Cytochrome *c*: Occurrence and Functions

Ivano Bertini,*[,†,‡] Gabriele Cavallaro,[†,‡] and Antonio Rosato[†,‡]

Magnetic Resonance Center (CERM), University of Florence, Via L. Sacconi 6, 50019 Sesto Fiorentino, Italy, and Department of Chemistry, University of Florence, Via della Lastruccia 3, 50019 Sesto Fiorentino, Italy

Received January 4, 2005

Contents

* Corresponding Author. Mailing address: Magnetic Resonance Center, University of Florence, Via Luigi Sacconi 6, 50019 Sesto Fiorentino (Italy). Tel: +39 055 4574272. Fax: +39 055 4574271. E-mail: bertini@cerm.unifi.it.
† Magnetic Resonance Center (CERM).
‡ Department of Chemistry.

1. Introduction

Cytochromes *c* are among the most studied proteins. This is possibly due to their relatively high thermodynamic stability and their red color, which makes protein purification easier.[1] The three-dimensional structure of mitochondrial cytochrome *c* has been solved in the 1970s,[2] followed by a number of other structures from various sources. The small size, high solubility and high helical content and the presence of the heme cofactor have allowed mitochondrial as well as some bacterial cytochromes *c* to be studied through a variety of spectroscopic techniques. These features have contributed to making cytochrome *c* a very popular protein among biochemists and biophysicists.

Cytochromes *c* can bind one or several *c*-type hemes through two thioether bonds involving the sulfydryl groups of two cysteine residues. The heme iron ion is always axially coordinated by a histidine side chain. In his classical work,[3] Ambler identified four classes of cytochromes *c*, depending on the number of hemes, the type and the position of the axial iron ligands, and the redox potential. In the present review, we focus on mono-heme cytochrome *c* domains, defined by the property of a conserved structural fold (cytochrome *c* fold) and by the presence of a single Cys-Xaa-Xaa-Cys-His (CXXCH) signature for heme attachment. The cytochrome *c* fold has been already described in detail in the literature[4] and corresponds to the definition of protein superfamilies implemented in widely used protein classification tools such as CATH[5] or SCOP.[6] This definition also broadly corresponds to Ambler's class I. The minimal requirement for the cytochrome *c* fold is the presence of the three structural elements that are found in all cytochrome *c* experimental structures, that is, the N- and C-terminal α-helices (respectively, helix α1 and α5 in mitochondrial cytochromes), as well as the long helix (helix α3, also called the 60's helix in mitochondrial cytochromes) preceding the short helix and the loop containing the second axial ligand to the heme iron, which is nearly always a methionine[7−10] (Figure 1). In exceptional cases (which are mentioned in this review when relevant), the second ligand can be a different amino acid, such as asparagine or histidine, or even be absent. It is important to note that this cytochrome *c* "core" can be found embedded in a variety of different proteins: in these cases, the cytochrome *c* domain can be fused to other domains (even itself).

The function of cytochrome *c* is essentially that of an electron transfer protein, mainly involved in aerobic as well

10.1021/cr050241v CCC: $59.00
Published on Web 11/23/2005

Ivano Bertini was born in 1940, obtained the Italian degree of Doctor in Chemistry at the University of Florence (Italy) in 1964 and then the Libera Docenza in 1969. He became full professor in General and Inorganic Chemistry in 1975 at the University of Florence, where he is now. He received the Laurea Honoris Causa from the University of Stockholm in 1998, Ioannina in 2002, and Siena in 2003. He is a physical inorganic chemist who has cultivated NMR since 1965. Since 1974, he has been working with metals in life sciences. He has solved the solution structure of almost 100 proteins, has discussed the mobility of some of them, and has studied a number of protein–protein interactions. He is founder and director of the Magnetic Resonance Center (CERM), an NMR-based infrastructure operating in the field of life sciences.

Gabriele Cavallaro was born September 14, 1973, in Florence, Italy. He graduated in chemistry from the University of Florence (Italy), where he received also his Ph.D. in structural biology recognized by the Universities of Frankfurt (Germany) and Utrecht (Netherlands). He is now a postdoctoral fellow at the Magnetic Resonance Center (CERM) of Florence. His research interests include several areas of computational biology ranging from molecular dynamics simulations to protein structure modeling, with particular reference to the development of methods and computer programs for the calculation of NMR structures of paramagnetic metalloproteins. At the present, he is mainly involved in the implementation of bioinformatic tools for the analysis of genome sequences with emphasis on metal-binding proteins.

as anaerobic respiration. In mammalian cells, cytochrome *c* is also involved in apoptosis (for a review, see ref 11). Owing to the more recent discovery of the involvement of cytochrome *c* in the latter process, there are still several important questions to be clarified, such as how widespread is the cytochrome *c*-dependent pathway of cell death. Very recently, it has been reported that in mammalian mitochondria the enzyme $p66^{Shc}$ can oxidize cytochrome *c* to generate reactive oxygen species, which act as signaling molecules for apoptosis.[12] Mitochondrial cytochrome *c* appears to be necessary also for the assembly of cytochrome *c* oxidase.[13] Finally, a minor but interesting role for cytochrome *c* in Eukaryota is in the pathway of hydrogen peroxide scaveng-

Antonio Rosato obtained his Ph.D. degree from the University of Florence (Italy) in 1998. He is now Associate Professor of Chemistry at the same University. His research interests focus on the mechanisms of folding/unfolding and the factors controlling the stability of metalloproteins, on the theory and practice of NMR spectroscopy applied to paramagnetic systems, and on the application and development of bioinformatic tools to browse gene banks and genome sequences, with a particular interest in metalloproteins.

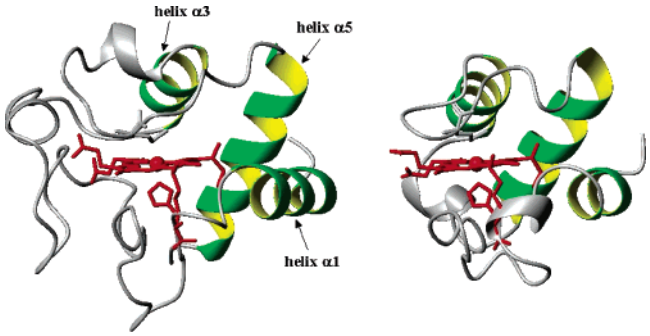

Figure 1. Comparison of the three-dimensional structures of two different cytochromes *c*: left, crystal structure of tuna cytochrome *c* (first crystal structure obtained for this class of proteins[2]); right, solution structure of oxidized cytochrome *c* from *Bacillus pasteurii*, determined by NMR.[27] The cytochrome *c* fold characteristics are colored in green and yellow; the heme and the CXXCH sequence signature are in red.

ing.[14] Cytochrome *c* is also widespread in the bacterial world, where it takes part in biochemical processes such as respiration and H_2O_2 scavenging, as well as in a number of other pathways. In particular, it is not uncommon that cytochrome *c* is fused to redox enzymes and constitutes an entry/exit point for electrons in the catalytic cycle of the enzyme.

The availability of complete genome sequences in a variety of organisms across all kingdoms of life has pushed us toward the compilation of a list of cytochrome *c* domains to provide a survey of (i) cytochrome *c* availability in each organism and (ii) the variety of proteins having a cytochrome *c* domain. These domains have been retrieved using bioinformatic methods and have been analyzed in terms of sequence, co-occurrence in operons of prokaryotes or gene fusion events, and three-dimensional structure, when available. Hints on the function have been obtained through the analysis of the genomic context together with the literature data. The result is a comprehensive and accurate data set of cytochrome *c* domains, ordered by organism and grouped according to the (proved or proposed) function, which provides a starting point for further biochemical and biophysical studies in the frame of the comprehension of the molecular mechanisms of life. In this review, we discuss the

occurrence and the biological functions of cytochrome *c* domains in all the three kingdoms of life, Eukaryota, Bacteria (separately Gram-negative and Gram-positive), and Archaea, considering several biochemical processes and metabolic and interaction pathways at different levels of knowledge.

An extended bioinformatic analysis of mitochondrial cytochromes *c* was performed in our laboratory in 1999.[9] That analysis allowed the identification of functionally relevant residues through multiple sequence alignments on mitochondrial proteins as well as the mappping of their position within the protein structure. This is important to identify features such as the location and size of intermolecular recognition patches or the intramolecular contacts determining the protein core (and thus protein stability). In the present work, the focus is shifted onto a much larger ensemble of quite distantly related sequences (as opposed to the high sequence conservation observed in mitochondrial cytochromes *c*). This diversity prevents us from reaching the same level of detail as in our former work but in return provides a much broader overview on the various functional roles of the cytochrome *c* domains in different organisms.

2. Methods

The first goal of the present work was that of collecting all the sequences of cytochrome *c* domains in available genome sequences. To this end, it is necessary to define the "sequence" of a cytochrome *c* domain, so that appropriate search criteria and filters can be used. The criteria selected were (i) a conserved protein fold (cytochrome *c* fold) and (ii) presence of the CXXCH signature for covalent attachment of the heme. Initially, results of genome analyses available from the Superfamily[15] server (version 1.65, http://supfam.mrc-lmb.cam.ac.uk/SUPERFAMILY) were used to gather proteins containing at least one domain fulfilling criterion i. The approach implemented in Superfamily consists of creating a library of profiles (hidden Markov models, HMMs[16]) representing all the known protein superfamilies.[17] These profiles are derived from the protein superfamilies identified in SCOP,[6] which is a database providing a description of the structural and evolutionary relationships of all proteins of known structure. In particular, superfamilies are families of proteins with low sequence identities whose structures and, in many cases, functional features suggest that a common evolutionary origin is probable.[6] The sequences contained in the Superfamily results were filtered by criterion ii, which results from common bioinorganic chemistry notions that *c*-type cytochromes possess a well-characterized signature for heme attachment. Note that when one analyzes protein primary sequences obtained from genome sequencing data there is no guarantee that posttranslational modifications such as covalent heme attachment are actually carried out by the organism. In addition, it cannot be excluded that the protoporphyrin IX moiety undergoes some further chemical modification, as found in cytochrome P460 or hydroxylamine oxidoreductase.[18,19] This kind of problem should however apply at most to a very small minority of the proteins analyzed. A final caveat that must be added is that proteins that are not cytochromes but have a similar topology (at least locally) and contain an occasional CXXCH stretch of residues can be selected by the above procedure. Such contaminations can be removed by checking the ensemble of retrieved sequences against the Pfam database,[20] which is a curated collection of multiple sequence alignments of many common protein domains. Finally, sequences with uncertain assignments to cytochrome *c* domains were confirmed or discarded by detecting similarities to other known cytochromes *c* through BLAST[21] searches in all nonredundant protein sequence databases.

The sequences identified in Proteobacteria and Cyanobacteria were grouped according to the number of cytochrome *c* domains present within each sequence; then the sequences containing the same number of cytochrome *c* domains were clustered based on their similarities. For proteins containing one cytochrome *c* domain, only the sequences of this domain were used for clustering (i.e., all protein segments corresponding to other domains were removed from the sequences prior to clustering). The clustering procedure was based on all-against-all BLAST searches within each group of sequences. We defined a minimal cluster as composed by three sequences such that each sequence had BLAST matches below a threshold *E*-value with the other two. All the minimal clusters were found; then those with two sequences in common were merged until no new minimal clusters could be joined. This approach, which is somehow reminiscent of the method employed by Tatusov et al. to calculate their clusters of orthologous groups (COGs),[22] was implemented in the program CYTCLUST (available on request from the authors). Clusters were initially determined applying a BLAST cutoff of $E = 10^{-10}$; then calculations were repeated using higher, less stringent threshold *E*-values (up to 10^{-3}) to broaden the coverage of clusters. Finally, BLAST results were analyzed in detail for sequences not included in any cluster to find pairs of significantly related protein sequences. Subsequently, all the pairwise alignments among the sequences included in any cluster were built with the CLUSTALW[23] program (version 1.60), and the average sequence identities within each cluster as well as between pairs of different clusters were calculated as a measurement of the degree of sequence similarity.

Multiple alignments of the same sequences used as input in the clustering procedure described above were also constructed with the program CLUSTALW. The resulting alignments for the proteins containing one, two, and three cytochrome *c* domains (included in Supporting Information Tables S1, S2, and S3, respectively) were used to generate three sequence identity dendrograms by means of the tree building option of CLUSTALW. Inspection of these dendrograms (not shown) revealed that closely related sequences are grouped in distinct branches in a way that is substantially identical to that obtained by the clustering procedure.

Figures 4 and 5 were obtained with the program CLANS,[24] which generates graphs representing pairwise sequence similarities established by BLAST matches and is available at the web site http://protevo.eb.tuebingen.mpg.de/miscpages/clans.

Homology modeling of *Bacillus subtilis* cytochrome *c* domains was performed with the program MODELLER[25] (version 6.2). The domain within the *bc* complex was modeled upon the structure of cytochrome c_{553} from *Bacillus pasteurii*[8] (PDB code 1B7V, sequence identity 38%), while the structure of cytochrome c_H from *Methylobacterium extorquens*[26] (PDB code 1QN2, sequence identity 30%) was used as the template for modeling the domain within the caa_3 oxidase. Structural models for the single-domain cytochromes *c* were calculated in a previous work.[27]

3. Results

All the complete genome sequences available as of August 2004 at the Superfamily web site were scanned for proteins containing at least one domain with the cytochrome *c* fold. The application of the strategy described in the Methods section (section 2) resulted in a list of 736 proteins in 112 out of 188 genome sequences analyzed. In detail, a total of 966 cytochrome *c* domains were detected in 32 Eukaryota, 59 Gram-negative bacteria, 17 Gram-positive bacteria, and 4 Archaea (Table 1 and Supporting Information Tables S4−S8). Seventy-six organisms do not appear to possess any cytochrome *c*, as judged from the absence of proteins both satisfying the fold criterion and having the required CXXCH (Table 2). Note that Table 2 counts all the strains available for each organism.

The following analysis is performed per type of organism, that is, Eukaryota, Gram-negative bacteria, Gram-positive bacteria, and Archaea. Note that this division of bacteria into Gram-negative and Gram-positive was done purposely for this discussion and is not a phylogeny-based partition, while Eukaryota and Archaea are indeed phylogenetic groups.

4. Eukaryota

Thirty-four genome sequences of Eukaryota have been analyzed. In 32 of these, 85 proteins containing one cytochrome *c* domain were detected. All of these proteins are single-domain cytochromes *c*, either of the "canonical" mitochondrial type (i.e., a soluble single-domain protein of about 100−110 residues that is located in the intermembrane space of the mitochondrion) or of the c_1 type, which is a part of the membrane-bound cytochrome bc_1 complex (also called ubiquinol/cytochrome *c* oxidoreductase, QCR) and is a physiological partner of mitochondrial cytochrome *c* (Figure 2 and Supporting Information Table S4). The bc_1 complex additionally contains a Rieske protein and a cytochrome *b*.[28]

If we look in detail at the soluble mitochondrial cytochromes *c*, *Homo sapiens* (as well as other primates[29]) has one cytochrome *c*, 105 residues long (including the initial methionine), which is encoded in chromosome 7 and is somatically expressed. A number of related pseudogenes (pseudogenes are genomic DNA sequences similar to normal genes but nonfunctional; they are regarded as defunct relatives of functional genes) have been identified.[29] Rat (*Rattus norvegicus*) and mouse (*Mus musculus*) possess a second cytochrome *c*, which is known to be testis-specific and only expressed during spermatogenesis.[30] Two cytochromes *c* are found also in the fruit fly *Drosophila melanogaster*, in the African malaria mosquito *Anopheles gambiae*, in the *Caenorhabditis* worms, in the transparent sea squirt *Ciona intestinalis*, in the parasites of the *Plasmodium* genus, and in the fungi of the *Saccharomyces* genus, where the expression of cytochromes *c* is regulated by partial oxygen pressure.[31] It has also been proposed that yeast cytochrome *c* is a sequence-specific DNA-binding protein and might have a regulatory role in the nucleus.[32] The plants *Arabidopsis thaliana* and *Oryza sativa* (rice) each contain three cytochromes *c*, one of which is of the so-called c_6 type (see also section 5.1.11). Plant cytochromes c_6 are targeted to the thylakoid lumen of chloroplasts[33] but do not function alongside plastocyanin in photosynthetic electron flow, and a regulatory or accessory role was proposed[34] for these proteins. The Japanese pufferfish *Takifugu rubripes* is the only other eukaryotic organism sequenced so far encoding three cytochromes *c* in its genome. Mitochondrial cytochromes *c* have been analyzed by us in a preceding paper, including homology modeling of 113 proteins from animals as well as plants.[9] The level of sequence identity among these proteins is higher than 45%, with sequence lengths varying between 100 and 120 residues. All mitochondrial cytochromes *c* are positively charged at physiological pH.[9]

During aerobic respiration, mitochondrial cytochrome *c* shuttles electrons from the bc_1 complex to cytochrome *c* oxidase (CCO) in the mitochondrion. Mitochondrial CCOs are of the so-called aa_3 type (see also section 5.1.1). The bc_1 complexes are oligomeric membrane protein complexes that transfer electrons from a relatively low potential quinol in the lipid phase to an acceptor protein in the aqueous phase; the electron transfer is coupled to the generation of a proton gradient across the membrane, which drives ATP synthesis.[35] Cytochrome c_1 is generally a much larger protein than cytochrome *c*, comprising a soluble domain of about 200 residues and one transmembrane helix.[36−38] On the other hand, the fish *Takifugu rubripes* has a predicted cytochrome c_1 domain of only about 100 residues. Sequences of cytochrome c_1 include an additional N-terminal hydrophobic segment, which is predicted to be transmembrane and is however absent in the mature protein, thus possibly constituting a signal sequence. Cytochrome c_1 possesses the fundamental features of the cytochrome *c* fold, that is, the three helices mentioned in the Introduction (section 1) with the conserved reciprocal orientation (Figure 3). Evolutionary studies suggest that the expanded sequence of cytochrome c_1 with respect to cytochrome *c* may be due to cytochrome c_1 arising from the structural collapse of a c_4-type dicytochrome *c* after corruption or deletion of its C-terminal CXXCH motif.[39] Within a given organism, the level of sequence identity between cytochrome *c* and cytochrome c_1 is of the order of 15−30%, with several long insertions in the latter sequences. The degree of sequence identity between cytochrome c_1 pairs from different organisms is 30% or higher. Conserved regions are distributed over the whole sequence length. While most Eukaryota contain one cytochrome c_1, *Drosophila melanogaster* and the plants *Arabidopsis thaliana* and *Oryza sativa* contain two proteins with a "canonical" 220 residue long cytochrome c_1 domain.

As far as the interaction between cytochrome *c* and cytochrome bc_1 is concerned, from the structural and functional point of view the most prominent information is derived from the crystal structure of the yeast cytochrome bc_1 complex with its bound substrate cytochrome *c*.[40] The soluble domain of cytochrome c_1 is negatively charged at neutral pH, thus complementary to the positively charged soluble cytochrome *c*. However, the structure of the complex suggests that despite the high electrostatic charge present on both proteins and the good complementary character in this respect between the two partners, the most stable configuration of the adduct is predominantly determined by nonpolar contacts. Electron transfer between the two partners is proposed to occur through a short-distance heme−heme direct contact.[40]

As mentioned, the other physiological partner of cytochrome *c* is CCO. In particular, cytochrome *c* interacts with subunit II of CCO, which contains the so-called Cu_A site. The electrostatic features of this region of CCO show a good complementary character to cytochrome *c*, as was the case for QCR. The region surrounding the Cu_A site, where

Table 1. List and Taxonomy of the 112 Genomes Where at Least One Cytochrome *c* Domain Was Detected

Archaea (Four Organisms)

	Phylum	Class	Order	Family	NS[a]	ND[b]
total for group					7	9
Aeropyrum pernix	Crenarchaeota	Thermoprotei	Desulfurococcales	Desulfurococcaceae	1	1
Pyrobaculum aerophilum IM2	Crenarchaeota	Thermoprotei	Thermoproteales	Thermoproteaceae	2	2
Archaeoglobus fulgidus DSM 4304	Euryarchaeota	Archaeoglobi	Archaeoglobales	Archaeoglobaceae	1	1
Methanosarcina acetivorans C2A	Euryarchaeota	Methanomicrobia	Methanosarcinales	Methanosarcinaceae	3	5

Eukaryota (32 Organisms)

	Kingdom	Phylum	Class	Order	NS[a]	ND[b]
total for group					85	85
Dictyostelium discoideum 2				Dictyosteliida	2	2
Plasmodium falciparum 1		Apicomplexa		Haemosporida	3	3
Plasmodium yoelii ssp. yoelii 1		Apicomplexa		Haemosporida	3	3
Aspergillus nidulans 1 r3.1	Fungi	Ascomycota	Eurotiomycetes	Eurotiales	2	2
Ashbya gossypii 1.0	Fungi	Ascomycota	Saccharomycetes	Saccharomycetales	2	2
Candida albicans	Fungi	Ascomycota	Saccharomycetes	Saccharomycetales	2	2
Candida glabrata	Fungi	Ascomycota	Saccharomycetes	Saccharomycetales	2	2
Debaromyces hansenii	Fungi	Ascomycota	Saccharomycetes	Saccharomycetales	2	2
Kluyveromyces lactis	Fungi	Ascomycota	Saccharomycetes	Saccharomycetales	2	2
Kluyveromyces waltii	Fungi	Ascomycota	Saccharomycetes	Saccharomycetales	2	2
Saccharomyces cerevisiae	Fungi	Ascomycota	Saccharomycetes	Saccharomycetales	3	3
Saccharomyces mikatae MIT	Fungi	Ascomycota	Saccharomycetes	Saccharomycetales	3	3
Saccharomyces paradoxus MIT	Fungi	Ascomycota	Saccharomycetes	Saccharomycetales	3	3
Yarrowia lipolytica	Fungi	Ascomycota	Saccharomycetes	Saccharomycetales	2	2
Schizosaccharomyces pombe	Fungi	Ascomycota	Schizosaccharo-mycetes	Schizosaccharo-mycetales	2	2
Fusarium graminearum 1	Fungi	Ascomycota	Sordariomycetes	Hypocreales	2	2
Magnaporthe grisea 7 r2.3	Fungi	Ascomycota	Sordariomycetes		2	2
Neurospora crassa 3	Fungi	Ascomycota	Sordariomycetes	Sordariales	2	2
Ustilago maydis 1 r2	Fungi	Basidiomycota	Ustilaginomycetes	Ustilaginales	2	2
Anopheles gambiae 22.2b	Metazoa	Arthropoda	Insecta	Diptera	3	3
Drosophila melanogaster 3.2	Metazoa	Arthropoda	Insecta	Diptera	4	4
Danio rerio 22.3b	Metazoa	Chordata	Actinopterygii	Cypriniformes	2	2
Takifugu rubripes 22.2c	Metazoa	Chordata	Actinopterygii	Tetraodontiformes	4	4
Ciona intestinalis 1.0	Metazoa	Chordata	Ascidiacea	Enterogona	3	3
Homo sapiens 22.34d	Metazoa	Chordata	Mammalia	Primates	2	2
Mus musculus 22.32b	Metazoa	Chordata	Mammalia	Rodentia	3	3
Pan troglodytes 22.1	Metazoa	Chordata	Mammalia	Primates	2	2
Rattus norvegicus 22.3b	Metazoa	Chordata	Mammalia	Rodentia	3	3
Caenorhabditis briggsae Aug03	Metazoa	Nematoda	Chromadorea	Rhabditida	3	3
Caenorhabditis elegans WS123	Metazoa	Nematoda	Chromadorea	Rhabditida	3	3
Arabidopsis thaliana 5	Viridiplantae	Streptophyta		Brassicales	5	5
Oryza sativa ssp. japonica 2.0	Viridiplantae	Streptophyta	Liliopsida	Poales	5	5

Table 1. (Continued)

Gram-Negative Bacteria (59 Organisms)						
	Phylum	Class	Order	Family	NS[a]	ND[b]
total for group					602	818
Aquifex aeolicus VF5	Aquificae	Aquificae	Aquificales	Aquificaceae	7	9
Bacteroides thetaiotaomicron VPI-5482	Bacteroidetes	Bacteroides	Bacteroidales	Bacteroidaceae	2	4
Parachlamydia sp. UWE25	Chlamydiae	Chlamydiae	Chlamydiales	Parachlamydiaceae	1	2
Chlorobium tepidum TLS	Chlorobi	Chlorobia	Chlorobiales	Chlorobiaceae	10	10
Gloeobacter violaceus	Cyanobacteria		Chroococcales		5	5
Synechococcus sp. WH 8102	Cyanobacteria		Chroococcales		4	4
Synechocystis sp. PCC 6803	Cyanobacteria		Chroococcales		4	5
Thermosynechococcus elongatus BP-1	Cyanobacteria		Chroococcales		4	4
Nostoc sp. PCC 7120	Cyanobacteria		Nostocales	Nostocaceae	4	4
Prochlorococcus marinus ssp. marinus CCMP1375	Cyanobacteria		Prochlorophytes	Prochlorococcaceae	2	2
Thermus thermophilus HB27	Deinococcus-Thermus	Deinococci	Thermales	Thermaceae	12	13
Rhodopirellula baltica sp. 1	Planctomycetes	Planctomycetacia	Planctomycetales	Planctomycetaceae	79	96
Caulobacter crescentus CB15	Proteobacteria	Alphaproteobacteria	Caulobacterales	Caulobacteraceae	11	12
Agrobacterium tumefaciens C58	Proteobacteria	Alphaproteobacteria	Rhizobiales	Rhizobiaceae	9	12
Bartonella henselae Houston-1	Proteobacteria	Alphaproteobacteria	Rhizobiales	Bartonellaceae	1	1
Bartonella quintana	Proteobacteria	Alphaproteobacteria	Rhizobiales	Bartonellaceae	1	1
Bradyrhizobium japonicum USDA 110	Proteobacteria	Alphaproteobacteria	Rhizobiales	Bradyrhizobiaceae	42	62
Brucella melitensis 16M	Proteobacteria	Alphaproteobacteria	Rhizobiales	Brucellaceae	7	9
Brucella suis 1330	Proteobacteria	Alphaproteobacteria	Rhizobiales	Brucellaceae	7	9
Mesorhizobium loti	Proteobacteria	Alphaproteobacteria	Rhizobiales	Phyllobacteriaceae	13	19
Rhodopseudomonas palustris CGA009	Proteobacteria	Alphaproteobacteria	Rhizobiales	Bradyrhizobiaceae	17	24
Sinorhizobium meliloti	Proteobacteria	Alphaproteobacteria	Rhizobiales	Rhizobiaceae	5	6
Rhodobacter sphaeroides	Proteobacteria	Alphaproteobacteria	Rhodobacterales	Rhodobacteraceae	18	25
Rickettsia conorii	Proteobacteria	Alphaproteobacteria	Rickettsiales	Rickettsiaceae	2	2
Rickettsia prowazekii	Proteobacteria	Alphaproteobacteria	Rickettsiales	Rickettsiaceae	2	2
Wolbachia	Proteobacteria	Alphaproteobacteria	Rickettsiales	Rickettsiaceae	2	2
Bordetella bronchiseptica RB50	Proteobacteria	Betaproteobacteria	Burkholderiales	Alcaligenaceae	17	29
Bordetella parapertussis 12822	Proteobacteria	Betaproteobacteria	Burkholderiales	Alcaligenaceae	16	26
Bordetella pertussis Tohama I	Proteobacteria	Betaproteobacteria	Burkholderiales	Alcaligenaceae	12	18
Ralstonia solanacearum	Proteobacteria	Betaproteobacteria	Burkholderiales	Burkholderiaceae	19	26
Chromobacterium violaceum ATCC 12472	Proteobacteria	Betaproteobacteria	Neisseriales	Neisseriaceae	15	23
Neisseria meningitidis MC58	Proteobacteria	Betaproteobacteria	Neisseriales	Neisseriaceae	6	9
Nitrosomonas europaea ATCC 19718	Proteobacteria	Betaproteobacteria	Nitrosomonadales	Nitrosomonadaceae	19	26

Table 1. (Continued)

Gram-Negative Bacteria (59 Organisms), Continued						
	Phylum	Class	Order	Family	NS[a]	ND[b]
Bdellovibrio bacteriovorus HD100	Proteobacteria	Deltaproteobacteria	Bdellovibrionales	Bdellovibrionaceae	10	12
Desulfovibrio vulgaris ssp. vulgaris Hildenborough	Proteobacteria	Deltaproteobacteria	Desulfovibrionales	Desulfovibrionaceae	6	7
Geobacter sulfurreducens PCA	Proteobacteria	Deltaproteobacteria	Desulfuromonadales	Geobacteraceae	16	21
Campylobacter jejuni ssp. jejuni NCTC 11168	Proteobacteria	Epsilonproteobacteria	Campylobacterales	Campylobacteraceae	13	19
Helicobacter hepaticus ATCC 51449	Proteobacteria	Epsilonproteobacteria	Campylobacterales	Helicobacteraceae	8	11
Helicobacter pylori 26695	Proteobacteria	Epsilonproteobacteria	Campylobacterales	Helicobacteraceae	6	9
Wolinella succinogenes	Proteobacteria	Epsilonproteobacteria	Campylobacterales	Helicobacteraceae	15	19
Shewanella oneidensis MR-1	Proteobacteria	Gammaproteobacteria	Alteromonadales	Alteromonadaceae	15	21
Escherichia coli K12	Proteobacteria	Gammaproteobacteria	Enterobacteriales	Enterobacteriaceae	3	4
Salmonella enterica ssp. enterica ser. Typhi	Proteobacteria	Gammaproteobacteria	Enterobacteriales	Enterobacteriaceae	1	2
Salmonella typhimurium LT2	Proteobacteria	Gammaproteobacteria	Enterobacteriales	Enterobacteriaceae	2	3
Shigella flexneri 2a 301	Proteobacteria	Gammaproteobacteria	Enterobacteriales	Enterobacteriaceae	3	4
Yersinia pestis CO92	Proteobacteria	Gammaproteobacteria	Enterobacteriales	Enterobacteriaceae	3	4
Haemophilus ducreyi 35000HP	Proteobacteria	Gammaproteobacteria	Pasteurellales	Pasteurellaceae	1	1
Haemophilus influenzae Rd KW20	Proteobacteria	Gammaproteobacteria	Pasteurellales	Pasteurellaceae	1	1
Pasteurella multocida	Proteobacteria	Gammaproteobacteria	Pasteurellales	Pasteurellaceae	3	4
Pseudomonas aeruginosa PAO1	Proteobacteria	Gammaproteobacteria	Pseudomonadales	Pseudomonadaceae	27	44
Pseudomonas putida KT2440	Proteobacteria	Gammaproteobacteria	Pseudomonadales	Pseudomonadaceae	24	40
Pseudomonas syringae pv. tomato DC3000	Proteobacteria	Gammaproteobacteria	Pseudomonadales	Pseudomonadaceae	6	7
Vibrio cholerae O1 biovar eltor N16961	Proteobacteria	Gammaproteobacteria	Vibrionales	Vibrionaceae	10	13
Vibrio parahaemolyticus RIMD 2210633	Proteobacteria	Gammaproteobacteria	Vibrionales	Vibrionaceae	16	23
Vibrio vulnificus YJ016	Proteobacteria	Gammaproteobacteria	Vibrionales	Vibrionaceae	11	15
Xanthomonas axonopodis pv. citri 306	Proteobacteria	Gammaproteobacteria	Xanthomonadales	Xanthomonadaceae	8	9
Xanthomonas campestris pv. campestris ATCC 33913	Proteobacteria	Gammaproteobacteria	Xanthomonadales	Xanthomonadaceae	7	8
Xylella fastidiosa 9a5c	Proteobacteria	Gammaproteobacteria	Xanthomonadales	Xanthomonadaceae	1	1
Leptospira interrogans ser. lai 56601	Spirochaetes	Spirochaetes	Spirochaetales	Leptospiraceae	11	15

electrons from cytochrome *c* are transferred, indeed is also suitable for interaction with the partner from the structural point of view. The importance of electrostatic interactions in the formation of the electron-transfer complex between the two partners has been demonstrated.[41,42] However, it has been recently shown that there has been a clear evolutionary trend to reduce the electrostatic charges at the protein−protein interface as part of the adaptive evolution of anthropoid primates.[43] High-resolution structural data for the interaction of CCO with cytochrome *c* are lacking, especially

Table 1. (Continued)

Gram-Positive Bacteria (17 Organisms)

	Phylum	Class	Order	Family	NS[a]	ND[b]
total for group					42	54
Corynebacterium diphtheriae	Actinobacteria	Actinobacteria	Actinomycetales	Corynebacteriaceae	1	2
Corynebacterium efficiens YS-314	Actinobacteria	Actinobacteria	Actinomycetales	Corynebacteriaceae	1	2
Corynebacterium glutamicum ATCC 13032	Actinobacteria	Actinobacteria	Actinomycetales	Corynebacteriaceae	1	2
Mycobacterium avium ssp. paratuberculosis k10	Actinobacteria	Actinobacteria	Actinomycetales	Mycobacteriaceae	1	2
Mycobacterium bovis AF2122−97	Actinobacteria	Actinobacteria	Actinomycetales	Mycobacteriaceae	1	2
Mycobacterium leprae	Actinobacteria	Actinobacteria	Actinomycetales	Mycobacteriaceae	1	2
Mycobacterium tuberculosis H37Rv	Actinobacteria	Actinobacteria	Actinomycetales	Mycobacteriaceae	1	2
Streptomyces avermitilis MA-4680	Actinobacteria	Actinobacteria	Actinomycetales	Streptomycetaceae	1	2
Streptomyces coelicolor A3−2	Actinobacteria	Actinobacteria	Actinomycetales	Streptomycetaceae	1	2
Tropheryma whipplei Twist	Actinobacteria	Actinobacteria	Actinomycetales	Cellulomonadaceae	1	2
Deinococcus radiodurans R1	Deinococcus-Thermus	Deinococci	Deinococcales	Deinococcaceae	8	10
Bacillus anthracis Ames	Firmicutes	Bacilli	Bacillales	Bacillaceae	4	4
Bacillus cereus ATCC 14579	Firmicutes	Bacilli	Bacillales	Bacillaceae	4	4
Bacillus halodurans	Firmicutes	Bacilli	Bacillales	Bacillaceae	4	4
Bacillus subtilis ssp. subtilis 168	Firmicutes	Bacilli	Bacillales	Bacillaceae	4	4
Bacillus thuringiensis ser. konkukian 97−27	Firmicutes	Bacilli	Bacillales	Bacillaceae	4	4
Oceanobacillus iheyensis HTE831	Firmicutes	Bacilli	Bacillales	Bacillaceae	4	4

Total (112 Organisms)

NS[a]	ND[b]
736	966

[a] Number of sequences containing at least one cytochrome *c* domain. [b] Total number of cytochrome *c* domains detected.

as far as the CCO interface region is concerned. It is to be noted that the mechanism of O_2 reduction in mitochondria is likely to involve so-called supercomplexes between several of the membrane enzymatic complexes. In particular, evidence is available for the formation of QCR/CCO[44] and NADH−ubiquinone oxidoreductase/QCR[45] supercomplexes. A structural model for the latter has been recently obtained.[46]

In yeast, cytochrome *c* has an additional role with respect to higher Eukaryota such as mammals, in that it is involved in scavenging hydrogen peroxide by delivering electrons to the enzyme cytochrome *c* peroxidase (CCP), which uses them to reduce the substrate. A three-dimensional structure of the adduct between cytochrome *c* and CCP is available.[47] The region of cytochrome *c* in direct contact with CCP is relatively similar to that involved in the interaction with cytochrome c_1, described above. The most notable common feature of the adducts of cytochrome *c* with these two partners is the involvement of thioether 4 in intermolecular contacts, suggesting that this substituent of the porphyrin ring may be a common route on the electron-transfer path. Also in the CCP−cytochrome *c* structure, hydrophobic interactions are of crucial importance, even though interactions between polar groups are more significant than in the case of the adduct with cytochrome c_1.

In Eukaryota, cytochrome *c* is required not only for electron transfer but also for the production of a functional CCO. Indeed, in mitochondria lacking the folded and mature (heme-containing) form of cytochrome *c*, the CCO subunits are not properly assembled.[13,48] It is therefore likely that cytochrome *c* participates, together with the numerous proteins constituting a dedicated machinery, to CCO assembly through a still unknown mechanism.

In human, cytochrome *c* is a known activator of apoptosis.[49] The cytochrome *c*-dependent pathway to apoptosis is started by the interaction of cytochrome *c* molecules released from the mitochondrion with the C-terminal WD40 repeats of the apoptotic protease-activating factor 1 (Apaf-1). This complex, dubbed the apoptosome, in the presence of ATP or dATP nucleotides then recruits and activates the initiator caspase, caspase-9, starting a cascade of activation events involving other caspases, which eventually results in execution of apoptosis and cell death. Determination of the detailed organization of the apoptosome awaits a structure at atomic resolution. A structure of the apoptosome was determined

Table 2. List of the 76 Genomes Where No Cytochrome *c* Domain Was Detected

Archaea (16)	
Halobacterium sp. NRC-1	*Picrophilus torridus* DSM 9790
Methanobacterium thermoautotrophicum	*Pyrococcus abyssi*
Methanocaldococcus jannaschii	*Pyrococcus furiosus* DSM 3638
Methanococcus maripaludis	*Pyrococcus horikoshii*
Methanopyrus kandleri AV19	*Sulfolobus solfataricus*
Methanosarcina mazei Goe1	*Sulfolobus tokodaii*
Methanothermobacter thermautotrophicus Delta H	*Thermoplasma acidophilum*
Nanoarchaeum equitans Kin4-M	*Thermoplasma volcanium*
Eukaryota (2)	
Encephalitozoon cunicoli	*Trypanosoma brucei*
Gram-Positive Bacteria (36)	
Bifidobacterium longum NCC2705	Onion yellows phytoplasma
Clostridium acetobutylicum	*Staphylococcus aureus* ssp. aureus MRSA252
Clostridium perfringens 13	*Staphylococcus aureus* ssp. aureus MSSA476
Clostridium tetani E88	*Staphylococcus aureus* ssp. aureus Mu50
Enterococcus faecalis V583	*Staphylococcus aureus* ssp. aureus MW2
Lactobacillus johnsonii NCC 533	*Staphylococcus aureus* ssp. aureus N315
Lactobacillus plantarum WCFS1	*Staphylococcus epidermidis* ATCC 12228
Lactococcus lactis ssp. lactis	*Streptococcus agalactiae* 2603V−R
Listeria innocua	*Streptococcus agalactiae* NEM316
Listeria monocytogenes 4b *F*2365	*Streptococcus mutans* UA159
Listeria monocytogenes EGD-e	*Streptococcus pneumoniae* R6
Mycoplasma gallisepticum R	*Streptococcus pneumoniae* TIGR4
Mycoplasma genitalium	*Streptococcus pyogenes* M1 GAS
Mycoplasma mobile 163K	*Streptococcus pyogenes* MGAS315
Mycoplasma mycoides ssp. mycoides SC	*Streptococcus pyogenes* MGAS8232
Mycoplasma penetrans	*Streptococcus pyogenes* SSI-1
Mycoplasma pneumoniae	*Thermoanaerobacter tengcongensis*
Mycoplasma pulmonis	*Ureaplasma urealyticum*
Gram-Negative Bacteria (22)	
Acinetobacter sp. ADP1	*Chlamydophila pneumoniae* J138
Borrelia burgdorferi B31	*Chlamydophila pneumoniae* TW-183
Buchnera aphidicola APS	*Coxiella burnetii* RSA 493
Buchnera aphidicola Bp	*Erwinia carotovora* ssp. atroseptica SCRI 1043
Buchnera aphidicola Sg	*Fusobacterium nucleatum* ssp. nucleatum ATCC 25586
Candidatus Blochmannia floridanus	*Photorhabdus luminescens* ssp. Laumondii TT01
Chlamydia muridarum	*Porphyromonas gingivalis* W83
Chlamydia trachomatis	*Thermotoga maritima*
Chlamydophila caviae GPIC	*Treponema denticola* ATCC 35405
Chlamydophila pneumoniae AR39	*Treponema pallidum* ssp. pallidum Nichols
Chlamydophila pneumoniae CWL029	*Wigglesworthia glossinidia*

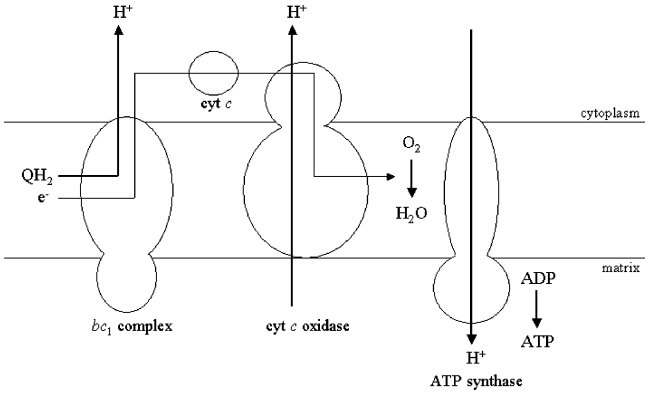

Figure 2. Schematic picture of the enzymes of the mitochondrial inner membrane involved in the terminal steps of aerobic respiration. Electrons are transferred from ubiquinol (QH_2) through the bc_1 complex and cytochrome *c* (cyt *c*) to cyt *c* oxidase, where they are used to reduce dioxygen into water. Both the bc_1 complex and cyt *c* oxidase translocate protons across the membrane. The proton gradient is required by ATP synthase to synthesize ATP. Figure adapted from ref 197. Reprinted with permission from *Science* (http://www.aaas.org), ref 197. Copyright 1999 The American Association for the Advancement of Science.

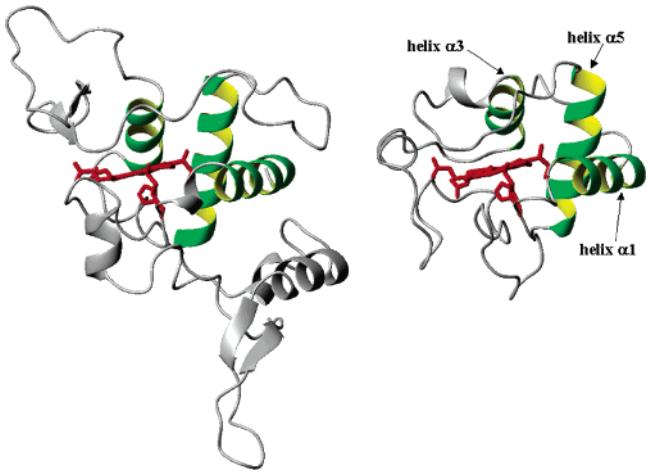

Figure 3. Comparison of the structures of the soluble domain of cytochrome c_1[37] (left) and of cytochrome c[2] (right). The cytochrome *c* fold characteristics are colored in green and yellow; the heme and the CXXCH sequence signature are in red.

at 27 Å resolution in 2002 by cryo-electron spectroscopy, revealing a wheel-like particle with 7-fold symmetry.[50] The 2.2-Å crystal structure of WD40-deleted Apaf-1 was recently determined,[51] shedding light on the molecular mechanism by which the binding and hydrolysis of nucleotides promote the formation of the apoptosome and the activation of caspase-9. This is only one of various different pathways

Table 3. Functional Classes Identified for Proteobacterial and Cyanobacterial Sequences Containing One Cytochrome *c* Domain[a]

matrix row or column	functional class	number of sequences	matrix row or column	functional class	number of sequences
1	alcohol dehydrogenases	4	15	iron uptake	5
2	caa_3 oxidases subunit II	9	16	multi-copper nitrite reductases	2
3	δ-caa_3 oxidases subunit II	2	17	NorC (nitric oxide reductases)	8
4	cbb_3 oxidases subunit II	29	18	PS-cyt c_{550}	7
5	cytochrome c_1	28	19	SHP	2
6	cytochrome c_2	39	20	solute transporters	4
7	cytochrome c_5	16	21	SorB (sulfite oxidation)	3
8	cytochrome c_{551}/c_{552}	8	22	SoxA (thiosulfate oxidation) (mono-heme)	3
9	cytochrome c_{552}/c_{554}	23	23	SoxD (sulfur oxidation)	3
10	cytochrome c_{553}	11	24	SoxX (thiosulfate oxidation)	4
11	cytochrome c_6	12	25	unknown I	4
12	cytochrome c_M	6	26	unknown II	4
13	DMSO reductases	18	27	unknown III	4
14	flavin-containing amine oxidases	4			

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27
1	**35.7**	21.1	17.0	16.6	16.6	17.3	19.5	21.0	20.8	20.4	21.2	22.5	17.9	23.9	22.4	20.2	18.0	17.1	17.2	21.0	18.9	17.6	20.4	20.6	21.6	19.4	22.8
2	21.1	**59.5**	21.9	15.1	19.7	19.6	21.4	18.3	23.7	20.5	23.9	21.0	19.7	36.8	24.7	31.6	21.9	18.1	15.6	18.9	18.8	19.1	24.5	23.9	21.7	18.8	30.8
3	17.0	21.9	**35.1**	16.5	17.1	22.1	21.5	25.9	20.2	21.0	22.3	19.6	17.8	19.7	19.9	18.5	20.5	17.0	15.1	17.7	17.4	16.3	24.3	22.5	22.3	22.5	23.7
4	16.6	15.1	16.5	**55.0**	16.4	16.8	15.9	15.0	15.3	14.7	15.8	17.3	14.0	16.5	15.1	16.2	16.7	16.2	14.2	16.2	13.4	17.4	18.6	16.5	17.2	14.9	20.7
5	16.6	19.7	17.1	16.4	**39.7**	18.3	17.6	18.2	17.3	16.5	17.9	17.0	15.4	18.4	19.3	18.3	17.5	17.5	17.6	16.9	20.0	16.2	17.8	18.1	18.9	14.9	24.6
6	17.3	19.6	22.1	16.8	18.3	**38.7**	22.9	23.7	22.0	20.1	19.7	17.5	17.8	21.0	21.5	18.5	18.6	17.7	16.4	18.8	18.7	17.5	21.3	20.8	20.7	15.3	21.2
7	19.5	21.4	21.5	15.9	17.6	22.9	**44.6**	22.4	19.3	21.0	21.8	18.9	14.4	24.1	21.5	23.8	21.1	19.8	19.1	19.4	19.0	17.8	20.5	20.9	20.6	19.0	26.0
8	21.0	18.3	25.9	15.0	18.2	23.7	22.4	**55.4**	23.6	22.7	22.2	22.8	15.9	21.3	20.1	25.7	23.4	18.6	18.0	21.1	20.7	18.0	19.9	21.9	21.8	17.2	25.5
9	20.8	23.7	20.2	15.3	17.3	22.0	19.3	23.6	**41.5**	26.9	22.3	22.0	16.8	24.2	23.0	23.0	20.7	19.9	19.2	22.2	20.5	19.8	22.8	20.9	22.0	19.2	26.4
10	20.4	20.5	21.0	14.7	16.5	20.1	21.0	22.7	26.9	**31.7**	21.0	20.8	16.4	21.3	22.0	21.8	20.5	20.0	18.4	21.3	17.0	18.6	23.3	18.7	19.4	22.2	24.7
11	21.2	23.9	22.3	15.8	17.9	19.7	21.8	22.2	22.3	21.0	**38.5**	23.2	18.1	22.7	23.1	23.3	20.3	23.7	18.4	22.3	20.0	18.4	22.7	22.6	20.9	20.0	23.4
12	22.5	21.0	19.6	17.3	17.0	17.5	18.9	22.8	22.0	20.8	23.2	**43.9**	19.4	24.6	21.8	23.6	22.8	16.5	17.7	24.6	18.9	19.1	22.2	23.7	21.4	18.7	20.1
13	17.9	19.7	17.8	14.0	15.4	17.8	14.4	15.9	16.8	16.4	18.1	19.4	**36.2**	17.4	17.1	20.0	17.2	17.2	19.0	17.2	18.6	19.4	18.4	16.9	20.2	19.6	20.3
14	23.9	36.8	19.7	16.5	18.4	21.0	24.1	21.3	24.2	21.3	22.7	24.6	17.4	**51.8**	26.9	30.4	22.7	17.7	17.5	22.5	20.9	19.2	26.0	22.4	25.0	19.5	27.0
15	22.4	24.7	19.9	15.1	19.3	21.5	21.5	20.1	23.0	22.0	23.1	21.8	17.1	26.9	**48.4**	19.6	22.6	19.2	18.2	21.7	19.2	18.2	21.8	20.6	23.5	19.9	29.3
16	20.2	31.6	18.5	16.2	18.3	18.5	23.8	25.7	23.0	21.8	23.3	23.6	20.0	30.4	19.6	**38.5**	22.7	18.4	20.4	20.5	20.3	17.5	21.2	21.1	22.8	18.1	28.7
17	18.0	21.9	20.5	16.7	17.5	18.6	21.1	23.4	20.7	20.5	20.3	22.8	17.2	22.7	22.6	22.7	**59.7**	19.4	16.7	20.5	18.5	17.2	19.8	20.0	23.3	20.6	17.6
18	17.1	18.1	17.0	16.2	17.5	17.7	19.8	18.6	19.9	20.0	23.7	16.5	17.2	17.7	19.2	18.4	19.4	**42.1**	17.3	18.3	16.7	17.9	20.2	18.2	18.7	15.2	17.1
19	17.2	15.6	15.1	14.2	17.6	16.4	19.1	18.0	19.2	18.4	18.4	17.7	19.0	17.5	18.2	20.4	16.7	17.3	**41.4**	17.1	17.6	19.6	20.3	18.7	17.6	16.0	16.5
20	21.0	18.9	17.7	16.2	16.9	18.8	19.4	21.1	22.2	21.3	22.3	24.6	17.2	22.5	21.7	20.5	20.5	18.3	17.1	**35.4**	19.0	19.5	20.2	22.1	20.7	18.7	22.2
21	18.9	18.8	17.4	13.4	20.0	18.7	19.0	20.7	20.5	17.0	20.0	18.9	18.6	20.9	19.2	20.3	18.5	16.7	17.6	19.0	**35.7**	17.0	19.7	18.8	19.3	16.1	23.8
22	17.6	19.1	16.3	17.4	16.2	17.5	17.8	18.0	19.8	18.6	18.4	19.1	19.4	19.2	18.2	17.5	17.2	17.9	19.6	19.5	17.0	**37.4**	20.6	19.0	19.5	19.3	17.0
23	20.4	24.5	24.3	18.6	17.8	21.3	20.5	19.9	22.8	23.3	22.7	22.2	18.4	26.0	21.8	21.2	19.8	20.2	20.3	20.2	19.7	20.6	**53.5**	22.7	24.0	22.4	27.1
24	20.6	23.9	22.5	16.5	18.1	20.8	20.9	21.9	20.9	18.7	22.6	23.7	16.9	22.4	20.6	21.1	20.0	18.2	18.7	22.1	18.8	19.0	22.7	**34.3**	22.8	20.9	22.1
25	21.6	21.7	22.3	17.2	18.9	20.7	20.6	21.8	22.0	19.4	20.9	21.4	20.2	25.0	23.5	22.8	23.3	18.7	17.6	20.7	19.3	19.5	24.0	22.8	**48.5**	17.8	27.4
26	19.4	18.8	22.5	14.9	14.9	15.3	19.0	17.2	19.2	22.2	20.0	18.7	19.6	19.5	19.9	18.1	20.6	15.2	16.0	18.7	16.1	19.3	22.4	20.9	17.8	**42.1**	19.4
27	22.8	30.8	23.7	20.7	24.6	21.2	26.0	25.5	26.4	24.7	23.4	20.1	20.3	27.0	29.3	28.7	17.6	17.1	16.5	22.2	23.8	17.0	27.1	22.1	27.4	19.4	**59.9**

[a] The number of sequences belonging to each class is reported (top section), as well as a matrix (bottom section) showing the average percentage values of sequence identity within each class (diagonal of the matrix) and between pairs of different classes (extradiagonal).

leading to apoptosis,[52] and it appears that it may not be common to all Eukaryota.[53]

5. Gram-Negative Bacteria

According to the traditional classification of prokaryotic genera, Gram-negative bacteria form a primary category of microbial types sharing a general feature of cell wall morphology, namely, they all have an inner membrane that encloses the cytoplasm and a second, outer membrane separated from the first by a space called periplasm. The periplasmic space, which is partly occupied by a peptidoglycan layer, is the site where cytochrome *c* molecules are both assembled and localized inside the cell.

The present work has produced a list of 818 cytochrome *c* domains detected in 602 protein sequences from 59 Gram-negative organisms (Supporting Information Tables S5 and S6), which, following the NCBI taxonomy database,[54] represent as many as nine distinct bacterial phyla (see Table 1). To gain information on the range of roles played by cytochrome *c* from this large amount of data, we have grouped the protein sequences according to the number of cytochrome *c* domains present within each sequence, and we have clustered sequences containing the same number of cytochrome *c* domains on the basis of their similarity, following the procedure described in the Methods section (section 2). For proteins containing one cytochrome *c* domain, only the sequences of this domain have been used for clustering. This approach, aimed at highlighting possible correlations between cytochrome *c* sequences and specific functional types, has been applied to the two phyla encompassing the large majority of bacteria that possess at least one cytochrome *c*, namely, Proteobacteria (46 organisms) and Cyanobacteria (6 organisms), thus covering a total of 669 cytochrome *c* domains in 480 sequences. Out of these 480 sequences, 319 contain a single cytochrome *c* domain, 137 contain two cytochrome *c* domains, 22 contain three cytochrome *c* domains, and 2 contain five cytochrome *c* domains. While being recognized as a coherent phylogenetic group, Proteobacteria are characterized by an extreme phenotypic diversity and comprise an enormous variety of morphological and physiological types scattered over five main classes known as Alpha-, Beta-, Gamma-, Delta-, and Epsilon- Proteobacteria (α, β, γ, δ and ϵ hereafter, respectively). Bacterial phyla having a single representative organ-

Table 4. Functional Classes Identified for Proteobacterial and Cyanobacterial Sequences Containing Two Cytochrome *c* Domains[a]

matrix row or column	functional class	number of sequences
1	CCP-MauG	35
2	Cbb_3 oxidases subunit III	28
3	Cytochrome c_{553} (di-heme)	8
4	Cytochrome c_4	36
5	Tetra-cyt *c* operons	12
6	ϵ-cytochrome *cc*	4
7	Cytochrome c_5 (di-heme)	5

	1	2	3	4	5	6	7
1	**33.8**	17.5	19.8	19.0	18.5	16.8	17.7
2	17.5	**40.2**	16.7	20.8	17.0	15.7	18.5
3	19.8	16.7	**39.3**	23.9	18.4	17.0	17.7
4	19.0	20.8	23.9	**33.4**	21.1	19.9	20.5
5	18.5	17.0	18.4	21.1	**42.0**	17.1	17.7
6	16.8	15.7	17.0	19.9	17.1	**51.5**	16.5
7	17.7	18.5	17.7	20.5	17.7	16.5	**52.9**

[a] The number of sequences belonging to each class is reported (top section), as well as a matrix (bottom section) showing the average percentage values of sequence identity within each class (diagonal of the matrix) and between pairs of different classes (extradiagonal).

Table 5. Functional Classes Identified for Proteobacterial and Cyanobacterial Sequences Containing Three Cytochrome *c* Domains[a]

matrix row or column	functional class	number of sequences
1	xanthine−GMC oxidoreductases	14
2	penta-cyt *c* operons	5

	1	2
1	**39.4**	38.0
2	38.0	**64.1**

[a] The number of sequences belonging to each class is reported (top section), as well as a matrix (bottom section) showing the average percentage values of sequence identity within each class (diagonal of the matrix) and between pairs of different classes (extradiagonal).

ism in Table 1, namely, Aquificae, Bacteroidetes, Chlamydiae, Chlorobi, Deinococcus-Thermus, Planctomycetes, and Spirochaetes (comprising a total of 149 cytochrome *c* domains in 122 proteins), have been analyzed separately on the basis of the results of standard BLAST searches in protein sequence databases.

5.1. Proteobacteria and Cyanobacteria

The analysis of cytochrome *c* sequences from Proteobacteria and Cyanobacteria hint at a sequence-dependent modulation of cytochrome *c* functional features within its general biological role as an electron-transfer protein, because a relationship appears between distinct known cytochrome *c* types and separate sequence clusters. On this basis, we have been able to identify a number of cytochrome *c* functional classes (Tables 3−5 and Figures 4 and 5), as well as to predict in favorable cases the pertinence to a certain class of several cytochromes *c* for which no detailed annotation was available. The inspection of clusters has been supported by more extensive analyses of the genes encoding the cytochrome *c* molecules presently identified. In particular, we have examined the occurrence of gene fusions between cytochrome *c* domains and other proteins, as well as the structure of the operons that possibly include cytochrome *c* genes, for the reason that even uncommon instances of fusion or coexpression may be useful to indicate a functional association between cytochrome *c* and other proteins, especially redox enzymes.[55] The results of this analysis are collected in Supporting Information Tables S5 and S6; in these tables, some sequences lack the assignment to a given class, either because they could not be unambiguously assigned to any cluster or because the assignment suggested by BLAST was not supported by the gene context.

The following subsections contain a broad overview of the functional classes identified in this way. Throughout the text, reference should be made to Tables 3−5 and to Figures 4 and 5. In addition, some singular cases of special interest are also discussed, although they may not appear in the figures; such sequences have specific annotations in Supporting Information Tables S5 and S6.

5.1.1. Cytochrome c Domains in Heme–Copper Oxygen Reductases: caa_3 and cbb_3 Oxidases

Several types of heme−copper terminal oxidases involved in the process of aerobic respiration are found in prokaryotes.[56] The two types that are relevant to the present work are caa_3 and cbb_3 oxidases, because they feature cytochrome *c* among their core subunits. In particular, the enzymes of the caa_3 type are similar to the mitochondrial aa_3 enzymes but have a distinctive cytochrome *c* domain at the C-terminus of their subunit II (called Cox2), which contains also the Cu_A binuclear copper center (Figure 6). This center is absent in the enzymes of the cbb_3 type, which contain both a monocytochrome *c* subunit (subunit II, also called FixO or CcoO) and a di-cytochrome *c* subunit (subunit III, also called FixP or CcoP) (see Figure 6).

The cbb_3 oxidases are enzymes with high substrate affinity found only in Gram-negative bacteria.[56,57] They are expressed in response to lower oxygen tensions, allowing the organism to survive and proliferate under micro-oxic conditions;[57−59] in particular, it has been suggested that their expression permits human pathogens such as *Campylobacter jejuni*, *Neisseria meningitidis*, and *Helicobacter pylori* to infect anoxic tissues.[59] Also, these enzymes have been observed to function as oxygen scavengers in diazotrophs of agricultural interest such as *Bradyrhizobium japonicum*, where they protect oxygen-labile nitrogenase during nitrogen fixation.[60] Through the present search, they have been found in all the classes of Proteobacteria, although they are not common to all the organisms (for instance, Enterobacteriaceae constitute a conspicuous exception). Subunit II has been retrieved as an approximately 200 residue long sequence containing one cytochrome *c* domain (cbb_3 oxidases subunit II class in Table 3 and cbb_3-II in Figure 4), while subunit III, which is encoded in the same operon, has been retrieved as an approximately 300 residue long sequence containing two cytochrome *c* domains (cbb_3 oxidases subunit III class in Table 4 and cbb_3−III in Figure 5). *Bdellovibrio bacteriovorus* (δ) is atypical in that its subunit III has only one such domain, and the sequence is correspondingly shorter (185 residues). The requirement for subunit III in the enzyme complex is not fully understood, since it is not essential for the assembly of the oxidase;[61] interestingly, it has been shown that one of its two hemes features histidine/histidine axial coordination rather than the more frequent histidine/methionine.[59]

The oxidases of the caa_3 type are also present in all the classes of Proteobacteria excluding ϵ, yet they are found in

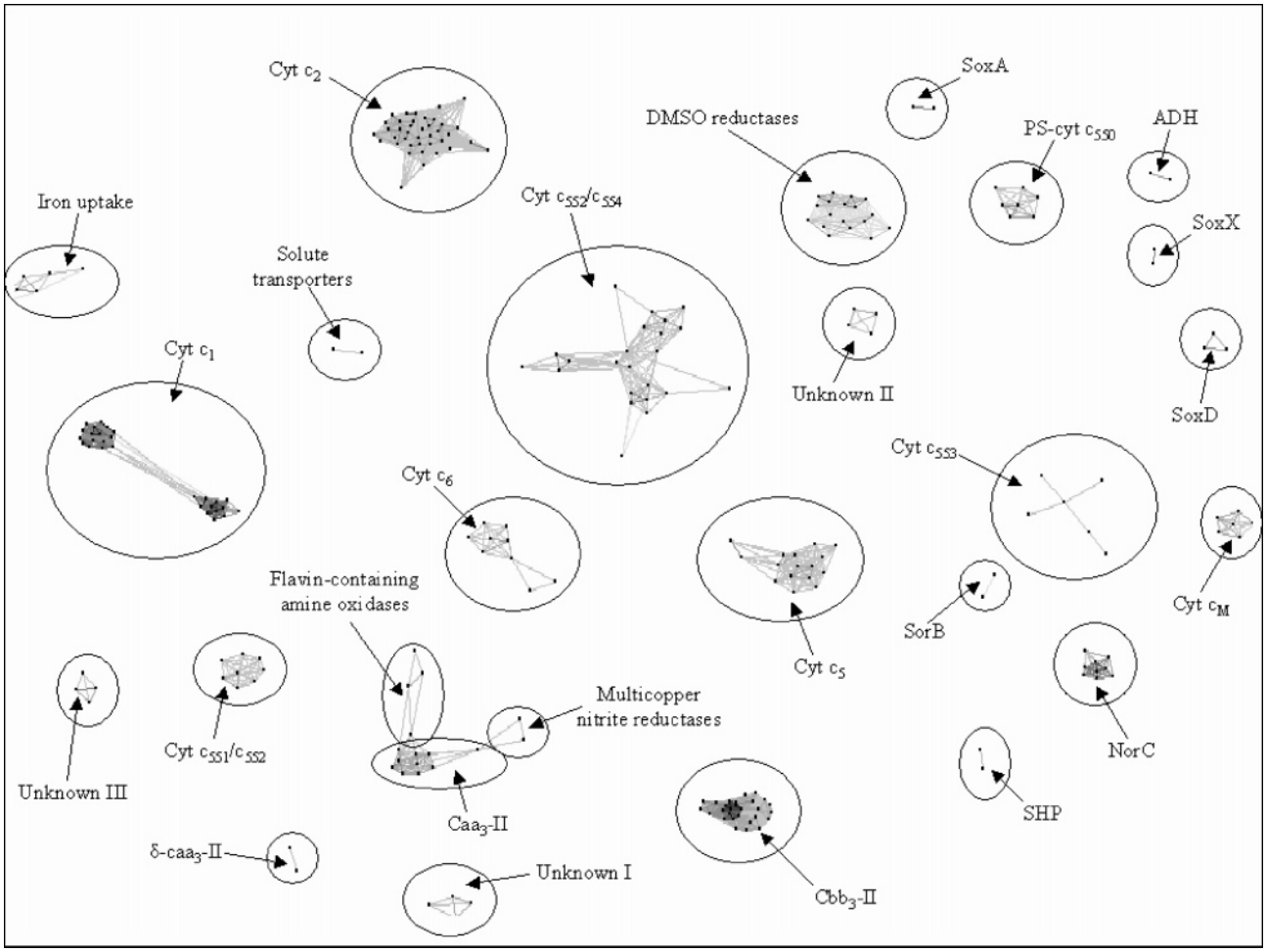

Figure 4. Two-dimensional CLANS graph visualizing the functional classes identified by clustering sequences containing one cytochrome *c* domain (abbreviations are explained in the text). Only the sequences of cytochrome *c* domains have been used for the clustering procedure. Sequences are represented by vertices in the graph, and BLAST matches below the threshold *E*-value of 10^{-10} are shown as edges connecting vertices.

relatively few organisms: for example, the only instance of this kind of enzyme in the α class occurs in *Rhodobacter sphaeroides*. As stated above, the characteristic cytochrome *c* domain is present at the C-terminus of subunit II sequences, which are approximately 400 residues long (*caa*$_3$ oxidases subunit II class in Table 3 and *caa*$_3$-II in Figure 4). The sequences from *Shewanella oneidensis* (γ) and *Desulfovibrio vulgaris* (δ) contain an additional cytochrome *c* domain. As shown in Figure 4, the majority of the sequences of cytochrome *c* domains of *caa*$_3$ oxidases are remarkably similar to those of cytochrome *c* domains found within the sequences of multi-copper nitrite reductases from *Chromobacterium violaceum* (β) and *Bdellovibrio bacteriovorus* (δ) (average pairwise identity between the classes is 31.6%), as well as to those of single-domain cytochromes *c* that are coexpressed with flavin-dependent monoamine oxidases in *Caulobacter crescentus* (α) and *Xanthomonas* (γ) (36.8% average identity), suggesting some possible mechanistic analogy among the three types of enzyme. On the other hand, the *caa*$_3$ sequences from *Rhodobacter sphaeroides* (α) and the δ-Proteobacteria *Bdellovibrio bacteriovorus* and *Geobacter sulfurreducens* are divergent, the latter two forming a small separate cluster (δ-*caa*$_3$ oxidases subunit II class in Table 3 and δ-*caa*$_3$-II in Figure 4).

5.1.2. Cytochrome c Domains in bc_1 Complexes

In Gram-negative bacteria, bc_1 complexes (QCR) are localized in the inner membrane and are common central elements of respiratory chains. The cytochrome c_1 subunit has been retrieved in all the Proteobacteria excluding *Bartonella*, Enterobacteriaceae, Pasteurellaceae, and the three members of the δ class. Bacterial cytochrome c_1 sequences are relatively long (ranging between 200 and 250 residues) and constitute a well-defined cluster (cytochrome c_1 class in Table 3 and cyt c_1 in Figure 4). They are typically encoded in a single operon together with the other two core subunits of the complex, namely, cytochrome *b* and a Rieske iron−sulfur protein (*pet*ABC or *fbc*FBC). Bradyrhizobiaceae are unusual because they have cytochromes *b* and c_1 expressed by a single gene encoding a precursor protein of almost 700 residues. A more striking exception to the typical cytochrome bc_1 architecture is provided by ε-proteobacteria, in which QCR actually lacks any cytochrome c_1. In these organisms, as previously described,[62] the functional equivalent of cytochrome c_1 in the complex is a di-heme cytochrome *c*, analogous to what occurs in the Gram-positive Actinobacteria (see section 6). Therefore, these sequences from ε-proteobacteria have been put together with those containing two cytochrome *c* domains and called ε-cytochromes *cc* in Table 4 (ε-cyt *cc* in Figure 5) by analogy to Actinobacteria. Such a similarity of names however does not imply any relationship of homology, in an evolutionary sense, between the ε-proteobacterial and the actinobacterial proteins.

There is only one experimental structure available for a bacterial bc_1 complex, which was determined for the α-pro-

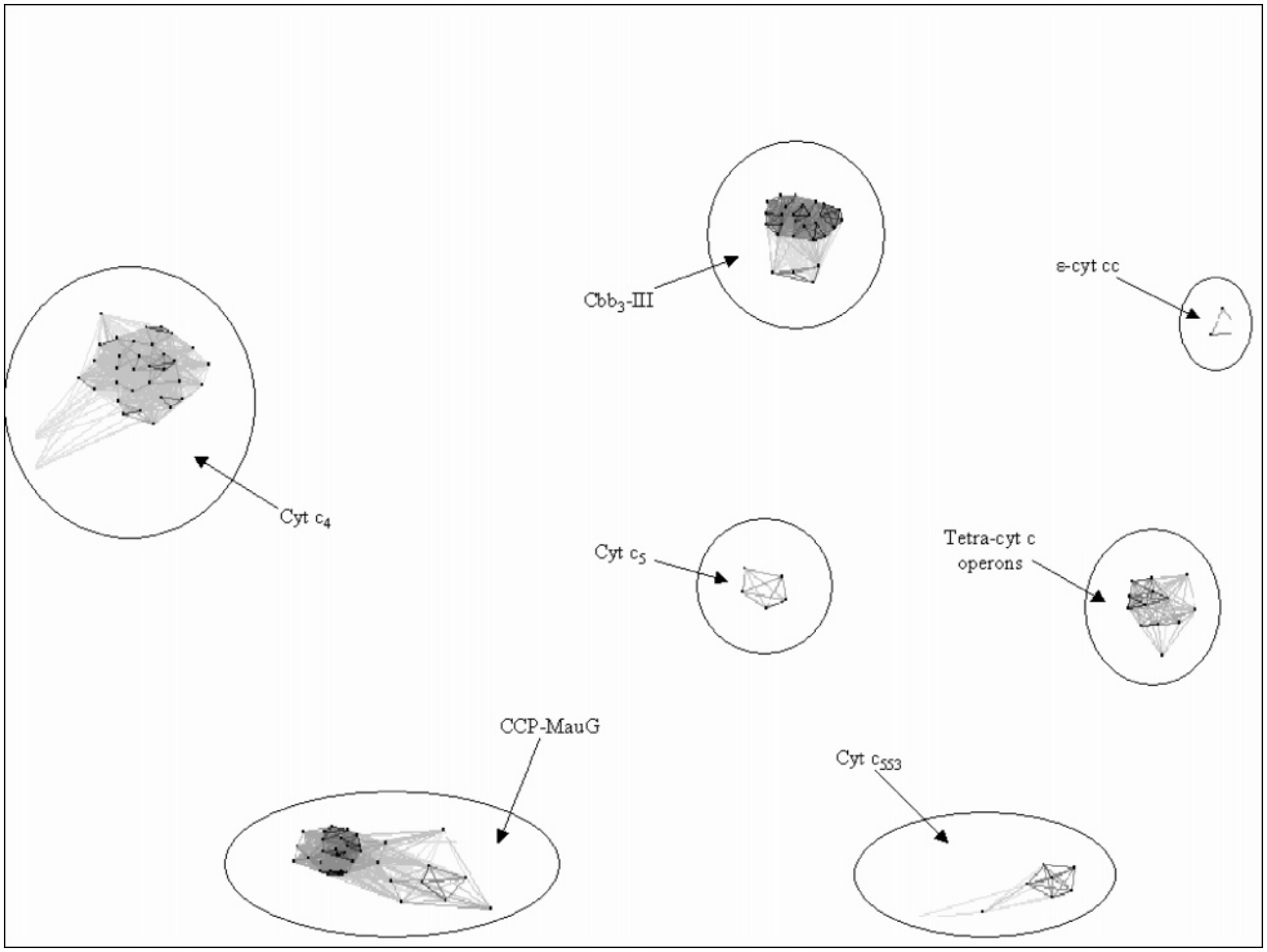

Figure 5. Two-dimensional CLANS graph visualizing the functional classes identified by clustering sequences containing two cytochrome *c* domains (abbreviations are explained in the text). Sequences are represented by vertices in the graph, and BLAST matches below the threshold *E*-value of 10^{-10} are shown as edges connecting vertices.

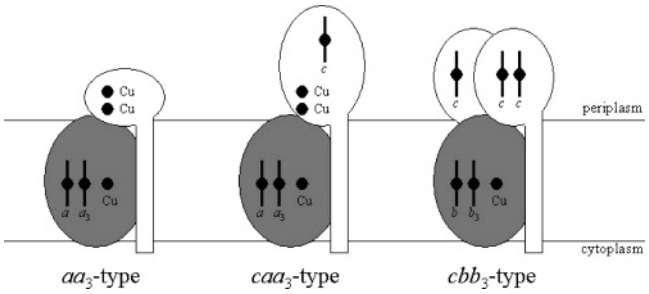

Figure 6. Schematic picture of the bacterial inner membrane heme−copper oxidases of the caa_3 and cbb_3 types. Mitochondrial-like aa_3 oxidase is also shown for comparison. All the enzymes have a subunit I (grey) containing a Cu ion (Cu_B) and two heme groups, whose types are used to designate the enzymes. Subunit II of aa_3 and caa_3 oxidases (white) contains a dinuclear copper center (Cu_A); subunit II of caa_3 oxidases includes also a *c*-type heme. cbb_3 enzymes do not have a Cu_A center; their subunits II and III (white) contain one and two *c*-type hemes, respectively. Figure adapted from ref 56. Reprinted with permission from ref 56. Copyright 2001 Elsevier).

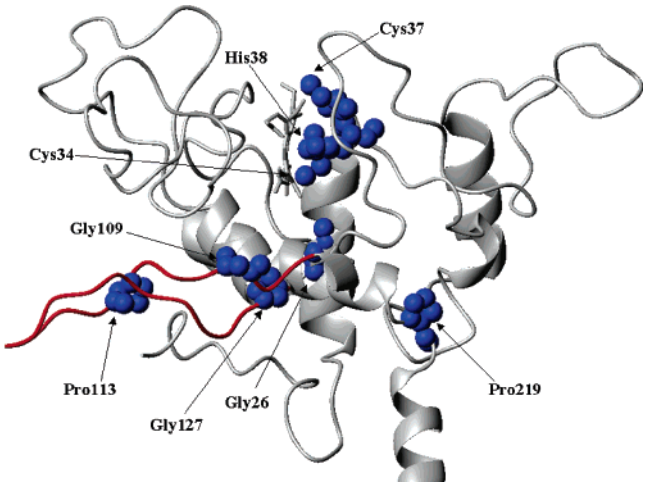

Figure 7. Cytochrome c_1 residues predicted by TRACE to have a significant functional role mapped onto the structure from *Rhodobacter capsulatus*.[63] Residue numbering refers to the latter structure. Cys34, Cys37, and His38 constitute the CXXCH signature for heme attachment. Gly109 and Pro113 are inserted in a long loop (shown in red) that is absent in mitochondrial bc_1 complexes.

teobacterium *Rhodobacter capsulatus* by Berry et al.[63] We have mapped onto this structure (see Figure 7) the cytochrome c_1 residues identified as functionally important by the program TRACE, which implements the so-called evolutionary trace method described in ref 64. Using as input the CLUSTALW alignment of the 28 sequences comprised in the cytochrome c_1 class, TRACE predicted a significant functional role for a glycine and a proline inserted in a long loop, which is absent in mitochondrial bc_1 complexes. In their work, Berry et al. suggested that the presence of this loop represents an intermediate evolutionary state between mitochondrial cytochromes *c*, where an equivalent loop covers the heme propionates, and mitochondrial cytochromes

c_1, where the loop is absent and the heme is exposed.[63] The results of TRACE point to a functional role of the loop, possibly in docking with the partner(s), as already proposed by Berry et al. for α-proteobacteria.[63]

5.1.3. Cytochrome c Domains in Cytochrome c Peroxidases and Methylamine-Utilization Proteins (CCP-MauG)

Apparently, some Gram-negative bacteria are able to scavenge hydrogen peroxide, being endowed with a cytochrome *c* peroxidase. At variance with their eukaryotic counterparts, which contain one noncovalently bound heme, bacterial enzymes possess two *c*-type heme groups with different redox potentials and can reduce the substrate without the need to generate semistable free radicals. Electrons are transferred from a donor (typically either a mono-heme cytochrome *c* or a blue-copper protein[65]) through the high-potential heme to the low-potential heme, where the reaction takes place.[66] Prokaryotic cytochrome *c* peroxidases are about 300 residues long and fold into two distinct cytochrome *c* domains, which have been retrieved as such in the present search. They are especially widespread among γ-, δ- and ϵ-proteobacteria, where only few genera (i.e., the γ *Haemophilus*, *Xanthomonas*, and *Xylella* and the δ *Desulfovibrio*) do not encode such enzymes in their genomes, while they are relatively less frequent in the α and the β classes, where they are found in *Bradyrhizobium japonicum*, *Mesorhizobium loti*, *Rhodobacter sphaeroides* (α), *Chromobacterium violaceum*, and *Nitrosomonas europaea* (β). As we shall see (section 6), they are present also among Gram-positive bacteria. In Table 4 and Figure 5, the class of cytochrome *c* peroxidases (CCP-MauG) includes also a smaller group of approximately 400 residue long sequences, most of which were annotated as MauG proteins. MauG proteins are di-cytochrome *c* proteins similar to cytochrome *c* peroxidases that are involved in the maturation of methylamine dehydrogenase.[67] Methylamine dehydrogenase catalyses the oxidative deamination of methylamine at the initial step in the metabolism of this substrate, and contains a tryptophan tryptophylquinone (TTQ) cofactor, which is formed by posttranslational modification. MauG is required for TTQ biogenesis because of its atypical oxygenase-like properties.[68] Putative MauG sequences have been retrieved here in a few α- (*Agrobacterium tumefaciens*, *Bradyrhizobium japonicum*, and *Mesorhizobium loti*) and β-proteobacteria (*Chromobacterium violaceum*, *Nitrosomonas europaea*, and *Ralstonia solanacearum*), as well as in *Geobacter sulfurreducens* (δ). However, we have not distinguished these proteins from cytochrome *c* peroxidases, because their above-mentioned annotations are not generally supported by the operon structures, as it appears that the presumed *mauG* genes are not located in methylamine utilization (*mau*) gene clusters comprising the other products required for methylamine metabolism.

5.1.4. Cytochrome c Domains in Denitrification Enzymes: Nitrite, Nitric Oxide and Nitrous Oxide Reductases

Denitrification is a respiratory process in which oxidized nitrogen compounds are used as electron acceptors in place of oxygen. It consists of four reactions through which nitrate is reduced via nitrite, nitric oxide, and nitrous oxide to dinitrogen. The steps of this pathway are catalyzed by an array of enzymes comprising nitrate reductase, nitrite reductase, nitric oxide reductase, and nitrous oxide reductase,

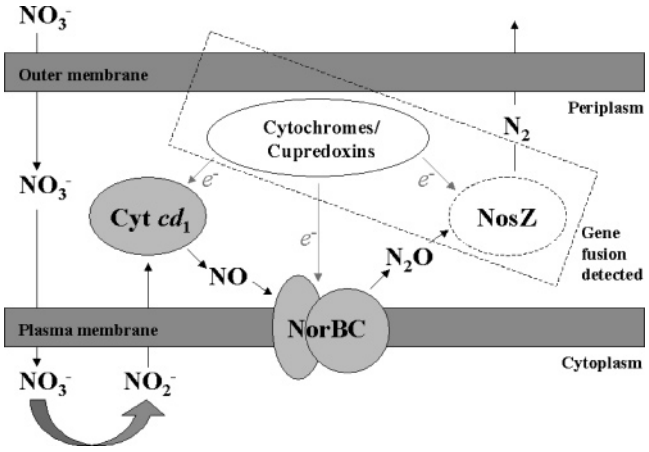

Figure 8. The role of cytochrome *c* domains in nitrate respiration. Abbreviations are explained in the text. Shaded blocks contain cytochrome *c* domains. A gene fusion between a cytochrome *c* domain and a NosZ sequence has been detected in *Wolinella succinogenes*.

which are induced sequentially under anaerobic conditions.[69−71] Figure 8 depicts the various roles of cytochrome *c* domains in the pathway, described below.

The denitrification step from nitrite to nitric oxide can proceed via one of two enzymes, which are dissimilar in structure and metal content: cytochrome cd_1 nitrite reductase is a periplasmic, soluble homodimer that contains one cytochrome *c* domain and one cytochrome d_1 domain in each subunit,[72,73] while multi-copper nitrite reductase is a trimeric enzyme characterized by the presence of a type 1 copper center as the electron acceptor site and a type 2 copper center as the catalytic site.[74,75] Among the available genomes, the cd_1 reductase occurs only in that from *Pseudomonas aeruginosa* (γ), where it is encoded by the *nirS* gene: our search has detected the cytochrome *c* domain at the N-terminus of the 568 residue long sequence containing also the cytochrome d_1 domain (see Table S5). The *nirS* gene is part of a complex operon that comprises a number of other genes implicated in the assembly of the active nitrite reductase.[76] Two of these genes (*nirM* and *nirC*), located immediately downstream from *nirS*, encode single-domain cytochromes *c*, both of which have been reported to be physiological electron donors for the enzyme.[77] Furthermore, we have retrieved one cytochrome *c* domain at the N-terminus of the 493 residue long protein encoded by the *nirN* gene. The role of the NirN product is unclear: it is strongly predicted to resemble the cytochrome *c*−cytochrome d_1 domain architecture of NirS; however no heme d_1 was detected in the purified protein;[77] indeed, the hypothesis that NirN may be an alternative cd_1 nitrite reductase is not supported by experimental data, which show that *Pseudomonas aeruginosa* mutant strains lacking the *nirS* gene do not have detectable nitrite-reducing activity.[76]

Unlike the cd_1 type, the retrieval of cytochrome *c* domains within multi-copper nitrite reductases is unexpected, since these enzymes have a cupredoxin-like fold and are thought to be redox partners of azurins and pseudo-azurins.[74] Nevertheless, we have found two instances of such a fusion in the genomes of *Chromobacterium violaceum* (β) and *Bdellovibrio bacteriovorus* (δ): in both cases, the cytochrome *c* domain is retrieved at the end of an approximately 500 residue long sequence; as mentioned, these cytochrome *c* sequences display a notable similarity with those of caa_3 oxidases. It is noteworthy that the occurrence of a cyto-

chrome *c* encoded within the gene for a multi-copper nitrite reductase appears to be correlated to the presence in the same operon of a gene encoding a protein of the SCO1 family (SCO1 proteins are also often annotated as SenC, or SCO1/SenC), which is known to be involved in the biogenesis of the Cu_A site in cytochrome *c* oxidase.[78,79] As a matter of fact, this co-occurrence is observed in *Chromobacterium violaceum* and *Bdellovibrio bacteriovorus*, whereas the SCO1/SenC gene is absent from operons encoding ordinary (i.e., without cytochrome *c*) multi-copper nitrite reductases in other organisms (e.g., *Brucella melitensis*, *Brucella suis*, *Rhodopseudomonas palustris* (α), and *Neisseria meningitidis* (β)). A similar example occurs also in *Pseudomonas putida* (γ), where cytochrome *c* is fused to a SCO1/SenC protein within a gene adjacent to one encoding a multi-copper oxidase (see Supporting Information Table S5). It is tempting to speculate that in these systems, the cytochrome *c* domain provides electrons that are used by SCO1/SenC in the process of maturation of the multi-copper nitrite reductase. A similar contention has been proposed in the context of CCO assembly.[80]

Nitric oxide reductase is a membrane-bound enzyme that is generally described as a complex of two subunits encoded by the *norB* and *norC* genes. The catalytic subunit (NorB) is an integral membrane protein that contains a dinuclear center composed by a *b*-type heme and a non-heme iron (Fe_B).[70] The other subunit (NorC) is a membrane-anchored cytochrome *c* that is lacking in the quinol-dependent enzyme isolated from *Ralstonia eutropha*[81] and also in *Chromobacterium violaceum* (β), where the *norB* gene is found in isolation. We have found most instances of *norC* genes in the α-proteobacteria (Rhizobiales and *Rhodobacter*), yet it occurs in *Nitrosomonas europaea* (β), *Pseudomonas aeruginosa* (γ), and *Bdellovibrio bacteriovorus* (δ) as well. The NorC sequences retrieved by the present approach are about 150 residues long and are nicely clustered (NorC class in Table 3 and Figure 4); *Bdellovibrio bacteriovorus* is atypical in that its NorC is 215 residues long and comprises two cytochrome *c* domains.

The final step in the denitrification process is catalyzed by nitrous oxide reductase, a homodimeric, periplasmic enzyme encoded by the *nosZ* gene. It contains two copper centers, Cu_A and Cu_Z, the former being the entry site for electrons and the latter being the catalytic site.[82−84] The nitrous oxide reductase from the ϵ-proteobacterium *Wolinella succinogenes* (Table S5) is unique because it has a C-terminal cytochrome *c* domain that was suggested to be the electron donor to the Cu_A center.[85] It was also noted that the entire *nos* operon of *Wolinella succinogenes* is atypical in comparison with those present in other organisms, and it was hypothesized that it encodes a complete electron transport chain catalyzing the reduction of nitrous oxide by menaquinol:[85] thus, the two single-domain cytochromes *c* that we have further retrieved within this cluster might be involved in transferring electrons along this chain.

5.1.5. Cytochrome c Domains in Molybdenum Enzymes and GMC Oxidoreductases

There are several respiratory enzymes that contain molybdenum coordinated by a complex pterin cofactor. Depending on the structure of the active site, they can be classified in one of three main groups, namely, dimethyl sulfoxide (DMSO) reductases, xanthine oxidases, and sulfite oxidases.[86] The family of DMSO reductases comprises nitrate reductases and many other enzymes, which are able to reduce a variety of N- and S-oxides and exhibit a diverse range of substrate specificity and cell localization.[87] They may be involved in anaerobic respiratory processes, such as the trimethylamine *N*-oxide (TMAO) reductases encoded in the *torA* and *torZ* genes from *Escherichia coli*,[88,89] or not, such as the biotin sulfoxide (BSO) reductases from *Escherichia coli* and *Rhodobacter sphaeroides*, which are mainly implicated in the recycling of biotin from BSO.[90,91] Respiratory DMSO reductases located in the periplasm appear to use membrane-anchored multi-heme cytochromes to transfer electrons from the quinone pool into the periplasmic space.[92] These multi-heme proteins are about 400 residues long and are encoded in the same operon of the enzyme (e.g., the *torCAD* and *torYZ* systems of *Escherichia coli*[88,89]). They have been identified through the present search in some γ-proteobacteria (Enterobacteriaceae but *Yersinia pestis*, Pasteurellaceae, *Vibrio*, and *Shewanella*) and in the α-proteobacterium *Rhodobacter sphaeroides*, where they are formed by a tetra-heme cytochrome fused to a C-terminal cytochrome *c* domain. In ϵ-proteobacteria, we have identified three instances (*Campylobacter jejuni*, *Helicobacter hepaticus*, and *Wolinella succinogenes*) of DMSO reductases coexpressed with a single-domain cytochrome *c*, suggesting that this protein is the electron donor to the enzyme, instead of a multi-heme cytochrome. Despite this difference in the mechanism of electron transfer, single-domain cytochrome *c* sequences from ϵ-proteobacteria are clustered with those of the cytochrome *c* domains fused to the tetra-heme cytochromes of γ-proteobacteria (DMSO reductases class in Table 3 and Figure 4), suggesting functional similarity. Indeed, it has been proposed that in *Escherichia coli*, the cytochrome *c* domain of TorC is responsible for electron transfer to the enzyme TorA, receiving electrons from the menaquinone pool through the tetra-heme domain.[93] The latter domain is also important for docking between TorA and TorC.[93] Analogously, the ϵ-proteobacterial single-domain cytochrome *c* can transfer electrons to the DMSO reductase, receiving them from other electron-transfer proteins.

The members of the xanthine oxidase family are molibdoenzymes catalyzing the oxidative hydroxylation of a diverse range of aldehydes and aromatic heterocycles.[94] They are coexpressed, or expressed in a single transcriptional unit, with three cytochrome *c* domains in *Bradyrhizobium japonicum* (α), *Bordetella bronchiseptica* (β), *Pseudomonas aeruginosa*, and *Pseudomonas putida* (γ), where these operons are especially frequent (four cases). Sequences containing three cytochrome *c* domains are all largely similar to each other (overall, the average pairwise identity is around 40%), and the clustering procedure does not distinguish separate groups within this ensemble when using 10^{-10} as the BLAST cutoff. Therefore, the functional classification for these proteins was based on their gene context. Sequences associated with xanthine oxidases thus have been grouped together with sequences associated with glucose−methanol−choline (GMC) oxidoreductases. These systems display a similar genome organization, where a gene encoding a protein with three adjacent cytochrome *c* domains is located in close proximity to a gene encoding the oxidoreductase. We refer to this class as the xanthine−GMC oxidoreductase class (Table 5). GMC oxidoreductases comprise a wide variety of both prokaryotic and eukaryotic enzymes that contain flavin adenine dinucleotide (FAD) as a prosthetic group and catalyze an amazingly diverse range of reactions.[95] An

association between the FAD-binding dehydrogenase domain and the cytochrome *c* domains has been found in the genera *Bordetella*, *Ralstonia* (β), and *Pseudomonas* (γ) and suggests that the active form of the enzyme might be a flavocytochrome *c*, although GMC enzymes do not generally have a multidomain organization.[96] However, a *b*-type cytochrome domain has been recently proposed to be important for the catalytic function of fungal cellobiose dehydrogenase.[96] Other protein sequences containing three cytochrome *c* domains are instead included in the class of penta-cyt *c* operons on the basis of the proximity of the three cytochrome *c* domains to a c_4-type cytochrome (see section 5.1.9).

Bacterial sulfite oxidases catalyze the final step in the pathway of oxidation of reduced sulfur compounds, such as thiosulfate or sulfide, which can be used as an energy source by a variety of microorganisms supporting autotrophic growth.[97] The mechanism of this complex process has not been fully understood yet, but it is thought that sulfite can be oxidized either by multicomponent enzymatic complexes called TOMES (i.e., thiosulfate-oxidizing multi-enzyme system), which are encoded in *sox* gene clusters, or by free enzymes called Sor, which are encoded in isolated *sor* operons.[98−100] Sor enzymes are periplasmic, heterodimeric proteins formed by a catalytic subunit containing a molybdopterin-type cofactor (SorA) and a *c*-type cytochrome subunit (SorB).[101] Very recently, the crystal structure of the heterodimeric SorAB complex from the soil bacterium *Starkeya novella* has been determined.[102] We have identified *sorB* genes only in three organisms, namely, *Bradyrhizobium japonicum* (α), *Chromobacterium violaceum*, and *Ralstonia solanacearum* (β): these cytochrome *c* sequences, approximately 100 residues long, appear to be unrelated to any other type and form a small, separate cluster (SorB class in Table 3 and Figure 4).

According to domain structure, the analogous of Sor within TOMES complexes is the SoxCD enzyme, which was reported to be a heterotetramer comprising a subunit with a Mo center at the active site (SoxC) and a cytochrome *c* subunit (SoxD).[103] Nevertheless, its function as a sulfite oxidase was excluded by experimental data, and it was suggested that SoxCD is instead a sulfur oxidase.[98,103] We have found *soxD* genes in *sox* clusters from the Bradyrhizobiaceae *Bradyrhizobium japonicum* and *Rhodopseudomonas palustris* (α), but we have retrieved also an isolated *soxCD* operon in the genome of *Bradyrhizobium japonicum* itself: these three cytochrome *c* sequences are indeed highly similar and form the SoxD class (Table 3 and Figure 4). Furthermore, we have found two instances of *sox* clusters (in *Bradyrhizobium japonicum* and *Ralstonia solanacearum*) lacking the *soxCD* genes.

A few other cytochrome *c* proteins are encoded within operons that include genes for putative molybdopterin-binding oxidoreductases (see Table S5); however, their sequences do not group to any specific cluster, and no definite functional role can be inferred for them. It is anyway worthwhile mentioning the unique case of *Shewanella oneidensis* (γ), which has an operon encoding three cytochromes *c* together with a molybdoenzyme.

5.1.6. SoxAX Complexes

The SoxAX complex is the best characterized among the proteins that constitute the TOMES (see above) and plays a major role in thiosulfate oxidation.[104] The crystal structure from *Rhodovulum sulfidophilum* showed that it is a heterodimeric protein formed by a mono-heme cytochrome *c* (SoxX) and a di-heme cytochrome *c* (SoxA) featuring a novel type of domain packing.[105] We have detected the presence of both *soxX* and *soxA* genes within *sox* clusters from *Bradyrhizobium japonicum* (two instances), *Rhodopseudomonas palustris* (α), and *Ralstonia solanacearum* (β). Nevertheless, only one of the two SoxAX enzymes retrieved in *Bradyrhizobium japonicum* is actually a tri-heme system; the other three are instead di-heme enzymes, where the SoxA subunit binds only one heme moiety. Irrespective of the number of heme groups, all the SoxA sequences are homogeneous in length (around 280 residues), and the three sequences of the mono-heme form are closely related (SoxA class in Table 3 and Figure 4). SoxX sequences, conversely, exhibit larger length variation (from 167 to 247 residues); however, they also represent a distinct cytochrome *c* group (SoxX in Table 3 and Figure 4). It was suggested that the differing sizes of SoxX subunits may have implications for their function in the SoxAX dimer,[105] but no evidence of such a dependence has been found yet.

For the sake of completeness, we notice here that the *sox* clusters from *Rhodopseudomonas palustris* and *Ralstonia solanacearum* encompass one further gene encoding a cytochrome *c*. In both cases, this gene is adjacent to one encoding a flavoprotein (*soxF*), suggesting that the two products may be associated to form flavocytochrome *c*. It was shown that related proteins isolated from other sources have sulfide dehydrogenase activity in vitro, but they are not required for bacterial growth with hydrogen sulfide;[106,107] therefore their function in vivo is unclear.

5.1.7. Cytochrome c Domains in Alcohol Dehydrogenases

Bacterial respiratory processes may involve the donation of electrons by a variety of alcohols, which are oxidized by specific dehydrogenases. In particular, type II quinohemoprotein alcohol dehydrogenases are periplasmic enzymes that transfer electrons from the substrate first to pyrroloquinoline quinone (PQQ) and then to an internal heme group, which is found within a *c*-type cytochrome domain.[108] Such cytochrome subunits (about 100 residues long) have been retrieved in *Bradyrhizobium japonicum* (α) (two instances) and *Pseudomonas* (γ) and form the class referred to in Table 3 as that of alcohol dehydrogenases (ADH in Figure 4). Analogous associations of single cytochrome *c* domains with PQQ-dependent enzymes have been found also in *Rhodopseudomonas palustris* (α) and *Xanthomonas* (γ), yet their sequences do not cluster with those mentioned above.

5.1.8. Cytochrome c Domains in Iron Uptake and Solute Transport

A remarkable gene fusion is observed in *Nitrosomonas europaea* (β) and the three organisms of the *Pseudomonas* genus (γ) between cytochrome *c* and a putative iron permease of the FTR1 type. FTR1 permeases are known to be involved in high-affinity uptake systems for iron, which were first described in the yeast *Saccharyomyces cerevisiae*:[109] the mechanism of acquisition implies the oxidation of Fe(II) to Fe(III) by a multi-copper oxidase and was suggested to be important also in bacterial aerobic growth, as well as in infection of animal hosts by pathogens.[110−112] The cytochrome *c* domains fused to FTR1 permeases are contained in approximately 650 residue long proteins, and their sequences are included in the cluster labeled "iron uptake" in Table 3 and Figure 4. Nevertheless, these genes lack an

indicative organization in operons that may support such a functional annotation, or others. Whatever molecular machinery for iron transport involves FTR1, the possibility that cytochrome *c* might actively participate in iron uptake has not been considered until now and may deserve further studies. Such a speculation is also inferred and somehow extended by the singular case of *Nitrosomonas europaea* (β), where one gene encodes cytochrome *c* fused to a putative CopD protein within an operon that comprises also a putative CopC protein. CopD and CopC are involved in copper homeostasis: the former is thought to function as an inner membrane transporter conveying copper from the periplasm to the cytoplasm, while the latter is probably involved in copper mobilization in the periplasmic space.[113] A similar association (Supporting Information Table S5) occurs in *Pseudomonas putida* (γ), where one cytochrome *c* is coexpressed with a putative CopB protein, which is an integral outer membrane protein also contributing to protection from copper toxicity. Copper levels in cells are tightly controlled, since free copper ions can participate in redox reactions generating highly reactive, harmful radical species: the precise role of the different components underlying this sophisticated mechanism is still an open question.[113] Notably, the sequence of the above-mentioned *Nitrosomonas europaea* cytochrome *c* is clustered with those fused to FTR1 in the "iron uptake" class, hinting at some possible role for these cytochromes *c* in ion homeostasis, at least in some organisms. In this regard, it is worth mentioning that we have also found a number of cytochromes *c* fused or coexpressed with solute transporter proteins, although their sequences are weakly related: the cluster called "solute transporters" in Table 3 and Figure 4 actually comprises only proteins from α-proteobacteria (*Caulobacter crescentus* and Bradyrhizobiaceae), even if instances of this association occur also in γ-(*Pseudomonas putida*) and δ-proteobacteria (*Desulfovibrio vulgaris*, *Geobacter sulfurreducens*). Most of these solute transporters belong to the ATP-binding cassette (ABC) superfamily,[114] but members of the major facilitator superfamily (MFS)[115] and of the multidrug and toxic compound extrusion (MATE) family[116] have been found as well. Gram-negative bacteria use these systems to pump a wide variety of macromolecules, substrates, and metabolites across the two membranes of their envelope, as well as to export small toxic molecules such as drugs and heavy metals.[117,118] It is now well-known that multidrug and drug-specific efflux systems are responsible for resistance to chemotherapeutic agents in pathogenic bacteria;[118] therefore the understanding of the mechanism and the specificity of these molecular machines is crucial in the struggle against bacterial diseases.

5.1.9. Cytochrome c_4 and Multi-Cytochrome c Operons

Cytochrome c_4 is an approximately 200 residue long periplasmic electron carrier that is structured into two strongly symmetric cytochrome *c* domains.[119,120] Genes encoding such proteins (cytochrome c_4 class in Table 3 and cyt c_4 in Figure 5) are encountered, usually in multiple instances, in the genomes of all the β-proteobacteria, some genera of γ-proteobacteria (*Pseudomonas*, *Shewanella*, *Vibrio*, and *Xanthomonas*) and a few α organisms (*Bradyrhizobium japonicum*, *Rhodopseudomonas palustris*, and *Rhodobacter sphaeroides*). It was suggested that these proteins may use one heme as the entrance and the other heme as the exit for the electron transfer, thus functioning as an electron wire that connects the donor and the acceptor proteins.[119] The donor protein has been identified in the blue-copper protein rusticyanin.[121] In particular, it was hypothesized that cytochrome c_4 may be tightly bound to cytochrome *c* oxidase in a membrane complex whose arrangement would resemble the organization of cbb_3 oxidases.[119] Nevertheless, a specific role for cytochrome c_4 has not yet been directly revealed, and the analysis of the genomes does not show any obvious association with other genes that might indicate its functional interactions. An exception in this respect is provided by the *pvcD* gene of *Pseudomonas aeruginosa* (γ) (see Supporting Information Table S5), which encodes a c_4-type cytochrome and is part of the *pvc* gene cluster involved in the biosynthesis of the chromophore moiety of the siderophore pyoverdine.[122] Also, one of the genes retrieved in *Vibrio parahaemolyticus* (γ) was annotated as the di-heme cytochrome subunit of a flavocytochrome *c* sulfide dehydrogenase (FCSD, see Table S5), but this attribution is not supported by genome examination.

In the same *Vibrio parahaemolyticus*, as well as in *Bordetella* (β), *Pseudomonas*, and *Shewanella* (γ), one of the cytochromes c_4 is coexpressed with another di-cytochrome *c* protein, whose sequence is unrelated to that of c_4. In fact, these proteins share relatively larger similarity with other uncharacterized cytochromes from *Brucella* (α), *Campylobacter jejuni*, and *Wolinella succinogenes* (ϵ) and form a separate cluster lacking a definite classification (tetra-cyt *c* operons in Table 4 and Figure 5). Furthermore, cytochrome c_4 in *Bradyrhizobium japonicum* (α), *Bordetella*, and *Nitrosomonas europaea* (β) is coexpressed with tri-cytochrome *c* proteins, thus constituting operons encoding no fewer than five cytochrome *c* domains (penta-cyt *c* operons in Table 5, see also section 5.1.5). The fusion of such operons into single genes is presumably also at the origin of the sequences found in *Rhodopseudomonas palustris* (α) and *Pseudomonas aeruginosa* (γ) containing five domains (see Supporting Information Table S5). The rationale for the presence of these tetra- and penta-cytochrome *c* blocks in several genomes is not obvious: it may involve a direct functional interaction of the proteins encoded within complexes or even as a single polypeptide chain, but it may also represent a convenient way to produce separate components of a metabolic system. A characterization of these cytochromes *c* is needed to elucidate their role.

5.1.10. Single-Domain Cytochromes c: Cytochromes c_2, c_5, c_{551}/c_{552}, c_{553}, and c_{552}/c_{554}

Cytochrome c_2 is the closest bacterial homologue of mitochondrial cytochrome *c*. It mediates electron transfer between bc_1 complexes and cytochrome *c* oxidases during aerobic growth, as well as between bc_1 complexes and photosynthetic reaction centers in some phototrophic bacteria, such as *Rhodobacter sphaeroides*.[123] All the α-proteobacteria taken into account in the present work possess at least one cytochrome c_2, which we have retrieved also in *Bordetella* (β), *Xanthomonas*, *Vibrio parahaemolyticus*, and *Pseudomonas syringae* (γ). Several α species exhibit two or three isoforms, which may be soluble or membrane-bound and might be specialized for reaction with different oxidases or for photosynthetic electron transfer, presumably depending on their redox potentials.[124] All the c_2 domains detected are about 100 residues long and constitute a well-defined class (cytochrome c_2 in Table 3 and cyt c_2 in Figure 4). For the most part, proteins of this class fit into class IB of the traditional subdivision by Ambler,[3] which includes eukaryotic cytochromes *c* and prokaryotic "short" cytochromes c_2

exemplified by the protein from *Rhodopseudomonas globiformis*.[125] Instances of "long" (due to some additional loops) cytochromes c_2 included in Ambler's class IA and represented by the protein from *Rhodospirillum rubrum*[126] appear to be restricted to *Rhodobacter sphaeroides* and *Rhodopseudomonas palustris*.

Cytochromes of the c_5 class (Table 3, cyt c_5 in Figure 4) have been retrieved mainly in γ-proteobacteria (*Pseudomonas*, *Shewanella*, *Vibrio*, and *Xanthomonas*), even if single instances occur in α (*Caulobacter crescentus*) as well as in β (*Chromobacterium violaceum*). They are characterized by a smaller size with respect to c_2 (about 80 residues) and by the presence of an uncommon disulfide bridge,[3] which however is lacking in the *Xanthomonas* proteins. These proteins fit into Ambler's class IE, illustrated by cytochrome c_5 from *Azotobacter vinelandii*, whose structure was determined 20 years ago.[127] The organisms of the genus *Pseudomonas* have two or three isoforms, out of which at least one is soluble and one is membrane-anchored. β-Proteobacteria (except for *Nitrosomonas europaea*) are the only possessors of a di-heme form (cytochrome c_5 (di-heme) in Table 4 and cyt c_5 in Figure 5), presumably originated by gene duplication.

Cytochromes of the c_{551}/c_{552} class (Table 3, cyt c_{551}/c_{552} in Figure 4) are found mainly in the equipment of electron carriers of β-proteobacteria, all of which (except *Neisseria meningitidis*) make use of these 80 residue long proteins in their respiratory network. This group includes also the above-mentioned cytochrome c_{551} from *Pseudomonas aeruginosa* (γ), which functions as an electron donor for the cd_1 nitrite reductase of this organism. In Ambler's classification, it corresponds to class ID, for which he proposed also the name cytochrome c_8: sequences included in this class, exemplified by *Hydrogenobacter thermophilus* cytochrome c_{552},[128] have several proline residues around the sixth ligand methionine and a trytophan residue near the C-terminus.[3]

On the other hand, mono-heme cytochrome c_{553} (Table 3, cyt c_{553} in Figure 4) appears to be a characteristic element of the electron transport chains in ϵ-proteobacteria, where it may be found in one (*Helicobacter pylori*) or two different genes (*Campylobacter jejuni*, *Helicobacter hepaticus*, and *Wolinella succinogenes*). These proteins are found also in γ- (*Pasteurella multocida* and *Yersinia pestis*) and δ-proteobacteria (*Desulfovibrio vulgaris*), where the sequence is predicted to be slightly longer than the typical 80 residues. They generally fit into Ambler's class IC, whose member proteins possess a widened or split α-band of lowered absorptivity.[3] A representative structure was determined for the protein from *Desulfovibrio vulgaris*.[129] The di-heme variant of cytochrome c_{553} (cytochrome c_{553} (di-heme) in Table 4 and cyt c_{553} in Figure 5) appears to be restricted to some α-proteobacteria, since it is encoded only in Bradyrhizobiaceae, Rhizobiaceae, *Mesorhizobium loti*, and *Rhodobacter sphaeroides*.

The c_{552}/c_{554} class (Table 3, cyt c_{552}/c_{554} in Figure 4) is formed by an ensemble of approximately 80 residue long sequences retrieved mainly in β- (*Bordetella*, *Nitrosomonas europaea*, and *Ralstonia solanacearum*) and γ-proteobacteria (*Shewanella*, *Vibrio*, *Xanthomonas*, and *Yersinia pestis*), along with a single occurrence in *Rhodopseudomonas palustris* (α). Similarly to c_{553}, these sequences generally fit into Ambler's class IC. The cytochromes of this group are remarkable in that they are encoded by two different genes in all the aforesaid β and γ organisms excluding *Xanthomonas campestris* and *Yersinia pestis*, and the two genes are organized in the same operon with the exception of those from *Vibrio*, which are distant from each other.

5.1.11. Photosynthesis-Related Single-Domain Cytochromes c: Cytochromes PS-c_{550}, c_6, and c_M

Photosynthesis is the conversion of light energy into chemical energy by means of two large membrane-integral protein complexes: photosystem I (PSI) and photosystem II (PSII). The catalytic core of each system is generally referred to as the reaction center (RC), which is classified according to the kind of its terminal electron acceptor, namely, Fe_4S_4 clusters for type I and quinones for type II.[130] Like eukaryotic algae and higher plants, cyanobacteria possess electrochemically linked PSI and PSII and are able to produce molecular oxygen through water oxidation. All the other phototrophic bacteria have only one RC, either of type I (Heliobacteria, green sulfur bacteria) or of type II (purple bacteria, green filamentous bacteria), and are anoxygenic.[131]

Cytochrome PS-c_{550} is a subunit of PSII of cyanobacteria with no apparent redox role: it is one of three extrinsic proteins located on the lumenal (i.e., interior) surface of the enzyme, which serve to insulate the catalytic site from reductive attack and contribute to stabilization of the structure of the complex.[132,133] Cytochrome c_6 is known to function interchangeably with the copper-containing plastocyanin as the electron donor to PSI in cyanobacteria, the relative synthesis of the two proteins being regulated by copper availability.[134−136] This circumstance does not occur in green plants, where plastocyanin is the exclusive electron donor to PSI; it was hypothesized that cytochrome c_6 may be the older evolutionary donor, which was then replaced by plastocyanin in response to iron limitations in the environment.[137] Instances of cytochrome PS-c_{550} have been identified exclusively in the genomes of cyanobacteria, which encode one (*Nostoc*, *Prochlorococcus marinus*, *Synechococcus*, and *Synechocystis*) or two (*Gloeobacter violaceus* and *Thermosynechococcus elongatus*) such proteins, with sequence lengths ranging from 160 to 180 residues. Likewise, one (*Prochlorococcus marinus*, *Synechocystis*, and *Thermosynechococcus elongatus*) or two (*Gloeobacter violaceus*, *Nostoc*, and *Synechococcus*) genes encoding cytochrome c_6, which has a typical size of about 80 residues, have been found in all the cyanobacteria; PS-c_{550} and c_6 sequences form separate clusters (see Table 3), called PS-cyt c_{550} and cyt c_6, respectively, in Figure 4. However, we have included in the c_6 class also one protein from *Nitrosomonas europaea* (β) and two proteins from *Geobacter sulfurreducens* (δ) that are not photosynthetic. These proteins share relatively high sequence similarity to cyanobacterial c_6 (about 50%) and lack a functional characterization.

We mention here also cytochrome c_M because it has been found solely in cyanobacteria, but its physiological function is still unknown, and its involvement in photosynthesis is questioned. It is a soluble protein that in *Synechocystis* is expressed only under stress conditions such as low temperature or exposure to high-intensity light, when the synthesis of both plastocyanin and cytochrome c_6 is suppressed.[138] Hence, it was supposed that cytochrome c_M could be a third stress-induced electron donor to PSI, but evidence was provided against this hypothesis.[139] A single cytochrome c_M gene has been found for each of the available cyanobacteria, and the related protein sequences cluster separately from

those of PS-c_{550} and c_6 (cytochrome c_M class in Table 3 and cyt c_M in Figure 4).

5.1.12. An Oxygen-Binding Single-Domain Cytochrome c: SHP

SHP (Sphaeroides heme protein) is an unusual *c*-type cytochrome with a high-spin heme, which was discovered in *Rhodobacter sphaeroides* (α) and is capable of transiently binding oxygen during auto-oxidation.[140] The crystal structure of the oxidized protein from *Rhodobacter sphaeroides* revealed an atypical histidine/asparagine coordination for iron(III). The asparagine ligand moves away from iron upon reduction or binding of small molecules such as cyanide or nitric oxide. It was also observed that the distal pocket of the heme bears a notable resemblance to other heme proteins that bind gaseous compounds.[141] We have retrieved one gene encoding the SHP protein in *Shewanella oneidensis* (γ) in addition to that from *Rhodobacter sphaeroides*: the two sequences are about 110 residues long and are the unique members of the SHP class (Table 3 and Figure 4). SHP was suggested to be the terminal electron acceptor of an electron-transfer pathway in which it could reduce a small ligand like peroxide or hydroxylamine.[141]

5.1.13. An Intriguing Gene Association between Cytochrome c and WD40 Repeats

It is quite interesting that fusion of a cytochrome *c* domain with one or more WD40 repeats is found in *Bradyrhizobium japonicum* (α), as well as in the planctomycete *Rhodopirellula baltica* (see Supporting Information Table S5 and below). Unfortunately, the analysis of the genomic context of these genes does not provide hints about their role in these organisms. As described in section 4, the interaction between cytochrome *c* and WD40 repeats is a key step for apoptosis in Eukaryota. The present finding may suggest that the recruitment of cytochrome *c* to regulate biochemical processes through interaction with WD40 repeats may have been inherited by Eukaryota from their ancestors. Because apoptosis is absent in bacteria, in these organisms the same interaction between cytochrome *c* and WD40 could regulate other processes, not known at present.

5.1.14. "Unknown" Functional Classes

Not all the sequences retrieved in the present search could be assigned to a specific functional class, either because the similarity to other sequences was not significant enough to allow unambiguous assignment to a given cluster or because the classification determined by the clustering procedure was not supported by the gene context. As a consequence, 69 proteobacterial and cyanobacterial proteins in Supporting Information Table S5 lack such an assignment. Nevertheless, a functional prediction could be given for 26 of these 69 proteins on the basis of the operon structures and the available gene annotations. Even if most of the unclassified sequences are likely examples of divergent evolution restricted to single lineages, clustering of single cytochrome *c* domains highlighted three groups of proteins, which have been annotated here as "unknown I", "unknown II", and "unknown III" (Table 3 and Figure 4). Members of each group share relatively high sequence similarity only with the other proteins of the same group, and none of them exhibits a gene context that can suggest a possible role. Therefore, cytochromes *c* belonging to these classes could be interesting targets for structural and biochemical characterization.

5.2. Other Gram-Negative Bacteria

Aquifex aeolicus is a hyperthermophilic organism (it grows optimally at 85 °C[142]) that belongs to the phylum Aquificae and is able of both oxidizing molecular hydrogen and reducing molecular oxygen.[143] Its battery of cytochromes *c* (Table S5) includes functional types already encountered in Proteobacteria, namely, a bc_1 complex (which was experimentally characterized[144]), a di-heme cytochrome *c* peroxidase, a SoxAX complex with a di-heme SoxA subunit,[145] and three single-domain cytochromes, two of which have been characterized: they are highly similar in sequence and biochemical properties and are both located in the periplasm, although one is soluble and one is membrane-bound.[146,147]

A similar set of cytochromes *c* is found (Supporting Information Table S5) in the aerobic obligate heterotroph *Thermus thermophilus*, which is also an extreme thermophile with a maximum growth temperature of about 85 °C:[148] indeed, *Aquifex aeolicus* and *Thermus thermophilus* are the only examples among the genome sequences analyzed here of eubacterial species showing such an adaptation to extreme environments, a feature that is usually associated with Archaea. The genome of *Thermus thermophilus* encodes a large *sox* cluster comprising two SoxAX complexes, both with a mono-heme SoxA subunit, and a SoxCD complex. Five single-domain mono-heme cytochromes *c* and one di-cytochrome *c* protein have been also retrieved: one of the mono-heme proteins is a highly thermostable cytochrome c_{552}, which is the specific electron donor to a ba_3-type cytochrome *c* oxidase[149] and whose structure was solved.[150,151] At variance with *Aquifex aeolicus*, no cytochrome *c* peroxidase is present, but a caa_3 oxidase is.[152] This oxidase can also catalyze the reduction of nitric oxide to nitrous oxide under anaerobic conditions.[153]

The genome of *Chlorobium tepidum* is the only one available for a photosynthetic green sulfur bacterium. These bacteria possess a type I reaction center (RC) whose PscC subunit is a membrane-bound cytochrome c_{551}.[154] It is unclear whether this cytochrome is the physiological electron donor to the RC or this role is played by a smaller, water-soluble species.[155] In addition to PscC (indicated in Supporting Information Table S5 as PS-cyt c_{551}), *Chlorobium tepidum* encodes nine other mono-heme cytochromes *c*, including a SoxAX complex, a subunit of a flavocytochrome *c* with sulfide dehydrogenase activity,[156] and a soluble cytochrome c_{555}, which was suggested to transfer electrons between SoxA and the photosystem.[124]

Leptospira interrogans is an obligate aerobic spirochaete responsible for leptospirosis, which has emerged in the past decade as a globally important infectious disease.[157] We have retrieved (Supporting Information Table S5) four genes encoding putative cytochrome *c* peroxidases, one for a caa_3 oxidase and six other for single-domain cytochromes *c*, one of which is coexpressed with a molybdopterin-containing oxidoreductase. No literature data are available on the characterization of these proteins.

Bacteroides thetaiotaomicron and *Parachlamydia UWE25* are both symbionts, the former being a dominant member of the intestinal microflora of humans and other mammals[158] and the latter being an obligate intracellular symbiont of free-living amoeba.[159] *Parachlamydia* is nonpathogenic and provides an instance of the recent recognition of the diversity of chlamydiae, in opposition to the previously established belief that these organisms were exclusively pathogens of mammals.[148] The presence of cytochromes *c* within the

genomes of these two bacteria is very limited (Supporting Information Table S5): *Bacteroides thetaiotaomicron* encodes a cytochrome *c* peroxidase fused to a third cytochrome *c* domain and a putative surface layer protein (about 600 residues long) containing a cytochrome *c* domain at the C-terminus, while *Parachlamydia* has only one gene for a di-cytochrome *c* protein; this gene is located in near proximity to one encoding a protein similar to the DoxD subunit of a terminal quinol oxidase isolated from the archaeon *Acidianus ambivalens*.[160] Generally speaking, such a poor apparatus of cytochromes *c* is not surprising in symbiotic as well as in pathogenic species, which can develop metabolism and genome reduction to adapt within specific niches.[161]

A completely different scenario is provided by the genome sequence of the marine planctomycete *Rhodopirellula baltica* (formerly called *Pirellula*), which is representative of an abundant and environmentally important phylum.[162] This aerobic, heterotrophic bacterium displays an amazingly vast array of cytochromes *c*, which amounts to as many as 96 domains within 79 proteins (see Supporting Information Table S5). Thirty-eight of these hits correspond to novel, planctomycete-specific cytochrome domains referred to as PSC3 in the Pfam database: such domains have been recently identified by Studholme et al.,[163] who observed that no function could be reliably predicted for an unusually high proportion of the proteins encoded in the *Rhodopirellula baltica* genome (over 60%) and thus searched for conserved sequence motifs and domains in those proteins. In 16 cases, cytochrome *c* is found within genes encoding also a putative glucose/sorbosone dehydrogenase, while in three cases it is associated with WD40 repeats (see section 5.1.13). Three instances of a cytochrome *c* peroxidase have been retrieved, as well as two caa_3-type and two cbb_3 oxidases: one of the two cbb_3 enzymes has both the mono-heme and the di-heme subunit encoded in the same gene, whereas the other has two mono-heme subunits, similar to *Bdellovibrio bacteriovorus*. Research on *Rhodopirellula baltica* and planctomycetes is just at its beginning, and much experimental work will be needed to understand the biology of these unique organisms.

6. Gram-Positive Bacteria

The cell wall of Gram-positive bacteria is composed of a single membrane, which delimits the cytoplasmic space (therefore called cytoplasmic membrane), enclosed by a thick peptidoglycan layer. Consequently, the homologues of the proteins that would perform their function in an intermembrane space, such as the periplasm of mitochondria or of Gram-negative bacteria, are generally exposed to the extracellular medium in Gram-positive bacteria. Some way of anchoring these proteins to the cell is thus needed to avoid their dispersion in the medium. There are different ways to achieve membrane anchoring, such as fusing one protein terminus to a lipophilic molecule, which "solubilizes" in the membrane, or adding one or more transmembrane segments (again, typically, at one terminus). Fusion with a membrane protein is also observed. The use of one of these means for membrane anchoring is observed for all cytochromes *c* from Gram-positive bacteria identified by our search (54 domains in 42 sequences, see Supporting Information Table S7). In Actinobacteria (10 instances) and *Bacillus* (six instances), we have found, respectively, one and four such proteins. The one protein detected in Actinobacteria contains two cytochrome *c* domains, while each of the four *Bacillus* proteins contains one. On the other hand, no cytochrome *c* domains could be retrieved in anaerobic Gram-positive organisms, such as *Staphylococcus* or *Streptococcus*.

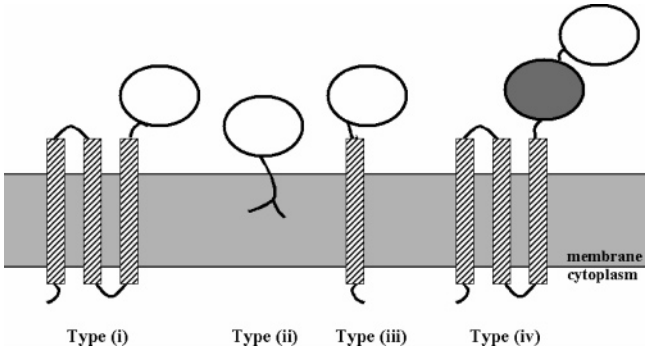

Figure 9. Organization of the four cytochrome *c* domains (white blocks) in a bacterium of the genus *Bacillus*. The gray block is the Cu_A-containing domain of subunit II of caa_3 oxidase. Transmembrane regions are all depicted as striped rectangles regardless of any sequence/functional relationship.

The four instances of *Bacillus* cytochrome *c* domains, depicted in Figure 9, are as follows: (i) a cytochrome *c* protein fused to three N-terminal transmembrane domains, respectively constituting the cytochrome *c* subunit and the fourth subunit (SUIV) of a *bc* complex (see below), (ii) a cytochrome c_{551} anchored to the membrane via linkage to a lipophilic molecule, (iii) a cytochrome c_{550} with an N-terminal transmembrane segment, and (iv) a cytochrome *c* domain fused to subunit II of a caa_3-type oxidase. The caa_3 oxidases have been already described in section 5.1.1. Several single-domain cytochromes *c* belonging to the above-mentioned types ii and iii (see Figure 9) have been structurally characterized.[27] The role of these proteins, in particular their relative levels of expression under different environmental conditions, is not known in detail. Even though the fact that cytochromes *c* are not detected in anaerobic Gram-positives could suggest a role exclusively in aerobic respiration for the Actinobacteria and *Bacillus* proteins, it is worth noting that *Bacillus* organisms can grow also anaerobically. Consequently, it cannot be excluded that their cytochromes may be involved in different pathways, for example, depending on O_2 pressure.

The *Bacillus* homologue of QCR (see sections 4 and 5.1.2) has been proposed to be a three-protein complex, carrying out reduction of either of the two smaller cytochromes *c* at the expenses of menaquinone.[164] The structure of the operon shows that it is formed by a Rieske protein, a b_6-type cytochrome, and a fused protein (QcrC), about 250 residues long, consisting of a three-helix transmembrane subunit and a cytochrome *c* domain spanning less than 100 residues (type i in Figure 9). With respect to previously discussed bc_1 complexes (Section 5.1.2), two major differences thus emerge: first, the cytochrome *c* domain is not of the c_1 type; second, the cytochrome *b* is split into a cytochrome b_6 subunit (corresponding to the N-terminal part of *b*) and a transmembrane subunit called SUIV (corresponding to the C-terminal part of *b*). The presence of cytochrome *b* as two separate proteins resembles the organization of the cyanobacterial b_6f complexes involved both in respiration and in photosynthesis[165,166] and was suggested to be due to cleavage of the gene encoding cytochrome *b*.[62] All *Bacillus* QcrC sequences are very similar to each other, with sequence identity values ranging from 50% to 70%. In each organism, QcrC is similar

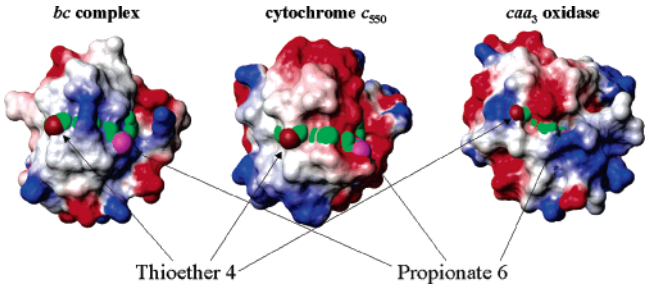

Figure 10. Electrostatic surface potentials (red = negative charge, blue = positive charge) based on structural models of the cytochrome *c* domains of *Bacillus subtilis* cytochrome *bc* complex, cytochrome c_{550}, and caa_3 oxidase. Heme atoms are shown as green spheres, except for the carbon CBC of thioether 4, depicted as a brown sphere, and the oxygens O1D and O2D of propionate 6, depicted as magenta spheres.

in sequence also to the single-domain cytochromes *c* with sequence identity values ranging from 30% to 40%. The final recipient of electrons from menaquinone is the terminal caa_3 oxidase, via interaction of the donor with the cytochrome *c* domain of subunit II (type iv in Figure 9) of the latter.[167] Note that caa_3 is not the only terminal oxidase of *Bacillus* bacteria: the caa_3 oxidase is actually expressed under a relatively small number of environmental conditions, such as nutrient limitation.[168]

To obtain some hints on their possible interaction, we have attempted to perform homology modeling of the cytochrome *c* domains of *Bacillus subtilis*, although the cytochrome *c* domain of the caa_3 oxidase has poor sequence similarity to any cytochrome of known structure (for details see Methods, section 2). On the other hand, the use of template structures with the same cytochrome *c* fold should provide models with a reasonable reliability, at least in regard to general structural features such as the charge distribution over the protein surface. In this respect, the obtained structural models show a poor complementary character of the electrostatic potential surface at the putative interaction regions (Figure 10), based on the configuration seen in the eukaryotic complex between bc_1 and cytochrome *c*[40] and the configuration predicted for the complex between cytochrome *c* and CCO.[169] This may suggest that interprotein interaction is mainly driven by nonpolar contacts. In this respect, it is noteworthy that horse heart cytochrome *c* (positively charged) can transfer electrons to *Bacillus subtilis* caa_3, albeit less efficiently than cytochrome c_{550} of the same organism.[167] Given all the above observations, one could also speculate that electron transfer in *Bacillus* may occur directly between the *bc* complex and the caa_3 oxidase without the intervention of single-domain cytochromes *c* in between. In this case, the electron transfer would take place through a direct contact of the cytochrome *c* domain of caa_3 with the QcrC subunit of the partner.

Actinobacteria only contain a single instance of a di-cytochrome *c* protein (each of the two cytochrome *c* domains being 75−80 residues long), which is encoded in an operon containing also a Rieske protein and a cytochrome *b*. This occurrence is comparable to what is seen for ϵ-proteobacteria (see section 5.1.2). For *Corynebacterium glutamicum*, it has been demonstrated that these three proteins form an adduct that is analogous to the bc_1 complex;[170] the above three proteins thus correspond to QcrABC, with the di-cytochrome *c* protein, called cytochrome *cc* after a suggestion of Sone et al.,[170] having the role of cytochrome c_1. The topology of these QcrC variants is also analogous to that of cytochrome c_1 (and thus different from that of the cytochrome *c* subunit of *Bacillus* QCR), in that the two soluble cytochrome *c* domains are between two predicted transmembrane regions: similar to eukaryotic cytochrome c_1 precursors (see Section 4), the N-terminal transmembrane segment is presumably a signal sequence not present in the mature protein.[170] Adjacent to the QCR operon there is always a gene encoding a protein similar to subunit III of CCO. In various instances, proteins related to other CCO subunits are also found close to the QCR operon. A "supercomplex" between QCR and CCO (of the aa_3 kind rather than the caa_3 observed in *Bacillus*) was purified from *Corynebacterium glutamicum* and showed quinol oxidase activity,[171] indicating that the second cytochrome *c* domain of the QcrC subunit effectively mediates electron transfer from the first domain to CCO. The experimental observation of a super-complex in Actinobacteria somehow reinforces the possibility that an analogous adduct is formed in *Bacillus*, where the bridging cytochrome *c* domain is fused to CCO rather than to QCR.

In the strict aerobe *Deinococcus radiodurans*, the situation is different from all the other organisms discussed above, with eight different cytochrome *c*-containing proteins identified. These comprise different combinations of a single-domain cytochrome with membrane anchors, as well as a putative di-cytochrome *c* peroxidase (see section 5.1.3). A three-gene operon similar to QCR can also be detected, formed by a Rieske protein, a cytochrome *b* and a cytochrome *c*. The latter is quite atypical: it is 335 residues long and contains one N-terminal and three C-terminal predicted transmembrane helices with the soluble cytochrome *c* domain spanning residues 95−174. Unfortunately, the analysis of neither gene neighbors nor the literature provides more functional clues on the *Deinococcus* proteins.

7. Archaea

Despite their importance to the biosphere, Archaea are the most poorly understood domain of life. All of them can be regarded as extremophilic by virtue of their adaptation to extreme environmental conditions such as temperature, pH, or salt concentration. Most known species are strictly anaerobic, but several organisms can carry out oxygen respiration, and some are obligate aerobes.[172]

The adoption of cytochrome *c* in bioenergetic routes as well as in other metabolic pathways does not appear to be popular in Archaea, since we have not retrieved any such protein in 16 out of the 20 genomes analyzed (see Tables 1 and 2). The largest number of cytochrome *c* instances occurs in *Methanosarcina acetivorans*, a versatile methanogen that can use no fewer than nine methanogenic substrates, including acetate:[173] its genome encodes a mono-heme cytochrome *c* within a 200 residue long sequence and two cytochrome *c* peroxidases. Cytochrome *c* peroxidases presumably function like the MauG proteins in the metabolism of methylamine, which can be used as a methanogenic substrate. Remarkably, one of these two enzymes (20091744 in Supporting Information Table S8) represents the only apparent example of horizontal gene transfer for all the cytochromes *c* identified in the present work. Although a detailed treatment of this topic is beyond our scope, we have found that the above-mentioned CCP-Maug protein yields the best BLAST match with a CCP-Maug protein (39996638 in Supporting Information Table S5) from the δ-proteobacterium *Geobacter sulfurreducens*, and the opposite is also true. Assuming such bidirectional BLAST matches as indicators of orthologous

relationships, the emerging picture is that horizontal transfer of cytochrome *c* genes between phylogenetically distant organisms is a rare event, where the import of a CCP-Maug by *Methanosarcina acetivorans* represents a singular exception.

Archaeoglobus fulgidus is an anaerobic hyperthermophile that can use lactate, pyruvate, or hydrogen to grow;[174] *Archaeoglobus* is also the only known archaeal genus capable of sulfate reduction to obtain energy.[175] Its genome encodes only one cytochrome *c* (Supporting Information Table S8), which was suggested to serve as an electron acceptor for a NADH oxidase possibly involved in sulfate respiration.[176]

Pyrobaculum aerophilum and *Aeropyrum pernix* are atypical crenarchaeota, because the members of this phylum of Archaea are almost exclusively anaerobes, whereas *Pyrobaculum aerophilum* is capable of growing aerobically[177] and *Aeropyrum pernix* has been reported to be strictly aerobic;[178] *Pyrobaculum aerophilum* can also use nitrate reduction to produce energy.[179,180] The genome of *Pyrobaculum aerophilum* encodes two mono-heme cytochromes *c* (Table S8): one is found within an operon containing also a Rieske protein and a *b*-type cytochrome, suggesting that it might be a component of a bc_1-like complex, while the other is annotated as one subunit of a putative cd_1 nitrite reductase. This attribution, however, is in contrast with experimental data excluding the presence of *c*-type cytochromes among the components of the denitrification pathway.[181] *Aeropyrum pernix*, finally, has a single cytochrome *c* (Table S8), whose gene is located in near proximity to one encoding a putative cytochrome *c* oxidase subunit: this organism has been shown to have both a ba_3-type and an aa_3-type terminal oxidase,[182] thus it is presumable that the cytochrome *c* retrieved is involved in electron transport in aerobic respiratory chains.

8. A Comment on Multi-Heme Cytochromes *c*

In section 5.1.5, we have mentioned that periplasmic DMSO reductases make use of multi-heme cytochromes anchored to the membrane to accomplish the task of moving electrons from the quinone pool to the periplasmic compartment. These proteins bind *c*-type hemes, and are therefore referred to as cytochromes *c*. Nevertheless, they do not have a uniquely defined structural fold and do not fit into the definition of cytochrome *c* domain that has been adopted here. As a matter of fact, the recruitment of such multi-heme cytochromes to mediate electron transfer from the membrane to soluble periplasmic enzymes, as well as within redox enzymes, is common to several bacterial respiratory systems. Some examples of processes involving multi-heme cytochromes *c*, which can be both soluble and membrane-anchored, include nitrate reduction in *Paracoccus denitrificans*, where a tetra-heme cytochrome *c* (NapC) is implicated,[183] fumarate reduction in *Shewanella frigidimarina*, involving an iron-induced tetra-heme flavocytochrome c_3 (Ifc_3),[184] hydroxylamine oxidation in *Nitrosomonas europaea*, carried out by a hydroxylamine oxidoreductase (HAO) containing eight heme groups,[185] and nitrite reduction in *Sulfurospirillum deleyianum*, which is due to a penta-heme nitrite reductase (NrfA).[186] It was pointed out that the above-mentioned Ifc_3, HAO, and NrfA share a conserved arrangement of the heme groups, suggesting an evolutionary relationship among these proteins.[92] This arrangement is common also to the tetra-heme cytochrome c_{554} from *Nitrosomonas europaea*,[187] and is presumably imposed by the requirements of the electron-transfer process. On the other hand, different heme organizations are observed in the tetra-heme cytochrome subunit of the photosynthetic reaction center from *Rhodopseudomonas viridis*[188] and in the members of the multi-heme cytochrome c_3 family. This family includes cytochrome c_3 itself, which binds four heme groups,[189] the tri-heme cytochromes c_7,[190] nona-heme cytochromes,[191] and the HmcA protein from *Desulfovibrio vulgaris*, which binds 16 heme groups.[192] This variability in the number of heme groups is due to different combinations of tandem repeats, originated by gene duplication, of the cytochrome c_3 tetra-heme unit, deletion (as in the case of c_7) or addition (as in the case of nona-heme proteins) of heme binding sites. Proteins of the cytochrome c_3 family are found in many sulfate-reducing bacteria, where they are hypothesized to transfer electrons from periplasmic hydrogenases to membrane proteins within the process of sulfite reduction.[193] A further example of the employment of multi-heme cytochromes in bacterial systems is provided by Fe(III) respiration. Since Fe(III) is insoluble at pH values higher than 2, anaerobic growth of a Gram-negative bacterium using Fe(III) as the terminal electron acceptor implies that the electrons produced by the metabolic enzymes located in the cytoplasm are transferred through the two cell membranes and the periplasm to an outer-membrane protein, where iron is reduced. In *Shewanella frigidimarina*, it appears that this far-reaching electron transfer is achieved by several multi-heme cytochromes connected to form an electron "wire", which are expressed under conditions inducing Fe(III) respiration.[184] A related protein from *Shewanella oneidensis* was recently characterized.[194]

9. Conclusions

In the present work we have browsed genome sequences with the aim of compiling an extensive list of proteins containing at least one mono-heme cytochrome *c* domain, based on the conserved fold of the latter and the primary sequence requirements for heme attachment. 736 proteins were identified in a total of 188 genomes scanned, which should yield a wide coverage of the possible physiological roles of a large share of cytochromes *c*. The analysis of the genome context through operon structure and gene fusion events also proves insightful in this respect.

The analysis of bacterial genomes reveals an astonishing variety in the number and the types of cytochromes *c* encoded by the different organisms. Cytochrome *c* is important in aerobic as well as other respiration mechanisms. In these processes, it provides electrons to the sites where they are accepted by O_2 or oxidized compounds. A relatively large number of cytochromes *c* encoded by a single bacterial genome is generally correlated to a high degree of flexibility in respiration, but the presence of multiple genes may also correspond to the capability of the organism to adapt quickly to environmental changes. Sequence similarities between cytochrome *c* domains from different bacteria indicate that there is often a good correlation between sequence and functional features, especially if the genomic context in which cytochrome *c* domains are found is explicitly considered. Speaking more generally, we see evidence for cytochrome *c* serving as an all-purpose electron transporter, capable of interacting with a variety of redox enzymes, such as hydrogenases, peroxidases, reductases of different kinds, etc. There are roles for cytochrome *c* even beyond electron transfer. This is suggested in one case by the occurrence of fusions with a CCO assembly factor, together with available

Table 6. Conventional Cytochrome *c* Subgroups Recommended To Be Retained in Cytochrome *c* Nomenclature by the Nomenclature Committee of the International Union of Biochemistry and Molecular Biology in 1989[196a]

name	reference section	number of hemes	iron ligands	Ambler's classification
cytochrome *c*	4	1	His, Met	class IB
cytochrome c_1	4, 5.1.2	1	His, Met	not included
cytochrome c_2	5.1.10	1	His, Met	class IA, IB
cytochrome c_3	8	4	His, His (all hemes)	class III
cytochrome c_4	5.1.9	2	His, Met (both hemes)	class IC
cytochrome c_5	5.1.10	1	His, Met	class IE
cytochrome c_6	5.1.11	1	His, Met	class IC
Pseudomonas cytochrome c_{551}	5.1.10	1	His, Met	class ID
bacterial photosystem cytochrome *c*	8	4	His, Met (three hemes) His, His (one heme)	not included
Chlorobium cytochrome c_{555}	5.2	1	His, Met	class IE
cytochrome *c*′	*b*	1	His	class II

[a] For each subgroup, the relevant section of the present work, the number of heme groups bound, the axial ligands of the iron ion, and the Ambler's classification[3] are reported. For cytochrome *c*′, not discussed here, the description given in ref 196 is reported. [b] High-spin cytochrome *c*, widely distributed in bacteria, folding as a four-α-helix bundle. It usually exists as a dimer. The heme is pentacoordinated with a single histidine ligand.

studies indicating that the lack of cytochrome *c* in mitochondria prevents the formation of a functional CCO. In another case, the identification of cytochrome *c* fused with WD40 repeats indicates that in bacteria there are interactions between these two domains. Such an interaction triggers apoptosis in humans, and thus the present finding indicates that cytochrome *c* might act as a regulator also in bacteria (presumably not regulating cell death, but some other unknown cellular process).

Cytochrome *c* appears to be an extremely flexible unit, which may change dramatically its sequence (up to almost 100%) while still retaining its fold and its main function, that is, electron transfer. It is thus tempting to speculate that the role of the protein moiety is mainly to wrap the heme, which is ultimately the only constant in cytochromes *c*, to enable selectivity in partner recognition as well as tune the reduction potential of the iron ion. Therefore evolution of cytochrome *c* is strongly driven by co-evolution: that is, the interacting enzymes change, and the protein consequently changes to optimize the interaction, the electron transfer, or both. Thus, cytochrome *c*, especially in bacteria, looks like an adaptable module of general use that can be shaped around a conserved, compact structural core purposely to interact with a number of other proteins. The situation in Eukaryota is (almost) opposite, since cytochrome *c* is (almost) exclusively used in the respiratory chain. This is consistent with the much higher level of regulation of cellular processes in Eukaryota, and in higher organisms in particular, where an extreme specialization of proteins occurs. On the other hand, the use of cytochrome *c* in triggering apoptosis is a striking example of the flexibility of its biological role. Whereas in Bacteria the tuning of cytochrome *c* function is mainly achieved through the use of multiple specialized cytochromes *c* with different sequences, in more complex organisms the role of a single cytochrome *c* can be modulated by varying its environment (e.g., intermolecular interactions, cell compartment), a possibility that is obviously larger than in microbial organisms.

10. A Note on Cytochrome *c* Nomenclature

The term "cytochrome" dates back to 1925, when Keilin introduced it to describe a group of heme proteins undergoing oxidation/reduction reactions, characterized in the reduced form by intense absorption bands in the 510−615 nm range.[195] Since then, the great build-up of knowledge on cytochromes increasingly highlighted that their impressive diversity would make it extremely difficult to find a simple basis for a complete classification of these proteins. As a result, no systematic nomenclature of cytochromes has been fixed till today, beyond their deep-rooted designation as cytochrome *a*, *b*, *c*, or *d* depending on the type and the binding mode of the heme moiety. Within the cytochrome *c* group, two main naming systems have been used to label experimentally characterized subgroups, using subscripts. In the first system, a progressive numbering of the subscripts has been used to identify different functional classes (such as cytochrome c_1, c_2, and so on), but this has not been applied consistently over the years. In the second system, subscripts are assigned based on the experimental wavelength (in nm) of the so-called α-band in the visible absorption spectrum of the reduced protein (such as cytochrome c_{550}, c_{551}, and so on). At present, the above approaches have resulted in a possibly confusing mixture of names based upon either criterion (see also next paragraph and Table 6).

To our knowledge, the most recent recommendations for the nomenclature of cytochromes were issued by the Nomenclature Committee of the International Union of Biochemistry and Molecular Biology (IUBMB) in 1989.[196] In that document, it was suggested to retain the traditional names of a number of well-established cytochrome *c* subgroups (Table 6), while naming all newly characterized cytochromes that would not fit in these subgroups through the second above-mentioned criterion. In our opinion, this proposed approach mainly reflects a pregenomic outlook, where new *c*-type cytochromes were identified through chromatography of cell lysates thanks to their bright color, and primary sequence determination was a complex task. In the present postgenomic era, where sequences are determined first and experimental characterization is carried out only later (or, sometimes, never), a naming convention entirely based on experimental features is unpractical. Thus, a refinement and update of the recommendations of the IUBMB would be useful and could incorporate criteria based on primary sequence features, which, as shown also in this work, can be quite informative with respect to the function.

11. List of Abbreviations

ABC = ATP-binding cassette
ADH = alcohol dehydrogenase
CCO = cytochrome *c* oxidase
CCP = cytochrome *c* peroxidase
DMSO = dimethyl sulfoxide
FAD = flavin adenine dinucleotide
FCSD = flavocytochrome *c* sulfide dehydrogenase
GMC = glucose−methanol−choline
HMM = hidden Markov model
MATE = multidrug and toxic compound extrusion
MFS = major facilitator superfamily
NADH = nicotinamide adenine dinucleotide
NMR = nuclear magnetic resonance
PQQ = pyrroloquinoline quinone
QCR = ubiquinol/cytochrome *c* oxidoreductase
TOMES = thiosulfate-oxidizing multi-enzyme system
TMAO = trimethylamine *N*-oxide
TTQ = tryptophan tryptophylquinone

12. Acknowledgment

We thank Julian Gough for providing us with the table of the sequences identified by superfamily in the genomes. This review was stimulated by an invitation of I.B. as a plenary lecturer at the 3rd International Conference on Porphyrins and Phthalocyanines. Financial support from the Italian Ministry of Education, University and Research (MIUR) is acknowledged (projects FIRB-RBAU01HYZ8 and COFIN-2003).

13. Supporting Information Available

CLUSTALW alignment of sequences from Proteobacteria and Cyanobacteria containing one, two, and three cytochrome *c* domains (Tables S1−3), lists of cytochrome *c*-containing proteins detected in Eukaryota, Gram-negative bacteria, Gram-positive bacteria, and Archaea (Tables S4, S5, S7, and S8), and a list of cytochrome *c*-containing proteins detected in Gram-negative bacteria sorted by functional class (Table S6). This material is available free of charge via the Internet at http://pubs.acs.org.

14. References

(1) Butt, W. D.; Keilin, D. *Proc. R. Soc. London* **1962**, *B512*, 429.
(2) Swanson, R.; Trus, B. L.; Mandel, N.; Mandel, G.; Kallai, O. B.; Dickerson, R. E. *J. Biol. Chem.* **1977**, *252*, 759.
(3) Ambler, R. P. *Biochim. Biophys. Acta* **1991**, *1058*, 42.
(4) Scott, R. A.; Mauk, A. G. *Cytochrome c. A multidisciplinary approach*; University Science Books: Sausalito, CA, 1996.
(5) Orengo, C. A.; Michie, A. D.; Jones, S.; Jones, D. T.; Swindells, M. B.; Thornton, J. M. *Structure* **1997**, *5*, 1093.
(6) Lo Conte, L.; Ailey, B.; Hubbard, T. J.; Brenner, S. E.; Murzin, A. G.; Chothia, C. *Nucleic Acids Res.* **2000**, *28*, 257.
(7) Chothia, C.; Lesk, A. M. *J. Mol. Biol.* **1985**, *182*, 151.
(8) Benini, S.; Rypniewski, W.; Wilson, K. S.; Van Beeumen, J.; Ciurli, S. *Biochemistry* **2000**, *39*, 13115.
(9) Banci, L.; Bertini, I.; Rosato, A.; Varani, G. *J. Biol. Inorg. Chem.* **1999**, *4*, 824.
(10) Bartalesi, I.; Rosato, A.; Zhang, W. *Biochemistry* **2003**, *42*, 10923.
(11) Jiang, X.; Wang, X. *Annu. Rev. Biochem.* **2004**, *73*, 87.
(12) Giorgio, M.; Migliaccio, E.; Orsini, F.; Paolucci, D.; Moroni, M.; Contursi, C.; Pelliccia, G.; Luzi, L.; Minucci, S.; Marcaccio, M.; Pinton, P.; Rizzuto, R.; Bernardi, P.; Paolucci, F.; Pelicci, P. G. *Cell* **2005**, *122*, 221.
(13) Pearce, D. A.; Sherman, F. *J. Biol. Chem.* **1995**, *270*, 20879.
(14) Pettigrew, G. W.; Moore, G. R. *Cytochromes c; Biological Aspects*; Springer-Verlag: Berlin, 1987.
(15) Gough, J.; Chothia, C. *Nucleic Acids Res.* **2002**, *30*, 268.
(16) Eddy, S. R. *Bioinformatics* **1998**, *14*, 755.
(17) Gough, J.; Karplus, K.; Hughey, R.; Chothia, C. *J. Mol. Biol.* **2001**, *313*, 903.
(18) Arciero, D. M.; Hooper, A. B. *FEBS Lett.* **1997**, *410*, 457.
(19) Arciero, D. M.; Hooper, A. B.; Cai, M.; Timkovich, R. *Biochemistry* **1993**, *32*, 9370.
(20) Bateman, A.; Coin, L.; Durbin, R.; Finn, R. D.; Hollich, V.; Griffiths-Jones, S.; Khanna, A.; Marshall, M.; Moxon, S.; Sonnhammer, E. L.; Studholme, D. J.; Yeats, C.; Eddy, S. R. *Nucleic Acids Res.* **2004**, *32 (*Database issue), D138.
(21) McGinnis, S.; Madden, T. L. *Nucleic Acids Res.* **2004**, *32*, W20.
(22) Tatusov, R. L.; Koonin, E. V.; Lipman, D. J. *Science* **1997**, *278*, 631.
(23) Thompson, J. D.; Higgins, D. G.; Gibson, T. J. *Nucleic Acids Res.* **1994**, *22*, 4673.
(24) Frickey, T.; Lupas, A. *Bioinformatics* **2004**, *20*, 3702.
(25) Sali, A.; Blundell, T. L. *J. Mol. Biol.* **1993**, *234*, 779.
(26) Read, J.; Gill, R.; Dales, S. L.; Cooper, J. B.; Wood, S. P.; Anthony, C. *Protein Sci.* **1999**, *8*, 1232.
(27) Banci, L.; Bertini, I.; Ciurli, S.; Dikiy, A.; Dittmer, J.; Rosato, A.; Sciara, G.; Thompsett, A. *ChemBioChem* **2002**, *3*, 299.
(28) Berry, E. A.; Guergova-Kuras, M.; Huang, L. S.; Crofts, A. R. *Annu. Rev. Biochem.* **2000**, *69*, 1005.
(29) Zhang, Z.; Gerstein, M. *Gene* **2003**, *312*, 61.
(30) Virbasius, J. V.; Scarpulla, R. C. *J. Biol. Chem.* **1988**, *263*, 6791.
(31) Burke, P. V.; Raitt, D. C.; Allen, L. A.; Kellogg, E. A.; Poyton, R. O. *J. Biol. Chem.* **1997**, *272*, 14705.
(32) Bhatnagar, A.; Raghavendra, P. R.; Kranthi, B. V.; Rangarajan, P. N. *Biochem. Biophys. Res. Commun.* **2004**, *321*, 900.
(33) Gupta, R.; He, Z.; Luan, S. *Nature* **2002**, *417*, 567.
(34) Weigel, M.; Pesaresi, P.; Leister, D. *Trends Plant Sci.* **2003**, *8*, 513.
(35) Crofts, A. R. *Annu. Rev. Physiol.* **2004**, *66*, 689.
(36) Zhang, Z.; Huang, L.; Shulmeister, V. M.; Chi, Y. I.; Kim, K. K.; Hung, L. W.; Crofts, A. R.; Berry, E. A.; Kim, S. H. *Nature* **1998**, *392*, 677.
(37) Iwata, S.; Lee, J. W.; Okada, K.; Lee, J. K.; Iwata, M.; Rasmussen, B.; Link, T. A.; Ramaswamy, S.; Jap, B. K. *Science* **1998**, *281*, 64.
(38) Hunte, C.; Koepke, J.; Lange, C.; Rossmanith, T.; Michel, H. *Struct. Folding Des.* **2000**, *8*, 669.
(39) Baymann, F.; Lebrun, E.; Nitschke, W. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 17737.
(40) Lange, C.; Hunte, C. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 2800.
(41) Maneg, O.; Ludwig, B.; Malatesta, F. *J. Biol. Chem.* **2003**, *278*, 46734.
(42) Witt, H.; Malatesta, F.; Nicoletti, F.; Brunori, M.; Ludwig, B. *J. Biol. Chem.* **1998**, *273*, 5132.
(43) Schmidt, T. R.; Wildman, D. E.; Uddin, M.; Opazo, J. C.; Goodman, M.; Grossman, L. I. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 6379.
(44) Boumans, H.; Grivell, L. A.; Berden, J. A. *J. Biol. Chem.* **1998**, *273*, 4872.
(45) Bianchi, C.; Genova, M. L.; Parenti, C. G.; Lenaz, G. *J. Biol. Chem.* **2004**, *279*, 36562.
(46) Dudkina, N. V.; Eubel, H.; Keegstra, W.; Boekema, E. J.; Braun, H. P. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 3225.
(47) Pelletier, H.; Kraut, J. *Science* **1992**, *258*, 1748.
(48) Barrientos, A.; Pierre, D.; Lee, J.; Tzagoloff, A. *J. Biol. Chem.* **2003**, *278*, 8881.
(49) Kluck, R. M.; Bossy-Wetzel, E.; Green, D. R.; Newmeyer, D. D. *Science* **1997**, *275*, 1132.
(50) Acehan, D.; Jiang, X.; Morgan, D. G.; Heuser, J. E.; Wang, X.; Akey, C. W. *Mol. Cell* **2002**, *9*, 423.
(51) Riedl, S. J.; Li, W.; Chao, Y.; Schwarzenbacher, R.; Shi, Y. *Nature* **2005**, *434*, 926.
(52) Adams, J. M. *Genes Dev.* **2003**, *17*, 2481.
(53) Dorstyn, L.; Mills, K.; Lazebnik, Y.; Kumar, S. *J. Cell Biol.* **2004**, *167*, 405.
(54) Wheeler, D. L.; Chappey, C.; Lash, A. E.; Leipe, D. D.; Madden, T. L.; Schuler, G. D.; Tatusova, T. A.; Rapp, B. A. *Nucleic Acids Res.* **2000**, *28*, 10.
(55) Snel, B.; Bork, P.; Huynen, M. A. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 5890.
(56) Pereira, M. M.; Santana, M.; Teixeira, M. *Biochim. Biophys. Acta* **2001**, *1505*, 185.
(57) Myllykallio, H.; Liebl, U. *Trends Microbiol.* **2000**, *8*, 542.
(58) Pitcher, R. S.; Brittain, T.; Watmough, N. J. *Biochem. Soc. Trans.* **2002**, *30*, 653.
(59) Pitcher, R. S.; Watmough, N. J. *Biochim. Biophys. Acta* **2004**, *1655*, 388.
(60) Preisig, O.; Anthamatten, D.; Hennecke, H. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *90*, 3309.
(61) Zufferey, R.; Preisig, O.; Hennecke, H.; Thöny-Meyer, L. *J. Biol. Chem.* **1996**, *271*, 9114.

(62) Schutz, M.; Brugna, M.; Lebrun, E.; Baymann, F.; Huber, R.; Stetter, K. O.; Hauska, G.; Toci, R.; Lemesle-Meunier, D.; Tron, P.; Schmidt, C.; Nitschke, W. *J. Mol. Biol.* **2000**, *300*, 663.
(63) Berry, E. A.; Huang, L.-S.; Saechao, L. K.; Pon, N. G.; Valkova-Valchanova, M.; Daldal, F. *Photosynth. Res.* **2004**, *81*, 251.
(64) Lichtarge, O.; Bourne, H. R.; Cohen, F. E. *J. Mol. Biol.* **1996**, *257*, 342.
(65) Pauleta, S. R.; Guerlesquin, F.; Goodhew, C. F.; Devreese, B.; Van Beeumen, J.; Pereira, A. S.; Moura, I.; Pettigrew, G. W. *Biochemistry* **2004**, *43*, 11214.
(66) Shimizu, H.; Schuller, D. J.; Lanzilotta, W. N.; Sundaramoorthy, M.; Arciero, D. M.; Hooper, A. B.; Poulos, T. L. *Biochemistry* **2001**, *40*, 13483.
(67) Wang, Y.; Graichen, M. E.; Liu, A.; Pearson, A. R.; Wilmot, C. M.; Davidson, V. L. *Biochemistry* **2003**, *42*, 7318.
(68) Pearson, A. R.; De la Mora-Rey, T.; Graichen, M. E.; Wang, Y.; Jones, L. H.; Marimanikkupam, S.; Agger, S. A.; Grimsrud, P. A.; Davidson, V. L.; Wilmot, C. M. *Biochemistry* **2004**, *43*, 5494.
(69) Berks, B. C.; Ferguson, S. J.; Moir, J. W.; Richardson, D. J. *Biochim. Biophys. Acta* **1995**, *1232*, 97.
(70) Ferguson, S. J. *Curr. Opin. Chem. Biol.* **1998**, *2*, 182.
(71) Philippot, L. *Biochim. Biophys. Acta* **2002**, *1577*, 355.
(72) Cheesman, M.; Ferguson, S. J.; Moir, J. W. B.; Richardson, D. J.; Zumft, W. G.; Thompson, A. J. *Biochemistry* **1997**, *36*, 16267.
(73) Nurizzo, D.; Silvestrini, M. C.; Mathieu, M.; Cutruzzola, F.; Bourgeois, D.; Fulop, V.; Hajdu, J.; Brunori, M.; Tegoni, M.; Cambillau, C. *Structure* **1997**, *5*, 1157.
(74) Dodd, F. E.; Van Beeumen, J.; Eady, R. R.; Hasnain, S. S. *J. Mol. Biol.* **1998**, *282*, 369.
(75) Murphy, M. E. P.; Turley, S.; Adman, E. T. *J. Biol. Chem.* **1997**, *272*, 28455.
(76) Kawasaki, S.; Arai, H.; Kodama, T.; Igarashi, Y. *J. Bacteriol.* **1997**, *179*, 235.
(77) Hasegawa, N.; Arai, H.; Igarashi, Y. *Biochem. Biophys. Res. Commun.* **2001**, *288*, 1223.
(78) Carr, H. S.; Winge, D. R. *Acc. Chem. Res.* **2003**, *36*, 309.
(79) Balatri, E.; Banci, L.; Bertini, I.; Cantini, F.; Ciofi-Baffoni, S. *Structure* **2003**, *11*, 1431.
(80) Arnesano, F.; Banci, L.; Bertini, I.; Martinelli, M. *J. Proteome Res.* **2005**, *4*, 63.
(81) Cramm, R.; Siddiqui, R. A.; Friedrich, B. *J. Bacteriol.* **1997**, *179*, 6769.
(82) Brown, K.; Tegoni, M.; Prudencio, M.; Pereira, A. S.; Besson, S.; Moura, J. J.; Moura, I.; Cambillau, C. *Nat. Struct. Biol.* **2000**, *7*, 191.
(83) Haltia, T.; Brown, K.; Tegoni, M.; Cambillau, C.; Saraste, M.; Mattila, K.; Djinovic-Carugo, K. *Biochem. J.* **2003**, *369*, 77.
(84) Rasmussen, T.; Berks, B. C.; Sanders-Loehr, J.; Dooley, D. M.; Zumft, W. G.; Thomson, A. J. *Biochemistry* **2000**, *39*, 12753.
(85) Simon, J.; Einsle, O.; Kroneck, P. M. H.; Zumft, W. G. *FEBS Lett.* **2004**, *569*, 7.
(86) Hille, R. *Chem. Rev.* **1996**, *96*, 2757.
(87) Basu, P.; Stolz, J. F.; Smith, T. *Curr. Sci.* **2003**, *84*, 1412.
(88) Gon, S.; Patte, J.-C.; Mejean, V.; Iobbi-Nivol, C. *J. Bacteriol.* **2000**, *182*, 5779.
(89) Mejean, V.; Iobbi-Nivol, C.; Lepelletier, M.; Giordano, G.; Chippaux, M.; Pascal, M. C. *Mol. Microbiol.* **1994**, *11*, 1169.
(90) Pierson, D. E.; Campbell, A. *J. Bacteriol.* **1990**, *172*, 2194.
(91) Pollock, V. V.; Barber, M. J. *J. Biol. Chem.* **1997**, *272*, 3355.
(92) Richardson, D. J. *Microbiology* **2000**, *146*, 551.
(93) Gon, S.; Giudici-Orticoni, M. T.; Mejean, V.; Iobbi-Nivol, C. *J. Biol. Chem.* **2001**, *276*, 11545.
(94) Rebelo, J. M.; Dias, J. M.; Huber, R.; Moura, J. J. G.; Romão, M. J. *J. Biol. Inorg. Chem.* **2001**, *6*, 791.
(95) Cavener, D. R. *J. Mol. Biol.* **1992**, *223*, 811.
(96) Zamocky, M.; Hallberg, M.; Ludwig, R.; Divne, C.; Haltrich, D. *Gene* **2004**, *338*, 1.
(97) Kelly, D. P.; Shergill, J. K.; Lu, W. P.; Wood, A. P. *Antonie Van Leeuwenhoek* **1997**, *71*, 95.
(98) Friedrich, C. G.; Rother, D.; Bardischewsky, F.; Quentmeier, A.; Fischer, J. *Appl. Environ. Microbiol.* **2001**, *67*, 2873.
(99) Kappler, U.; Dahl, C. *FEMS Microbiol. Lett.* **2001**, *203*, 1.
(100) Rother, D.; Henrich, H.-J.; Quentmeier, A.; Bardischewsky, F.; Friedrich, C. G. *J. Bacteriol.* **2001**, *183*, 4499.
(101) Kappler, U.; Bennett, B.; Rethmeier, J.; Schwarz, G.; Deutzmann, R.; McEwan, A. G.; Dahl, C. *J. Biol. Chem.* **2000**, *275*, 13202.
(102) Kappler, U.; Bailey, S. *J. Biol. Chem.* **2005**, *280*, 24999.
(103) Quentmeier, A.; Kraft, R.; Kostka, S.; Klockenkämper, R.; Friedrich, C. G. *Arch. Microbiol.* **2000**, *173*, 117.
(104) Kappler, U.; Aguey-Zinsou, K.-F.; Hanson, G. R.; Bernhardt, P. V.; McEwan, A. G. *J. Biol. Chem.* **2004**, *279*, 6252.
(105) Bamford, V. A.; Bruno, S.; Rasmussen, T.; Appia-Ayme, C.; Cheesman, M. R.; Berks, B. C.; Hemmings, A. M. *EMBO J.* **2002**, *21*, 5599.
(106) Reinartz, M. J.; Tschäpe, T.; Brüser, T.; Trüper, H. G.; Dahl, C. *Arch. Microbiol.* **1998**, *170*, 59.
(107) Visser, J. M.; de Jong, G. A. H.; Robertson, L. A.; Kuenen, J. G. *Arch. Microbiol.* **1997**, *167*, 295.
(108) Chen, Z.-W.; Matsushita, K.; Yamashita, T.; Fujii, T.-A.; Toyama, H.; Adachi, O.; Bellamy, H. D.; Mathews, F. S. *Structure* **2002**, *10*, 837.
(109) Stearman, R.; Yuan, D.; Yamaguchi-Iwan, Y.; Klausner, R. D.; Dancis, A. *Science* **1996**, *271*, 1552.
(110) Andrews, S. C.; Robinson, A. K.; Rodriguez-Quiñones, F. *FEMS Microbiol. Rev.* **2003**, *27*, 215.
(111) Dubbels, B. L.; DiSpirito, A. A.; Morton, J. D.; Semrau, J. D.; Neto, J. N. E.; Bazylinski, D. A. *Microbiology* **2004**, *150*, 2931.
(112) Huston, W. M.; Jennings, M. P.; McEwan, A. G. *Mol. Microbiol.* **2002**, *45*, 1741.
(113) Puig, S.; Rees, E.; Thiele, D. J. *Structure* **2002**, *10*, 1292.
(114) Higgins, C. F. *Res. Microbiol.* **2001**, *152*, 205.
(115) Pao, S. S.; Paulsen, I. T.; Saier, M. H. *Microbiol. Mol. Biol. Rev.* **1998**, *62*, 1.
(116) He, G.-X.; Kuroda, T.; Mima, T.; Morita, Y.; Mizushima, T.; Tsuchiya, T. *J. Bacteriol.* **2004**, *186*, 262.
(117) Nies, D. H. *FEMS Microbiol. Rev.* **2003**, *27*, 313.
(118) Saier, M. H.; Paulsen, I. T. *Semin. Cell Dev. Biol.* **2001**, *12*, 205.
(119) Abergel, C.; Nitschke, W.; Malarte, G.; Bruschi, M.; Claverie, J.-M.; Giudici-Orticoni, M. T. *Structure* **2003**, *11*, 547.
(120) Brown, K.; Nurizzo, D.; Besson, S.; Shepard, W.; Moura, J. J.; Moura, I.; Tegoni, M.; Cambillau, C. *J. Mol. Biol.* **1999**, *289*, 1017.
(121) Giudici-Orticoni, M. T.; Guerlesquin, F.; Bruschi, M.; Nitschke, W. *J. Biol. Chem.* **1999**, *274*, 30365.
(122) Stintzi, A.; Johnson, Z.; Stonehouse, M.; Ochsner, U.; Meyer, J.-M.; Vasil, M. L.; Poole, K. *J. Bacteriol.* **1999**, *181*, 4118.
(123) Donohue, T. J.; McEwan, A. G.; Van Doren, S.; Crofts, A. R.; Kaplan, S. *Biochemistry* **1988**, *27*, 1918.
(124) Meyer, T. E.; Cusanovich, M. A. *Photosynth. Res.* **2003**, *76*, 111.
(125) Ambler, R. P.; Meyer, T. E.; Cusanovich, M. A.; Kamen, M. D. *Biochem. J.* **1987**, *246*, 115.
(126) Dus, K.; Sletten, K.; Kamen, M. D. *J. Biol. Chem.* **1968**, *243*, 5507.
(127) Carter, D. C.; Melis, K. A.; O'Donnell, S. E.; Burgess, B. K.; Furey, W. F.; Wang, B. C. *J. Mol. Biol.* **1985**, *184*, 279.
(128) Sanbongi, Y.; Ishii, M.; Igarashi, Y.; Kodama, T. *J. Bacteriol.* **1989**, *171*, 65.
(129) Nakagawa, A.; Higuchi, Y.; Yasuoka, N.; Katsube, Y.; Yagi, T. *J. Biochem. (Tokyo)* **1990**, *108*, 701.
(130) Nitschke, W.; Rutherford, A. W. *Trends Biochem. Sci.* **1991**, *16*, 241.
(131) Schubert, W. D.; Klukas, O.; Saenger, W.; Witt, H. T.; Fromme, P.; Krauss, N. *J. Mol. Biol.* **1998**, *280*, 297.
(132) Ferreira, K. N.; Iverson, T. M.; Maghlaoui, K.; Barber, J.; Iwata, S. *Science* **2004**, *303*, 1831.
(133) Goussias, C.; Boussac, A.; Rutherford, A. W. *Philos. Trans. R. Soc. London B* **2002**, *357*, 1369.
(134) Chitnis, P. R.; Xu, Q.; Chitnis, V. P.; Nechustai, R. *Photosynth. Res.* **1995**, *44*, 23.
(135) Ho, K. K.; Krogmann, D. W. *Biochim. Biophys. Acta* **1984**, *766*, 310.
(136) Navarro, J. A.; Hervás, M.; De la Rosa, M. A. *J. Biol. Inorg. Chem.* **1997**, *2*, 11.
(137) De la Rosa, M. A.; Navarro, J. A.; Diaz-Quintana, A.; De la Cerda, B.; Molina-Heredia, F. P.; Balme, A.; Murdoch Pdel, S.; Diaz-Moreno, I.; Duran, R. V.; Hervás, M. *Bioelectrochemistry* **2002**, *55*, 41.
(138) Malakhov, M. P.; Malakhova, O. A.; Murata, N. *FEBS Lett.* **1999**, *444*, 281.
(139) Molina-Heredia, F. P.; Balme, A.; Hervás, M.; Navarro, J. A.; De la Rosa, M. A. *FEBS Lett.* **2002**, *517*, 50.
(140) Klarskov, K.; Van Driessche, G.; Backers, K.; Dumortier, C.; Meyer, T. E.; Tollin, G.; Cusanovich, M. A.; Van Beeumen, J. J. *Biochemistry* **1998**, *37*, 5995.
(141) Leys, D.; Backers, K.; Meyer, T. E.; Hagen, W. R.; Cusanovich, M. A.; Van Beeumen, J. *J. Biol. Chem.* **2000**, *275* (6), 16050.
(142) Huber, R.; Wilharm, T.; Huber, D.; Trincone, A.; Burggraf, S.; König, H.; Rachel, R.; Rockinger, I.; Fricke, H.; Stetter, K. O. *Syst. Appl. Microbiol.* **1992**, *15*, 340.
(143) Brugna-Guiral, M.; Tron, P.; Nitschke, W.; Stetter, K. O.; Burlat, B.; Guigliarelli, B.; Bruschi, M.; Giudici-Orticoni, M. T. *Extremophiles* **2003**, *7*, 145.
(144) Schutz, M.; Schoepp-Cothenet, B.; Lojou, E.; Woodstra, M.; Lexa, D.; Tron, P.; Dolla, A.; Durand, M. C.; Stetter, K. O.; Baymann, F. *Biochemistry* **2003**, *42*, 10800.
(145) Mukhopadhyaya, P. N.; Deb, C.; Lahiri, C.; Roy, P. *J. Bacteriol.* **2000**, *182*, 4278.

(146) Aubert, C.; Guerlesquin, F.; Bianco, P.; Leroy, G.; Tron, P.; Stetter, K. O.; Bruschi, M. *Biochemistry* **2001**, *40*, 13690.
(147) Baymann, F.; Tron, P.; Schoepp-Cothenet, B.; Aubert, C.; Bianco, P.; Stetter, K. O.; Nitschke, W.; Schutz, M. *Biochemistry* **2001**, *40*, 13681.
(148) Myers, G. S.; Fraser, C. M. *Trends Microbiol.* **2004**, *12*, 303.
(149) Soulimane, T.; von Walter, M.; Hof, P.; Than, M. E.; Huber, R.; Buse, G. *Biochem. Biophys. Res. Commun.* **1997**, *237*, 572.
(150) Fee, J. A.; Chen, Y.; Todaro, T. R.; Bren, K. L.; Patel, K. M.; Hill, M. G.; Gomez-Moran, E.; Loehr, T. M.; Ai, J.; Thony-Meyer, L.; Williams, P. A.; Stura, E.; Sridhar, V.; McRee, D. E. *Protein Sci.* **2000**, *9*, 2074.
(151) Than, M. E.; Hof, P.; Huber, R.; Bourenkov, G. P.; Bartunik, H. D.; Buse, G.; Soulimane, T. *J. Mol. Biol.* **1997**, *271*, 629.
(152) Hellwig, P.; Soulimane, T.; Mantele, W. *Eur. J. Biochem.* **2002**, *269*, 4830.
(153) Giuffre, A.; Stubauer, G.; Sarti, P.; Brunori, M.; Zumft, W. G.; Buse, G.; Soulimane, T. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 14718.
(154) Okkels, J. S.; Kjaer, B.; Hansson, O.; Svendsen, I.; Moller, B. L.; Scheller, H. V. *J. Biol. Chem.* **1992**, *267*, 21139.
(155) Hauska, G.; Schoedl, T.; Remigy, H.; Tsiotis, G. *Biochim. Biophys. Acta* **2001**, *1507*, 260.
(156) Verte, F.; Kostanjevecki, V.; De Smet, L.; Meyer, T. E.; Cusanovich, M. A.; Van Beeumen, J. J. *Biochemistry* **2002**, *41*, 2932.
(157) Bharti, A. R.; Nally, J. E.; Ricaldi, J. N.; Matthias, M. A.; Diaz, M. M.; Lovett, M. A.; Levett, P. N.; Gilman, R. H.; Willig, M. R.; Gotuzzo, E.; Vinetz, J. M. *Lancet Infect. Dis.* **2003**, *3*, 757.
(158) Comstock, L. E.; Coyne, M. J. *Bioessays* **2003**, *25*, 926.
(159) Horn, M.; Collingro, A.; Schmitz-Esser, S.; Beier, C. L.; Purkhold, U.; Fartmann, B.; Brandt, P.; Nyakatura, G. J.; Droege, M.; Frishman, D.; Rattei, T.; Mewes, H. W.; Wagner, M. *Science* **2004**, *304*, 728.
(160) Purschke, W. G.; Schmidt, C. L.; Petersen, A.; Schafer, G. *J. Bacteriol.* **1997**, *179*, 1344.
(161) Moran, N. A. *Cell* **2002**, *108*, 583.
(162) Glockner, F. O.; Kube, M.; Bauer, M.; Teeling, H.; Lombardot, T.; Ludwig, W.; Gade, D.; Beck, A.; Borzym, K.; Heitmann, K.; Rabus, R.; Schlesner, H.; Amann, R.; Reinhardt, R. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 8298.
(163) Studholme, D. J.; Fuerst, J. A.; Bateman, A. *FEMS Microbiol. Lett.* **2004**, *236*, 333.
(164) Yu, J.; Hederstedt, L.; Piggot, P. J. *J. Bacteriol.* **1995**, *177*, 6751.
(165) Stroebel, D.; Choquet, Y.; Popot, J. L.; Picot, D. *Nature* **2003**, *426*, 413.
(166) Kurisu, G.; Zhang, H.; Smith, J. L.; Cramer, W. A. *Science* **2003**, *302*, 1009.
(167) Assempour, M.; Lim, D.; Hill, B. C. *Biochemistry* **1998**, *37*, 9991.
(168) Sun, G.; Sharkova, E.; Chesnut, R.; Birkey, S.; Duggan, M. F.; Sorokin, A.; Pujic, P.; Ehrlich, S. D.; Hulett, F. M. *J. Bacteriol.* **1996**, *178*, 1374.
(169) Maneg, O.; Malatesta, F.; Ludwig, B.; Drosou, V. *Biochim. Biophys. Acta* **2004**, *1655*, 274.
(170) Sone, N.; Nagata, K.; Kojima, H.; Tajima, J.; Kodera, Y.; Kanamaru, T.; Noguchi, S.; Sakamoto, J. *Biochim. Biophys. Acta* **2001**, *1503*, 279.
(171) Niebisch, A.; Bott, M. *J. Biol. Chem.* **2003**, *278*, 4339.
(172) Schafer, G.; Purschke, W.; Schmidt, C. L. *FEMS Microbiol. Rev.* **1996**, *18*, 173.
(173) Galagan, J. E.; Nusbaum, C.; Roy, A.; Endrizzi, M. G.; Macdonald, P.; FitzHugh, W.; Calvo, S.; Engels, R.; Smirnov, S.; Atnoor, D.; Brown, A.; Allen, N.; Naylor, J.; Stange-Thomann, N.; DeArellano, K.; Johnson, R.; Linton, L.; McEwan, P.; McKernan, K.; Talamas, J.; Tirrell, A.; Ye, W.; Zimmer, A.; Barber, R. D.; Cann, I.; Graham, D. E.; Grahame, D. A.; Guss, A. M.; Hedderich, R.; Ingram-Smith, C.; Kuettner, H. C.; Krzycki, J. A.; Leigh, J. A.; Li, W.; Liu, J.; Mukhopadhyay, B.; Reeve, J. N.; Smith, K.; Springer, T. A.; Umayam, L. A.; White, O.; White, R. H.; Conway, D. M.; Ferry, J. G.; Jarrell, K. F.; Jing, H.; Macario, A. J.; Paulsen, I.; Pritchett, M.; Sowers, K. R.; Swanson, R. V.; Zinder, S. H.; Lander, E.; Metcalf, W. W.; Birren, B. *Genome Res.* **2002**, *12*, 532.
(174) Stetter, K. O.; Lauerer, G.; Thomm, M.; Neuner, A. *Science* **1987**, *236*, 822.
(175) Ding, Y. H.; Ferry, J. G. *J. Bacteriol.* **2004**, *186*, 90.
(176) Reed, D. W.; Millstein, J.; Hartzell, P. L. *J. Bacteriol.* **2001**, *183*, 7007.
(177) Fitz-Gibbon, S. T.; Ladner, H.; Kim, U. J.; Stetter, K. O.; Simon, M. I.; Miller, J. H. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 984.
(178) Sako, Y.; Nomura, N.; Uchida, A.; Ishida, Y.; Morii, H.; Koga, Y.; Hoaki, T.; Maruyama, T. *Int. J. Syst. Bacteriol.* **1996**, *46*, 1070.
(179) Afshar, S.; Kim, C.; Monbouquette, H. G.; Schroder, I., I *Appl. Environ. Microbiol.* **1998**, *64*, 3004.
(180) Volkl, P.; Huber, R.; Drobner, E.; Rachel, R.; Burggraf, S.; Trincone, A.; Stetter, K. O. *Appl. Environ. Microbiol.* **1993**, *59*, 2918.
(181) de Vries, S.; Schroder, I. *Biochem. Soc. Trans.* **2002**, *30*, 662.
(182) Ishikawa, R.; Ishido, Y.; Tachikawa, A.; Kawasaki, H.; Matsuzawa, H.; Wakagi, T. *Arch. Microbiol.* **2002**, *179*, 42.
(183) Roldan, M. D.; Sears, H. J.; Cheesman, M. R.; Ferguson, S. J.; Thomson, A. J.; Berks, B. C.; Richardson, D. J. *J. Biol. Chem.* **1998**, *273*, 28785.
(184) Dobbin, P. S.; Butt, J. N.; Powell, A. K.; Reid, G. A.; Richardson, D. J. *Biochem. J.* **1999**, *342* (Part 2), 439.
(185) Igarashi, N.; Moriyama, H.; Fujiwara, T.; Fukumori, Y.; Tanaka, N. *Nat. Struct. Biol.* **1997**, *4*, 276.
(186) Einsle, O.; Messerschmidt, A.; Stach, P.; Bourenkov, G. P.; Bartunik, H. D.; Huber, R.; Kroneck, P. M. H. *Nature* **1999**, *400*, 476.
(187) Iverson, T. M.; Arciero, D. M.; Hsu, B. T.; Logan, M. S.; Hooper, A. B.; Rees, D. C. *Nat. Struct. Biol.* **1998**, *5*, 1005.
(188) Deisenhofer, J.; Epp, O.; Sinning, I.; Michel, H. *J. Mol. Biol.* **1989**, *246*, 429.
(189) Morais, J.; Palma, P. N.; Frazao, C.; Caldeira, J.; LeGall, J.; Moura, I.; Moura, J. J. G.; Carrondo, M. A. *Biochemistry* **1995**, *34*, 12830.
(190) Pokkuluri, P. R.; Londer, Y. Y.; Duke, N. E.; Erickson, J.; Pessanha, M.; Salgueiro, C. A.; Schiffer, M. *Protein Sci.* **2004**, *13*, 1684.
(191) Matias, P. M.; Coelho, R.; Pereira, I. A.; Coelho, A. V.; Thompson, A. W.; Sieker, L. C.; Gall, J. L.; Carrondo, M. A. *Struct. Folding Des.* **1999**, *7*, 119.
(192) Matias, P. M.; Coelho, A. V.; Valente, F. M.; Placido, D.; LeGall, J.; Xavier, A. V.; Pereira, I. A.; Carrondo, M. A. *J. Biol. Chem.* **2002**, *277*, 47907.
(193) Rossi, M.; Pollock, W. B.; Reij, M. W.; Keon, R. G.; Fu, R.; Voordouw, G. *J. Bacteriol.* **1993**, *175*, 4699.
(194) Pitts, K. E.; Dobbin, P. S.; Reyes-Ramirez, F.; Thomson, A. J.; Richardson, D. J.; Seward, H. E. *J. Biol. Chem.* **2003**, *278*, 27758.
(195) Keilin, D. *Proc. R Soc. London, Ser. B: Biol. Sci.* **1925**, *98*, 312.
(196) Palmer, G.; Reedijk, J. *Biochim. Biophys. Acta* **1991**, *1060*, 599.
(197) Saraste, M. *Science* **1999**, *283*, 1488.

CR050241V