*, 1994), form the lowest branch within the group containing the Rieske proteins of cytochrome $b_6 f$ complexes. However, the complexes isolated from these organisms display only limited (*Bacillus*) (Sone *et al.*, 1996) or very low (*Chlorobium*) (Schütz *et al.*, 1994) similarity to the $b_6 f$ complexes. Both organisms belong to rather different lineages and the positioning of these sequences is not in complete agreement with the phylogenetic tree based on comparison of 16 S rRNA sequences (Olsen *et al.*, 1994).

Archaeal and Bacterial Rieske Proteins

Archaeal as well as bacterial Rieske protein sequences cluster together on a third branch of the dendrogram shown in Fig. 3. These sequences can be further divided into a purely archaeal group which, in agreement with the rRNA-based tree (Olsen *et al.*, 1994), contains the sequences from the crenarchaea *Sulfolobus acidocaldarius*, *S. solfataricus*, and *Aeropyrum pernix*. The other sequences of the third group mainly originate from very distantly related organisms (Olsen *et al.*, 1994). *Mycobacterium* and *Streptomyces* are exceptions and both belong to the Actinomycetales.

The *Thermus* Rieske protein was grouped together with the *Bacillus* protein in a previously published dendrogram (Henninger *et al.*, 1999). However, according to current ideas on bacterial systematics, these two organisms are only distantly related, a situation which is clearly reflected by the low degree of similarity between their Rieske protein sequences depicted in Fig. 3. A second deviation from a previously calculated dendrogram (Henninger *et al.*, 1999) is the position of the *Pyrobaculum* protein. Like *Sulfolobus* and *Aeropyrum*, *Pyrobaculum* belongs to the Crenarchaeota, suggesting that its Rieske protein sequence ought to segregate within the cluster formed by the proteins from the former organisms. However, this could not be verified using any reasonable set of alignment parameters in this analysis.

Why Does the Rieske Protein Phylogeny Presented here Deviate from the rRNA–Based Phylogeny and Previously Published Dendrograms of the Rieske Proteins?

The dendrogram calculated for the Rieske proteins presented in this work exhibits several deviations from the phylogeny based on rRNA sequence comparisons. These are: (1) The positioning of the sequences from *Aquifex*, *Anabena* sp. (PCC7120), *Chlorobium*, *Bacillus*, and *Pyrobaculum* and, most significantly, (2) the appearance of the third, mixed group of archaeal and bacterial sequences.

Deviations between protein- and rRNA-based phylogenetic trees are frequently observed (Doolittle, 1999). One possible explanation is offered by the now widely discussed idea of a frequently occurring interspecies transfer of genetic information. Other explanations might include convergent evolution or an unusually strong divergence due to evolutionary pressure imposed by environmental conditions or components interacting with the protein. However, these effects do not convincingly explain the occurrence of the third branch consisting of the archaeal and bacterial sequences. Compared to previously published analyses (Schmidt *et al.*, 1996; Castresana *et al.*, 1995), the appearance of this branch can be attributed to two factors. The first is the introduction of several new sequences. This resulted in the positioning of the *Pyrobaculum* sequence at a distance from the other archaeal sequences, an observation that, in principle, could be explained by the effects mentioned above. The second factor is the rooting of the tree. Without the additional information derived from Fig. 2, the resulting dendrogram could still be arranged in a way very similar to the previously published dendrograms (Schmidt *et al.*, 1996; Castresana *et al.*, 1995), *i.e.*, with the cluster of the *Sulfolobus* and *Aeropyrum* sequences forming the deepest branch. Thus, the most significant change to the phylogenetic tree of the Rieske proteins is a direct consequence of our assumption that at least the [2Fe–2S] binding domains of the Rieske and the Rieske-type proteins are homologous, an assumption without which the calculation of the tree shown in Fig. 2 would be meaningless. However, a careful inspection of the alignment, together with the structural similarities (Link, 1999) observed between the Rieske proteins from the beef heart mitochondrial cytochrome bc_1 complex (Iwata *et al.*, 1996), the cytochrome $b_6 f$ complex of spinach chloroplasts (Carrell *et al.*, 1997), and even the Rieske-type protein naphthalene dioxygenase (Kauppi *et al.*, 1998) convinced us of the validity of this assumption. Consequently, the branching pattern of the dendrogram in Fig. 2 can be most simply explained by postulating an early duplication of an ancient Rieske protein gene, preceding the split into the lineages of the Archaea and the Bacteria. One of these genes (*Rieske A* in Fig. 3) is presumably the ancestor of the genes currently known in the mitochondria, chloroplasts, cyanobacteria, and purple bacteria, whereas the descendants of the other one (*Rieske B* in Fig. 3) are only known in the Archaea and the other Bacteria included in this analysis.

Rieske-Type Proteins

This class of proteins consists of the Rieske oxygenases and the Rieske ferredoxins. In the alignment shown in Fig. 2, the branching at the root of the dendrogram is rather indistinct. However, three main groups of Rieske-type proteins are recognizable, these being the ferredoxins and two groups of oxygenases, all of which seem to share a common ancestor separate from the Rieske proteins. The relationship between the various Rieske-type proteins is clearer and more complete in Fig. 4 and this dendrogram forms the basis of the following discussion.

Rieske Ferredoxins

Regarding the Rieske ferredoxins, the alignment clearly separates the ferredoxins into two main groups. The first encompasses the Rieske ferredoxins associated with the class IIB and class III oxygenase systems. This group also includes two ferredoxins belonging to incompletely characterized enzyme systems. The ferredoxin *mocE* from the *moc* gene cluster of *Rhizobium leguminosarum* may form a part of an electron transfer chain, coupling a potential FAD-reductase (*mocF*) with a putative μ-oxodi-iron oxygenase (*mocD*) (Bahar *et al.*, 1998). PSEsp.phnR belongs to an uncharacterized oxygenase system possibly functioning as a biphenyl or phenanthrene dioxygenase in *Pseudomonas* DJ77 (Kim and Park, 1998).

The second group of ferredoxins form part of the electron transfer chains associated with the enzymes toluene-4-monooxygenase from *Pseudomonas mendocina* (PSEmen tmoC) and alkene monooxygenase from *Xanthobacter* strain Py2 (XANsp. xamoC). Both enzyme systems draw electrons from NADH using an [FAD–2Fe–2S] reductase and transfer them via the Rieske-ferredoxins to the respective oxygenase components, which contain a μ-oxodi-iron center (Pikus *et al.*, 1996; Small and Ensign, 1997). A small (approximately 11 kDa), non-redox coupling protein is also involved in both systems. Interestingly, these enzymes are closely related to the soluble methane monooxygenase from *Methylococcus capsulatus* (Bath) (Lipscomb, 1994), with the exception that the latter enzyme system uses a Cys_4[2Fe–2S]-ferredoxin.

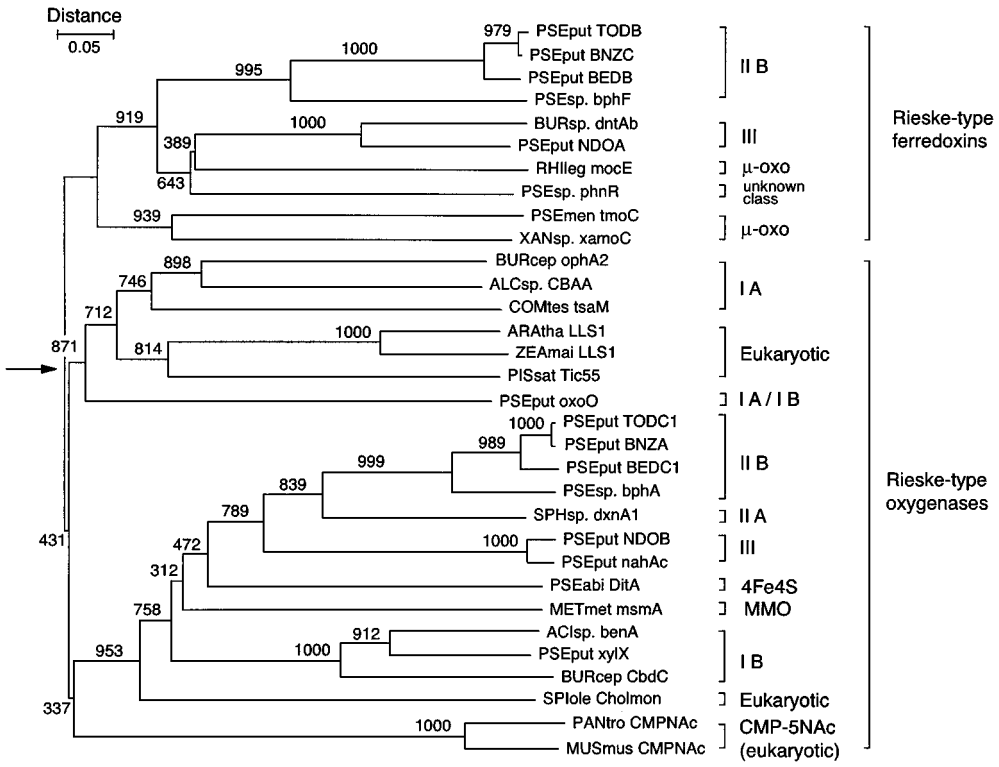

Fig. 4. Dendrogram constructed from the alignment of the Rieske-type protein sequences given in the section Materials and Methods. The bootstrap values (out of 1000 trials) are shown. The arrow indicates the possible root of this tree as discussed in the text. In addition to the usual classification (*i.e.*, type IA, IB, IIA, IIB, and III), the following abbreviations are used: μoxo, Rieske ferredoxins coupling with an oxygenase of the μ-oxodi-iron type; eukaryotic, eukaryotic oxygenase; 4Fe4S, oxygenase receiving electrons from a ferredoxin containing a [4Fe–4S] prosthetic group; MMO, oxygenase with a ferredoxin related to that occurring in methane monooxygenase.

The phylogenetic relationship between the type-III and -IIB Rieske ferredoxins found in this work was also observed by other workers in an alignment of a partially extended set of sequences (Sylvestre *et al.*, 1996). However, as far as the authors are aware, the alignment presented in this article is the first to show that diverse Rieske ferredoxin types functioning in a variety of oxygenase systems apparently share a common ancestor.

Rieske Oxygenases

Until recently, oxygenases possessing Rieske centers had only been reported in bacteria where they catalyze the oxidation of a variety of hydrophobic, mainly aromatic substances by the insertion of one or two hydroxyl groups (Mason and Cammack, 1992; Butler and Mason, 1997). The phylogenetic tree describing the relationships between the Rieske oxygenases is more complicated than that of the ferredoxins, presumably as a result of the fact that oxygenases exhibit a greater diversity in their substrate specificity, quaternary structure, and components of the respective redox chains. Nevertheless, the segregation of the various bacterial oxygenases belonging to classes IA, IB, IIA, IIB, and III is readily discernible in Fig. 4. Notable is the fact that the class IA oxygenases seem to form a lineage separate from the class IB–III oxygenases, as was suggested by Nakatsu *et al.* (1995). This is probably a reflection of the fact that the class IA enzymes possess a homomultimeric quaternary structure, while the type-IB to III oxygenases are generally composed of α and β subunit types, the α subunits bearing the catalytic center of these oxygenases. Moreover, within the main branch containing the latter oxygenase sequences, a clear separation of the class IB from the class IIA, IIB, and III oxygenases is visible. The overall branching pattern of this dendrogram (Fig. 4) compares well with previous studies performed on different sets of sequences and using different alignment algorithms (Nakatsu *et al.*, 1995; Werlen *et al.*, 1996; Sylvestre *et al.*, 1996; Rosche *et al.*, 1997), confirming the validity of the approach adopted in this work.

The segregation of bacterial oxygenases not included in the above classification scheme is also worthy of note. The enzyme 2-oxo-1,2-dihydroquinoline 8-monooxygenase from *Pseudomonas putida* (abbreviated PSEput oxoO), for example, functions with an electron transport chain characteristic of the type-IB oxygenases. Nevertheless the sequence of this enzyme is related, albeit distantly, to those of the type-IA oxygenases, probably reflecting the homomultimeric quaternary structure of this enzyme (Rosche *et al.*, 1997). Two further enzymes, which represent variants of the established classification, are also included in the alignment. The first, the α subunit of 7-oxodehydroabietic acid dioxygenase, which is involved in diterpenoid degradation in *Pseudomonas abietaniphila* BKME-9 (abbreviation PSEabi DitA1; Martin and Mohn, 1999), shows a weak homology to the type-IIB and -III oxygenases. In common with these enzymes, the 7-oxodehydroabietic acid dioxygenase is composed of α and β subunits. However, the ferredoxin in the associated redox chain is of the [4Fe–4S] or [3Fe–4S] type, and not the usual Rieske ferredoxin found in the class IIB or III oxygenases. The second unusual bacterial oxygenase dealt with here is the α subunit of methanesulfonic acid monooxygenase from *Methylosulfonomonas methylovora*. Although the sequence of this protein seems to be distantly related to the class IIA, IIB, and III oxygenases, as indicated by its possessing α and β subunits, it nevertheless segregates on a branch separate from the former oxygenase classes. This difference is underlined by the involvement of a ferredoxin component, which is similar to that present in methane monooxygenase (de Marco *et al.*, 1999).

An interesting new insight emerging from this analysis is the phylogeny of the Rieske-type proteins of eukaryotic origin, *i.e.*, choline monooxygenase, CMP-Neu5Ac hydroxylase, Tic55 and LIS1. These proteins segregate into the oxygenase branch of the Rieske-type proteins, suggesting that they are related to the bacterial Rieske oxygenases. In the case of choline monooxygenase, its segregation reflects its known catalytic function as an oxygenase (Rathinasabapathi *et al.*, 1997). In this alignment, choline monooxygenase shows no unequivocal phylogenetic relationship with any specific bacterial oxygenase class, although it does appear to more associated with the type-IB, -IIA, -IIB, and -III oxygenases. Based on the homodi- or trimeric structure suggested for the choline monooxygenase (Rathinasabapathi *et al.*, 1997), a distant relationship to these oxygenases may be postulated, since certain members of the class IB oxygenases have a homomultimeric structure (Butler and Mason, 1997).

The segregation of the eukaryotic sequences LIS1 and Tic55 are particularly interesting from an evolutionary point of view. All three proteins are distantly related to the class IA bacterial oxygenases. Although the catalytic functions of LIS1 and Tic55 are unknown, they were originally postulated to be Rieske oxygenases on the basis of their possessing a sequence motif homologous to the binding site of the mononuclear iron cofactor present in the bacterial oxygenases (Caliebe *et al.*, 1997; Gray *et al.*, 1997; Butler and Mason, 1997) which is the postulated site of oxygen activation (Twilfer *et al.*, 1985; Bill *et al.*, 1985). Their segregation into the group of the bacterial Rieske oxygenases, therefore, supports this assumption. From biochemical studies, there is no evidence for the presence of a β-type subunit in LIS1 or Tic55 (Caliebe *et al.*, 1997; Gray *et al.*, 1997), providing a further basis for their relatedness to the class IA oxygenases, which also consist of one subunit type.

Perhaps the most intriguing of the eukaryotic Rieske oxygenases is the mammalian CMP-Neu5Ac hydroxylase. Although this enzyme is clearly a Rieske-type protein, its positioning in the dendrogram in Fig. 4 suggests that it is only distantly related to the other Rieske oxygenases. Indeed, in the dendrogram constructed from a global alignment (Fig. 2), the CMP-Neu5Ac hydroxylase seems to be more related to the ferredoxins than to the oxygenases, although this conclusion must be treated with care, because of the diversity of protein types in this alignment. The proximity of the branch point of CMP-Neu5Ac hydroxylase to the root of the Rieske-type proteins suggests that it may have evolved along a lineage that separated early from the other eukaryotic oxygenases mentioned above. In spite of the apparent ancient origin of this enzyme, several lines of evidence are consistent with CMP-Neu5Ac hydroxylase being a relatively recent development in evolution. First, the product of the reaction catalyzed by this enzyme, the activated sugar-nucleotide cytidine monophosphate-*N*-glycolylneuraminic acid (CMP-Neu5Gc), is a precoursor for the glycoconjugate-bound sialic acid *N*-glycolylneuraminic acid (Neu5Gc). Thus far, this sialic acid has only been detected in deuterostomes, a lineage that encompasses animals from the echinoderms (*e.g.*, starfish and sea urchins) through to the mammals (Warren, 1963; Schauer *et al.*, 1999). Although sialic acid-containing glycoconjugates do occur in bacteria, they are restricted to the nonhydroxylated form *N*-acetylneuraminic acid and are only found in a limited number of species (Schauer *et al.*, 1995).

A further interesting characteristic of this enzyme is its subcellular localization. Whereas the plant oxygenases choline monooxygenase and Tic55 described above are associated with chloroplasts, CMP-Neu5Ac hydroxylase is cytosolic (Schneckenburger *et al.*, 1994; Schlenzka *et al.*, 1994; Kawano *et al.*, 1994). Whether the association of

the plant oxygenases with an organelle proposed to be of prokaryotic origin has any relevance to their apparent phylogenetic relationship with the bacterial enzymes is open to debate. It is, however, tempting to speculate that the cytosolic location of the CMP-Neu5Ac hydroxylase is a further manifestation of its rather unexpected phylogeny.

The monomeric nature of the CMP-Neu5Ac hydroxylase has potentially far-reaching implications for the structure and mechanism of this enzyme (Schneckenburger *et al.*, 1994; Schlenzka *et al.*, 1994; Kawano *et al.*, 1994). From the recently solved three-dimensional structure of the naphthalene dioxygenase from *Pseudomonas putida* (Kauppi *et al.*, 1998), it is clear that the quaternary structure of this enzyme is fundamental to its function. This molecule has a mushroomlike structure, the three α subunits forming the head and the β subunits the stalk of the mushroom. While the β subunits may be involved in influencing the substrate specificity (Hurtubise *et al.*, 1998), the α subunits possess the catalytic Rieske and mononuclear iron centers necessary for activity. From the geometry of these centers in the crystal structure, it has emerged that the mononuclear iron atom on one subunit is situated such that it could only receive electrons from the Rieske center on the adjacent subunit, strongly suggesting intersubunit cooperation in catalysis (Kauppi *et al.*, 1998). Since all bacterial Rieske oxygenases so far reported consist of several subunits, ranging from the type-IA oxygenases, which are generally homodimers, -trimers, or -tetramers (Mason and Cammack, 1992; Butler and Mason, 1997), to the other oxygenases, which generally seem to have an $\alpha_3\beta_3$ quaternary structure, this mode of electron transfer within the molecule may be generally applicable in the bacterial oxygenases. Whether a similar mechanism occurs in the oxygenases of plant origin is unknown. Choline monooxygenase does have a di- or trimeric structure, raising the possibility of such a mechanism (Rathinasabapathi *et al.*, 1997). As yet, the subunit composition of the other hypothetical plant oxygenases Tic55 and LIS1 is unknown. Evidently, the monomeric structure of CMP-Neu5Ac hydroxylase rules out intersubunit electron transfer being involved in its catalytic mechanism.

One further feature of CMP-Neu5Ac hydroxylase, making it unique among the Rieske oxygenases, is its electron donor. While bacterial Rieske oxygenases (Butler and Mason, 1997) and the plant choline monooxygenase (Rathinasabapathi *et al.*, 1997) are known to receive electrons from either a ferredoxin or, in the case of the class I bacterial oxygenases, a [2Fe-2S] iron-sulfur flavoprotein reductase possessing a ferredoxin-type domain, CMP-Neu5Ac hydroxylase receives its reducing equivalents from cytochrome b_5 (Kozutsumi *et al.*, 1990; Shaw *et al.*, 1994). Cytochrome b_5 is a membrane-bound electron donor which, in conjunction with an NADH-dependent flavoprotein reductase, supports the activity of a number of enzymes associated with the cytosolic face of the endoplasmic reticulum, including lipid desaturases and certain forms of cytochrome P450 (Vergères and Waskell, 1995). Although this general purpose electron donor is ubiquitous among the eukaryotes a prokaryotic cytochrome b_5-like protein of unknown function was recently reported in *Ectothiorhodospira vacuolata* (Kostanjevecki *et al.*, 1999). This raises the possibility that the ancestor of CMP-Neu5Ac hydroxylase already existed at the separation of the prokaryotes and eukaryotes, estimated to have occurred about 2×10^9 years ago (Doolittle *et al.*, 1996), a suggestion consistent with the very deep branch point observed for the hydroxylase in the dendrograms (Figs. 3 and 4). To date no sequences even remotely related to the CMP-Neu5Ac hydroxylase can be found in the GenBank so that the nature of the evolutionary precursors of this enzyme remains a mystery. One candidate might be the UDP-*N*-acetylmuramic acid hydroxylase, a sugar nucleotide oxygenase postulated to be responsible for the biosynthesis of *N*-glycolylmuramic acid present in the cell wall of certain bacteria (Gateau *et al.*, 1976). This suggestion is, however, highly speculative since the enzyme has been neither purified nor cloned and sequenced.

Despite these uncertainties, the significant functional and structural differences between CMP-Neu5Ac hydroxylase and the remaining Rieske oxygenases provide a basis for the proposed separate evolutionary lineage of this enzyme.

Functional Significance of the Phylogeny of Proteins Possessing a Rieske [2Fe–2S] Center

Since the sequence both within and surrounding the Rieske motif is the common factor in the alignments presented here, it is this region that dominates the resulting phylogeny. These relationships are presumably the result of functional and structural factors (Link, 1999). Although the Rieske centers of the various proteins discussed in this article all mediate a one-electron transfer reaction, the nature of the electron donors and acceptors are variable, depending on the type of protein in question. Table I gives a brief overview of the possible electron donor–acceptor combinations currently known for the various proteins with a Rieske cluster. For optimal functioning, the redox potential of the Rieske center in question must be poised at a value dependent on the nature of the redox reaction partners with which it must interact. Although the liganding of

Table I. Summary of Proteins and Cofactors Serving as Electron Donors and Acceptors for Rieske Centers

Rieske protein type	Electron donor	Electron acceptor
A. Rieske proteins		
Cytochrome bc_1	Reduced quinones	Cytochrome c_1
Cytochrome $b_6 f$	Reduced plastoquinone	Cytochrome f
Archaeal Rieske proteins	Reduced quinones	Not known
Bacillus-type Rieske proteins	Reduced menaquinone	Cytochrome c'
B. Rieske-type proteins		
Bacterial Rieske ferredoxins		
Ferredoxins of bacterial type-IIB and -III Rieske oxygenases	Type IIB: FAD-reductase Type III: FAD-cys$_4$[2Fe–2S]-reductase	Rieske center of oxygenase Rieske center of oxygenase
Rieske ferredoxins of bacterial toluene-4-monooxygenase and alkene monooxygenase	FAD-cys$_4$[2Fe–2S]-reductase	μ-Oxodi-iron center of oxygenase components
Rieske ferredoxin of oxygenase in rhizopine degradation (MocE)	FAD-reductase	μ-Oxodi-iron center of oxygenase component
Bacterial Rieske oxygenases		
Type IA	FMN-cys$_4$[2Fe–2S]-reductase	Mononuclear iron center
Type IB	FAD-cys$_4$[2Fe–2S]-reductase	Mononuclear iron center
Type IIA	Cys$_4$[2Fe–2S]-ferredoxin	Mononuclear iron center
Type IIB	Rieske ferredoxin	Mononuclear iron center
Type III	Rieske ferredoxin	Mononuclear iron center
Methane sulfonate monooxgenase	Cys$_4$[2Fe–2S]-ferredoxin	Mononuclear iron center
Oxygenase of diterpenoid degradation (DitA1)	Cys$_4$[4Fe–4S]-ferredoxin	Mononuclear iron center
Eukaryotic Rieske oxygenases		
Plant choline monooxygenase	Cys$_4$[2Fe–2S]-ferredoxin	Mononuclear iron center[a]
Tic55/LIS1	Unknown	Mononuclear iron center[a]
CMP-Neu5Ac hydroxylase	Cytochrome b_5	Mononuclear iron center[b]

[a] Postulated on the basis of sequence comparisons, although not unequivocally proved in structural studies.
[b] Postulated on the basis of enzymic studies and weak sequence homologies.

the redox-active iron ions in all Rieske centers is identical, the surrounding milieu, which is largely dictated by the spatially adjacent amino acid side chains, may have a profound influence on the redox properties of these [2Fe–2S] centers. An examination of the three-dimensional structures of the Rieske center-containing proteins so far solved reveals that the amino acid residues both within and immediately surrounding the Rieske sequence motif make a significant contribution to the pocket in which the [2Fe–2S] cluster is situated (Link, 1999). This is most evident in the Rieske domains of cytochromes bc_1 (Iwata *et al.*, 1996) and $b_6 f$ (Carrell *et al.*, 1997) where the environment of the iron–sulfur center is dominated by the amino acid residues of this primary structure region. In the α subunit of naphthalene dioxygenase, however, the additional close contact of the Rieske center with the catalytic site of an adjacent subunit may also influence the redox chemistry of the cluster (Kauppi *et al.*, 1998).

Apart from directly liganding the [2Fe–2S] cluster via the His and Cys residues, a number of main-chain and side-chain atoms both within and adjacent to the Rieske motif also form a second ligand shell which interacts not only with the iron-binding side chains, but also with the sulfide ions in the cluster. Mutational studies on residues within the Rieske motif of cytochrome bc_1 from *Paracoccus denitrificans* (Schröter *et al.*, 1998) and *Rhodobacter capsulatus* (Liebl *et al.*, 1997) underline their importance not only in determining the redox potential of a Rieske center, but also for its stability and interaction with substrates. Furthermore, the overall polarity of the cluster's environment, which is also influenced by spatially adjacent side chains may also have some bearing on its chemical properties.

These considerations all underscore the importance of the amino acid residues of the Rieske motif and thus provide a functional basis to explain the phylogenetic relationships between Rieske [2Fe–2S] proteins.

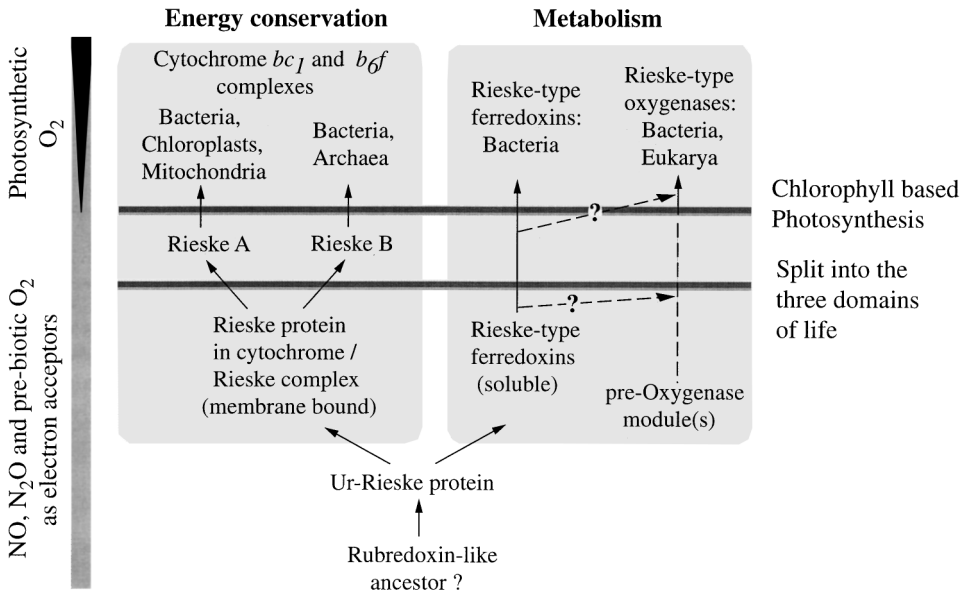

Fig. 5. Possible evolutionary pathway leading to the various groups of proteins in the Rieske protein superfamily.

Relevance of Phylogeny to the Evolution of Rieske Redox Proteins

This study also demonstrates that all sequences currently known can be clearly allocated to one of these two protein families. These form a protein superfamily whose members are involved in redox events of energy conservation and metabolism. The fact that proteins possessing Rieske centers are found in the Eukarya, Bacteria, and the Archaea, suggests that the Rieske [2Fe–2S] center, as such, appeared before the separation of the three domains of life. Since Rieske-type proteins are known in Eukarya and the Bacteria and the Rieske proteins are found in all three domains, although their occurrence in the Eukarya is restricted to the mitochondria and chloroplasts, it can be concluded that both both Rieske and Rieske-type proteins were present in the last common ancestor of all living organisms. Furthermore, the close phylogenetic relationship between the Rieske proteins of the bacterial cytochrome b_6f or bc_1 complexes and the corresponding eukaryotic proteins, lends further weight to the theory of the endosymbiotic origin of mitochondria and chloroplasts.

To what extent endosymbiotic transfer of Rieske-type proteins into eukaryotes occurred cannot be deduced with any certainty. As discussed above, the association of the Tic55 and choline monooxygenase with chloroplasts hints at an endosymbiotic origin of these proteins. However, the deep branch points of these proteins from their bacterial homologs could equally well have occurred by progressive evolution from a common ancestor without invoking endosymbiosis.

The Early Evolution of the Rieske Protein Superfamily under Anoxic Conditions and the Emergence of Present-Day Representatives

The results of this analysis and other studies provide a basis for speculation on the events leading to the appearance of the fundamental forms of the Rieske protein superfamily and their subsequent evolution into the currently known proteins. The scheme shown in Fig. 5 depicts a possible model for this evolutionary pathway.

The presence of Rieske proteins in organisms of the three accepted domains of life, the Bacteria, Archaea, and the Eukarya, suggests that the early evolution of the Rieske center occurred under the largely anoxic conditions prevailing in the time between the appearance of the first cellular life forms some 3.5×10^9 years ago and the divergence of the prokayotes from the eukaryotes and Archaea about 2×10^9 years ago (Doolittle *et al.*, 1996).

Similarities in the topology of the simple rubredoxin iron center with the [2Fe–2S] cluster-binding subdomains of the Rieske proteins from cytochromes b_6f and bc_1 as well as of naphthalene dioxygenase, led to the suggestion that the former protein may be an ancestor of the Rieske center-binding domain (Link, 1999). From this, one could postulate the appearance of an archaic "Ur–Rieske protein," which might have been a component in some sort of redox chain. The fact that Rieske [2Fe–2S]

centers generally possess a higher redox potential than the corresponding Cys_4 clusters may have allowed them to occupy unique niches in early metabolism, so contributing to their persistence in evolution.

In a decisive step, the duplication of this protein and subsequent evolution possibly gave rise to the basic Rieske and Rieske-type proteins. In the former case, the Rieske domain would have gained a membrane anchor and was incorporated into a primitive cytochrome–Rieske complex. A possible duplication of this forerunner of the Rieske proteins prior to the separation of the three domains of life may have given rise to the type-A and -B Rieske proteins, discussed above (see also Fig. 3). Even at this point, it is probable that anaerobic conditions were still prevailing. The existence of bacterial cytochrome bc_1 complexes involved in the anaerobic respiratory chains, for example, in the reduction of nitric oxide (Itoh *et al.*, 1989; Carr *et al.*, 1989; Zumft, 1997), suggests that cytochrome bc_1-type complexes could have existed before the appearance of oxygen.

The advent of oxygenic photosynthesis presumably provided an impetus for the further evolution of the Rieske proteins within their respective complexes. In the case of the type-B Rieske proteins, these probably persisted in the "mixed-group" bacteria and the Archaea, while the type-A Rieske protein–cytochrome complexes evolved into the cytochrome bc_1 complexes of present-day bacterial aerobic respiratory chains. After endosymbiotic migration, the Rieske proteins of the bacterial cytochrome bc_1 complex and the cyanobacterial cytochrome $b_6 f$ complex became integral components of eukaryotic mitochondria and chloroplasts, respectively.

The second evolutionary lineage leading from the duplication of the postulated "Ur–Rieske protein" probably gave rise to the present-day Rieske-type proteins. Since the Rieske ferredoxins are structurally and functionally very simple, they probably resemble the ancestral Rieske-type protein more closely than the oxygenases. Indeed, an examination of the alignment of several Rieske-type protein sequences (Fig. 6) suggests that the oxygenases may

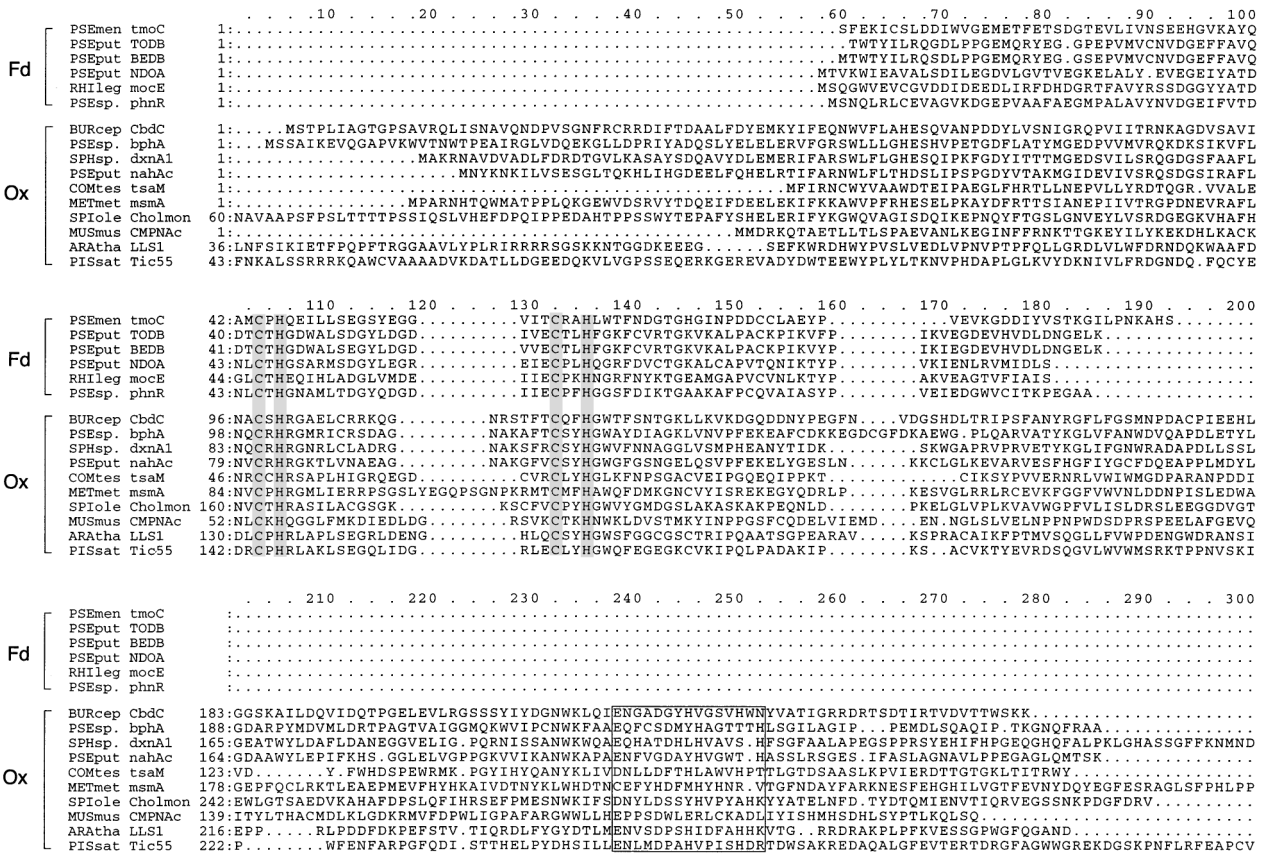

Fig. 6. Partial alignment of representative Rieske-type protein sequences showing a Rieske ferredoxin-like motif at the N-terminus of the Rieske oxygenases. The boxed region contains the residues either known to be or, in the case of CMP-Neu5Ac hydroxylase, postulated to be responsible for the coordination of the mononuclear iron center. Fd, Rieske-type ferredoxins; Ox, Rieske-type oxygenases.

have evolved by the fusion of a Rieske ferredoxin gene with a gene encoding a domain for substrate recognition and oxygen activation. The dendrogram in Fig. 4 suggests that the Rieske oxygenases may be derived from one common ancestor. However, this conclusion must be treated with caution, because of the deep branching points in this region. These pre-oxygenase modules, indicated in Fig. 5, may represent proteins containing the highly conserved mononuclear iron-binding sequence found in the bacterial and plant Rieske oxygenases. Although this sequence, as such, cannot be found in CMP-Neu5Ac hydroxylase, candidate motifs have been identified (Schlenzka *et al.*, 1996). This discrepancy might have arisen from extensive mutation of a typical mononuclear iron-binding motif, or by fusion of a Rieske ferredoxin with an oxygenase module different from that of the above oxygenases. Both mechanisms might explain the unique phylogeny of CMP-Neu5Ac hydroxylase within the Rieske-type proteins.

As yet, one can only speculate on the time point of this postulated gene fusion event. Because of endosymbiotic transfer, the eukaryotic Rieske proteins of energy metabolism are closely related to their bacterial counterparts (Fig. 3). The eukaryotic Rieske-type oxygenases are, however, more distantly related to their bacterial relatives, suggesting that they were not acquired by endosymbiosis. Nevertheless, their distribution indicates that the last common ancestor of all domains of life, which existed about 2×10^9 years ago, may well have possessed a Rieske oxygenase-like protein. Since oxygen is a substrate for the Rieske oxygenases, it is not unreasonable to suggest that they appeared either after or during the increase in the level of this gas in the atmosphere. According to recent estimates, this occurred between $1.85–2.5 \times 10^9$ years ago (Deutsch *et al.*, 1998; Kerr, 1999), possibly coinciding with the separation of the three domains of life. However, since oxygen-evolving (*i.e.*, chlorophyll-based) photosynthesis is solely known in the bacteria and derived endosymbionts, we assume that this capability evolved subsequently to the separation of the Bacteria and the Archaea, as depicted in Fig. 5. This raises the distinct possibility that Rieske-type proteins also occur in the Archaea. Although no archaeal Rieske-type proteins have been unequivocally demonstrated, sulredoxin, recently described in *Sulfolobus* sp. strain 7 as a small (11-kDa), soluble protein with electrochemical and spectroscopic properties reminiscent of a Rieske-type ferredoxin, is a potential candidate (Iwasaki *et al.*, 1995, 1996). No putative archaeal Rieske oxygenases are known. Clearly, the identification of archaeal Rieske-type proteins would validate this hypothetical evolutionary scheme. Because of differential expression and difficulties in protein purification, the classical approaches used in uncovering such proteins might be meaningfully augmented by computer searches of the archaeal genomes (*e.g.*, *Sulfolobus* or *Thermoplasma*), which are currently being studied.

ACKNOWLEDGMENTS

The authors wish to thank Prof. Günter Schäfer for his critical reading of this manuscript. We would also like to express our gratitude to Dr. Wolfgang Nitschke (Marseille) and Dirk Schneider (Bochum) for stimulating discussions.

REFERENCES

Alm, R. A., Ling, L. S., Moir, D. T., King, B. L., Brown, E. D., Doig, P. C., Smith, D. R., Noonan, B., Guild, B. C., de Jonge, B. L., Carmel, G.,Tummino, P. J., Caruso, A., Uria-Nickelsen, M., Mills, D. M., Ives, C., Gibson, R., Merberg, D., Mills, S. D., Jiang, Q., Taylor, D. E., Vovis, G. F., and Trust, T. J. (1999). *Nature* (*London*) **397**, 176–180.

Armengaud, J., Happe, B., and Timmis, K. N. (1998). *J. Bacteriol.* **180**, 3954–3966.

Arnold, M., (2000). Direct submission to GenBank.

Asturias, J. A., Diaz, E., and Timmis, K. N. (1995). *Gene* **156**, 11–18.

Bahar, M., de Majnik, J., Wexler, M., Fry, J., Poole, P. S., and Murphy, P. J. (1998). *Mol. Plant-Microbe Interactions* **11**, 1057–1068.

Batie, C. J., Ballou, D. P., and Correll, C. C. (1991). In *Chemistry and Biochemistry of Flavoenzymes* (Müller, F., ed.), CRC Press, Boca Raton, IL, pp. 544–556.

Beckmann, J. D., Ljungdahl, P. O., Lopez, J. L., and Trumpower, B. L. (1987). *J. Biol. Chem.* **262**, 8901–8909.

Beinert, H., Holm, R. H., and Münck, E. (1997). *Science* **277**, 653–659.

Bill, E., Bernhardt, F.-H., Trautwein, A. X., and Winkler, H. (1985). *Eur. J. Biochem.* **147**, 177–182.

Blöchl, E., Keller, M., Wächtershäuser, G., and Stetter, K.-O. (1992). *Proc. Natl. Acad. Sci. USA* **89**, 8117–8120.

Brandt, U., Yu, L., Yu, C. A., and Trumpower, B. L. (1993). *J. Biol. Chem.* **268**, 8387–8390.

Butler, C. S. and Mason, J. R. (1997). *Advan. Microbial Physiol.* **38**, 47–83.

Caliebe, A., Grimm, R., Kaiser, G., Lübeck, J., Soll., J., and Heins, L. (1997). *EMBO J.* **16**, 7342–7350.

Cammack, R. (1992). *Advan. Inorg. Chem.* **38**, 281–321.

Carr, G. J., Page, M. D., and Ferguson, S. J. (1989). *Eur. J. Biochem.* **179**, 683–692.

Carrell, C. J., Zhang, H., Cramer, W. A., and Smith, J. L. (1997). *Structure* **5**, 1613–1625.

Castresana, J., Lübben, M., and Saraste, M. (1995). *J. Mol. Biol.* **250**, 202–210.

Chan, M. K., Kim, J., and Rees, D. C. (1993). *Science* **260**, 792–794.

Chang, H. K. and Zylstra, G. J. (1998). *J. Bacteriol.* **180**, 6529–6537.

Chen, Y. L., Dincturk, H. B., Qin, H., and Knaff, D. B. (1997). Direct submission to GenBank.

Chou, H. H., Takematsu, H., Diaz, S., Iber, J., Nickerson, E., Wright, K. L., Muchmore, E. A., Nelson, D. L., Warren, S. T., and Varki, A. (1998). *Proc. Natl. Acad. Sci. USA* **95**, 11751–11756.

Cole, S. T., Brosch, R., Parkhill, J., Garnier, T., Churcher, C., Harris, D., Gordon, S. V., Eiglmeier, K., Gas, S., Barry, C. E. III., Tekaia, F., Badcock, K., Basham, D., Brown, D., Chillingworth, T., Connor, R., Davies, R., Devlin, K., Feltwell, T., Gentles, S., Hamlin, N., Holroyd, S., Hornsby, T., Jagels, K., Krogh, A., McLean, J., Moule,

S., Murphy, L., Oliver, S., Osborne, J., Quail, M. A., Rajandream, M. A., Rogers, J., Rutter, S., Seeger, K., Skelton, S., Squares, S., Sqares, R., Sulston, J. E., Taylor, K., Whitehead, S., and Barrell, B. G. (1998). *Nature (London)* **393**, 537–544.
Davidson, E. and Daldal, F. (1987a). *J. Mol. Biol.* **195**, 13–24.
Davidson, E. and Daldal, F. (1987b). *J. Mol. Biol.* **195**, 25–29.
Davidson, E., Ohnishi, T., Atta-Asafo-Adjei, E., and Daldal, F. (1992). *Biochemistry* **31**, 3342–3351.
Deckert, G., Warren, P. V., Gaasterland, T., Young, W. G., Lenox, A. L., Graham, D. E., Overbeek, R., Snead, M. A., Keller, M., Aujay, M., Huber, R., Feldman, R. A., Short, J. M., Olsen, G. J., and Swanson, R.V. (1998). *Nature (London)* **392**, 353–358.
de Marco, P., Moradas-Ferreira, P., Higgins, T. P., McDonald, I., Kenna, E. M., and Murrell, J. C. (1999). *J. Bacteriol.* **181**, 2244–2251.
Deutsch, A., Greshake, A., Pesonen, L. J., and Pihlaja, P. (1998). *Nature (London)* **395**, 146–148.
de Vitry, (1994). *J. Biol. Chem.* **269**, 7603–7609.
Doolittle, R. F., Feng, D., Tsang, S., Cho, G., and Little E. (1996). *Science* **271**, 470–477.
Doolittle, W. F. (1999). *Science* **284**, 2124–2128.
Emmermann, M., Clericus, M., Braun, H. P., Mozo, T., Heins, L., Kruft, V., and Schmitz, U. K. (1994). *Plant Mol. Biol.* **25**, 271–281.
Erickson, B. D. and Mondello, F. J. (1992). *J. Bacteriol.* **174**, 2903–2912.
Gateau, O., Bordet, C., and Michel, G. (1976). *Biochim. Biophys. Acta* **421**, 395–405.
Gatti, D. L., Tarr, G., Fee, J. A., and Ackerman, S. H. (1997). Direct submission to GenBank.
Gray, J., Close, P., Briggs, S. P., and Johal, G. S. (1997). *Cell* **89**, 25–31.
Gurbiel, R. J., Ohnishi, T., Robertson, D. E., Daldal, F., and Hoffman, B. M. (1991). *Biochemistry* **30**, 11579–11584.
Gurbiel, R. J., Doan, P. E., Gassner, G. T., Macke, T. J., Case, D. A., Ohnishi, T., Fee, J. A., Ballou, D. P., and Hoffmann, B. M. (1996). *Biochemistry* **35**, 7834–7845.
Haak, B., Fetzner, S., and Lingens, F. (1995). *J. Bacteriol.* **177**, 667–675.
Harayama, S., Rekik, M., Bairoch, A., Neidle, E. L., and Ornston, L. N. (1991). *J. Bacteriol.* **173**, 7540–7548.
Harnisch, U., Weiss, H., and Sebald, W. (1985). *Eur. J. Biochem.* **149**, 95–99.
Henninger, T., Anemüller, S., Fitz-Gibbon, S., Miller, J. H., Schäfer, G., and Schmidt, C. L. (1999). *J. Bioenerg. Biomembr.* **31**, 119–128.
Huang, J. T., Struck, F., Matzinger, D. F., and Levings, C. S., III (1991). *Proc. Natl. Acad. Sci. USA* **88**, 10716–10720.
Hurtubise, Y., Barriault, D., and Sylvestre (1998). *J. Bacteriol.* **180**, 5828–5835.
Irie, S., Doi, S., Yorifuji, T., Takagi, M., and Yano, K. (1987). *J. Bacteriol.* **169**, 5174–5179.
Itoh, M. Mizukami, S., Matsuura, K., and Satoh, T. (1989). *FEBS Lett.* **244**, 81–84.
Iwasaki, T., Isogai, Y., Iizuka, T., and Oshima, T. (1995). *J. Bacteriol.* **177**, 2576–2582.
Iwasaki, T., Imai, T., Urushima, A., and Oshima, T. (1996). *J. Biol. Chem.* **271**, 27659–27663.
Iwata, S., Saynovits, M., Link, T. A., and Michel, H. (1996). *Structure* **4**, 567–579.
Junker, F., Kiewitz, R., and Cook, A. M. (1997). *J. Bacteriol.* **179**, 919–927.
Kallas, T., Spiller, S., and Malkin, R. (1988). *Proc. Natl. Acad. Sci. USA* **85**, 5794–5798.
Kauppi, B., Lee, K., Carredano, E., Parales, R. E., Gibson, D. T., Eklund, H., and Ramaswamy, S. (1998). *Structure* **6**, 571–586.
Kawano, T., Kozutsumi, Y., Kawasaki, T., and Suzuki, A. (1994). *J. Biol. Chem.* **269**, 9024–9029.
Kawarabayasi, Y., Hino, Y., Horikawa, H., Yamazaki, S., Haikawa, Y., Jin-no, K., Takahashi, M., Sekine, M., Baba, S., Ankai, A., Kosugi, H., Hosoyama, A., Fukui, S., Nagai, Y., Nishijima, K., Nakazawa, H., Takamiya, M., Masuda, S., Funahashi, T., Tanaka, T., Kudoh, Y., Yamazaki, J., Kushida, N., Oguchi, A., Aoki, K., Kubota, K., Nakamura, Y., Nomura, N., Sako, Y., and Kikuchi, H. (1999). *DNA Res.* **6**, 83–101.
Kerr, R. A. (1999). *Science* **284**, 2111–2113.
Kim, S. and Park, Y.-C. (1998). Direct submission to GenBank.
Kostanjevecki, V., Leys, D., van Driessche, G., Meyer, T. E., Cusanovich, M. A., Fischer, U., Guisez, Y., and van Beeumen, J. (1999). *J. Biol. Chem.* **274**, 35614–35620.
Kozutsumi, Y, Kawano, T., Yamakawa, T., and Suzuki, A. (1990). *J. Biochem.* **108**, 704–706.
Kuila, D., Schoonover, J. R., Dyer, R. B., Batie, C. J., Ballou, D. P., Fee, J. A., and Woodruff, W. H. (1992). *Biochim. Biophys. Acta* **1140**, 175–183.
Kurkela, S., Lehvaslaiho, H., Palva, E. T., and Teeri, T. H. (1988). *Gene* **73**, 355–362.
Kurowski, B. and Ludwig, B. (1987). *J. Biol. Chem.* **262**, 13805–13811.
Liebl, U., Sled, V., Brasseur, G., Ohnishi, T., and Daldal, F. (1997). *Biochemistry* **36**, 11675–11684.
Link, T. A. (1999). *Advan. Inorg. Chem.* **47**, 83–157.
Lipscomb, J. D. (1994). *Annu. Rev. Microbiol.* **48**, 371–399.
Martin, V. J. J. and Mohn, W. W. (1999). *J. Bacteriol.* **181**, 2675–2682.
Mason, J. R. and Cammack, R. (1992). *Annu. Rev. Microbiol.* **46**, 277–305.
Majewski, C. and Trebst, A. (1990). *Mol. Gen. Genet.* **224**, 373–382.
Mayes, S. R. and Barber, J. (1991). *Plant Mol. Biol.* **17**, 289–293.
Meissner, M., Stark, K., Cresnar, B., Kirk, D. L., and Schmitt, R. (1998). Direct submission to GenBank.
Nakatsu, C. H., Straus, N. A., and Wyndham, C. (1995). *Microbiology* **141**, 485–495.
Neidle, E. L., Hartnett, C., Ornston, N., Bairoch, A., Rekik, M., and Harayama, S. (1991). *J. Bacteriol.* **173**, 5385–5395.
Nishikimi, M., Hosokawa, Y., Toda, H., Suzuki, H., and Ozawa, T. (1989). *Biochem. Biophys. Res. Commun.* **159**, 19–25.
Olsen, G. J., Woese, C. R., and Overbeek, R. (1994). *J. Bacteriol.* **176**, 1–6.
Palmer, G. (1975). The Enzymes, 3rd ed., Vol. XII, part B, pp. 1–56.
Palomares, R., Herrmann, R. G., and Oelmueller, R. (1991). *J. Photochem. Photobiol.* **B11**, 151–162.
Parro, V. and Mellado, R. P. (1998). Direct submission to Genbank.
Perriére, G. and Gouy, M. (1996). *Biochimie* **78**, 364–369.
Pikus, J. D., Studts, J. M., Achim, C., Kauffmann, K. E., Münck, E., Steffan, R. J., McClay, K., and Fox, B. G. (1996). *Biochemistry* **35**, 9106–9119.
Ramaswamy, K. S., Carrasco, C. D., Fatma, T., and Golden, J. W. (1996). Direct submission to GenBank.
Rathinasabapathi, B., Burnet, M., Russell, B., Gage, D. A., Liao, P.-C., Nye, G. J., Scott, P., Golbeck, J. H., and Hanson, A. D. (1997). *Proc. Natl. Acad. Sci. USA* **94**, 3454–3458.
Riedel, A., Rutherford, A. W., Hauska, G., Müller, A., and Nitschke, W. (1991). *J. Biol. Chem.* **266**, 17838–17844.
Riedel, A., Fetzner, S., Rampp, M., Lingens, F., Liebl, U., Zimmermann, J.-L., and Nitschke, W. (1995). *J. Biol. Chem.* **270**, 30869–30873.
Rieske, J. S., MacLennan, D. H., and Coleman, R. (1964). *Biochem. Biophys. Res. Commun.* **15**, 338–344.
Rosche, B., Tshisuaka, B., Hauer, B., Lingens, F., and Fetzner, S. (1997). *J. Bacteriol.* **179**, 3549–3554.
Roualt, T. A. and Klausner, R. D. (1996). *Trends Biochem. Sci.* **21**, 174–180.
Russell, M. J., Daniel, R. M., Hall, A. J., and Sherringham, J. A. (1994). *J. Mol. Evol.* **39**, 231–243.
Salter, A. H., Newman, B. J., Napier, J. A. and Gray, J. C. (1992). *Plant Mol. Biol.* **20**, 569–574.
Schäfer, G., Purschke, W., and Schmidt, C. L. (1996). *FEMS Microbiol. Rev.* **18**, 173–188.
Schauer, R., Kelm, S., Reuter, G., Roggentin, P., and Shaw, L. (1995). In *Biology of the Sialic Acids: Biochemistry and Role of Sialic Acids* (Rosenberg, A. ed.), Plenum Press, New York, pp. 7–67.
Schauer, R., Malykh, Y. N., Krisch, B., Gollub, M., and Shaw, L. (1999). In *Sialobiology and Other Forms of Glycosylation: Biosynthesis*

and Biology of N-Glycolylneuraminic Acid (Inoue, Y., Lee, Y. C., and Troy, F. A., II, eds.), Gakushin, Japan, pp. 17–27.
Schlenzka, W., Shaw, L., Schneckenburger, P., and Schauer, R. (1994). *Glycobiology* **4**, 675–683.
Schlenzka, W., Shaw, L., Kelm, S., Schmidt, C. L., Bill, E., Trautwein, A. X., Lottspeich, F., and Schauer, R. (1996). *FEBS Lett.* **385**, 197–200.
Schmidt, C. L., Anemüller, S., and Schäfer, G. (1996). *FEBS Lett.* **388**, 43–46.
Schneckenburger, P., Shaw, L., and Schauer, R. (1994). *Glycoconjugate J.* **11**, 194–203.
Schröter, T., Hatzfeld, O. M., Gemeinhardt, S., Korn, M., Friedrich, T., Ludwig, B., and Link, T. A. (1998). *Eur. J. Biochem.* **255**, 100–106.
Schütz, M., Zirngibl, S., le Coutre, J., Büttner, M., Xie, D.-L., Nelson, N., Deutzmann, R., and Hauska, G. (1994). *Photosyn. Res.* **39**, 163–174.
Schütz, M., Brubna, M., Lebrun, E., Baymann, F., Huber, R., Stetter, K.-O., Hauska, G., Toci, R., Lemesle-Meunier, D., Tron, P., Schmidt, C. L., and Nietschke, W. (2000). *J. Mol. Biol.,* **300**, 663–675.
Shaw, L., Schneckenburger, P., Schlenzka, W., Carlsen, J., Christiansen, K., Jürgensen, D., and Schauer, R. (1994). *Eur. J. Biochem.* **219**, 1001–1011.
Simon, M. J., Osslund, T. D., Saunders R., Ensley, B. D., Suggs, S., Harcourt, A., Suen, W. C., Cruden, D. L., Gibson, D. T., and Zylstra, G. J. (1993). *Gene* **15**, 31–37.
Small, F. J. and Ensign, S. A. (1997). *J. Biol. Chem.* **272**, 24913–24920.
Sone, N., Tsuchiya, N., Inoue, M., and Noguchi, S. (1996). *J. Biol. Chem.* **271**, 12457–12462.
Steppuhn, J., Rother, C., Hermans, J., Jansen, T., Salnikow, J., Hauska, G., and Herrmann, R. G. (1987). *Mol. Gen. Genet.* **210**, 171–177.
Suen, W. C., Haigler, B. E., and Spain, J. C. (1996). *J. Bacteriol.* **178**, 4926–4934.
Sylvestre, M., Sirois, M., Hurtubise, Y., Bergeron, J., Ahmad, D., Shareck, F., Barriault, D., Guillemette, I., and Juteau, J. M. (1996). *Gene* **174**, 195–202.
Tan, H. M., Tang, H. Y., Joannou, C. L., Abdel-Wahab, N. H., and Mason, J. R. (1993). *Gene* **130**, 33–39.
Thompson, J. D., Gibson, T. J., Plewniak, F., Jeanmougin, F., and Higgins, D. G. (1997). *Nucl. Acids Res.* **24**, 4876–4882.
Trumpower, B. L., and Gennis, R. B. (1994). *Annu. Rev. Biochem.* **63**, 675–716.
Twilfer, H., Bernhardt, F.-H., and Gersonde, K. (1985). *Eur. J. Biochem.* **147**, 171–176.
Vèrgeres, G. and Waskell, L. (1995). *Biochemie* **77**, 604–620.
Wächtershäuser, G. (1992). *Progr, Biophys. Mol. Biol.* **58**, 85–201.
Wagner, M. J., Packer, J. C., Howe, C. J., and Bendall, D. S. (1996). *Biochim. Biophys. Acta* **1276**, 246–252.
Warren, L. (1963). *Comp. Biochem. Physiol.* **10**, 153–171.
Werlen, C., Kohler, H.-P. E., and van der Meer J. R. (1996). *J. Biol. Chem.* **271**, 4009–4016.
Xiong, J., Inoue, K., and Bauer, C. E. (1998). *Proc. Natl. Acad. Sci. USA* **95**, 14851–14856.
Yen, K.-M., Karl, M. R., Blatt, L. M., Simon, M. J., Winter, R. B., Fausset, P. R., Lu, H. S., Harcourt, A. A., and Chen, K. (1991). *J. Bacteriol.* **173**, 5315–5327.
Yu, J., Hederstedt, L., and Piggot, P. J. (1995). *J. Bacteriol.* **177**, 6751–6760.
Zhou, N. Y., Jenkins, A., Chan Kwo Chion, C. K., and Leak, D. J. (1999). *Appl. Environ. Microbiol.* **65**, 1589–1595.
Zumft, W. G. (1997). *Microbiol. Mol. Biol. Rev.* **61**, 533–616.
Zylstra, G. J. and Gibson, D. T. (1989). *J. Biol. Chem.* **264**, 14940–14946.