

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/51151693>

The Evolution of Respiratory Chain Complex I from a Smaller Last Common Ancestor Consisting of 11 Protein Subunits

ARTICLE *in* JOURNAL OF MOLECULAR EVOLUTION · MAY 2011

Impact Factor: 1.68 · DOI: 10.1007/s00239-011-9447-2 · Source: PubMed

CITATIONS

30

READS

25

2 AUTHORS:



[vamsi krishna Moparthi](#)

Uppsala University

14 PUBLICATIONS 130 CITATIONS

SEE PROFILE



[Cecilia Hägerhäll](#)

Lund University

40 PUBLICATIONS 1,423 CITATIONS

SEE PROFILE

The Evolution of Respiratory Chain Complex I from a Smaller Last Common Ancestor Consisting of 11 Protein Subunits

Vamsi K. Moparthi · Cecilia Hägerhäll

Received: 24 January 2011 / Accepted: 4 May 2011 / Published online: 20 May 2011
© The Author(s) 2011. This article is published with open access at Springerlink.com

Abstract The NADH:quinone oxidoreductase (complex I) has evolved from a combination of smaller functional building blocks. Chloroplasts and cyanobacteria contain a complex I-like enzyme having only 11 subunits. This enzyme lacks the N-module which harbors the NADH binding site and the flavin and iron–sulfur cluster prosthetic groups. A complex I-homologous enzyme found in some archaea contains an F₄₂₀ dehydrogenase subunit denoted as FpoF rather than the N-module. In the present study, all currently available whole genome sequences were used to survey the occurrence of the different types of complex I in the different kingdoms of life. Notably, the 11-subunit version of complex I was found to be widely distributed, both in the archaeal and in the eubacterial kingdoms, whereas the 14-subunit classical complex I was found only in certain eubacterial phyla. The FpoF-containing complex I was present in Euryarchaeota but not in Crenarchaeota, which contained the 11-subunit complex I. The 11-subunit enzymes showed a primary sequence variability as great or greater than the full-size 14-subunit complex I, but differed distinctly from the membrane-bound hydrogenases. We conclude that this type of compact 11-subunit complex I is ancestral to all present-day complex I enzymes. No designated partner protein, acting as an electron delivery device, could be found for the compact version of complex I. We propose that the primordial complex I, and many of

the present-day 11-subunit versions of it, operate without a designated partner protein but are capable of interaction with several different electron donor or acceptor proteins.

Keywords NADH:quinone oxidoreductase · NiFe-hydrogenase · Formate:hydrogen lyase · Protein phylogeny · Functional modules · Antiporter

Introduction

NADH:ubiquinone oxidoreductase or complex I (EC 1.6.99.3) is the largest enzyme in the respiratory chain of mitochondria and bacteria, where it catalyzes the oxidation of NADH and the reduction of quinone, coupled with the translocation of protons across the membrane by a hitherto unknown molecular mechanism. The enzyme complex is also capable of $\Delta\mu\text{H}^+$ supported NAD^+ reduction (Brandt 2006). The core of the enzyme complex comprises 14 protein subunits that are conserved between the prokaryote and eukaryote enzyme complexes, whereas the eukaryote complex I enzymes contain up to 31 additional supernumerary or accessory subunits (Carroll et al. 2006). Recently, three of these accessory subunits were also found in complex I from an α -proteobacteria (Yip et al. 2011), the branch in which the ancestor of mitochondria is thought to originate (Gray et al. 1999). Seven of the 14 core subunits included in the minimal functional unit are membrane spanning and seven of them protrude from the membrane into the bacterial cytoplasm or the mitochondrial matrix (Fig. 1a). The seven membrane-spanning proteins are in eukaryotes encoded by mitochondrial DNA. The structure of the seven soluble protein subunits from *Thermus thermophilus* have been solved at high resolution (Sazanov and Hinchliffe 2006). Since then, additional structures in the oxidized and reduced states have

Electronic supplementary material The online version of this article (doi:10.1007/s00239-011-9447-2) contains supplementary material, which is available to authorized users.

V. K. Moparthi · C. Hägerhäll (✉)
Department of Biochemistry and Structural Biology, Center
for Molecular Protein Science, Lund University, Box 124,
22100 Lund, Sweden
e-mail: Cecilia.Hagerhall@biochemistry.lu.se

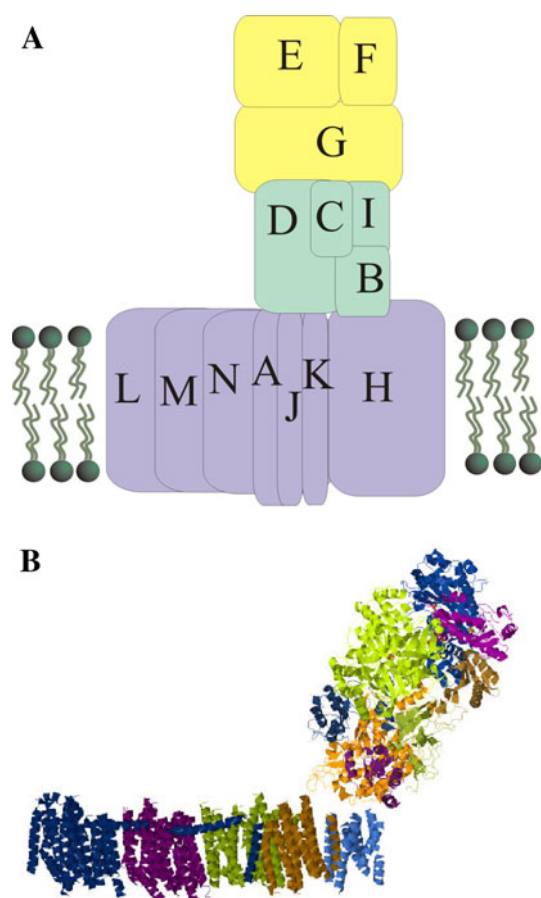


Fig. 1 **a** Schematic drawing of the complex I protein subunits. The proteins are labeled with letters as in the *E. coli* nomenclature, where A stands for NuoA, etc. The subunits in the N-module are NuoE, NuoF and NuoG, the Q-module is comprised of NuoB, NuoC, NuoD and NuoI, and the P-module make up the remaining seven membrane-spanning subunits. The exact position of the NuoA, NuoJ and NuoK relative to other subunits is not clear from the currently available structures (see also **b**). **b** The structure of all three complex I modules from *T. thermophilus*, drawn using the coordinates from PDB file 3M9S. Note that this complex I contains an additional protein subunit, Nqo15, located at the interface between the N- and Q-modules (Sazanov and Hinchliffe 2006). This protein is generally not present in complex I. Nqo15 resembles the iron chaperone frataxin, and may therefore be involved in iron–sulfur cluster regeneration or may simply be used to stabilize the complex at high growth temperatures

been revealed (Berrisford and Sazanov 2009). Recently, structural information regarding the membrane-spanning domain has also become available, both from the prokaryotes *T. thermophilus* (Fig. 1b) and *Escherichia coli* (Efremov et al. 2010) and the eukaryote *Yarrowia lipolytica* (Hunte et al. 2010), although at lower resolution.

It has long been recognized that complex I arose through the combining of smaller functional building blocks (Friedrich et al. 1993; Friedrich and Weiss 1997). Obtaining a better understanding of how evolutionary driving forces brought these building blocks together, could provide key information regarding the functional mechanisms of the

present-day complex I. The so-called NADH dehydrogenase module (N-module, Fig. 1a) of complex I consists of three proteins: NuoE, NuoF, and NuoG. The NuoE and NuoF subunits (the so-called FP fragment) contain FMN and FeS clusters and harbor the NADH binding site (Yano et al. 1996). These subunits show NADH dehydrogenase activity to various artificial electron acceptors. The NuoG subunit, also part of the N-module, resembles Fe-only hydrogenase and such molybdopterin-containing enzymes as formate dehydrogenase and nitrate reductase. It contributes FeS clusters to the electron transfer chain of the present-day complex I (Rothery et al. 2008; Sazanov and Hinchliffe 2006; Yano et al. 1995). The C-terminal end, corresponding to the part where the H_2 binding or formate binding site is located in the homologous smaller enzymes, has lost all of its primary sequence conservation in complex I. There are only a few complex I enzymes in which an additional FeS cluster is retained (Sazanov and Hinchliffe 2006). The quinone module (Q-module, Fig. 1a) of complex I is composed of NuoC, the ferredoxin-like NuoI and two proteins resembling the small and the large subunit of soluble NiFe-hydrogenases, which in complex I correspond to the NuoB and NuoD subunits, respectively. The Q-module accepts electrons from the N-module and transfers them via iron–sulfur clusters to quinone. Interestingly enough, the quinone-binding site in complex I appears to correspond to the NiFe-active site in hydrogenase (Brandt 2006; Darrouzet et al. 1998; Kersch et al. 2001; Tocilescu et al. 2010). The modular evolution of hydrogenases has been extensively reviewed elsewhere (Vignais and Billoud 2007; Vignais et al. 2001). Finally, the proton translocation module (P-module, Fig. 1a) is composed of the seven membrane-spanning subunits NuoA, H, J, K, L, M, and N (Brandt 2006). Each of the three complex I subunits NuoL, NuoM, and NuoN is homologous to the protein subunits in one particular type of Na^+/H^+ antiporter, denoted as Mrp/Pha/Sha or Mnh in various organisms (Swartz et al. 2005). Antiporters of this type are built up of seven proteins, MrpA, B, C, D, E, F, and G, MrpA having been shown to possess a higher sequence similarity to NuoL and MrpD a greater similarity to NuoM and NuoN (Mathiesen and Hägerhäll 2002). In addition, the MrpC subunit has been shown to be homologous to NuoK, indicating that the entire Mrp-antiporter-derived module NuoKLM was recruited to complex I (Mathiesen and Hägerhäll 2003). In an alternative terminology, aimed at dividing complex I up into modules related to smaller present-day enzymes, the hydrogenase module comprise the Q-module and two additional membrane-spanning proteins, NuoH and one Mrp-antiporter-derived protein. The hydrogenase module has a composition equivalent to that of present-day Ech and Hyc hydrogenases (Friedrich and Scheide 2000; Hedderich 2004; Vignais et al. 2001). In this nomenclature, the transporter-module contains

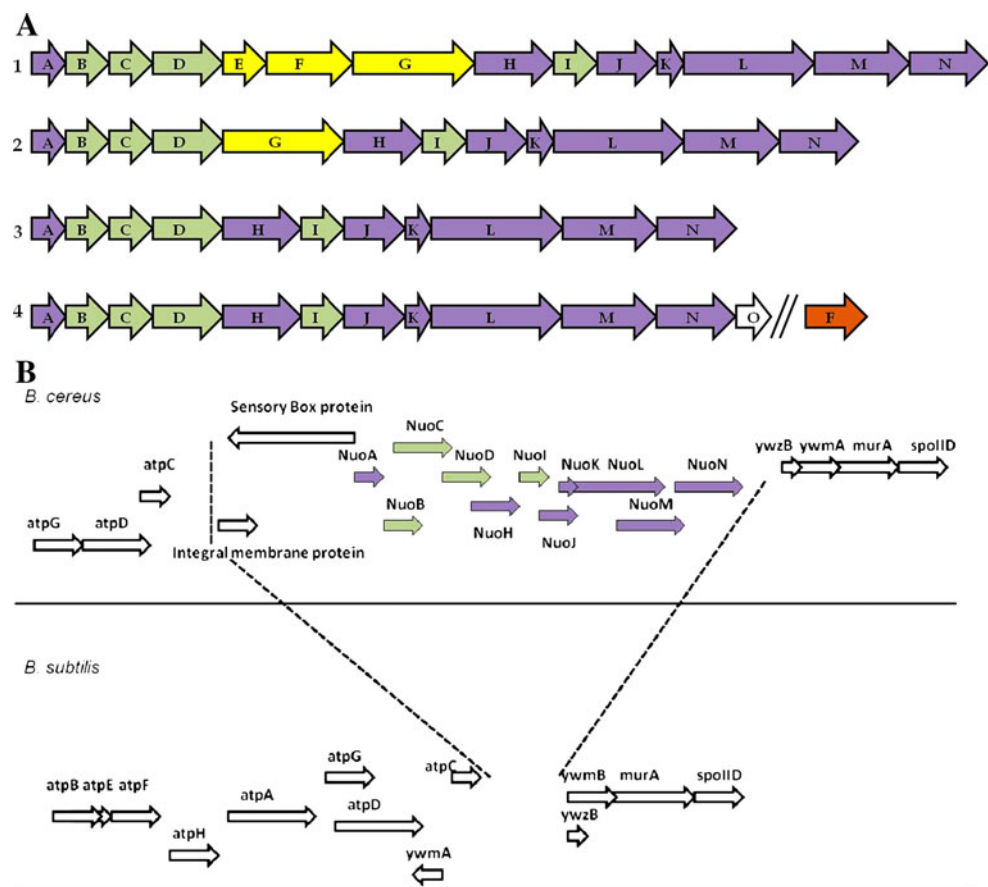
the remaining membrane-spanning proteins (Friedrich 2001; Friedrich and Scheide 2000).

The N-module appears to be the latest addition to complex I, acting as an electron-input device connecting the citric acid cycle with the aerobic respiratory chain. Some methanogenic archaea have been shown to contain a smaller complex I enzyme in which the NADH dehydrogenase module is replaced by another module analogous in function. This so-called F₄₂₀ dehydrogenase module consists of only one protein subunit, FpoF (Bäumer et al. 2000). In the microaerophilic organisms *Helicobacter pylori* and *Campylobacter jejuni*, only the FP fragment NuoE and F is absent from complex I, but the enzymes contain NuoG (Finel 1998), suggesting that the formation of the N-module may have involved two consecutive steps. It has been proposed that in the latter organism a flavodoxin donates electrons to the 12-subunit complex I (Weerakoon and Olson 2008). The markedly conserved operon structure of the complex I genes found in most bacteria also support the idea of a late addition of the N-module (Fig. 2). A strange complex I-like protein complex lacking the entire N-module was first noticed in chloroplast (Ohyama et al. 1986; Sazanov et al. 1998). Friedrich and Weiss later termed it the “alien” complex I (Friedrich and Scheide 2000; Friedrich et al. 1995). The

same type of 11-subunit complex I is also present in cyanobacteria, in which it exists in multiple versions that use different sets of antiporter-like NuoL, NuoM, and NuoN proteins under different growth conditions. The different forms of complex I have been shown to be involved in cyclic electron flow around photosystem I, chlororespiration, and CO₂ acquisition (Battchikova et al. 2011; Peng et al. 2011). The N-module equivalent remains elusive, however, no additional partner protein having been found to date, neither in plant chloroplasts nor in cyanobacteria (Battchikova et al. 2011; Martin and Sabater 2010; Peng et al. 2011; Suorsa et al. 2009). At the same time, there are four additional proteins, unique to organisms performing oxygenic photosynthesis that are present in this version of complex I (Birungi et al. 2010).

As already mentioned, bacteria often have the complex I encoding genes organized in an operon (Fig. 2), although there are many examples of prokaryotes for which this is not the case, or that have two or three dispersed gene clusters for complex I genes. Therefore, without a whole genome sequence or a pure and isolated enzyme complex, it is not possible to account for all the protein subunits with certainty. With the abundance of whole genome sequences that are presently available, it has become possible to differentiate between different types of complex I with a high

Fig. 2 **a** Typical genomic organization of the complex I of various kinds obtained from bacteria: **1** An operon encoding a classical, 14-subunit version of complex I comprising *nuaA*, *B*, *C*, *D*, *E*, *F*, *G*, *H*, *I*, *J*, *K*, *L*, *M* and *N* genes, exemplified by *Bradyrhizobium japonicum*. **2** An operon encoding a 12-subunit version of complex I as found in *Campylobacter jejuni* and *Helicobacter pylori*. **3** An operon encoding an 11-subunit version of complex I as found in *Chlorobium tepidum*. **4** The genes encoding the 12-subunit version of complex I containing an F₄₂₀ dehydrogenase, as exemplified by *Methanosarcina mazei*. **b** The *nua*-operon gene context of *B. cereus* (above) is compared with that of *B. subtilis* (below), which lacks *nua* genes



degree of accuracy. In the present study, we investigated the presence of such compact 11-subunit versions of complex I enzymes, consisting of only two modules, in the prokaryote genomes currently available. We found this enzyme to by no means be a rare exception. Our finding instead shows such 2-module complex I enzymes scattered rather widely throughout the phylogenetic tree of life. The novel complex I-like 11-subunit enzymes showed a primary sequence variability as great as or greater than that of the standard, full-size 14-subunit complex I. Yet all the 11-subunit enzymes found differed distinctly from the membrane-bound hydrogenases. We thus conclude that this version of compact 2-module complex I is ancestral to all present-day complex I enzymes.

Materials and Methods

Distribution of 11-Subunit Complex I

Searching for Complex I Proteins in the Sequenced Whole Genomes Available

Whole genome sequences from 656 prokaryote organisms were found to be available in the Comprehensive Microbial Resource (CMR) database (Peterson et al. 2001), from 1,715 organisms in the National Center for Biotechnology Information (Sayers et al. 2009) and from 1,317 prokaryote organisms in the integrated microbial genomes database (Markowitz et al. 2010) from the DOE Joint Genome Institute. Of these, 1,516 were unique organisms, 1,426 were eubacteria and 90 were archaea. We used the COGs of each of the complex I subunits to screen the genomic data. These are NuoA (COG:0838), NuoB (COG:0377), NuoC (COG:0852), NuoD (COG:0649), NuoE (COG:1905), NuoF (COG:1894), NuoG (COG:1034), NuoH (COG:1005), NuoI (COG:1143), NuoJ (COG:0839), NuoK (COG:0713), NuoL (COG:1009), NuoM (COG:1008), and NuoN (COG:1007). During screening, the strongly conserved NuoH sequence was used as an initial bait, followed by NuoE, NuoF, and NuoG (that are quite often located outside the typical *nuc*-operon structure) for identifying the organisms that contain a standard, full-size, 14-subunit complex I and the organisms that lack complex I. The organisms suspected of containing a smaller, 11- or 12-subunit complex I were then screened for the presence of the remaining subunits, and the chromosomal gene context was inspected manually. FpoF subunits were identified using COG:1035. Proteins for which misannotation or other irregularities were suspected, were checked by BLAST search or by alignment of the sequence with known gene sequences using the ClustalW (Thompson et al. 1994). The complete set of organisms investigated is

listed in Table S1, which is included in Supplementary material.

Construction of the 16S rRNA Phylogenetic Tree

To construct the 16S rRNA phylogenetic tree, a representative sample was first chosen from the organisms listed in Table S1. The organisms were selected so as to include members of each class of each phyla contained in the archaeal and eubacterial kingdoms. The organisms chosen were marked, the 16S rRNA accession numbers being listed in Table S1. The 16S rRNA sequences from the organisms selected were collected from the CMR database (Peterson et al. 2001) at TIGR or from the NCBI genome resource (Sayers et al. 2009) and were aligned using ClustalW (Thompson et al. 1994). The aligned data set was analyzed in Data Analysis in Molecular Biology and Evolution (DAMBE) ver 4.13, and was converted into MEGA format. Unrooted phylogenetic trees were created using MEGA version 4.1 (Tamura et al. 2007) and neighbor-joining method (Saitou and Nei 1987) with the bootstrap support of 120 replicates. Creating the phylogenetic tree, the parameters used were: complete deletion of gaps/missing data, distance model set to applying the nucleotide kimura-2-parameter, homogeneous pattern among lineages and uniform rates among sites and using the maximum composite likelihood model. The scale unit is number of substitutions per site.

Analysis of Polypeptide Primary Sequences

Construction of Phylogenetic Trees of NuoH, and NuoBCDI, Subunits of Complex I

A representative sample of protein subunits was selected, each type of complex I and members of each of the phyla of each kingdom both of the archaea and the eubacteria being represented in it. The organisms selected and their classification are listed in Supplementary material contained in Table S2. The primary protein sequences of NuoH, NuoB, NuoC, NuoD, and NuoI from the organisms selected were collected from the CMR and NCBI databases, as before. The NuoH subunit sequences were used without modification for construction of the NuoH tree, whereas the NuoB, NuoC, NuoD, and NuoI sequences were joined manually for constructing of a joint Q-module NuoBCDI phylogenetic tree. The peculiar N-terminal extension of the NuoC subunit found only in *Firmicutes* was omitted in constructing the NuoBCDI tree. The primary protein sequences were aligned using ClustalW with default settings. The aligned data set was analyzed in DAMBE ver 4.13, the misaligned data was deleted, and the data set was converted into MEGA format. Unrooted phylogenetic trees were created using MEGA version

5.0.3, bootstrap support of 100 replicates being provided in the neighbor-joining method and the maximum likelihood method, respectively. When using the neighbor-joining method, the trees were drawn to scale, branch lengths being expressed in the same units as those for the evolutionary distances used to infer the phylogenetic tree. The evolutionary distances were computed using the Poisson correction method and are expressed in terms of the number of amino acid substitutions per site. The Whelan and Goldman (WAG) model was used (Whelan and Goldman 2001) for the maximum likelihood method. All positions containing gaps or missing data were eliminated, and the tree heuristic method being employed for obtaining the nearest neighbor interchange. Calculations of distance between protein sequences were carried out using MEGA ver 4.0.1, settings of pairwise distance calculations being employed for purpose of analysis. Evolutionary distances in terms of the number of amino acid substitutions per site, were computed using the Poisson correction method. Calculations of distances between protein sequences were performed using MEGA ver 4.0.1, settings of pairwise distance calculations being used for purpose of analysis. In computing distances only, Gaps/Missing Data was set to complete deletion, the substitution model used was amino: poisson correction (Zuckerandl and Pauling 1965), homogeneous pattern was used among lineages and uniform rates among sites and the maximum composite likelihood model was used. The complete data set was used for including substitutions. The names of organisms are abbreviated by use of four character designations, for example *Escherichia coli* is abbreviated as E.col. The protein accession numbers are listed in Supplementary material.

Construction of Phylogenetic Trees Comparing the NuoD and NuoH Subunits of Complex I with the Homologous Subunits of the Membrane Hydrogenases

To obtain a representative sample of membrane-bound hydrogenases, use was made both of the smaller hydrogenase-3 and the larger hydrogenase-4 type enzymes. The CMR database was used as before. FHL-1(Hyc operon), FHL-2 (Hyf operon), and Ech hydrogenases were screened for in the 656 organisms by use of the COG search option in the CMR tools. The query proteins used were HycE (COG:3261), HyfC (COG:0650), and Ech B (COG:3260), respectively. The primary protein sequences of the HyfC (NuoH homologue) and HycE (NuoD homologue) subunits of FHL and the EchB (NuoH homologue) and EchE (NuoD homologue) subunits of Ech hydrogenase from both archaea and eubacteria were then collected. This first protein data set was added to the NuoH data set previously collected, this combined data set then being aligned using

ClustalW at default settings. The second data set consisted of NuoD sequences that had been used in the earlier, combined tree to which the newly obtained hydrogenase sequences were added. The aligned data sets were analyzed as before using DAMBE ver 4.13 and were converted into MEGA format. Unrooted phylogenetic trees were created using MEGA version 4.0 using the neighbor-joining method with bootstrap support of 100 replicates. The tree is drawn to scale, branch lengths being expressed in the same units as those for the evolutionary distances used to infer the phylogenetic tree. The evolutionary distances were computed as before. The protein accession numbers that apply are listed in Supplementary material.

Searching for a Putative Partner Protein for 11-Subunit Complex I

Genome Versus Genome Comparisons

The attempt to find a putative partner protein for the 11-subunit complex I was made through use of the multi genome comparison tool in the CMR data base (Peterson et al. 2001). Genome comparisons were carried out using *Bacillus cereus* ATCC 14579 as the reference molecule and *Bacillus licheniformis*, *Bacillus subtilis*, *Bacillus halodurans*, *Bacillus clausii*, and *Oceanobacillus iheyensis* as excluded comparison molecules, and *Bacillus thuringiensis*, *Bacillus anthracis*, and *Geobacillus kaustophilus* as included comparison molecules. The minimum percentage identity accepted was set to 50% as the cut-off value during the search.

BLAST and PSI-BLAST Searches

BLAST (Altschul et al. 1990) and PSI-BLAST (Altschul and Koonin 1998) searches for candidate protein sequences were carried out using the BlastP program with default settings. For running the PSI-BLAST, the settings were the same, except that the statistical threshold was set to 0.005. In searching for a NADH dehydrogenase module homologue, the protein subunits of the complex I of *E. coli* were used as query proteins, the subunits being NuoE (GenBank ID: AAC75345.1), NuoF (GenBank ID: AAC75344.1), and NuoG (GenBank ID: AAC75343.1). The *Methanosarcina mazei* FpoF sequence (GenBank ID: AAM30323.1) was used in searching for FpoF homologues.

Searching for a Putative Transcription Factor for 11-Subunit Complex I

Searching for a transcription factor, controlling the expression of the *nuo* genes encoding the 11-subunit complex I and its hypothetical partner protein(s) was

attempted by using the Prodonet tool with default settings in Prodoric ver 3.0 (Munch et al. 2003). In *Bacillus subtilis*, DBTBS release.5 database (Sierro et al. 2008) was used for searching of transcription factors regulating the expression of the *Qcr* and *Cta* operons (encoding the *bc*₁ complex and cytochrome c oxidase, respectively) with default settings.

Results

Distribution of Different Types of Complex I in Nature

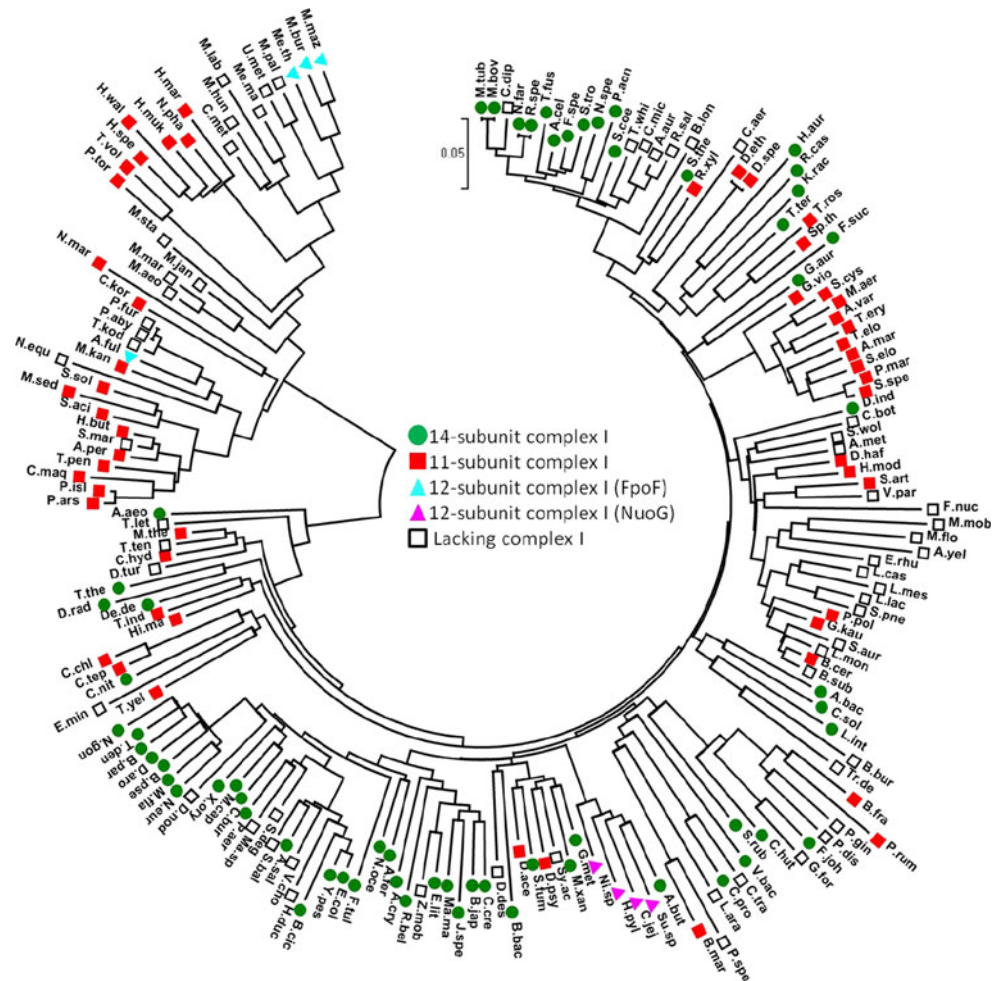
The distribution of the different versions of complex I was first investigated by searching for complex I encoding proteins throughout the whole genome sequences in the CMR, NCBI, and DOE Joint Genome Institute databases. The 1,426 unique bacterial and 90 archaeal whole genomes available at the time were searched for genes encoding for complex I as described in “Materials and Methods” section. All together, 625 organisms were found to contain the “full-size” 14-subunit complex I, 41.2% of the sample thus containing this classical, standard complex I. At the same time, 670 of the organisms were found to lack complex I, and to make use instead of other NADH-metabolizing enzymes, such as the non-energy-coupled NdhII enzyme (Jaworowski et al. 1981). The organisms lacking complex I comprised 44.1% of the sample. In addition, 40 of the organisms were found to contain a complex I that consisted of 12 subunits, either with FpoF-type subunits or NuoG-like subunits, but lacking NuoE and NuoF subunits. These organisms made up 2.6% of the sample. Finally, there were 181 organisms, or 11.9% of the sample, that contained the 2-module version of complex I that contained 11 subunits. A complete list of the organisms investigated, and the types of complex I they utilized, is given in Supplementary material (Table S1). Note, however, that since the choice of organisms to be subjected to whole genome sequencing was determined by factors other than that of whether they contained complex I, the percentages obtained do not exactly reflect how common the various complex I types are in nature. The sample can be said to nevertheless be representative for the tree of life in the sense that it contains at least some members of all major branches of the tree. Thus, the information presented concerns the distribution rather than the abundance of the 11-subunit, 2-module enzyme complex. To visualize the distribution of the different types of complex I involved, a sample representing at least some members of each of the branches was selected from Table S1, this being used to construct a standard phylogenetic tree based on 16S rRNA from these organisms (Fig. 3). In the archaea, 11-subunit complex I enzymes were found in the crenarchaeota (specifically in Sulfolobales, Thermoproteales, and Desulfurococcales),

the euryarchaeota (in Halobacteriales and Thermoplasmales), the korarchaeota (one organism only), the thumarchaeota (in Cenarchaea and Nitrosopumilales) and in one unclassified organism in the archaeal category, *Caldiarchaeum subterraneum*. The 12-subunit version of complex I, in which an F₄₂₀ dehydrogenase subunit (FpoF) substitutes for the N-module, is only found in the euryarchaeota. In the eubacteria, the 11-subunit complex I was found in the actinobacteria, δ -proteobacteria, bacteroidetes, chlorobi, chloroflexi, prochlorales, oscillatoriales, plancomycetales, nitrospirae, thermodesulfobacteria, and firmicutes. It can be concluded that the 11-subunit complex I is quite abundant, both in the archaea and the eubacteria. It is thus much more likely to represent a last common ancestor than its being a modified enzyme version that has repeatedly lost the N-module.

The Relationship Between the Compact 11-Subunit Complex I-Like Enzymes and the Standard, Full-Size 14-Subunit Complex I

To evaluate the properties of the compact 11-subunit enzymes as compared with those of the other complex I-type enzymes, sequence alignments of protein subunits from the 11-, 12-, and 14-subunit complex I, respectively, were performed, unrooted phylogenetic trees being drawn from the datasets obtained. The NuoH protein was selected for the initial comparisons. Although NuoH is a membrane-spanning protein, it is strongly conserved and it aligns well without gaps or problematic sequence stretches being frequent at all. Also, in all the enzyme complexes it presumably resides in the same environment and has the same protein neighbors. As can be seen in the phylogenetic tree of NuoH (Fig. 4), the sequences are distributed about as would be expected on the basis of phylogenetic relationships between the organisms involved. For example, the same phyletic members, *Chlorobium tepidum*, *Chlorobium chlorochromatii*, *Prosthecochloris vibrioformis*, and *Pelodictyon luteolum*, form separate branches within the Chlorobi group. There are very few exceptions to this pattern, suggesting that the NuoH protein has essentially the same function in all the organisms investigated, irrespective of whether they contain the 11-, 12-, or 14-subunit version of complex I. Trees constructed using the neighbor-joining method and trees constructed using the maximum likelihood method gave very similar results (not shown). One interesting exception should be noted. The ϵ -proteobacteria from *Arcobacter*, which contains the full-size 14-subunit version of complex I, are not paring with the other ϵ -proteobacteria members *C. jejuni* and *H. pylori* that harbor a peculiar 12-subunit complex I containing an NuoG-like subunit but lacking the NuoE and F subunits. This could possibly have some sort of functional implication.

Fig. 3 Distribution of different versions of complex I in an unrooted phylogenetic tree constructed on the basis of 16S rRNA sequences. Organisms that contain a full size, standard 14-subunit version of complex I are marked with circles and organisms that contain an 11-subunit version of complex I with filled squares. Organisms containing a 12-subunit version of complex I are marked with triangles. The methanogens *Methanosarcina mazei* (M.maz), *Methanococcoides burtonii* (M.bur) and *Methanosaeta thermophila* (Me.th) and *Archaeoglobus fulgidus* (A.ful) have FpoF as the 12th subunit whereas the *Campylobacter jejuni* (C.jej), *Helicobacter pylori* (H.pyl), *Sulfurovum* sp. (Su.sp) and *Nitratiruptor* sp. (Ni.sp) have a NuoG-like protein that make up the 12th subunit. Organisms that lack complex I are marked by open squares. A complete list of the organisms investigated is provided in the supplementary material, with those used in the figure and the corresponding accession numbers being marked with an asterisk in the list (Table S1)



Phylogenetic trees of NuoB, NuoC, NuoD, and NuoI subunits comprising the Q-module were also constructed (not shown), to investigate to what extent any functional specialization appear to exist in this part of the enzyme complex. A combined tree, based on a sequence alignment of all four subunits was also constructed (Fig. 5). Interestingly enough, the Q-module tree was found to be very similar to the NuoH tree, strongly suggesting that the core subunits common to the compact 11-subunit version of complex I and to the full-size 14-subunit version of complex I, respectively, are very similar in their functions. The very few differences found between the two trees (Figs. 4, 5) could very well result from experimental artifacts, since NuoH is a membrane protein and the Q-module members are soluble proteins. In any case, it is obvious from the complex I structure (Sazanov and Hinchliffe 2006) that the surface contact area between the N-module and the Q-module is in fact quite small. In the thermophile *T. thermophilus*, an additional, frataxin-like subunit (denoted as Nqo15) is present at the interface between the N and Q modules (see Fig. 1b), presumably to stabilize the enzyme complex (Hinchliffe et al. 2006; Sazanov and

Hinchliffe 2006). Since few amino acid residues are involved in intermodule interaction in the full-size 14-subunit version of complex I, it is perhaps not surprising that the differences between the membrane-spanning NuoH subunits and the Q-module subunits were very modest in the two types of complex I.

The individual Q-module subunit trees showed no particular features other than what could be observed in the combined tree, except for one thing, that of a peculiar N-terminal extension being present in the NuoC subunit of 11-subunit version of complex I in *Bacillus cereus*, *Bacillus anthracis*, *Bacillus thuringiensis*, and *Geobacillus kaustophilus*. This extra domain is not present in any the other 11-subunit version of complex I outside the realm of the *Bacillaceae* (Fig. 6). The structure of the full-size version of complex I suggests NuoC to have an important role in stabilizing the Q-module, since the subunit is in contact with both NuoD and NuoI in the Q-module. In *T. thermophilus*, the C-terminal loop of NuoC interacts with NuoG in the N-module (Sazanov and Hinchliffe 2006). A shorter and unrelated N-terminal extension is also to be seen in NuoC from *A. fulgidus*, in which a 12-subunit, FpoF-containing

Fig. 4 An unrooted phylogenetic tree of the NuoH subunit with a bootstrap support of 100 replicates. Organisms containing standard, 14-subunit versions of complex I are marked with circles, organisms containing 11-subunit versions of complex I with squares, and organisms containing 12-subunit versions of complex I with triangles. *M. mazei* (*M. maz*), *Methanosarcina barkeri* (*M. bar*) and *A. fulgidis* (*A. ful*) represent such enzymes having FpoF and *C. jejuni* (*C. jej*) and *H. pylori* (*H. pyl*) are examples of enzymes having a NuoG-like 12th subunit. A list of the full names of the organisms and the sequence accession numbers involved are contained in the supplementary material

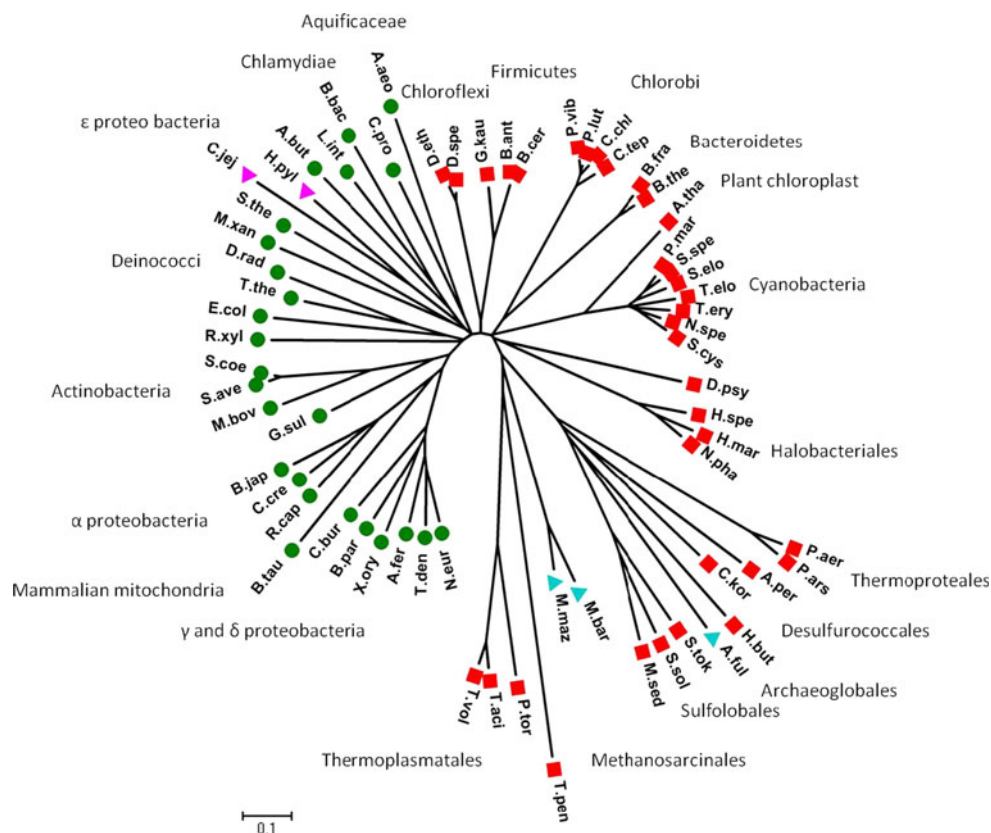
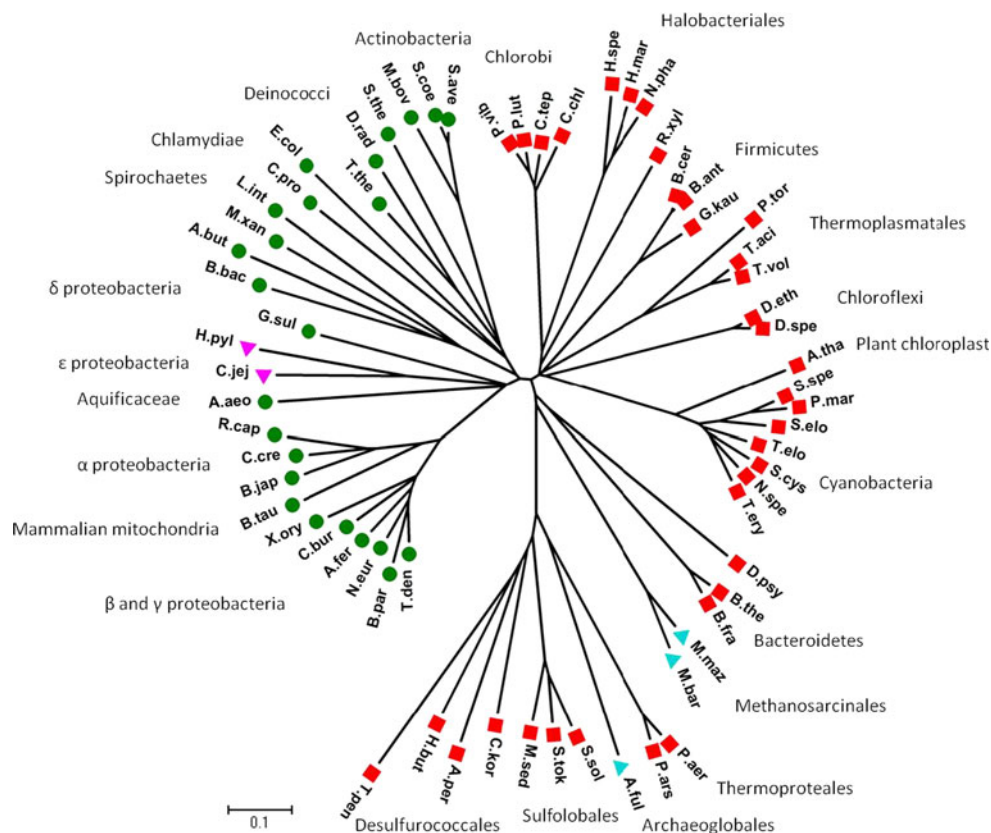


Fig. 5 An unrooted phylogenetic tree with a bootstrap support of 100 replicates of the combined NuoBCDI subunits comprising the Q-module of complex I. The N-terminal extension, found only in the *Bacillaceae* NuoC subunits, was omitted from the alignment, so as not to obfuscate the overall comparison. The organisms are labeled as they are in Fig. 4. As before, the full names of organisms and sequence accession numbers are listed in the supplementary material



NuoC Ba	*MSDPNKDLEDLKREARRAKEEARKRLVAKHGVEISKLEENREKEKALPKNDMDTIEEAKRRAAAAAKAKAALAKQKREGIEEVTEEEKVKAKAKAAAAKAKAALAKQKREGIEEV	120
NuoC Bt	*MSDPNKDLEDLKREARRAKEEARKRLVAKHGVEISKLEENREKEKALPKNDMDTIEEAKRRAAAAAKAKAALAKQKREGIEEVTEEEKAKAKAKAAAAKAKAALAKQKREGIEEV	120
NuoC Bc	*MSNPKNKLEDLKREARRAKEEARKRLVAKHEAISELEEDQEKEKALPKNDITIEEAKRRAAAAAKAKAALAKQKREGTEEVTEEEKAKAKAKAAAAKAKAALAKQKREGTEEV	120
FpoC Af		*MISMDEFIEFVRGGP 16
NuoC Ba	TEEE-----KAKAKAKAVAAKAKAALA	144
NuoC Bt	TEEEKVKAKAKAAASLAKQKREGTEEVTEEEKAKAKAAAAKAKAALAKQKREGTEEVTEEEKAKAKAKAVAAKAKAALA	240
NuoC Bc	TEEEK-----AKAKAKAAAAKAKAALAKQKREGTEEVTEEEKAKAKAAAAKAKAALAKQK-----REGTEEVTEEEKAKAKAKAALA	204
FpoC Af	LGPIKKGWTKWSLWPVHLVTACCGAELAHAFACGYDGERIGALNYGIARQTNLIIVEGAITRKMAKVLRIWEQMPDPKFVIMVGCAGLGGGIFWNGYHVMKPSDVPVDFPFGCPTPT	135
NuoC Ba	KQKASQNGDSDGEKAKAIAAAKAKAAAAARAKTKGAEKGKEELKQEEPSVNEPYNQYVEVIGKIGEDILVDIFYINKLSKDVPTLVVEPAKYIEVMELLRFHEGLAFDYMSELHATD	264
NuoC Bt	KQKASQNGDSDGEKAKAIAAAKAKAAAAARAKTKGAEKGKEEPPKQEEPSVNEPYNQYVEVIGKIGEDILVDIFYINKLSKDVPTLVVEPAKYIEVMELLRFHEGLAFDYMSELHATD	360
NuoC Bc	KQKASQNGDSDGEKAKAIAAAKAKAAAAARAKTKGAEKGKEDEPKQEEPSVNPQYLYVEAIREKLGEALVDSYINKLSKDVPTLVVDPEKYIEVMESLRFHEGLAFDYMSELHATD	324
NuoC Ct		*MEEAANQMSPAVQSKAAVDNIKERFGDAISEFDANPTMFFFEVLVDKSVVDIAYLMRDSLLQPNYLACLSGVD 75
NuoC Bf		*MEEIKYIEPAALHDEMLRLRNEKQMDFLESITGMDWGVAEEGD 43
NuoC Ss		*MGDEKMSALKSLVNDLQTKFKCAKIESDRRTIVIPDKKIIRDVASFLLKRRITVIVPDKKIIRDVASFLLKDFGDFDHVKAFTGID 65
NuoC Cj		*MRKYSDDKNVQIQNYKDRFYHAPHTIKSNVNESIFKDDFEVLKTEVEILNSYVELDFWVIEIKKEDNVKTQLMLKTGLYLSFTEASAI 90
FpoC Af	EALLRGIRQLQFKIETGEAKTSATFPEVPLEAGRKPRVLKPSKKISKAPAVIVNAPKEVDWFEFGQKLVEEIKAKVEAKSVITGKNRIAVKVESDYIPRAAMRLREMGDFHVKNNVID	255
FpoC Mm		*MDARTIIESLTGKFPPEAISEAGIESPIRIRAYVDKKAKEVCEYELKGSQDFDLCSVSGVD 61
NuoC Ec		*MTDLTAQEPAWQTRDHLDDPVTIGELRNRFGPDFTVQATRTGVPVVMWIKREQLLEVGFDFLKLKPKPYVMLFDLHGMDE 78
NuoC Bt	*MAAAVAAAAGCQWQLVGSAAAPARVAGRPSVLLLPVRRESSAADTRPTVRPRNDVAHKQLSAFGEYVAILPKYVQVQVSCFNELEICIHDPGIVPVLFLRDSHNAQPKSLADLTAVD	120
HycE Ec		*MSEELKGQHYLAALNEAFPVVLDHAWQTKDQLTVTVKNVYLPVEVFEYLYKQGGWLSVLFNGD 64
HyfG Ec		*MNVSSSNRGEAILAALKTKQPPGAVLDEERQTPQVITVKNINLLPQVQYLYYQHGDWLPVLFNGD 67
EchE Ma		*MDVKMNETKELLSTISSIAKEEPELLSARENEIYLIKEAGFEDITPALAERFSLIGLFCAE 64
NuoC Ba	FVTH-----MEVYVHLFSYG---KKQSVAVKVKLD-REAPQVESVTALWKGADWPEREAYDLLGIVFKGHPNL-SRLIMPDDWVG--HPLRKDYEPYDVEV*	353
NuoC Bt	FVTH-----MEVYVHLFSYG---KKQSVAVKVKLD-REAPQVESVTALWKGADWPEREAYDLLGIVFKGHPNL-SRLIMPDDWVG--HPLRKDYEPYDVEV*	449
NuoC Bc	FVTH-----MEVYVHLFSYG---KKQSVAVKVKLD-REAPQVESVTALWKGADWPEREAYDLLGIVFKGHPNL-TRILMPDDWVG--YPLRKDYEPYDVEV*	413
NuoC Ct	YPPEQK-----LGIVCNLEICIG---KYTHKIAVKVKCP-RDGGSIPIVSVSCVWHTANWHEREREYDMYGMVFEGHPDL-RRILCPDWTG--FPLRKDYQVQETYHGIKVPY*	173
NuoC Bf	APNVTRG-----LGVVYHLESTVT---GERIAVKTSTDNRETPEIPSVSDIWKADFNREYDFDYIGVIFIGHPDM-RRILYLRNDWVG--YPMRKDNNEKDNPLRMDNEETDITQ-	150
NuoC Ss	YPEHEK-----LEVYHISYSNLDLAKIILALRTTVS-YKDPISLPSLYSIFESAWTGERETIEMLGIRFEGHPDL-RRFLLDEDFEGY-YPLRKSYSIKLEGVFVDPKPTTE*	169
NuoC Cj	FIANKNG-----FEVFFYQLNLEK---KLRVRIKTFIG--VKERLQSVAHIFKCANWSEEREIYDMFGIFIVGHPNL-KRILMPDDWVG--HPLKITYPLKGEFARWYEDIKIFGKE-	195
FpoC Af	VPNE-----DKFIVEYHFFSSYVKELMPVIVNVFADIP-RENPRVKSLASMFPSADYMEREMHDFGFIIFEGNPMGRKPELLAPDAPE--FPLRKD---FKLEEEVYVR*	353
FpoC Mm	YPRDE-----LEAVYHIASD---HPVVLMLKARLP-RDSPEIESVSVSVYWNANWYEREITYELGIFFKNHPDL-KPLVLPDDMLGE-WPLRKDYEGFPNRTARNLV*	158
NuoC Ec	RLRTHREGLPAADFVVFYHLISIDR---NRDIMLKVALA-ENDLHVPTTFKLFNPANWYEREITWDLFGITFDGHPNL-RRIMMPQTKWG--HPLRKDYPARATEFSPFELTKAKQDLEM	190
NuoC Bt	IPTRQNR-----FEIVYNLLSLR---FNSRIRVKTYTD-ELTPIESSVPVYKANWYEREITWDMFGVFFANHPDL-RRILTDYGFEG--HPLRKDFPLSGYVLELYRDEVKRVVAEP	225
HycE Ec	-ERKLNG-----HYAVYVLSMEK---GTKCWITVRVEVD-ANKPEYPSVTPRPVPAVWGEREVRDMLYGLIPVGLPDE-RRVLVLPDDWDELYLQKAFKNGFVHPVETPSKE--YKFR-	173
HyfG Ec	-ERTLNG-----HYAVYVLSMEG---AEKCIWVVKALVD-ADSRFFPSVTPRPVPAVWGEREIRDMYGLIPVGLPQD-RRVLVLPDDWDEPMHPLRKDAMDYRLRPEPTTDSITYFFIN-	176
EchE Ma	VFENTDG-----FTLFYAFKQAG---KASVLVLRREP-GKEKRVSVAGIFPSAWFEREIRDFGLDFAGALDN-RRFLHECYHEGHPLQKAFKNGFVHPVETPSKE--YKFR-	169

Fig. 6 Alignment of the NuoC subunit. The N-terminal extension of *B. cereus*, *B. anthracis* and *B. thuringiensis*. In order of appearance, the sequences shown are NuoC polypeptides from *B. anthracis*, *B. thuringiensis*, *B. cereus* (Firmicutes), *Chlorobium tepidum* (Chlorobi), *Bacteroides fragilis* (Bacteroidetes), *Sulfolobus solfataricus* (Crenarchaeota—Sulfolobales), *Campylobacter jejuni* (Proteobacteria—Campylobacteriales), *Archaeoglobus fulgidus* (Euryarchaeota—Archaeoglobales), *Methanosarcina mazei* (Euryarchaeota—Methanosarcinales), *E. coli* (Proteobacteria—Enterobacteriales), and *Bos*

taurus (mitochondria). The names of the respective polypeptides are followed by the two first letters in the name of the organism, i.e., the *E. coli* protein sequence is denoted as NuoC Ec, the *M. mazei* sequence is denoted as FpoC Mm and so forth. Conserved residues are marked in bold letters and the start and the end of a polypeptide are indicated by asterisks. The NuoC and NuoD subunits are fused in *E. coli* and *B. fragilis*. Likewise, the NuoC subunit is homologous to the N-terminal part of the HycE, HyfG, and EchE polypeptides in the membrane-bound hydrogenases from both bacteria and archaea

version of complex I is found (Fig. 6). It is possible that these NuoC extensions may have a role of some sort in the contact which is established with the putative partner proteins.

The Search for a Partner Protein

Initially we searched the genomes in question for proteins resembling either NADH dehydrogenases or FpoF, without finding any particularly promising candidates. Accordingly, and also to avoid being limited by preconceptions regarding the appearance that such a partner protein should have, we adopted a different search strategy. In firmicutes, there are several closely related species that either contain an 11-subunit version of complex I or lack complex I entirely. In *B. cereus*, *B. thuringiensis*, *B. anthracis*, and *G. kaustophilus* the operon encoding the 11-subunit version

of complex I is located between the gene clusters encoding ATP synthase and MurA, a protein involved in peptidoglycan biosynthesis (Fig. 2b). The chromosomal context is remarkably well conserved in other *Bacillaceae* that lack complex I, except of course that the *nuo* genes are missing (Fig. 2b). Because of the close relationship between these organisms, the attempt was made to identify a partner protein for the 11-subunit version of complex I by means of genome:genome comparisons. Using the *B. cereus* chromosome as the reference molecule, we included all proteins that matched *B. thuringiensis*, *B. anthracis*, and *G. kaustophilus*, i.e., those obtained from related organisms that contained the 11-subunit version of complex I. The criteria for matches of a given protein to the reference employed were set rather generously ($\geq 50\%$ identity) for such closely related bacteria. We then excluded all proteins

that matched the genomes of the closely related bacteria that lack complex I, i.e., those of *Bacillus licheniformis*, *Bacillus subtilis*, *Bacillus halodurans*, *Bacillus clausii*, and *Oceanobacillus iheyensis*. This resulted in 119 candidate proteins remaining. All of the 11 proteins comprising the compact 11-subunit version of complex I were included among these, indicating the sorting strategy to have indeed worked. Some of the 108 remaining proteins could be discarded immediately on the bases of their annotation. The next step was to conduct a BLAST search for candidate proteins in the genomes of other organisms having the 11-subunit version of complex I. A few additional candidate proteins were found using this approach, but none of them stood up to the final test of being uniquely associated with the 11-subunit version of complex I. The best putative candidate subunit in *Bacillaceae* (BC0791 found in *B. cereus*) that was also present in *Bacterioides* and *Chlorobium*, was for example also found in *Yersinia* (that contains a classical 14-subunit complex I) and in *Lactococcus lactis* (that is without complex I).

In another search strategy performed, databases of transcriptional regulation were used to search for genes that

were putatively co-expressed together with genes encoding the 11-subunit version of complex I, the *bc₁* complex and cytochrome oxidase (see “Materials and Methods” section). Its being assumed that transcription factors and recognition sequences in *B. subtilis* and the other *Bacillaceae* that contain the 11-subunit version of complex I are well conserved. No partner protein could be identified by means of this approach, but it should be emphasized that there is virtually no real data available concerning the transcriptional regulation or gene expression profiles of *B. cereus*, *B. thuringiensis*, or *B. anthracis*.

The Relationship to Present-Day Membrane-Bound Hydrogenases

The membrane-bound hydrogenases are of two basic types. The smaller ones, termed hydrogenase-3, contain six protein subunits corresponding to the Q-module and NuoH together with one antiporter-like subunit. The larger hydrogenase-4 enzymes contain additional membrane-spanning subunits, having three antiporter-like subunits and a homologue of NuoK. Since these subunits are poorly

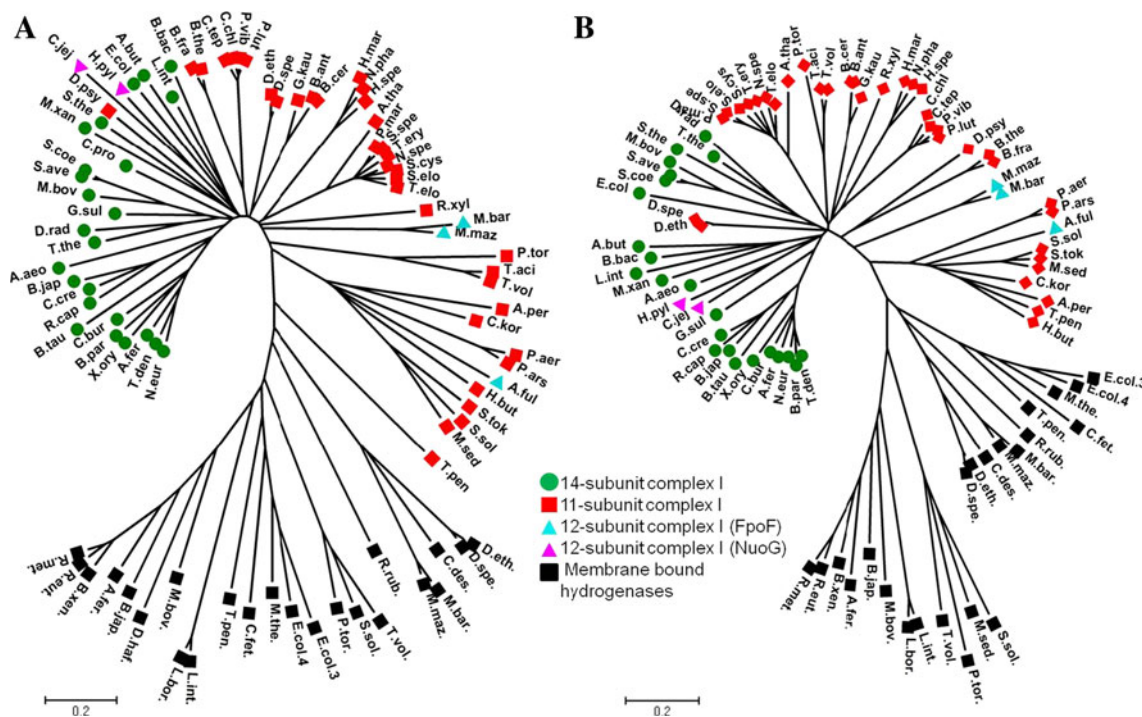


Fig. 7 a An unrooted phylogenetic tree of NuoH-homologous subunits with a bootstrap support of 100 replicates. Organisms containing 14-subunit versions of complex I are marked with circles (green in the electronic edition), organisms containing 11-subunit versions of complex I with grey squares (red in the electronic edition), organisms containing 12-subunit versions of complex I are shown with triangles; enzymes that contain NuoG are shown in pink in the electronic edition and enzymes containing FpoF with blue triangles. The NuoH-homologous subunits in membrane bound

hydrogenases, EchB/HycD/HyfC, are marked with black squares. Since some organisms contain both complex I and hydrogenases a dot after the abbreviation of the organism is used in the latter case. **b** An unrooted phylogenetic tree of NuoD-homologous subunits with a bootstrap support of 100 replicates. The symbols for complex I subunits are the same as in a. The NuoD-homologous subunits from membrane bound hydrogenases, EchE/HycE/HyfG, are marked with black squares. The full names of organisms and the sequence accession numbers are given in the supplement

differentiated in hydrogenases, it is not possible to determine which antiporter-like subunit that corresponds to the complex I subunits NuoL, NuoM, and NuoN or to the Mrp subunits MrpA and MrpD (Mathiesen and Hägerhäll 2002). NuoA and NuoJ do not seem to be present in any of the hydrogenase. The phylogenetic trees shown in Figs. 4 and 5 were redrawn so as to also include sequences from membrane-bound hydrogenases of the hydrogenase-3 and hydrogenase-4 type to be able to investigate the relationship between the novel 11-subunit versions of complex I and the membrane-bound hydrogenases. A representative sample of primary sequence was collected from both archaea and eubacteria. Insofar as possible, organisms already used that contained both hydrogenase and complex I were selected here. The hydrogenases were found to be consistent in forming distinct groups, separate from all of the different versions of complex I, both when the NuoH-homologous subunits (Fig. 7a) and the Q-module subunits (Fig. 7b) were compared with one another. The relative positions of the different complex I subunits are basically the same as in Figs. 4 and 5. This similarity of sequences points to complex I and membrane-bound hydrogenases having a common ancestor, but the present-day hydrogenases, whether small, as Ech and Hyc are, or of large, as Hyf is, are being distinctly different from all present-day versions of complex I, regardless of if 11-, 12-, or 14-protein subunit enzymes are involved.

The obvious demarcation between the complex I enzyme complexes and the hydrogenase enzyme complexes is the presence of the NiFe-active site. The crucial metal center is ligated by two cysteine pairs, one at the N-terminal and the other at the C-terminal of NuoD (Fig. 8). As already mentioned, residues at the C-terminal of NuoD has been indicated in quinone binding, both by inhibitor resistant mutants and site-directed mutants defect in quinone reductase activity. Thus, the absence or presence of conserved cysteines provides a strong indication of whether or not a given enzyme can function as a hydrogenase. All of the compact, 11-subunit versions of complex I enzymes were found to lack a cysteine pair at the C-terminal (Fig. 8 and not shown). Note, however, that some of the enzymes found in methanogens having the 12-subunit version of complex I, do in fact retain one of the cysteines (Fig. 8), suggesting that loss of the NiFe site could be a relatively recent event that occurred following the split from the last common ancestor.

Discussion

In the present study we could show that the 11-subunit version of complex I is considerably more common than had previously been thought. The compact, 2-module enzyme is widely distributed in the phylogenetic tree of life

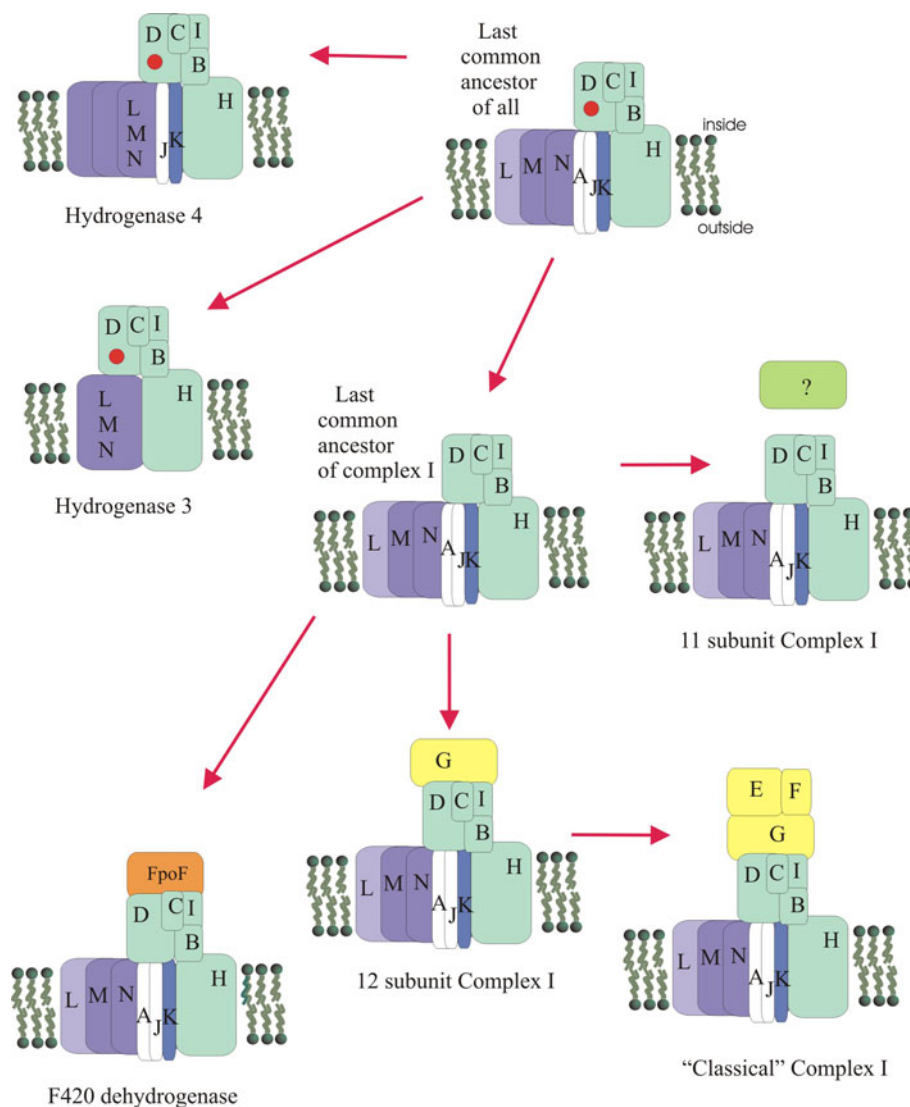
NuoD Bt	MLADVVAIIGTQDIVFGEVDR*----	463
NuoD Ec	LVSDLIVYLGSIDFVMSDVDR*----	433
NuoD Ae	TIADAIALLGSLDPVVGEDTR*----	446
NuoD Cj	YLADVVAIMGNVNIVLGEIDR*----	408
NuoD Bc	NIANLITILGGVDIVLGEVDG*----	366
NuoD Ct	LIPDLVAIIGSIDIVLGEVDR*----	368
NuoD Sc	KVADIMAILGSLDIIMGSVDR*----	394
NuoD Bf	KIADLIAIGGTLDYVVPDIDR*----	405
NuoD Np	YIPDLIASLGSLDIVLGEVDR*----	409
NuoD Ck	RVADIPAIYGSGLDYFPPEADR*----	389
NuoD Tv	MIADLISIGGSIDLVFGEVDR*----	367
FpoD Af	RLADLQAIYGSFGYFPPEADR*----	413
FpoD Mm	TVADMVSIISGSMGCTSEVDR*----	373
FpoD Mb	TVADVAAISGSMGCTSEADR*----	374
HycE Ec	TVSDAPLIIGSLDPCYSTDRTVVD-	403
HyfG Ec	TVSDAPLIIGSLDPCYSTDRVTLV-	409
HyfG Cf	TIADAALIVCSLDPCYSTERTLVN-	552
HycE De	ELADVPIVTLISIDPCISCTER*----	359
EchE Mb	KLADVPIVVLITIDPCVSCTER*----	358
EchE Mm	KLADVPIVLTIDPCISCTER*----	358
CooH Rr	RISDIPLIVNSIDPCISCTER*----	361

Fig. 8 The C-terminal end of the NuoD polypeptide and its homologues in membrane-bound hydrogenases, with the two conserved cysteines acting as NiFe site ligands marked in red. The polypeptide names are followed by the first two letters of the name of the organism, i.e., the *E. coli* protein sequence being denoted NuoD Ec, the *M. mazei* sequence being denoted FpoD Mm, and so forth. In order of appearance, the sequences shown are complex I polypeptides from *Bos taurus* (mitochondria), *E. coli* (Proteobacteria—Enterobacteriales), *Aquifex aeolicus* (Aquificae), *C. jejuni* (Proteobacteria—Campylobacteriales), *B. cereus* (Firmicutes), *C. tepidum* (Chlorobi), *Synechocystis* (Cyanobacteria), *B. fragilis* (Bacteroidetes), *Natronomonas pharaonis* (Euryarchaeota—Halobacteriales), *Candidatus korarchaeum cryptofilum* (Korarchaeota), *Thermoplasma volcanium* (Euryarchaeota—Thermoplasmatales), *A. fulgidus* (Euryarchaeota—Archaeoglobales), *M. mazei* and *Methanosarcina barkeri* (Euryarchaeota—Methanosarcinales), HycE polypeptides from *E. coli* and *Dehalococcoides ethenogenes* (Chloroflexi—Dehalococcoidetes) and HyfG polypeptides from *E. coli* and *Campylobacter* (Firmicutes—Clostridiales), EchE polypeptides from *M. mazei* and *M. barkeri* and a CooH polypeptide from *Rhodospirillum rubrum* (Proteobacteria—Rhodospirales). The end of the polypeptides are indicated by an asterisk

among both archaea and eubacteria (Fig. 3). Such an 11-subunit version of complex I thus appears to be the last common ancestor of all currently existing complex I-like type of enzymes. Analyses of the primary sequence showed unambiguously that the 11-subunit enzyme belongs to the complex I family, differing distinctly in this respect from the membrane-bound hydrogenases. Previously, the typical 14-subunit version of complex I, found in many prokaryotes, had been designated as representing the “minimal functional unit” of the enzyme. However, the 11-subunit version of complex I, which contains both the combined Q and P modules, represent the actual energy-coupling engine of the machine, whereas the N-module can be regarded as a mere electron delivery device.

No designated partner protein, acting as an electron delivery device, could be found for the 11-subunit version of complex I. This could of course be due to some flaw in

Fig. 9 Schematic representation of the evolution of complex I and of present-day membrane bound hydrogenases from an 11-subunit last common ancestor. The protein subunits are labeled as in Fig. 1. The NuoL subunit is more similar to the MrpA antiporter subunit whereas the NuoM and N subunits are more similar to MrpD, but the hydrogenase antiporter-like proteins appear to be undifferentiated. The NiFe active site is indicated by a circle



our search strategies (see “[The Search for a Partner Protein](#)” section). If the compact versions of the complex I enzymes differ in the partner proteins they use, one of the remaining 108 *B. cereus* proteins could still be a viable candidate for being a partner of only the *Bacillaceae* enzymes. A final answer to the question thus posed can only be obtained from further molecular biological and biochemical studies of the *B. cereus* enzyme. As already mentioned, however, no partner protein has been found in plant chloroplast or in cyanobacteria either, despite these enzymes having been studied quite extensively. In cyanobacteria, several varieties of the compact versions of the complex I enzymes exist, with quite versatile functions (Battchikova et al. 2011). On the bases of all of this, it is tempting to speculate that the primordial enzymes, and many of the present-day 11-subunit versions complex I, operate without a designated partner protein. Perhaps the Q-module in these enzymes should better be regarded as a

docking platform for different electron donor or acceptor proteins.

The Mrp antiporters consist of six or seven proteins typically encoded by one operon. The fact that these operons contain genes encoding two homologous proteins, MrpA and MrpD, indicates that these proteins are likely to differ somewhat in function. Phylogenetic analyses have revealed that NuoL can clearly be grouped together with MrpA, whereas NuoM and NuoN are more similar to MrpD (Mathiesen and Hägerhäll 2002). Recent complementation studies using *B. subtilis* MrpA and MrpD deletion strains have corroborated the functional similarity of MrpA to NuoL and of MrpD to NuoN (Moparthi et al. 2011). The homologous antiporter-like subunits that hydrogenases contain show a lesser degree of sequence similarity, and do not consistently group together with either MrpA or MrpD, irrespective of whether they originated from hydrogenase-3, and thus have one homologous

subunit, or from hydrogenase-4, which harbors three such subunits (Mathiesen and Hägerhäll 2002). This suggests that the hydrogenase subunits have lost some of the primordial functional specialization they once possessed. Therefore, it must be concluded that an enzyme like the present-day membrane-bound hydrogenase could not be the last common ancestor of respiratory chain complex I. The phylogenetic analyses performed in the present study (Fig. 7a, b) corroborate these earlier findings. Thus, the larger enzyme family can be split in two groups, the one consisting of both small and large membrane-bound hydrogenases and the other, containing 11, 12 and 14 subunits, representing the bona fide complex I (Fig. 9). It should also be emphasized that loss of the NiFe site need not necessarily have occurred as it is outlined in Fig. 9, prior to the split of the bona fide version of complex I from the present-day hydrogenases. The interplay with other transient partner proteins could have started as an alternative, making the enzyme more versatile in the absence of H₂, but retaining the NiFe site for use when H₂ was present. In terms of such a scenario, the NiFe site may have been lost at different points in time in the different 11-subunit versions of complex I.

To conclude, an 11-subunit version of the enzyme complex, resembling the present-day 11-subunit version of complex I but that harbored an NiFe-active site on NuoD, appears to have been the last common ancestor of membrane-bound NiFe-hydrogenases as well as of all the present-day versions of complex I (Fig. 9). During the evolution from this 11-subunit last common ancestor, some enzymes remained hydrogenases, but shrunk, gradually degenerated and lost subunits or subunit specialization, as was the case with the NuoL, M and N homologous proteins. Other enzymes grew, acquired more permanent electron donor partner proteins, and eventually evolved into the present-day full-size complex I.

Acknowledgments The financial support of Carl Tryggers Foundation, The Crafoord Foundation and the Swedish Research Council is gratefully acknowledged.

Open Access This article is distributed under the terms of the Creative Commons Attribution Noncommercial License which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

References

- Altschul SF, Koonin EV (1998) Iterated profile searches with PSI-BLAST—a tool for discovery in protein databases. *Trends Biochem Sci* 23:444
- Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. *J Mol Biol* 215:403
- Battchikova N, Eisenhut M, Aro E (2011) Cyanobacterial NDH-1 complexes: novel insights and remaining puzzles. *Biochim Biophys Acta Bioenerg*. doi:10.1016/j.bbabo.2010.10.017
- Bäumer S, Ide T, Jacobi C, Johann A, Gottschalk G, Deppenmeier U (2000) The F420H2 dehydrogenase from *Methanosarcina mazei* is a redox-driven proton pump closely related to NADH dehydrogenases. *J Biol Chem* 275:17968
- Berrisford JM, Sazanov LA (2009) Structural basis for the mechanism of respiratory complex I. *J Biol Chem* 284:29773
- Birungi M, Folea M, Battchikova N, Xu M, Mi HL, Ogawa T, Aro EM, Boekema EJ (2010) Possibilities of subunit localization with fluorescent protein tags and electron microscopy exemplified by a cyanobacterial NDH-1 study. *Biochim Biophys Acta* 1797:1681
- Brandt U (2006) Energy converting NADH:quinone oxidoreductase (complex I). *Annu Rev Biochem* 75:69
- Carroll J, Fearnley IM, Skehel JM, Shannon RJ, Hirst J, Walker JE (2006) Bovine complex I is a complex of 45 different subunits. *J Biol Chem* 281:32724
- Darrouzet E, Issartel JP, Lunardi J, Dupuis A (1998) The 49-kDa subunit of NADH-ubiquinone oxidoreductase (complex I) is involved in the binding of piericidin and rotenone, two quinone-related inhibitors. *FEBS Lett* 431:34
- Efremov RG, Baradaran R, Sazanov LA (2010) The architecture of respiratory complex I. *Nature* 465:441
- Finel M (1998) Does NADH play a central role in energy metabolism in *Helicobacter pylori*? *Trends Biochem Sci* 23:412
- Friedrich T (2001) Complex I: a chimaera of a redox and conformation-driven proton pump? *J Bioenerg Biomembr* 33:169
- Friedrich T, Scheide D (2000) The respiratory complex I of bacteria, archaea and eukarya and its module common with membrane-bound multisubunit hydrogenases. *FEBS Lett* 479:1
- Friedrich T, Weiss H (1997) Modular evolution of the respiratory NADH ubiquinone oxidoreductase and the origin of its modules. *J Theor Biol* 187:529
- Friedrich T, Weidner U, Nehls U, Fecke W, Schneider R, Weiss H (1993) Attempts to define distinct parts of NADH:ubiquinone oxidoreductase (complex I). *J Bioenerg Biomembr* 25:331
- Friedrich T, Steinmüller K, Weiss H (1995) The proton-pumping respiratory complex I of bacteria and mitochondria and its homologue in chloroplasts. *FEBS Lett* 367:107
- Gray MW, Burger G, Lang BF (1999) Mitochondrial evolution. *Science* 283:1476
- Hedderich R (2004) Energy-converting [NiFe] hydrogenases from archaea and extremophiles: ancestors of complex I. *J Bioenerg Biomembr* 36:65
- Hinchliffe P, Carroll J, Sazanov LA (2006) Identification of a novel subunit of respiratory complex I from *Thermus thermophilus*. *Biochemistry* 45:4413
- Hunte C, Zickermann V, Brandt U (2010) Functional modules and structural basis of conformational coupling in mitochondrial complex I. *Science* 329:448
- Jaworowski A, Mayo G, Shaw DC, Campbell HD, Young IG (1981) Characterization of the respiratory NADH dehydrogenase of *Escherichia coli* and reconstitution of NADH oxidase in ndh mutant membrane vesicles. *Biochemistry* 20:3621
- Kerscher S, Kashani-Poor N, Zwicker K, Zickermann V, Brandt U (2001) Exploring the catalytic core of complex I by *Yarrowia lipolytica* yeast genetics. *J Bioenerg Biomembr* 33:187
- Markowitz VM, Chen IMA, Palaniappan K, Chu K, Szeto E, Grechkin Y, Ratner A, Anderson I, Lykidis A, Mavromatis K, Ivanova NN, Kyrpides NC (2010) The integrated microbial genomes system: an expanding comparative analysis resource. *Nucleic Acids Res* 38:D382
- Martin M, Sabater B (2010) Plastid ndh genes in plant evolution. *Plant Physiol Biochem* 48:636

- Mathiesen C, Hägerhäll C (2003) The ‘antiporter module’ of respiratory chain complex I includes the MrpC/NuoK subunit—a revision of the modular evolution scheme. *FEBS Lett* 549:7
- Mathiesen C, Hägerhäll C (2002) Transmembrane topology of the NuoL, M and N subunits of NADH:quinone oxidoreductase and their homologues among membrane-bound hydrogenases and bona fide antiporters. *Biochim Biophys Acta* 1556:121
- Moparthy VK, Kumar B, Mathiesen C, Hägerhäll C (2011) Homologous protein subunits from *Escherichia coli* NADH:quinone oxidoreductase can functionally replace MrpA and MrpD in *Bacillus subtilis*. *Biochim Biophys Acta* 1807:427
- Munch R, Hiller K, Barg H, Heldt D, Linz S, Wingender E, Jahn D (2003) PRODORIC: prokaryotic database of gene regulation. *Nucleic Acids Res* 31:266
- Ohyama K, Fukuzawa H, Kohchi T, Shirai H, Sano T, Sano S, Umesono K, Shiki Y, Takeuchi M, Chang Z, Aota S, Inokuchi H, Ozeki H (1986) Chloroplast gene organization deduced from complete sequence of liverwort *Marchantia polymorpha* chloroplast DNA. *Nature* 322:572
- Peng L, Yamamoto H, Shikanai T (2011) Structure and biogenesis of the chloroplast NAD(P)H dehydrogenase complex. *Biochim Biophys Acta Bioenerg*. doi:10.1016/j.bbabi.2010.10.015
- Peterson JD, Umayam LA, Dickinson T, Hickey EK, White O (2001) The comprehensive microbial resource. *Nucleic Acids Res* 29:123
- Rothery RA, Workun GJ, Weiner JH (2008) The prokaryotic complex iron-sulfur molybdoenzyme family. *Biochim Biophys Acta* 1778:1897
- Saitou N, Nei M (1987) The neighbor-joining method—a new method for reconstructing phylogenetic trees. *Mol Biol Evol* 4:406
- Sayers EW, Barrett T, Benson DA, Bryant SH, Canese K, Chetvernin V, Church DM, DiCuccio M, Edgar R, Federhen S, Feolo M, Geer LY, Helmberg W, Kapustin Y, Landsman D, Lipman DJ, Madden TL, Maglott DR, Miller V, Mizrahi I, Ostell J, Pruitt KD, Schuler GD, Sequeira E, Sherry ST, Shumway M, Sirotkin K, Souvorov A, Starchenko G, Tatusova TA, Wagner L, Yaschenko E, Ye J (2009) Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res* 37:D5
- Sazanov LA, Hinchliffe P (2006) Structure of the hydrophilic domain of respiratory complex I from *Thermus thermophilus*. *Science* 311:1430
- Sazanov LA, Burrows PA, Nixon PJ (1998) The plastid ndh genes code for an NADH-specific dehydrogenase: isolation of a complex I analogue from pea thylakoid membranes. *Proc Natl Acad Sci* 95:1319
- Sierro N, Makita Y, de Hoon M, Nakai K (2008) DBTBS: a database of transcriptional regulation in *Bacillus subtilis* containing upstream intergenic conservation information. *Nucleic Acids Res* 36:D93
- Suorsa M, Sirpio S, Aro EM (2009) Towards characterization of the chloroplast NAD(P)H dehydrogenase complex. *Mol Plant* 2:1127
- Swartz TH, Ikewada S, Ishikawa O, Ito M, Krulwich TA (2005) The Mrp system: a giant among monovalent cation/proton antiporters? *Extremophiles* 9:345
- Tamura K, Dudley J, Nei M, Kumar S (2007) MEGA4: molecular evolutionary genetics analysis (MEGA) software version 4.0. *Mol Biol Evol* 24:1596
- Thompson JD, Higgins DG, Gibson TJ (1994) Clustal-W—improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* 22:4673
- Tocilescu MA, Zickermann V, Zwicker K, Brandt U (2010) Quinone binding and reduction by respiratory complex I. *Biochim Biophys Acta* 1797:1883
- Vignais PM, Billoud B (2007) Occurrence, classification, and biological function of hydrogenases: an overview. *Chem Rev* 107:4206
- Vignais PM, Billoud B, Meyer J (2001) Classification and phylogeny of hydrogenases. *FEMS Microbiol Rev* 25:455
- Weerakoon DR, Olson JW (2008) The *Campylobacter jejuni* NADH:ubiquinone oxidoreductase (complex I) utilizes flavodoxin rather than NADH. *J Bacteriol* 190:915
- Whelan S, Goldman N (2001) A general empirical model of protein evolution derived from multiple protein families using a maximum-likelihood approach. *Mol Biol Evol* 18:691
- Yano T, Yagi T, Sled VD, Ohnishi T (1995) Expression and characterization of the 66-kilodalton (Nqo3) iron-sulfur subunit of the proton-translocating NADH-quinone oxidoreductase of *Paracoccus denitrificans*. *J Biol Chem* 270:18264
- Yano T, Sled VD, Ohnishi T, Yagi T (1996) Expression and characterization of the flavoprotein subcomplex composed of 50-kDa (NQO1) and 25-kDa (NQO2) subunits of the proton-translocating NADH-quinone oxidoreductase of *Paracoccus denitrificans*. *J Biol Chem* 271:5907
- Yip CY, Harbour ME, Jayawardena K, Fearnley IM, Sazanov LA (2011) Evolution of respiratory complex I “supernumerary” subunits are present in the alpha-proteobacterial enzyme. *J Biol Chem* 286:5023
- Zuckerkandl E, Pauling L (1965) Evolutionary divergence and convergence in proteins. In: Bryson V, Vogel HJ (eds) *Evolving genes and proteins*. Academic Press, New York, pp 97–166