The \textit{cbb\textsubscript{3}} Oxidases Are an Ancient Innovation of the Domain Bacteria

Anne-Lise Ducluzeau,* Soufian Ouchane,† and Wolfgang Nitschke*

*Laboratoire de Bioénergétique et Ingénierie des Protéines (UPR 9036), Institut de Biologie Structurale et Microbiologie, Centre National de la Recherche Scientifique, Marseille, France; and †Centre de Génétique Moléculaire CNRS (UPR-2167) associé à l’Université Pierre et Marie Curie Bât. 24, Gif sur Yvette, France

A survey of genomes for the presence of gene clusters related to \textit{cbb\textsubscript{3}} oxidases detected bona fide members of the family in almost all phyla of the domain Bacteria. No archaeal representatives were found. The subunit composition was seen to vary substantially between clades observed on the phylogenetic tree of the catalytic subunit CcoN. The protein diade formed by CcoN and the monoheme cytochrome CcoO appears to constitute the functionally essential “core” of the enzyme conserved in all sampled \textit{cbb\textsubscript{3}} gene clusters. The topology of the phylogenetic tree contradicts the scenario of a recent origin of \textit{cbb\textsubscript{3}} oxidases and substantiates the status of this family as a phylogenetic entity on the same level as the other subgroups of the heme-copper superfamily (including nitric oxide reductase). This finding resuscitates and exacerbates the conundrum of the evolutionary origin of heme-copper oxidases.

Introduction

According to the paleogeochemical record, molecular oxygen had accumulated to detectable values on our planet by about 2.4 billion years ago (Canfield and Teske 1996; Rye and Holland 1998). This transition from an anoxic archaean to an increasingly oxidized proterozoic and phanerzoic environment entailed a dramatic change in ecology and metabolism of the biosphere. Many metabolic systems were more or less directly affected by this transition (Raymond and Segré 2007), but 2 specific enzymes are particularly intimately entangled with the history of dioxygen on Earth, that is 1) the penultimate biological generator of \textit{O\textsubscript{2}}, photosystem II of oxygenic photosynthesis and 2) its biological consumers, the respiratory \textit{O\textsubscript{2}} reductases. The latter enzymes function as terminal electron sinks in diverse energy converting electron transport chains and confer a very high energetic efficiency due to the strong oxidising power of \textit{O\textsubscript{2}}.

The evolutionary history of photosystem II is basically a mystery still today and controversial hypothetical scenarios abound in the literature (Blankenship and Hartman 1998; Baymann et al. 2001; Faller and Rutherford 2003; Mulkidjanian et al. 2006). Phylogenetic analysis of the parent organisms, however, suggests that the origin of the oxygen evolving photosystem precedes the actual appearance of \textit{O\textsubscript{2}} in the environment by a long geological time interval (Tomitani et al. 2006). Although net \textit{O\textsubscript{2}} accumulation in the atmosphere can thus be relatively well dated, the actual onset of photosynthetic \textit{O\textsubscript{2}} injection into the environment is still a matter of controversy. The complete absence of (bacterio)chlorophyll-based photosynthetic mechanisms in Archaea nevertheless strongly suggests that photosynthesis and a fortiori oxygenic photosynthesis appeared after the divergence of Archaea and Bacteria more than 3 billion years ago.

The origin and evolution of respiratory \textit{O\textsubscript{2}} reductases necessarily is linked to the history of \textit{O\textsubscript{2}} in the biosphere. Attempts to use enzyme phylogenies to gather information on this link date back to the mid 1990s when the group of Matti Saraste started to analyze the phylogeny of the major class of \textit{O\textsubscript{2}} reductases, the so-called heme-copper oxidases (HCOs) (Castresana et al. 1994, 1995; Castresana and Saraste 1995). These early studies identified 3 distinct subfamilies of HCOs and the obtained phylogenetic trees suggested the presence of HCOs in the common ancestor of Bacteria and Archaea, that is, prior to photosynthetic oxygen release. The implication that HCOs already existed over 3 billion years ago violently clashed with the total absence of free \textit{O\textsubscript{2}} in the archaean era taken for granted by the majority of the geochemical community. A more recent phylogenetic study (Pereira et al. 2001) of HCOs reported evidence for extensive horizontal gene transfer (HGT) and concluded on a late origin of HCOs followed by horizontal distribution over the 2 prokaryotic domains.

Within the HCO superfamily, the subgroup of the so-called \textit{cbb\textsubscript{3}} oxidases stands out. The \textit{cbb\textsubscript{3}} oxidases feature a distinctly different subunit composition as compared with all other \textit{O\textsubscript{2}} reductases of the HCO family (Garcia-Horsman et al. 1994; Gray et al. 1994), and the sequence position of their redox active tyrosine residue deviates from that of members of other studied HCO (Hemp et al. 2005, 2006; Rauhamäki et al. 2006). Already the earliest phylogenetic trees of the HCO superfamily suggested that the \textit{cbb\textsubscript{3}} subtree branches close to that of an exotic fourth subgroup of the tree, the nitric oxide (NO) reductases (van der Oost et al. 1994). Together with the observation that \textit{cbb\textsubscript{3}} oxidases feature with extraordinarily low \textit{O\textsubscript{2}} concentrations, this led to the proposal that they were closest to the ancestral uro oxidase (Castresana and Saraste 1995). This scenario was later challenged based on the observation that \textit{cbb\textsubscript{3}} oxidases seemed to exist only in proteobacteria (Myllykallio and Liebl 2000; Pereira et al. 2001), and they were correspondingly proposed to represent a relatively recent evolutionary radiation from 1 of the 2 other major groups of oxidases (Myllykallio and Liebl 2000). More recently, members of the \textit{cbb\textsubscript{3}} family not belonging to proteobacteria were reported (Cosseau and Batut 2004; Pitcher and W atmough 2004) weakening the evidence for a modern origin of the enzyme.

Although the evolutionary history of HCOs is expected to provide profound insights into that of dioxygen in the biosphere, many earlier attempts to decipher this history were hampered by limited sets of sequences and therefore produced controversial results. This is particularly true for the subgroup of the \textit{cbb\textsubscript{3}} oxidases. With the large number of sequenced bacterial genomes having become available.
during recent years, we therefore initiated a large-scale study of cbb3 gene clusters throughout prokaryotic genomes. Our data indicate the existence of cbb3-type cytochrome oxidases all over the eubacteria to the exception of Thermotogales, Deinococcales, and Firmicutes. With increasing phylogenetic distance from the previously characterized α-proteobacterial examples, new patterns of operon organization unfolded devoid of some of the “classic” subunits but containing new, hitherto unknown genes. Comparison of these variable gene clusters suggests the existence of an irreducible and invariable “core” of the enzyme consisting of the CcoN and CcoO subunits to which additional, substantially differing subunits are added. Phylogenies reconstructed on the catalytic and the extrinsic heme subunits provide a more detailed picture of the contributions of vertical descent and HGT in generating the extant distribution of the enzyme. The results presented in this work demonstrate that neither of the 2 previously proposed scenarios on the evolution of the HCO superfamily is applicable thus calling for a new paradigm to explain this evolutionary puzzle.

Materials and Methods

Sequences were retrieved via the NCBI server (http://www.ncbi.nlm.nih.gov/) and the TIGR CMR server (http://www.cmr.tigr.org/tigr-scripts/CMR/cmRHomePage.cgi). Open reading frames (ORFs) of Aquificales were analyzed using http://alrlab.research.pdx.edu/Aquificales/.

The obtained sample showed a very pronounced preponderance of proteobacterial sequences due to the high number of proteobacterial genomes sequenced to date. For the sake of clarity, gene clusters were omitted from the presented analysis if they did not add to the diversity of gene cluster organization and if at the same time the phylogenetic position of their parent species was close to that of one of the selected organisms.

Multiple sequence alignments were performed using ClustalX (Thompson et al. 1997), and prediction of secondary structure was carried out using the pSAAM package for Windows, a program for protein sequence analysis and modeling (University of Illinois). In places, the multiple alignment arrived at by ClustalX was optimized by hand based on results of secondary structure prediction. Phylogenetic relationships were calculated using the Neighbor-Joining (NJ) approach (ClustalX) of Saitou and Nei, as well as Kimura’s correction for multiple substitutions. Gap positions were excluded. Bootstrap values indicated in phylogenograms correspond to the frequency of occurrence of nodes in 1,000 bootstrap replicates.

Putative signal sequences were predicted using the SignalP 3.0 server (Bendtsen et al. 2004) at http://www.cbs.dtu.dk/services/SignalP/.

Results

The data set used for this study was obtained by screening available genomes for the presence of gene clusters related to the known representatives of the cbb3 family. To this end, Blast searches using the sequence of the catalytic subunit CcoN of the cbb3 oxidase from *Rubrivivax* (Rv.) *gelatinosus* as seed query were performed on the major genome databases. In subsequent steps, phylogenetically distant sequences still clustering with the cbb3 clade on NJ trees were used as queries to further explore the full space of cbb3-related sequences. Sets of adjacent ORFs which were situated on the same strand and contained the ccoN gene together with at least one more gene encoding a c-type cytochrome devoid of a CuA-binding motif were considered as potential cbb3 gene clusters. All ccoN genes retrieved in this way were found to phylogenetically cluster with the known proteobacterial cbb3 oxidases on trees containing representatives from the whole HCO superfamily. The genomic vicinity of ccoN genes was subsequently analyzed for the presence of genes related to cbb3 oxidases.

Phylogeny of the Catalytic Subunit

Figure 1A shows an unrooted NJ tree of the catalytic subunit CcoN. A very likely root for this tree is represented by the branching point to the sequence of the NO reductases’ catalytic subunit. The same rooting is suggested in phylogenies encompassing the whole HCO superfamily (Ducluzeau AL, unpublished data). The choice of the NO reductase from *Desulfitobacterium* (D) *haifniense* as out-group for the phylogenetic tree of figure 1 was guided by the observation that the *D. haifniense* enzyme branches closest to the cbb3 subtree in full trees of the HCO superfamily (Ducluzeau AL, unpublished data; Zumft 2005). The tree depicted in figure 1 contains substantially more deeply branching clades than the previously reported ones (Castresana and Saraste 1995; Pereira et al. 2001).

This tree features a prominent cleavage into 2 major subtrees situated on both sides of the putative root, which will in the following be denoted as “proximal” (encompassing the previously studied proteobacterial cbb3 oxidases; top in fig. 1A) and “distal” (lower part of fig. 1A), respectively. The proximal subtree consists of 5 well-defined clades and a single branch. The *γβγ*-proteobacterial clade, where the “standard” cbb3 oxidases from *Rhodobacter* (Rb.) *sphaeroides*, *Rhodobacter capsulatus*, *Pseudomonas* (Ps.) *stutzeri*, *Bradyrhizobium japonicum*, and *Rv. gelatinosus* are found, forms a tight clade which closely resembles the 16S rRNA phylogeny of the parent species. Among ε-proteobacteria, only the *Helicobacter pylori* enzyme has been biochemically and spectroscopically characterized in the past (Tsukita et al. 1999). CcoN sequences from this phylum form a clade that branches substantially deeper than that comprising the *γβγ* subgroups. In between these 2 clades, 2 compact clusters and a single branch are observed.

One of these groups is formed by gene clusters found in a subgroup of the Vibrionales (γ-proteobacteria). These enzymes are not only distinguished by the phylogenetic position of their catalytic subunit but also by a differing subunit composition (see below). All the Vibrionales containing this aberrant cbb3 oxidase except *Vibrio angustum* additionally contain a “normal” cbb3 oxidase positioned within the γ-proteobacterial clade discussed above. This redundancy suggests that the aberrant form results from an earlier duplication event followed by divergence toward a specialized function. Several of the parent species in this
clade are bioluminescent, and an involvement of respiratory enzymes in this process has been proposed (Bourgois et al. 2001). An involvement in bioluminescence therefore is a possible candidate for such a specialized function. However, not all the species containing this enzyme have been reported to be bioluminescent and a straightforward correlation of the presence of this enzyme and bioluminescence seems premature to us.

The second group is phylogenetically heterogeneous and encompasses Bacteroidetes, predatory δ-proteobacteria, Planctomycetes, and Verrucomicrobia. The isolated branch is represented by *Leptospira* (*Ls.*), a member of the Spirochaetes.

The lowest lying clade on this side of the putative root contains members of the (nonpredatory) δ-proteobacteria and of Chlorobiaceae.

On the other side of the root, that is, in the distal subtree, (bottom in fig. 1A), mainly deeply branching phyla of the 16S rRNA tree such as Aquificales, Actinobacteria, Acidobacteria, Cyanobacteria, Chlamydiae, and Planctomycetes are found in addition to 2 notable exceptions, that is, a β-/γ- and a δ-proteobacterial clusters. Sequences of *ccoN* genes from the distal subtree are strongly divergent from those of the characterized proteobacterial representatives. They nevertheless contain all crucial amino acids (6 histidines ligating the redox cofactors and 1 redox active tyrosine) and the full set of transmembrane helices of typical *cbb*3-type oxidases. In order to assay functional characteristics of the gene products from this group, we have initiated the biochemical and biophysical study of the enzyme from the Aquificais *Sulfurihydrogenibium azorense* (Aguiar et al. 2004). This enzyme was purified and indeed turned out to correspond to an oxygen reductase showing a high affinity for its substrate just like typical *cbb*3-type oxidases do (Ducluzeau AL, Nitschke W, Schoepp B, unpublished data). We therefore are confident that also this divergent group of genes codes for genuine CcoN proteins serving as catalytic subunit of *cbb*3-type O2 reductases.

The grouping together of sequences on the distal subtree might be dismissed as being due to long-branch attraction and thus as artifactual. However, this grouping is fully corroborated by a phylogenetic marker sequence provided by an insertion of about 10 amino acid residues, situated in a loop connecting helices V and VI, which is only present in members of the proximal subtree (see alignment in Supplementary Material online).

**Organication of Gene Clusters**

The "Canonical" Proteobacterial Gene Cluster

Proteobacteria are phylogenetically divided into the so-called α, β, γ, δ, and ε subgroups. In the past, *cbb*1-type oxidases from all but the δ subgroup have been studied. These enzymes are encoded by an operon of genes with the common order *ccoN–ccoO–ccoQ–ccoP*. The *ccoN* gene...
product is the catalytic subunit of the enzyme complex, whereas ccoO and ccoP encode membrane-tethered periplasmic mono- and diheme cytochromes, respectively. The ccoO gene product is a small, membrane-spanning protein of about 70 residues length and its functional role in the enzyme has not yet been clearly established.

Adjacent to the ccoNOQP cluster of structural genes lies a set of genes, cccoGHIS, which has for the first time been characterized in Sinorhizobium meliloti (Kahn et al. 1989). These genes encode membrane-bound proteins that play a crucial role in enzyme maturation (Preisig et al. 1996; Kulajta et al. 2005).

Analysis of the cluster of structural genes surrounding ccoN shows that the canonical organization of the structural operon (fig. 1B, line A) is conserved in all members of the α, β, γ, and ε subgroups for which the phylogenetic positioning of the ccoN tree corresponds to that of a standard 16S rRNA tree. The cases of β- and γ-proteobacterial CcoN sequences found on the distal part of the tree show substantially different organizations (lines M and N in fig. 1B) of the cbb3 gene cluster (see below).

The γ-Proteobacterial Vibrionales/Photobacter Group

All members of this family contain a monoheme and a pentaheme cytochrome upstream and downstream, respectively, of the ccoN/ccoO diade (fig. 1B, line B). Sequence comparisons show that the 2 N-terminal heme domains of the pentaheme protein actually correspond to CcoP. In this clade, the ccoNOQP set of genes is thus actually conserved and a triheme cytochrome seems to be fused to the C-terminal end of ccoP. The genomic proximity of the additional gene coding for a monoheme cytochrome suggests that the gene product is part of the enzyme.

The Clade Containing Bacteroides, Planctomycetes, Verrucomicrobia, and Predatory δ-Proteobacteria

This clade is heterogeneous with respect to species phylogenies. However, its grouping suggested by the ccoN tree is confirmed by the conserved gene cluster organization (fig. 1B, line C). A particularity of this clade consists in the fusion of the ccoO gene to the 3’ end of the ccoN gene. We have initiated the biochemical and biophysical characterization of the δ-proteobacterium Myxococcus xanthus to further explore the functional consequences of this gene fusion (Ducluzeau AL, Nitschke W, unpublished data). Downstream of ccoQ, an ORF coding for a second monoheme c-type cytochrome is found. Sequence comparisons between CcoO and the translated sequence of this new gene do not yield any substantial similarities between these 2 proteins. Because this monoheme cytochrome is found in other clades of cbb3 oxidases (see below), we propose to call this gene and its product ccoR and CcoR, respectively. Structural differences between CcoO and CcoR, as well as a putative evolutionary relationship between CcoR and the diheme cytochrome CcoP will be addressed in the Discussion.

The Spirochaetes Branch

The arrangement of structural genes in the Spirochaetis Ls. borgpetersenii (fig. 1B, line D) strongly resembles that of the above discussed heterogeneous clade (line C) to the notable exception that ccoN and ccoO are separate genes. This cbb3 oxidase contains the ccoQ and ccoR genes just as the heterogeneous group does suggesting the arrangement of genes in line D in figure 1B as the ancestral configuration of the heterogeneous clade that subsequently was modified by the cccoN/ccoO fusion.

δ-Proteobacteria and Chlorobiaceae

The characteristic feature of the δ-proteobacterial representatives in this clade consists in the fusion of a gene coding for a class I cytochrome to the 3’ end of the ccoO gene (fig. 1B, lines E and F). Due to its short sequence length, the phylogenetic positioning of this cytochrome domain is ambiguous and we cannot distinguish whether this heme protein is related to CcoR or represents an independent recruitment of a monoheme cytochrome into the complex. So far, 4 gene clusters belonging to this clade have been identified in Anaeromyxobacter dehalogenans 2CP-C and the Deferribacter Geovibrio thiophilus (both corresponding to fig. 1B, line E) as well as the δ-proteobacterium MLMS-1 and Geobacter (Gb.) bemidijensis Bem (both represented by line F). The latter species harbors a second cbb3 gene cluster, the ccoN gene of which is located on the distal subtree (see below). The δ-proteobacterium MLMS-1 and Geobacter contain a clearly identifiable ccoO gene together with an ORF coding for a further class I cytochrome.

Several green sulphur bacteria were found to contain cbb3 gene clusters that are affiliated to this clade by virtue of their ccoN sequences. The gene cluster arrangement (fig. 1B, line G), however, deviates from that of the above discussed δ-proteobacteria and rather resembles that from Ls. borgpetersenii and from species in the heterogeneous clade in that it codes for CcoN, CcoO, and CcoR as sole redox subunits.

The Distal Subtree

Almost all branches of the distal subtree represent deeply branching phyla of the 16S rRNA tree. The gene clusters observed for these branches are depicted in figure 1B lines H–O and their phylogenetic affiliation is indicated in figure 1A. In contrast to the situation encountered for the proximal subtree (fig. 1B, lines A–G), the gene cluster organization on the distal subtree is very heterogeneous. The only conserved feature is the cccoN/ccoO diade. Both monoheme and diheme proteins appear to supplement this conserved core in the different clades as represented in figure 1B. The monoheme cytochrome genes frequently occupy the position found for ccoR in the proximal subtree. Sequence similarities between these monoheme proteins and ccoR will be analyzed below.

Two Phylogenetic Outlier Clades Are Observed in the Distal Subtree:

1. Several β- and γ-proteobacteria are found on a common clade with a Planctomycetes and Salinibacter ruber, a species belonging to the Flexibacter/Bacteroides/Cytophaga phylum. For all members of this clade except Methyllobacillus flagellatus KT, the ccoN gene contains an additional N-terminal helix that is predicted
to correspond to a signal peptide (Bendtsen et al. 2004). All species except *Ps. aeruginosa* strain C and strain 2192 as well as *Ralstonia metallidurans* CH34 contain a monoheme cytochrome followed by genes presumably involved in maturation (fig. 1B, line M). Instead of the monoheme cytochrome, those 3 species contain a tandem of diheme proteins (fig. 1B, line N). These diheme cytochromes share high sequence similarities with diheme proteins in other *Pseudomonas* species suggesting an origin via HGT rather than duplications of monoheme subunits in the *ccb3* oxidase (see Discussion).

2. A clade containing (so far) only 2 δ-proteobacterial representatives is formed by the second *ccb3* gene cluster in *Gb. benidjiensis* *Bem* and that of *Geobacter* sp. FRC-32. The cluster of structural genes (fig. 1B, line J) once more contains a fused ORF constituted by *ccoO* and a class I monoheme cytochrome encoding gene. The maturation gene *ccoS* is located downstream of the structural cluster immediately followed by 2 genes coding for membrane proteins with unknown function.

### Maturation Genes

The full set of previously recognized maturation genes, *ccoGHIS*, is genomically adjacent to the structural genes in all cases of *ccb3* enzymes situated on the proximal subtree to the exception of the vibriionales cluster (line B in fig. 1B). However, because the Vibriionales also contain a normal *ccb3*-type operon that is close to the maturation genes, the respective machinery is available for these “exotic” cases. In stark contrast to the structural genes, the order of maturation genes features a remarkable plasticity between the different cases studied (see fig. 1B; Ouchane S. unpublished data). The polyferredoxin-encoding gene *ccoG* appears to be the best conserved of all maturation genes. Only in 3 cases, Aquificales, cyanobacteria, and Chlamydiae, we were unable to find an ORF showing significant similarity to *ccoG* in the vicinity of the structural cluster. The *ccoH* is absent in numerous species. The presence of 2 genes (COG5456 or COG3198) in the cases lacking *ccoH* suggests that the respective gene products may functionally replace *ccoH*.

### Discussion

The CcoN/CcoO Diade Forms the Functional Core of the Enzyme

Figure 1 shows that the subunit composition of *ccb3* oxidases varies dramatically between phyla. Only 2 genes are common to all gene clusters detected so far, that is, *ccoN* and *ccoO*, coding for the catalytic subunit and the monoheme *c*-type cytochrome, respectively. These 2 proteins thus constitute the minimal functional unit of *ccb3* oxidases. The close association of these 2 subunits is emphasized by the fact that in some species the corresponding genes are even fused together into a single gene. The crucial role of the CcoO cytochrome was already indicated by the lack of assembly of a CcoO-deficient mutant in proteobacteria (Zufferey et al. 1996, 1997), whereas a CcoP-deficient complex is still formed (Ouchane S. unpublished data).

The strictly conserved presence of the *ccoO* gene suggests that its product has structural properties that render it unexchangeable by other monoheme cytochromes. This assumption is borne out by a closer inspection of its amino acid sequence as compared with that of typical class I cytochromes (see Supplementary Material online). Class I cytochromes are characterized by 3 α-helical stretches engaging the heme moiety in a structurally conserved way. In none of the CcoO proteins sampled in our study significant probabilities for α-helices upstream of the putative methionine ligand were obtained in secondary structure prediction algorithms. Only a single α-helix at the C-terminal end of the protein was consistently predicted. In class I cytochromes, by contrast, 2 of the 3 structurally conserved α-helices are located in the N-terminal 2/3 of the protein. It therefore seems premature to us to unambiguously attribute a class I cytochrome structure to CcoO as has been done in the past (Hendriks et al. 1998). However, even assuming that CcoO belongs to the class I family, multiple alignments emphasize the singular presence of 2 long insertions amounting to more than 30 amino acid residues in CcoO (Supplementary Material online; Hendriks et al. 1998). The structural particularities of CcoO resulting from these long stretches are likely to be involved in formation of the functional CcoN/CcoO entity because the pattern of insertions is conserved in all CcoO sequences.

It is striking that in all subgroups of the HCO superfamily described so far, the catalytic subunit is obligatorily associated to an extrinsic, membrane-tethered protein. This extrinsic subunit is a copper protein in all O2 reductases except the *ccb3* type and is a monoheme protein in cytochrome-oxidizing NO reductases. Both O2 and NO reductases in certain species/clades work as quinol oxidases in which cases the extrinsic subunits are devoid of redox cofactors but nevertheless still present (Abramson et al. 2000; Hendriks et al. 2000). In quinol-oxidizing NO reductases (qNORs), the gene for the extrinsic subunit is fused to the N-terminal end of the catalytic unit.

This ostensibly essential protein diade in all types of HCO is complemented by varying additional redox subunits or domains generating the full diversity of the enzyme superfamily.

Monoheme CcoR rather than Diheme CcoP Is Present in Many Clades

The diheme cytochrome CcoP has been observed to complement the CcoN/CcoO diade in α-, β-, γ-, and δ-proteobacteria (Peissig et al. 1993; Gray et al. 1994; Tsukita et al. 1999; Strom et al. 2001; Comolli and Donohue 2004). As is evident from figure 1, this pattern of redox cofactor carrying subunits cannot be generalized to the whole enzyme family. The presence of a diheme subunit in addition to CcoO is restricted to the above-mentioned clades and part of the Aquificales, whereas many other clades contain a monoheme cytochrome gene instead of *ccoP* in their *ccb3* gene clusters (fig. 1B). In order to determine whether this subunit represents *c*-type cytochromes independently recruited in the individual clades, we performed multiple alignments of this cytochrome, the monoheme cytochrome
CcoO and the heme subunit from NO reductases (Supplementary Material online). The multiple alignments together with secondary structure prediction suggest that, in contrast to CcoO, this monoheme protein features the typical fold of class I cytochromes. On phylogenetic trees reconstructed from this multiple alignment (fig. 2; supplementary figs. 1 and 2, Supplementary Material online), all these monoheme cytochromes form a clade clearly distinct from both CcoO and NO reductase cytochrome subunit clusters. The genes coding for these heme proteins therefore 1) do not correspond to repetitive duplications of ccoO and 2) appear to be phylogenetically related to each other. The above described findings thus warrant our coining of a new name for these genes, ccoR, as introduced above (Results).

The tree shown in figure 2A does not fully exclude the possibility that CcoR has arisen twice. A prominent cleavage is visible in the CcoR clade separating genes situated in clusters corresponding to the proximal and distal subtrees as defined in the CcoN tree of figure 1. Such a cleavage is to be expected if CcoR was already present in the ancestral $cbb_3$ oxidase and subsequently coevolved with the enzyme although being lost or replaced by other hemeproteins in certain clades. The same tree topology, however, could alternatively also result from a scenario where the ancestral $cbb_3$ oxidase only contained the ccoN/ccO diade and recruited twice independently class I $c$-type cytochromes in the proximal and the distal subtrees. For parsimony reasons, we presently favor the first scenario.

In several phyla, an ORF coding for a class I cytochrome is fused to ccoO (fig. 1B, lines E, F, H, and J). Following the reasoning outlined above, it would seem likely that these monoheme cytochromes derive from ccoR. To assess this possibility, we reconstructed combined phylogenies of the respective heme domains in ccoR and fused class I cytochromes. The obtained results are inconclusive. The fused heme domains cluster in fact with the ccoR subtrees and not with other outgroup class I cytochromes (data not shown). However, the fact that the fused domains all branch close to the root of the ccoR tree rather than at a position corresponding to that indicated by the ccoN tree and that furthermore the respective nodes are predicted with low bootstrap probabilities makes it impossible to decide in favor of either independent recruitment of a new class I heme protein or fusion of a ccoR gene already present in the structural gene cluster.

Diheme Cytochromes Equivalent to CcoP Have Arisen Several Times Independently

The occurrence of CcoP-like diheme proteins in 3 topologically disconnected clades of the CcoN tree, that is 1) the $\beta\gamma$-proteobacterial and Vibrionales clusters, 2) the $\varepsilon$-proteobacteria, and 3) the Aquificales (except Hydrogenivirga sp., where no diheme gene is present in the $cbb_3$ cluster), is puzzling. According to amino acid sequence, CcoP appears to be structurally made up by 2 monoheme domains, very much like diheme cyt $c_4$ for which crystal structures are available (Kadziola and Larsen 1997; Abergel et al. 2003). The presence of interdomain sequence similarities have prompted the proposal that CcoP has resulted from a gene duplication of a monoheme cytochrome (Pereira et al. 2001; Pitcher et al. 2002). In order to assess whether all diheme genes observed in our study are monophyletic or
arise from independent gene duplication events, we have reconstructed the phylogeny of the individual heme domains. A schematic representation of these results is shown in the inset of figure 2 (the alignment and the detailed tree are available in the Supplementary Material online). The obtained tree strongly suggests 3 independent origins of the diheme genes, giving rise to ccoP in \( \alpha\)-\( \beta\)-\( \gamma\)-proteobacteria and the Vibrionales group, a second diheme cytochrome encoding gene in the \( \varepsilon\)-proteobacteria and finally a third group in the Aquificales. Repeated independent formations of the diheme genes rationalize the topological disconnection-ness of the diheme-containing clades. The similarity of redox subunit composition in \( \varepsilon\)-proteobacterial cbb\(_3\) oxidases (Tsukita et al. 1999) to the \( \alpha\)- and \( \beta\)-proteobacterial counterparts (Gray et al. 1994; Ouchane S. unpublished data) therefore is fortuitous and the attribution of the term CcoP to the \( \varepsilon\)-proteobacterial diheme cytochrome is misleading. An independent origin of the diheme gene in Aquificales is furthermore supported by the observation that in the deepest branching species on the Aquificales clade, that is, *Hydrogenenforga*, the diheme gene is absent in the cbb\(_3\) gene cluster.

As detailed in the Results, the diheme cytochrome genes observed in the \( \gamma\)-proteobacterial cluster situated on the distal subtree are recruited as such from closely related species.

The diheme cytochromes observed in the species corresponding to figure 1B, lines E, F, H, and J clearly do not fall into the category discussed above because they arise from a fusion of CcoO to a class I cytochrome.

### Vertical versus Horizontal Descent

The question on the importance of HGT in the evolutionary history of HCOs constitutes a major controversy (Castresana and Saraste 1995; Myllykallio and Liebl 2000; Pereira et al. 2001). The trees shown in figures 1 and 2 allow to assess this problem to significantly more detail than in previously published phylogenies.

Clear-cut examples of HGT on the level of both species and phyla are evident on these trees as will be detailed below. However, in their overall topologies, these trees and in particular that for the catalytic subunit (fig. 1A) reflect species phylogenies based on 16S rRNA astonishingly well.

The proximal subtree features the conventional branching hierarchy of proteobacteria starting with \( \delta\)-proteobacteria and proceeding through \( \varepsilon\)-, \( \gamma\)/\( \beta\)- to the \( \alpha\)-proteobacteria (von Mering et al. 2007). As within species trees, \( \beta\)- and \( \gamma\)-proteobacterial sequences form sisterclades.

The separate Vibrionales clade represents a secondary cbb\(_3\) enzyme in the respective species and is thus not in disagreement with species topologies. The mixed CFB/Verrucomicrobia/Planctomycetes/(predatory)\( \delta\)-proteobacterial clade, however, clashes substantially with the topology of 16S rRNA trees. This mixed clade may therefore well have arisen from a number of HGT events. The predatory \( \delta\)-proteobacteria among the species forming this clade may be involved in the detailed HGT mechanism (Gophna et al. 2006).

A further incongruency with respect to species trees consists in the grouping together of \( \delta\)-proteobacteria and Chlorobiaceae in the lowest branching cluster of the proximal subtree. With the small number of sequences available for this group and in the light of the differing gene cluster organization (fig. 1B, lines E, F, and G), it seems premature to us to conclude on HGT rather than limited reliability of tree topology in this region as a reason for this seemingly deviant tree topology.

The distal subtree contains all early-branching phyla of the species tree, that is, *Aquificales*, *Actinobacteria*, *Planctomycetes*, *Acidobacteria*, *Chlamydiae*, and cyanobacteria. A conspicuous exception is provided by the clade containing several \( \beta\)- and \( \gamma\)-proteobacteria together with a Planctomyces and a species belonging to the FBC phylum. The puzzling phylogenetic structure of this latter clade may be rationalized by the observation that the representative genes in the \( \gamma\)-proteobacteria *Ps. aeruginosa* strains C and 2192 as well as *R. metallidurans* carry the cbb\(_3\) gene cluster within a genomic island related to the mobile "integrative and conjugative elements" (ICEs), an HGT vector (Larbig et al. 2002). This strongly indicates that the cbb\(_3\) oxidase in these 3 species has arrived from other species via HGT. We therefore have tried to identify a putative ICE also in the other members of this clade. The genomic environment of the cbb\(_3\) gene cluster in the \( \beta\)-proteobacterial members of this clade contains numerous mobility elements such as integrases, transposases, and resolvases. Well-defined margins of a typical ICE, however, could not be detected that may simply reflect the longer time span and resulting genome evolution since insertion of these elements into the respective genomes. The 3 cases where the ICE is still clearly identifiable constitute in fact the most recent radiation of this clade. These observations lead us to suggest that the cbb\(_3\) oxidase genes found in this clade within \( \gamma\)- and \( \beta\)-proteobacteria have been laterally transferred into these species from a deep-branching phylum, possibly the Planctomycetes or the FBC.

For the 2 \( \delta\)-proteobacterial cbb\(_3\) oxidases found in the distal part of the tree, no clear evidence for a genomic affiliation to an HGT vector has been found. However, in this context, it is important that *Gb. bemidjiensis Bem* contains 2 distinct cbb\(_3\) gene clusters, one of which is located in the proximal part of the tree. One of the 2 clusters therefore must result from HGT and species phylogeny suggests that the cluster found on the proximal subtree corresponds to the genuine one.

### Phylogeny of cbb\(_3\)-Type Oxidases and Relationship to Other HCO Families

The cbb\(_3\) family contains representatives from basically all deep-branching phyla of the domain Bacteria. The corresponding gene clusters clearly do not derive via HGT from more recent radiations of the Bacteria. A conspicuous example are the Aquificales, for which a number of enzymes involved in bioenergetic mechanisms have previously been shown to derive from \( \varepsilon\)-proteobacteria (Schütz et al. 2000; Brugna-Guiral et al. 2003; Baymann et al. 2004). Their cbb\(_3\) oxidase, by contrast, does not cluster with \( \varepsilon\)-proteobacteria nor is it closely affiliated to any other major phylum. The absence of HGT from \( \varepsilon\)-proteobacteria concerning the cbb\(_3\) oxidase has recently been corroborated by genome sequencing of deep-sea
representatives of the e-proteobacterial phylum, that is, of species sharing the same habitat as Aquificales. The CcoN sequences of these deep-sea organisms cluster together with their pathogenic family members and are phylogenetically distant from the Aquificales sequences (Nakagawa et al. 2007).

Unfortunately, the low resolution of branching order in the distal subtree, almost certainly due to the much lower species coverage as compared with the proximal subtree, does not yet allow to deduce a detailed scenario for the very early evolutionary history of this enzyme. A recent branching off from one of the other classes of HCO as previously proposed (Myllykallio and Liebl 2000), however, must be discarded on the basis of our data. The whole cbb3 family, down to its earliest branches, is pronouncedly distinct from other HCOs, not only by the sequence of its catalytic subunit but also by its characteristic subunit composition. All members of the family contain at least the monoheme cyt CcoO as second redox cofactor-containing subunit, and no sequence motifs indicative of copper binding as in subunit II from other HCOs have been detected in any of the structural genes in the studied clusters.

Although such arguing against a late appearance of cbb3-type oxidases, our results indicate an origin of the enzyme within the domain of the Bacteria, that is, after the divergence of domains more than 3 billion years ago. Despite intensive efforts, we were unable to detect any putative members of the family within Archaea.

The results presented in this study therefore demonstrate that neither the initial scenario of predivergence origins for cbb3 oxidases put forward by Castresana and Saraste (1995) nor those stipulating more recent divergence from other cytochrome oxidases followed by distribution via HGT (Myllykallio and Liebl 2000; Pereira et al. 2001) are applicable.

Our phylogenies of the catalytic subunits from all members of the HCO superfamily corroborate the close evolutionary relationship of cbb3 oxidases to NO reductases (see fig. 1; Ducluzeau AL, unpublished data), already postulated by Hendriks et al. (2000). The major subgroup of NO reductase, the qNORs, contains archaeal and bacterial representatives and features a pronounced cleavage into members from the 2 domains (Zumft WG 2005). The qNORs thus appear to have arisen prior to the Archaea/Bacteria split. Because cbb3 oxidases root qNORs in between the 2 prokaryotic domains, they would be expected to have been present at the time of divergence, in contradiction to what has been outlined above.

This ensemble of seemingly incompatible results concerning the phylogeny of cbb3 oxidases thus constitutes a major conundrum to our understanding of the evolution of bioenergetic pathways using O2 as terminal electron acceptor.

Supplemental Material

Supplementary figures 1 and 2 alignment files and entry numbers are available at Molecular Biology and Evolution online (http://www.mbe.oxfordjournals.org/).

Acknowledgments

The authors gratefully acknowledge stimulating discussions with B. Schoepp-Cothenet and R. van Lis (Marseille) as well as help in setting up the figures by A. Ducret. Part of this work was financially supported by the French Agence Nationale pour la Recherche (ANR-06-BLAN-0384).

Literature Cited


**cb-type cytochrome c oxidase without a CuA center. Biochemistry. 33:3120–3127.**


