Hypothesis

The 'antiporter module' of respiratory chain Complex I includes the MrpC/NuoK subunit – a revision of the modular evolution scheme

Cecilie Mathiesen, Cecilia Hägerhäll*

Department of Biochemistry, Lund University, P.O. Box 124, 22100 Lund, Sweden

Received 6 May 2003; revised 27 June 2003; accepted 28 June 2003

First published online 16 July 2003

Edited by Peter Brzezinski

Abstract Respiratory chain Complex I or NADH:quinone oxidoreductase catalyzes oxidation of NADH in the mitochondrial matrix or bacterial cytoplasm and reduction of quinone in the membrane, coupled to pumping of $4H^+/2e^-$ across the membrane. The same enzyme complex is also capable of the reverse reaction, i.e. $\Delta\mu_{H^+}$-supported NAD^+ reduction. The molecular mechanism that couples electron transfer to proton pumping is not understood. The Complex I enzyme, containing 14 protein subunits necessary for function, has evolved from smaller functional building blocks. Three Complex I protein subunits, NuoL, NuoM and NuoN, show primary sequence similarity to one particular class of antiporters, and are thus predicted to play a role in the proton translocation machinery. These antiporters, MrpA and MrpD are encoded by a conserved gene cluster, that contains seven genes. In previous work we have determined that these antiporters come in two subclasses, MrpA-type and MrpD-type, and that the Complex I subunit NuoL is more closely related to MrpA and NuoM and N are more closely related to the MrpD antiporter. This implied that both MrpA and MrpD had been recruited to Complex I, rather than arising from gene duplications of one antiporter encoding gene. In this work we show that MrpC and NuoK are homologous proteins. The most plausible explanation for these findings is that a multisubunit antiporter complex was recruited to the ancestral enzyme. We further conclude that the last common ancestor of the Complex I enzyme family and membrane bound NiFe hydrogenases of type 3 and 4 contained the NuoKLMN subunit module.

Key words: NADH:quinone oxidoreductase; Antiporter; NQO11; ND4L; NAD4L; NdhE

1. Introduction

NADH:quinone oxidoreductase, NADH dehydrogenase type I or Complex I is the largest and most complex of the respiratory chain enzymes. The enzyme catalyzes oxidation of NADH to NAD^+ and reduction of quinone to quinol, coupled to H^+ pumping across the membrane, and it is also capable of the reverse reaction, i.e. $\Delta\mu_{H^+}$-supported NAD^+ reduction. Only low resolution (about 20 Å) structural information is available for this enzyme and the mechanism of coupling of electron and proton transfer is unknown. For reviews see Biochim. Biophys Acta 1364, 85–296 or J. Bioenerg. Biomembr. 33, 155–266, special issues on Complex I. Bacteria contain the 'minimal functional unit' of Complex I generally consisting of 14 different protein subunits, denoted NuoA–N. Seven subunits are located in the promontory part of the enzyme complex, facing the bacterial cytoplasm, and harbor FMN and iron–sulfur cluster prosthetic groups. The remaining seven hydrophobic polypeptides (NuoA, H, J, K, L, M and N) form the membrane spanning part of the enzyme and do not contain any identified cofactor, but must have important roles in the quinone-binding and proton translocation reactions [1]. Mammalian Complex I contains 32 additional protein subunits with unknown function [2].

Primary sequence comparisons have shed light on the evolution of the large Complex I enzyme from smaller functional building blocks [3–7]. Since the smaller entities must have been recruited to Complex I based on their function, a better understanding of such evolutionary relationships is also helpful when trying to elucidate the functional mechanism of present day Complex I. Since 1994 it is well established that three Complex I subunits, NuoL, NuoM and NuoN, are homologous to one particular class of antiporters [8] and thus are likely to contain H^+ channels and probably play a role in the proton translocation machinery. These antiporter-like protein subunits are also found in Complex I-like enzymes such as chloroplast NADPH dehydrogenase and archaeal $F_{420}H_2$ oxidoreductase and in the complex membrane bound NiFe hydrogenases (hydrogenase-3 and hydrogenase-4) [4,5,7,9]. The functional relatedness of these subunits to antiporters may be particularly interesting since there is a small but growing set of studies that indicate that Na^+ or K^+ may be directly involved in the catalytic mechanism of Complex I [10–12] and possibly also in the complex, membrane bound NiFe hydrogenase [13], but it is presently not known if this is a general feature of all Complex I enzymes.

Unfortunately, not much is known about this particular class of antiporters. It was first discovered in the alkalophile *Bacillus halodurans* C-125, where the first member of the family was identified in a mutant unable to grow at alkaline pH

*Corresponding author. Fax: (46)-46-222 4534.
E-mail address: cecilia.hagerhall@biokem.lu.se (C. Hägerhäll).

Abbreviations: NAD(P)H, nicotinamide adenine dinucleotide (phosphate) – reduced form; Mrp, multiple resistance and pH (*mrp*) locus; PSI-BLAST, position specific iterated basic logical alignment search tool; PSSM, position specific scoring matrix; aa, amino acids; MCT, monocarboxylate transporters

0014-5793/03/$22.00 © 2003 Published by Elsevier B.V. on behalf of the Federation of European Biochemical Societies.
doi:10.1016/S0014-5793(03)00767-1

[8]. The targeted protein was found to be responsible for maintaining a neutral cytoplasmic pH at alkaline growth conditions. This type of Na^+ or K^+/H^+ antiporter has since been found in many other alkalophile and mesophile bacteria [14,15], where it appears in a conserved gene context or operon including seven open reading frames (Fig. 1, Table 1). In *Bacillus subtilis* the corresponding proteins are denoted Mrp for multiple resistance and pH (*mrp*) locus, and are labeled MrpA–G. In other organisms another nomenclature is used, such as Sha, Pha and Mnh. Both MrpA and MrpD have been shown to have a role in Na^+ resistance and Na^+ dependent pH homeostasis in *B. subtilis* [16–18]. A *B. halodurans* mutant carrying a G82 to E mutation in MrpC was alkalosensitive, but in contrast to the first described MrpA mutant, maintained wild-type levels of internal pH [19]. Taken together, studies of mutants and deletion strains suggest that monovalent cation/H^+ antiporter activity depends on the MrpA and MrpD subunits, but that the remaining proteins are required for some combination of antiporter activity, expression and/or assembly [14,16,17]. MrpF can, independently of the other Mrp subunits, function as a transporter for cholate and Na^+ efflux [18]. It remains to be established whether the other proteins encoded by the *mrp* gene cluster form a multisubunit transport complex or if they can function individually in monomeric or multimeric form.

In previous work we noticed that the MrpA antiporter seemed to be more closely related to the Complex I subunit NuoL and the MrpD antiporter more closely related to NuoM and N, implying that both MrpA and MrpD had been recruited to Complex I, rather than arising from gene duplications of one antiporter encoding gene [9]. In this work we use PSI-BLAST (position specific iterated basic logical alignment search tool), a bioinformatic research tool particularly well suited to identify weak, but significant homologies in proteins and protein families, to further investigate the relationship between the Mrp antiporter assembly and respiratory chain Complex I. The results demonstrate that MrpC and NuoK are homologous proteins, leading to a revision of the Complex I evolutionary scenario. We conclude that a multisubunit antiporter module consisting of the MrpA, C and D proteins was present in the last common ancestor of Complex I and Complex I-like enzymes. It seems more plausible that the antiporter module was recruited in one step to the ancestral Complex I, rather than in three successive events. Thus, this implies that the Mrp proteins most probably do form a multisubunit transport complex.

2. Materials and methods

2.1. PSI-BLAST

The PSI-BLAST procedure was generally started with the relevant Mrp sequence from *B. subtilis* 168A using PSI-BLAST with default settings. PSI-BLAST iteration2 is run from the results page (saved as a HTML-file). Returning to the 'Formating' page the results were displayed as NCBI-gi: position specific scoring matrix (PSSM). The inclusion threshold format for PSI-BLAST was set to an expect-(*E*)-value that only included protein sequences that for certain were of the same type as the initial protein, and thus a low *E*-value limit was used. After formating again the resulting matrix was saved as a text file. Then a new PSI-BLAST search was done using this matrix in the PSSM window. The 'expect', 'descriptions', 'alignments' and 'Inclusion threshold' values were now set higher to get a more complete picture of possible homologues. Again, the results page was saved as html, and the procedure was repeated running 'iteration2' as before. The search was repeated until a definite, convincing result was obtained; typically two to four times. All searches were filtered for hypothetical, unknown or putative protein sequences.

2.2. ClustalW alignments, hydrophobicity plots and helical wheel analysis

ClustalW alignments [42] were done using the online resource at http://bioweb.pasteur.fr/seqanal/interfaces/clustalw.html with Phylip alignment output format and default settings as reported in [9]. The resulting infiles were saved and opened in BioEdit [43]. A consensus sequence was obtained at a threshold frequency of 60% (marked with black letters in Fig. 2B). The alignments were adjusted manually and compared to the patterns found by the PSI-BLAST search. In addition to the primary sequences shown in Fig. 2A, the following sequences were used for the detailed analyses of MrpC and NuoK: MrpC proteins from: *B. halodurans* (NP_242183), *Oceanobacillus iheyensis* (NP_694029), *Corynebacterium glutamicum* ATCC13032 (NP_601925), *Rhodospirillum rubrum* (ZP_00015087), *Sinorhizobium meliloti* (X93358), *Vibrio cholerae* (NP_232556), *Xanthomonas campestris* pv. campestris str ATCC33913 (NP_635839), *Desulfitobacterium hafniense* Desu_p_628 (ZP_00097535), NuoK proteins from: *Rhodobacter capsulatus* B10 (AAC25002), *Rickettsia prowazekii* MadridE (NP_221141), *Buchnera aphidicola* (BU163), *Helicobacter pylori* J99 (NP_223909), *Mycobacterium tuberculosis* H37Rv (NP_217671), *Synechococcus* sp. PCC7002 (AF381034), *Hordeum vulgare* (trO98694), *Nicotina tabacum* (spP06261), *Scenedesmus obliquus* (NP_057977), *Neurospora crassa* (spP05509), *Sacrophyton glaucum* (O63851), *Ostrinia nibulalis* (NP_563592), *Acanthamoeba castellanii* (spQ37379), *Amoebidium parasiticum* (AAN04070), *Reclinomonas americana* (AF007261), *Methanosarcina acetivorans* str.C2 (AAAM04917), *Methanocarcina barkeri* (ZP_00077678), C-terminal of HyfE proteins: *R. rubrum* (ZP_00014217) and *Thermoanaerobacter tengcongensis* MBT4 (NP_623300). The Kyte–Doolittle scale mean hydrophobicity profile [33] of the aligned sequences (with scan-window size 13) was plotted with BioEdit. Helical wheel projections contained 18 amino acids and were plotted using Antheprot [36] and were modified using Paint and Coral Draw.

Table 1
mrp gene clusters in archaea and eubacteria

Class	Order	Representative organism	Accession numbers
Methanococci[a]	Methanosarcinales	*M. acetivorans*	NP_619431-25
Bacilli	Bacillales	*B. subtilis*	NP_391038-44
Actinobacterea[b]	Actinobacteridae	*S. coelicolor*	NP_631020-25
α-Proteobacteria	Rhizobiales	*S. meliloti*	CAC45564-70
			CAA63734-40[c]
α-Proteobacteria[b]	Rhizobiales	*B. melitensis*	NP_541748-43
α-Proteobacteria[b]	Rhodobacterales	*R. capsulatus*	AAC16120-25
γ-Proteobacteria[b]	Xanthomonadales	*X. axonopondis*	NP_640817-12
Deinococci	Deinococcales	*D. radiodurans*	NP_294604-10

A similar fused subunit was found in *B. japonicum* (NP_773958-59) but outside the *mrp* gene cluster context.
[a]Contain a gene cluster with AB fusions.
[b]Contain a gene cluster with AB fusions.
[c]The MrpB subunit includes the C-terminal end of MrpA and the 'normal' MrpB fused into one polypeptide.

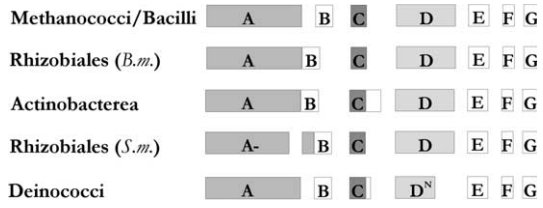

Fig. 1. Different *mrp* gene cluster organizations in archaea and bacteria. *B.m.* stands for *Brucella melitensis*, and *S.m.* for *S. meliloti*. The C-terminal extension of MrpC in *S. coelicolor* (Actinobacteria) and Deinococci is not homologous to the N-terminal extension of HyfE (see also Fig. 2A). The order Rhizobiales includes different *mrp* gene cluster organizations. *S. meliloti* contains the version where the C-terminal part of MrpA and MrpB has formed one separate polypeptide unit, corresponding to the protein that is no longer a part of NuoKLMN (see also Table 1).

3. Results and discussion

3.1. *The* mrp *gene cluster*

Since proteins homologous to MrpA and MrpD antiporters are found in several different enzyme complexes, and also in some with unknown function but with no relation to Complex I [9] we have chosen to regard only those gene products that result from the conserved *mrp* gene cluster as confirmed, bona fide antiporters. As illustrated in Fig. 1, the *mrp* gene clusters come in several slightly different variants, to a certain extent related to different bacterial orders (see also Table 1). The physiological role of the different Mrp proteins, as far as we know them today, was summarized in Section 1. Except for MrpA and MrpD, similarity to Complex I subunits has previously been reported for MrpB and MrpE, but in that study no further specification about the degree and nature of the detected homology was given [18]. A similarity of *B. halodurans* ORF3 (i.e. MrpC) to NuoK over a stretch of 22 amino acids was mentioned in [20]. However, a standard BLAST search [21–23], even with high set 'expect', 'descriptions' and 'alignments' values, is not able to identify any Complex I homologues among the Mrp proteins other than MrpA and MrpD.

3.2. *Using PSI-BLAST to identify homologous proteins*

Commonly used bioinformatic search methods such as BLAST have difficulty both to differentiate between highly similar protein subunits in different enzyme complexes and to identify distantly related proteins. PSI-BLAST [24] uses the power of several homologous sequences building up PSSMs and then uses this matrix as a basis for searching. The first round of second iteration should only include members of the same protein, for example MrpA, but in subsequent rounds the *E*-value limit can be extended such that other homologous proteins, for example NuoL, can be included in the search matrix. In this way, new previously unknown members of an extended protein family can be identified. However, the procedure must be carried out with great awareness of the danger of matrix migration [25], caused by non-relevant homologues being included in the matrix. A complete alignment of members of the proposed homologous protein families should always be made, to assess the significance of the hits obtained with the matrix search. A detailed description of the PSI-BLAST search procedure is given in Section 2.

As mentioned, MrpB has earlier been reported to be a Complex I homologue [18]. A first PSI-BLAST search with MrpB was carried out where the PSSM1 included *E*-limit 1.0, containing only MrpB and the MrpB part in AB fusions (see Fig. 1). A search with this matrix gave hits in EchA (a hydrogenase-3 protein similar to MrpA/NuoL) at amino acids 190–310 already at *E* 9e-8. In the subsequent PSSM2 all sequences with an *E*-value lower than 1.0 were included. A search with PSSM2 finds all MrpB proteins (including the MrpB part in A+B fusions, see Fig. 1), corresponding to 22% of the total number of hits. 12% of these 22% are hits to the N-terminal part of MprAB fusion proteins. 78% of subsequent hits are to the N-terminal part of the NdhF subunit (amino acids 150–290). NdhF is an NuoL homologue in the Complex I-like chloroplast enzyme that comes in a longer (approximately 700 amino acids) and a shorter version (about 420 amino acids) with a truncated C-terminus. The statistics include the first 100 hits. However, when the results were further investigated by ClustalW alignments of NuoL subunits and MrpA subunits versus MrpB subunits (but no MrpAB fusions), no significant conserved homology could be detected. This result is thus an example of matrix migration. The PSI-BLAST searches had been 'contaminated' by weak homology between MrpB and some single individual NuoL and MrpA subunits, a homology that is not representative for the whole protein family. We conclude that the previously reported homology between MrpB and NuoL is probably due to database hits in the relatively abundant various MrpAB fusion proteins (Fig. 1), that in databases often are labeled 'ND5', 'subunit 5' or 'NuoL'. This error is understandable, since the main part of the MrpA subunit is unmistakably an NuoL homologue.

3.3. *Mrp protein homologues correctly identified by PSI-BLAST*

In a new round of MrpB searches with PSSM1 limited at *E*-value 0.014 instead of 1.0, 27% of 100 hits were MrpB proteins or MrpAB fusion proteins. We also obtained one match to a protein of unknown function, consisting of 314 amino acids, i.e. larger than MrpB, but highly homologous (*E*-value 6e-4). A closer analysis of this *Bradyrhizobium japonicum* protein, denoted blr7318, using a standard BLAST search showed that the whole protein is similar to the C-terminal part of MrpA (the part lacking in NuoL subunits) and MrpB. Another such fusion protein consisting of the C-terminal part of MrpA and MrpB is found in the antiporter gene cluster context of some Rhizobiales (see Fig. 1). A neighboring open reading frame to blr7318 in *B. japonicum*, bsr7317 (NP_773957, 97 amino acids), showed about 60% sequence identity to MrpG over the first 50 N-terminal amino acids. No other homologues could be found in the vicinity of blr7318. This indicates the existence of a novel membrane bound protein or complex of unknown function, related to the *mrp* antiporter subunits but unrelated to Complex I. A majority of the hits using the MrpB matrix (48%) were to δ-endotoxins [26]. MrpB is homologous to the membrane spanning part of this three-domain toxic protein (see also Table 2). In summary, a 'clean' MrpB matrix search showed no sequence similarity to any Complex I protein, but revealed a homologous region in MrpB and δ-endotoxins that could be of structural and functional relevance. We conclude that there is no evidence for relatedness between NuoL and MrpB.

Several rounds of PSI-BLAST searches were also done with

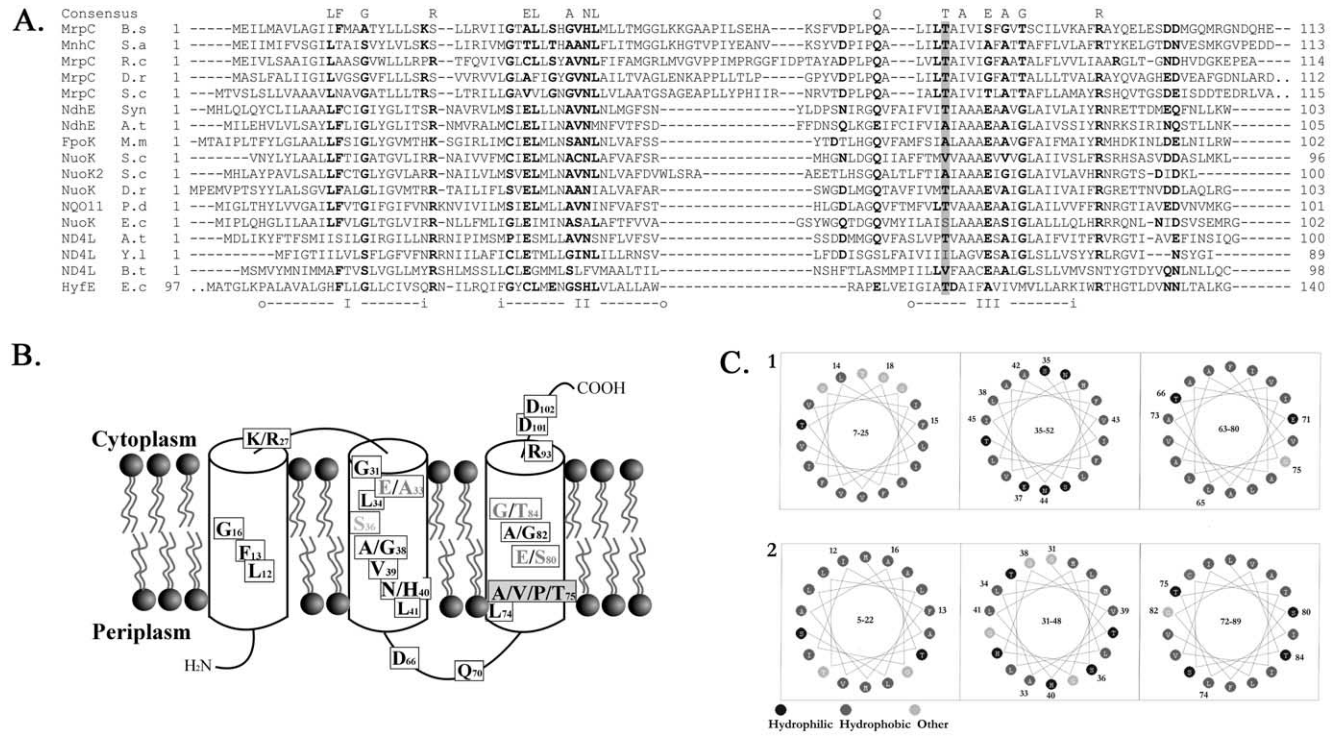

Fig. 2. A: Alignment of MrpC and NuoK from selected species, also including the HyfE protein from *E. coli*. A consensus sequence for the MrpC/NuoK protein family is shown at the top of the alignment, and is also plotted in B. The position of the primary LHON mutation 10663C discussed in the text is indicated with gray background. Transmembrane helices, 18 amino acids long, are marked with roman letters I–III and correspond to the residues analyzed in C. The sequences included in the alignment are MrpC proteins from bacteria and archaea: *B. subtilis* 168A BG12357, *S. aureus* BAA35097, *R. capsulatus* AAC16121, *D. radiodurans* NP_294606, *S. coelicolor* A3(2) NP_631021, NuoK from Complex I-like enzymes in synechosystis and chloroplasts: *Synechocystis* PCC 6803 P26524, *Arabidopsis thaliana* P26289, NuoK from various subgroups of bacteria: *S. coelicolor* A3(2) SCO4572 and SCO4605, *D. radiodurans* NP_295218, *P. denitrificans* P29923, *E. coli* P33606, and NuoK (ND4L) from various mitochondrial sources: *A. thaliana* Q04614, *Yarrowia lipolytica* CAC28106 and *Bos taurus* AF493542. HyfE is from *Escherichia coli* P77524 and is only found in membrane bound hydrogenases containing three antiporter-like subunits. B: A topology model for MrpC/NuoK showing residues conserved in both proteins in black font. Other positions indicated have conserved, similar but not identical residues. These are small amino acids important for helix packing (A, S, G, T, C, P), residues with positive charge (K, R), polarity, and thus capability to form one or two H-bonds (D, E, N and Q), or hydrophobic residues (A, I, L, V, M, F). If the position is conserved in both MrpC and NuoK the letter is shown in gray, but if the position is conserved in only one of the subfamilies the residue is shown in light gray. Residue numbering is from *B. subtilis* MrpC (Ac# BG12357). The NuoK subunit from *P. denitrificans* Complex I (NQO11, Ac# P29923), for which the topology has been experimentally determined, was used as a template for the model. C: Helical wheel projections of the central transmembrane domains in NQO11/NuoK (1) and MrpC (2) from *P. denitrificans* and *B. subtilis* respectively. Black background indicates hydrophilic, gray hydrophobic and light gray all other amino acid residues. Amino acid numbers are shown for conserved residues, also indicated in B.

the MrpE, MrpF and MrpG subunits, but no convincing homology between these proteins and subunits of Complex I could be detected (data not shown). The MrpE matrix did in a few cases hit a NuoJ subunit, but subsequent ClustalW alignment analysis did not support a significant homology between these two proteins. MrpF is, as mentioned earlier, reported to function independently of the other Mrp proteins, catalyzing cholate and Na^+ efflux in bacteria. Not surprisingly, the MrpF matrix hits Na^+-taurocholate cotransporters, a bile salt transporter. MrpF matrix searches subsequently hit voltage-gated Na^+ channels in various domains, explained by the fact that these channels contain four homologous domains [27]. Voltage-gated Na^+ channels are a member of a superfamily that also contain the voltage-gated K^+- and Ca^{2+} channels. The detected homology included the whole MrpF sequence and the center of each of the four Na^+ channel domains. The homologous region is rich in positive charges and corresponds roughly to the 'primary voltage sensor' in transmembrane helix S4 [28] (see also Table 2). MrpG was found to be homologous to the N-terminal part of Na^+K^+ transporting ATPases [29]. The homologous region is located around transmembrane helix one or two (see Table 2). Compared to the Ca^{2+} transporting ATPase, for which high resolution structural information is available, the Na^+K^+ transporting ATPase has an N-terminal extension of around 90 amino acids before the conserved TGES motif. The ATP-binding motif DKTG is located in the first part of a loop between transmembrane helix four and five and is thus not included in the MrpG homologous domain. Taken together, the homologies detected indicate that MrpB, MrpF and MrpG are transporter proteins, but their exact roles in the assembly of Mrp proteins remain to be determined.

For MrpC we did find a significant sequence similarity to a Complex I protein, the NuoK subunit. A search with PSSM3 against GenBank resulted in 106 hits, when including all sequences with *E*-values lower than 100. The first 11.3% were all MrpC proteins and the following 48.1% were NuoK proteins, from all the different kingdoms of life. HyfE, a membrane spanning subunit of hydrogenase-4, is also found by the matrix, but relatively late in the hit list (see also Fig. 2). The

relatively high content (40.6%) of non-relevant hits is due to the choice of method, where the matrix formation is based solely on an *E*-value limit. We could not substantiate the earlier claim of homology between MrpC and Na^+-coupled organic acid transporters [18]. The MrpC matrix only found two matches to monocarboxylate transporters (MCT) but the matching sequence was unrelated to the consensus sequence of the MCT family [30]. Likewise, a PSI-BLAST search using NuoK as input sequence (PSSM1 including NuoK with *E*-value below 10) found MrpC proteins. However, to prevent overflow of mitochondrially encoded ND4L protein sequences that are abundant in the databases, the search had to be limited to bacterial sequences.

3.4. A detailed comparative analysis of MrpC and NuoK

NuoK is the smallest of the membrane spanning Complex I subunits, consisting of approximately 100 amino acids, and has a size around 10 kDa. Other names for NuoK are NQO11, ND4L, NAD4L and NdhE. NuoK is certainly not the most strongly conserved Complex I subunit among different species. The bovine and the *R. capsulatus* ND4L/NuoK subunits show only 27% sequence similarity, compared to 90% similarity for the most highly conserved PSST/NuoB subunit. The other two small hydrophobic subunits, ND3/NuoA and ND6/NuoJ, show 60% and 18% similarity respectively [31]. After homologous NuoK and MrpC polypeptides were identified by PSI-BLAST the resulting sequence set was further investigated by alignments using ClustalW and by comparing hydrophobicity plots. A consensus sequence was obtained at a threshold frequency of 60% for inclusion. The alignments were then adjusted manually and compared to the conserved patterns found by the PSI-BLAST search. A smaller but representative alignment set, including MrpC and NuoK proteins from varying species and kingdoms of life respectively, is shown in Fig. 2A. The transmembrane topology of NuoK from *Paracoccus denitrificans* has been determined, revealing three transmembrane segments with the C-terminus facing the cytoplasm [32]. We subsequently compared the Kyte–Doolittle hydrophobicity profiles [33] of the set of NuoK and MrpC sequences. The analyses revealed highly similar profiles for NuoK and MrpC, further strengthening that the two proteins are indeed homologous. A sequence set of another Complex I transmembrane protein, NuoA, that is of similar size and that contains the same number of predicted transmembrane segments [34], was included in the analyses for comparison. The NuoA sequence hydropathy profiles differed significantly from those of MrpC and NuoK (data not shown). The conserved positive charges in NuoK and MrpC also fit well to the positive inside rule [35] and supports a common membrane orientation for MrpC and NuoK. A 2D model of the polypeptide, with positions of highly conserved amino acid residues indicated, is shown in Fig. 2B. Helical wheel plots of the putative transmembrane segments of NuoK and MrpC, containing 18 amino acids each, were done using Antheprot [36] and modified in Paint and Coral Draw (Fig. 2C). Conserved residues are indicated by position number in the plots. Experiments with α-helix transmembrane peptides show that amino acids capable of forming double hydrogen bonds (E, N, Q, D) are important for helix oligomerization [37]. Certain patterns of serine (S) or threonine (T) in membrane spanning peptides can also contribute to such multimer formation [38]. The highly conserved glutamates (E) and aspartate (D) in NuoK from Complex I and HycE from membrane bound hydrogenases might be of structural or functional importance. In some NuoK and MrpC polypeptides the conserved glutamates are substituted with serines [32] (Fig. 2A). The role of the conserved glutamates has been studied in *E. coli*, where a strain genetically deleted for NuoK was complemented by

Table 2
Homologues to the other Mrp proteins

Subunit	Previously reported [18]	Homologues found using PSI-BLAST	Homologous region (identified in this work)		Functional relevance	References
			in Mrp	in homologue		
MrpB	NuoL homologue[a]	δ-Endotoxins (~590 aa)	Typically aa 19–63 in a 140 aa long subunit I(4×)S(3×)FL(5×)PG×GFV×GL(13×)D(3×)V	50 first aa in membrane spanning, toxic part (Domain 1, aa 1–265)	Involved in channel formation in the membrane	Pdb:1CIY [26]
MrpC	Na^+-coupled organic acid transporter homologue[b]	NuoK	The whole peptide	The whole peptide	Sensor? Similar to NuoK	[20] [31]
MrpE	Complex I homologue[c]				Unknown	
MrpF	Cholate transporter homologue	Na^+-taurocholate cotransporter	Amino acids 1–57	Central region, aa 55–11 of 187 aa	Unknown	Ac# AAD53961
		Voltage-gated Na^+ channels *similar to* Voltage-gated K^+ channels	Most of the subunit (~90 aa)[d]	aa 214–279 in this Ac#. In the middle of each domain in voltage-gated Na^+ channels	Voltage sensor	[27] Ac# AF188679 [28]
MrpG		Na,K-ATPase	Typically aa 10–65 in a 120 aa long subunit[d]	Transmembrane helix 1–2[e]	Unknown	Pdb:1EUL [29]

[a]In this work found to be the MrpB part of a MrpAB fusion (the fusion protein is often denoted NuoL/ND5).
[b]The MrpC matrix used in this work only found two polypeptides identified as MCTs.
[c]The MrpE matrix did hit a few NuoJ/ND6 subunits, but further analysis with alignments and hydrophobicity plots did not support a true homology.
[d]Predicted to have three transmembrane helices.
[e]Not in the ATP-binding motif DKTG and before conserved TGES motif.

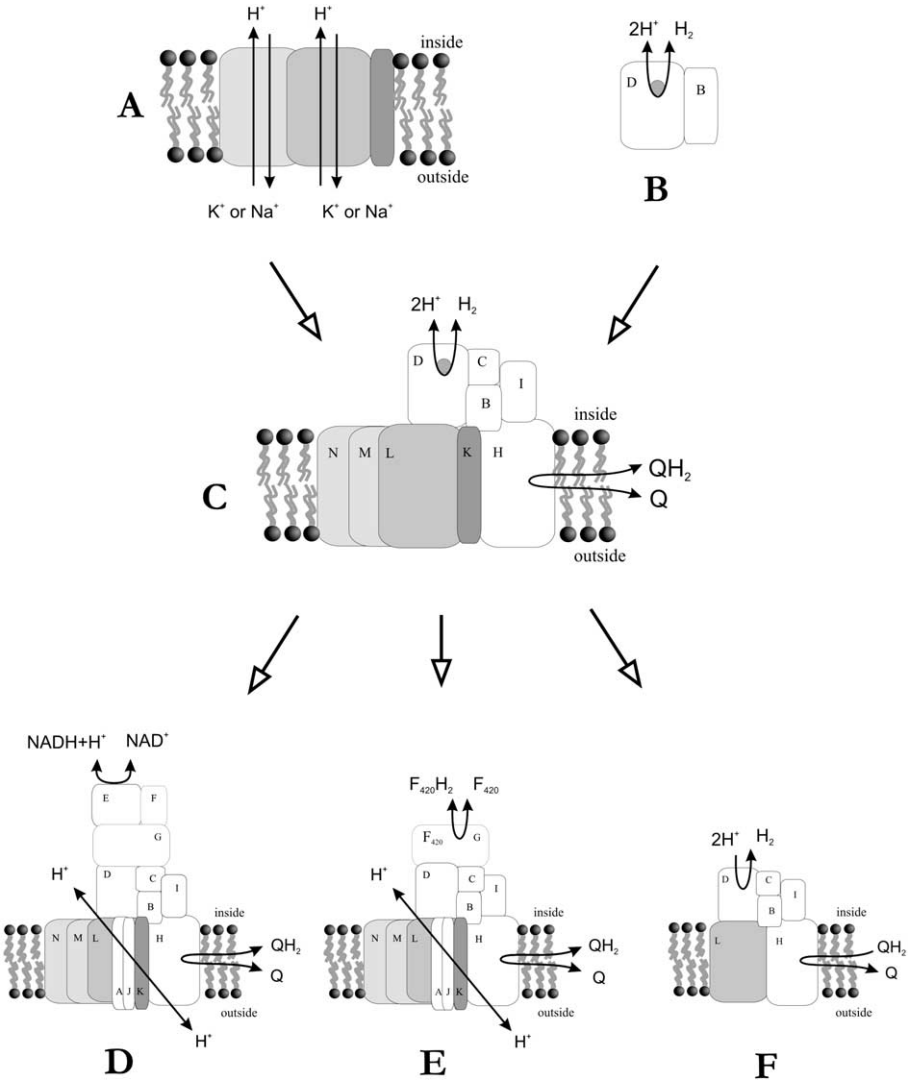

Fig. 3. A revised evolutionary scheme for Complex I. The three antiporter subunits MrpA, MrpC and MrpD (A) together with a soluble Ni/Fe hydrogenase (B) have formed an ancestral membrane bound hydrogenase (C). In this enzyme the B subunit is truncated and the C subunit added (both BC and CD fusion proteins can be found in present day enzymes from the Complex I family). The I and H subunits probably constitute a membrane bound electron donor. The ancestral enzyme depicted as (C) is the last common ancestor of Complex I (D), other Complex I-like enzymes such as F_{420}-dehydrogenase (E) and the present day membrane bound Ni–Fe hydrogenases, that in many cases have lost both the MrpC homologue and two of the MrpA/D homologues (F).

wild-type NuoK and NuoK with point mutations at the positions of the conserved glutamates [39]. This study indicated that the glutamates in NuoK were essential for the function of Complex I, but that they could be replaced by aspartate without loss of function. A primary mutation in mitochondrial DNA causing the maternally inherited disease Leber's hereditary optic neuropathy, or LHON [40], was found in NuoK (ND4L). The mutation, denoted 10663C (for nucleotide position), results in a valine (V) to alanine (A) substitution at a position following a highly conserved hydrophobic residue, see also Fig. 2A.

3.5. *Evolution of Complex I from smaller functional building blocks – a revision of the scheme*

As mentioned in Section 1, previous phylogenetic analyses demonstrated the MrpA antiporter group closer to NuoL and the MrpD antiporter closer to NuoM and N in the overall conserved protein family, implying that both MrpA and MrpD have been recruited to Complex I, rather than arising from gene duplications of one antiporter encoding gene [9]. Still, we could not rule out that the sequence similarity resulted from convergent evolution, caused by a functional relatedness of MrpA/NuoL and MrpD/NuoM. However, that MrpC and NuoK were found to be homologous proteins further strengthens the notion that a multisubunit 'antiporter module' was recruited to the ancestral Complex I rather than a single protein (Fig. 3). In this respect, the identification of a separate subunit that contains a fusion of the C-terminal end of MrpA (that is not present in NuoL) and MrpB in some *mrp* gene clusters (Fig. 1, Table 1) is particularly interesting. The loss of such a subunit from the 'antiporter module' recruited to Complex I would explain the appearance of the present day NuoKLMN domain of Complex I.

The earlier envisioned modular evolution scenario for Complex I is summarized in [6,7]. In this scheme the NuoB and D subunits, a common ancestor of soluble NiFe hydrogenases,

combined with NuoC, H, I and L to form a membrane bound hydrogenase enzyme. This enzyme is envisioned as the common ancestor of Complex I and Complex I-like enzymes and of hydrogenase-3 and hydrogenase-4. Then, after triplication of NuoL, and recruitment of NuoA, J and K the common ancestor of the Complex I enzyme family is formed. In the light of the new data presented here and in [6], this scheme must be revised (see Fig. 3). The last common ancestor of Complex I, Complex I-like enzymes and membrane bound hydrogenases contained the NuoK, L and M representing the MrpA, C and D antiporter module, the NuoB and D representing the NiFe hydrogenase module and NuoH and I probably comprising a membrane bound electron donor unit. Gene duplication of the NuoM encoding gene resulted in NuoN. From this last common ancestor, the Complex I enzyme family members arose from recruiting different electron input devices; the NADH dehydrogenase module NuoEFG as in Complex I, the F_{420} dehydrogenase module in the archaeal enzymes, and a yet unidentified module in the chloroplast enzyme [41]. The present day membrane bound hydrogenases seem to represent 'degenerate' forms of the last common ancestor depicted in Fig. 3. Only the enzymes containing three MprA/D-like polypeptides contain a subunit homologous to MrpC (Fig. 2A), whereas this polypeptide is absent from hydrogenases containing one or two MrpA/D-like polypeptides. Furthermore, the primary sequence of MrpA/D-like hydrogenase polypeptides, whether they come from enzymes containing one, two or three of the homologous polypeptides, do not fall into the MrpA or MrpD branches, but are found scattered all over the phylogenetic tree [9]. This may reflect a lesser need for specific functions associated with the H^+ translocating units in these enzymes. It is not possible to conclude from the present data whether the NuoA and NuoJ subunits were part of the last common ancestor, and were then lost from the membrane bound hydrogenases, just like MrpC was lost from many of them, or if these subunits were recruited to a later ancestor, common to the Complex I enzyme family only.

Acknowledgements: We thank Dr. Simon Gough for advice on using PSI-BLAST. This work was supported by a grant from the Crafoord Foundation. C.M. also acknowledges a pre-doctoral scholarship from The Sven and Lilly Lawski foundation.

References

[1] Dupuis, A., Darrouzet, E., Duborjal, H., Pierrard, B., Chevallet, M., van Belzen, R., Albracht, S.P. and Lunardi, J. (1998) Mol. Microbiol. 28, 531–541.
[2] Carroll, J., Fearnley, I.M., Shannon, R.J., Hirst, J. and Walker, J.E. (2003) Mol. Cell Proteomics 2, 117–126.
[3] Fearnley, I.M. and Walker, J.E. (1992) Biochim. Biophys. Acta 1140, 105–134.
[4] Friedrich, T. and Weiss, H. (1997) J. Theor. Biol. 187, 529–540.
[5] Finel, M. (1998) Biochim. Biophys. Acta 1364, 112–121.
[6] Friedrich, T. and Scheide, D. (2000) FEBS Lett. 479, 1–5.
[7] Friedrich, T. (2001) J. Bioenerg. Biomembr. 33, 169–177.
[8] Hamamoto, T., Hashimoto, M., Hino, M., Kitada, M., Seto, Y., Kudo, T. and Horikoshi, K. (1994) Mol. Microbiol. 14, 939–946.
[9] Mathiesen, C. and Hagerhall, C. (2002) Biochim. Biophys. Acta 1556, 121–132.
[10] Krebs, W., Steuber, J., Gemperli, A.C. and Dimroth, P. (1999) Mol. Microbiol. 33, 590–598.
[11] Steuber, J., Schmid, C., Rufibach, M. and Dimroth, P. (2000) Mol. Microbiol. 35, 428–434.
[12] Cho, K.H. and Kim, Y.J. (2000) Mol. Cells 10, 432–436.
[13] Trchounian, A., Bagramyan, K. and Poladian, A. (1997) Curr. Microbiol. 35, 201–206.
[14] Hiramatsu, T., Kodama, K., Kuroda, T., Mizushima, T. and Tsuchiya, T. (1998) J. Bacteriol. 180, 6642–6648.
[15] Putnoky, P., Kereszt, A., Nakamura, T., Endre, G., Grosskopf, E., Kiss, P. and Kondorosi, A. (1998) Mol. Microbiol. 28, 1091–1101.
[16] Ito, M., Guffanti, A.A., Oudega, B. and Krulwich, T.A. (1999) J. Bacteriol. 181, 2394–2402.
[17] Kosono, S., Morotomi, S., Kitada, M. and Kudo, T. (1999) Biochim. Biophys. Acta 1409, 171–175.
[18] Ito, M., Guffanti, A.A., Wang, W. and Krulwich, T.A. (2000) J. Bacteriol. 182, 5663–5670.
[19] Hashimoto, M., Hamamoto, T., Kitada, M., Hino, M., Kudo, T. and Horikoshi, K. (1994) Biosci. Biotech. Biochem. 58, 2090–2092.
[20] Seto, Y., Hashimoto, M., Usami, R., Hamamoto, T., Kudo, T. and Horikoshi, K. (1995) Biosci. Biotech. Biochem. 59, 1364–1366.
[21] Altschul, S.F., Gish, W., Miller, W., Myers, E.W. and Lipman, D.J. (1990) J. Mol. Biol. 215, 403–410.
[22] Gish, W. and States, D.J. (1993) Nat. Genet. 3, 266–272.
[23] Madden, T.L., Tatusov, R.L. and Zhang, J. (1996) Methods Enzymol. 266, 131–141.
[24] Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W. and Lipman, D.J. (1997) Nucleic Acids Res. 25, 3389–3402.
[25] George, R.A. and Heringa, J. (2002) Proteins 48, 672–681.
[26] Bravo, A. (1997) J. Bacteriol. 179, 2793–2801.
[27] Goldin, A.L. (2002) J. Exp. Biol. 205, 575–584.
[28] Ruta, V., Jiang, Y., Lee, A., Chen, J. and MacKinnon, R. (2003) Nature 422, 180–185.
[29] Kaplan, J.H. (2002) Annu. Rev. Biochem. 71, 511–535.
[30] Halestrap, A.P. and Price, N.T. (1999) Biochem. J. 343, 281–299.
[31] Dupuis, A., Chevallet, M., Darrouzet, E., Duborjal, H., Lunardi, J. and Issartel, J.P. (1998) Biochim. Biophys. Acta 1364, 147–165.
[32] Kao, M.C., Di Bernardo, S., Matsuno-Yagi, A. and Yagi, T. (2002) Biochemistry 41, 4377–4384.
[33] Kyte, J. and Doolittle, R.F. (1982) J. Mol. Biol. 157, 105–132.
[34] Bernardo, S.D., Yano, T. and Yagi, T. (2000) Biochemistry 39, 9411–9418.
[35] von Heijne, G. (1992) J. Mol. Biol. 225, 487–494.
[36] Deleage, G., Clerc, F.F., Roux, B. and Gautheron, D.C. (1988) Comput. Appl. Biosci. 4, 351–356.
[37] Gratkowski, H., Lear, J.D. and DeGrado, W.F. (2001) Proc. Natl. Acad. Sci. USA 98, 880–885.
[38] Dawson, J.P., Weinger, J.S. and Engelman, D.M. (2002) J. Mol. Biol. 316, 799–805.
[39] Kervinen, M., Pätsi, J., Finel, M. and Hassinen, I. (2002) EBEC Short Rep. 12, 188.
[40] Brown, M.D., Starikovskaya, E., Derbeneva, O., Hosseini, S., Allen, J.C., Mikhailovskaya, I.E., Sukernik, R.I. and Wallace, D.C. (2002) Hum. Genet. 110, 130–138.
[41] Friedrich, T., Steinmuller, K. and Weiss, H. (1995) FEBS Lett. 367, 107–111.
[42] Thompson, J.D., Higgins, D.G. and Gibson, T.J. (1994) Nucleic Acids Res. 22, 4673–4680.
[43] Hall, T.A. (1999) Nucl. Acids Symp. Ser. 41, 95–98.