Biochimica et Biophysica Acta 1857 (2016) 971–979

Contents lists available at ScienceDirect

Biochimica et Biophysica Acta

journal homepage: www.elsevier.com/locate/bbabio

The origin of the supernumerary subunits and assembly factors of complex I: A treasure trove of pathway evolution☆

Dei M. Elurbe, Martijn A. Huynen *

[a] Centre for Molecular and Biomolecular Informatics, Radboud Institute for Molecular Life Sciences, Radboud University Medical Center, Nijmegen, The Netherlands
[b] Nijmegen Centre for Mitochondrial Disorders, Radboud University Medical Center, Nijmegen, The Netherlands

ARTICLE INFO

Article history:
Received 18 January 2016
Received in revised form 16 March 2016
Accepted 18 March 2016
Available online 15 April 2016

Keywords:
Complex I
Assembly factors
Supernumerary subunits
Pathway evolution
ECSIT
Neofunctionalization

ABSTRACT

We review and document the evolutionary origin of all complex I assembly factors and nine supernumerary subunits from protein families. Based on experimental data and the conservation of critical residues we identify a spectrum of protein function conservation between the complex I representatives and their non-complex I homologs. This spectrum ranges from proteins that have retained their molecular function but in which the substrate specificity may have changed or have become more specific, like NDUFAF5, to proteins that have lost their original molecular function and critical catalytic residues like NDUFAF6. In between are proteins that have retained their molecular function, which however appears unrelated to complex I, like ACAD9, or proteins in which amino acids of the active site are conserved but for which no enzymatic activity has been reported, like NDUFA10. We interpret complex I evolution against the background of molecular evolution theory. Complex I supernumerary subunits and assembly factors appear to have been recruited from proteins that are mitochondrial and/or that are expressed when complex I is active. Within the evolution of complex I and its assembly there are many cases of neofunctionalization after gene duplication, like ACAD9 and TMEM126B, one case of subfunctionalization: ACPM1 and ACPM2 in *Yarrowia lipolytica*, and one case in which a complex I protein itself appears to have been the source of a new protein from another complex: NDUFS6 gave rise to cytochrome c oxidase subunit COX4/COX5b. Complex I and its assembly can therewith be regarded as a treasure trove for pathway evolution. This article is part of a Special Issue entitled Respiratory complex I, edited by Volker Zickermann and Ulrich Brandt.

1. Introduction

NADH:ubiquinone oxidoreductase, or complex I, the first and largest of the oxidative phosphorylation complexes, has, for a long time been a poster child for molecular evolutionary studies. There are a number of reasons for this: The alphaproteobacterial origin of complex I and that some of its proteins are still mitochondrially encoded, has allowed its use in establishing that mitochondria evolved from the alphaproteobacteria [1], while its slower evolving nuclear encoded genes have been used to pinpoint the mitochondrial origin in the Rickettsiales order [2]. The question why some hydrophobic proteins, including complex I subunits, are mitochondrially encoded has been the subject of a long debate [3]. In this debate, recent experimental results support the old hypothesis that those hydrophobic proteins would be mistargeted to the endoplasmic reticulum when nuclear encoded [3,4]. The determination of the composition of complex I in various bacteria and eukaryotes in combination with the high rate of evolution of some of its eukaryotic subunits, have spurred in depth studies into the gain and loss of complex I subunits along various evolutionary lineages [5,6]. The homology of separate parts of complex I to independently functioning protein complexes in bacteria has led to its proposed evolution out of three, independently functioning modules, which has enriched our conceptual understanding of the evolution of complex pathways [7]. Evolutionary sequence conservation of complex I proteins is used to derive hypotheses of how complex I couples electron transfer to proton translocation [8]. Complex I's frequent evolutionary loss has made it a suitable subject for co-evolution studies, and its intimate biochemical connection to central carbohydrate metabolism was recently confirmed by a large scale analysis of its co-evolution with metabolic pathways in bacteria [9]. In eukaryotes the coevolution of proteins with complex I has been used to predict new assembly factors [10,11], expanding the number of complex I proteins involved in congenital disease [12]. The evolution of complex I has also led to some surprises. One is its occurrence in anaerobic hydrogen producing mitochondria [13,14]. A second one is the documentation of a mitochondrial genome from an aerobic multicellular species that lacks complex I genes [15], strongly suggesting the absence of the complex from its mitochondria. Most of these aspects have already been addressed and sometimes debated in reviews, or are extensively discussed in the various papers describing new experimental

☆ This article is part of a Special Issue entitled Respiratory complex I, edited by Volker Zickermann and Ulrich Brandt.

* Corresponding author at: Centre for Molecular and Biomolecular Informatics, Radboud University Medical Center, PO Box 9101, 6500 HB Nijmegen, The Netherlands.
E-mail address: huynen@cmbi.ru.nl (M.A. Huynen).

http://dx.doi.org/10.1016/j.bbabio.2016.03.027
0005-2728/

or computational results. We therefore decided to focus on an aspect that, although addressed in one of the earliest publications discussing complex I proteins' evolutionary relationships [16], has not recently been reviewed: the evolution of complex I supernumerary subunits and the assembly factors from pre-existing protein families.

1.1. An accumulation of new subunits and assembly factors along multiple evolutionary lineages

Homology detection, and its daughter orthology detection that increases the likelihood of detecting functionally equivalent proteins [17] are one of the cornerstones of bioinformatics. Purely based on experimental data, the number of complex I subunits increased from 14 in proteobacteria [18], to 17 in alphaproteobacteria [19], 44 in mammals [20,21], 49 in plants [22], 42 in *Chlamydomonas* [23], and at least 46 in *Trypanosoma* [24]. Orthology detection, using increasingly sensitive sequence analysis tools based on sequence profiles [25], has been intensively applied to compare the subunit composition of complex I between species and has revealed that the large majority of the so-called "supernumerary" subunits are actually shared between the taxa [5,6,26]. In a recent example, the number of subunits that were deemed to be specific to *Trypanosoma* [24] has gone down by comparison with the proteome of *Euglena gracilis* [27]. As horizontal gene transfer does not appear to play a role in complex I evolution in eukaryotes, in contrast to prokaryotes [9], we can infer that most supernumerary subunits that are shared between the major evolutionary lineages were added before the radiation of the eukaryotes, in the so-called Last Eukaryotic Common Ancestor (LECA) [6]. This pattern of a dramatic increase in the number of subunits in LECA is not unique to complex I and has been observed in all mitochondrial protein complexes of bacterial origin [26]. In contrast, the loss of complex I subunits is rare and has only been documented for one subunit in the lineage leading to the mammals: the carbonic anhydrase that can be inferred to have been part of complex I in LECA but is absent from it in metazoa and fungi [28]. The only subunits that currently appear to be unique to the metazoa are NDUFA10, NDUFB6, NDUFV3, NDUFC1 and NDUFB1. The latter two have been suggested to be homologous to the fungal subunits NUUM and NUXM respectively [6]. Nevertheless, also with sequence profiles based on the latest sequence data they do not show significant levels of sequence similarity to those (data not shown) and their homology will have to be decided by structure data. The successful efforts in mapping the supernumerary subunits from the eukaryotes onto each other do however not explain where the supernumerary subunits came from in the first place. Many complex I supernumerary subunits appear in evolution "out of the blue": i.e. we cannot find homologs that are not also orthologs. However, all assembly factors and nine supernumerary subunits are part of larger protein families, providing sometimes revealing and often puzzling clues about their role in complex I. We document these relationships here (Table 1), complementing the published cases with a number of newly discovered ones that are based on sequence-profile based homology detection and on the conservation of sequence motifs documented to be involved in substrate binding and catalysis.

1.2. Recruitment of supernumerary subunits

1.2.1. NDUFAB1

NDUFAB1 is an example of a supernumerary complex I protein that has retained its original molecular function. It is homologous to the acyl carrier proteins from bacteria, chloroplasts and mitochondria that are involved in fatty acid synthesis [29]. Acyl carrier proteins contain a DSL motif of which the serine binds 4′ phosphopantetheine that, in turn, can carry acyl groups via a thioester linkage to its sulfhydryl group. This DSL motif is conserved in human NDUFAB1 and in its orthologs of *Neurospora crassa* and *Bos taurus* that bind 4′ phosphopantetheine and an acyl group [30,31]. *Y. lipolytica* has two NDUFAB1 homologs, ACPM1 and ACPM2, which are both bound to complex I [32]. The phosphopantetheine binding serine in ACPM1 is required for *Y. lipolytica* viability, while that serine in ACPM2 is required for complex I activity [32]. Mutagenesis aimed at just disturbing the binding of ACPM1 to complex I results in almost complete absence of ubiquinone reductase activity but is not lethal [33]. The more dramatic effect of the mutation in ACPM1 is therefore likely due to a role of ACPM1 in fatty acid synthesis, like its ortholog

Table 1
Homologies of complex I supernumerary subunits and assembly factors with protein families that also occur outside of complex I. Homologies were detected using HHpred [25] against PFAM and against the human genome, using maximally four iterations with HHblits, E < 1E–5. All homologies were checked by reciprocal searches. Only protein families that contain more proteins than just complex I subunits are indicated and NDUFB10 that might be related to the twin Cx(9)C family. The dating of the origin of the proteins is tentative. For FOXRED1, even though the protein family is older than the eukaryotes, the lack of co-evolution with complex I suggests a recent origin of its assembly function [84]. Name: human gene name. Synonyms: gene synonyms. LCA: last common ancestor. Paralogs: human paralogs, if any. Critical conserved residues: if the supernumerary subunit has a homolog with a characterized catalytic activity, binding site or similar, whether the motif or the residues are conserved. PFAM: significant PFAM domain hit.

Name	Synonyms	LCA	Paralogs	Conserved critical residues	PFAM
NDUFA2	B9	LECA	MRPL43, MRPL25, MRPL53		Mitochondrial ribosomal protein L51/S25/CI-B8 domain
NDUFA6	B8, LYRM6	LECA	LYRM2, LYRM3, ...	Yes (LYR)	Complex 1 LYR
NDUFB9	B22, LYRM3	LECA	LYRM4, LYRM6,...	Yes (LYR)	Complex 1 LYR
NDUFA8	PGIV, MGC793	LECA	CHCHD5	Yes (Cx(9)C)	CX9C
NDUFB7	B18, MGC2480	LECA		Yes (Cx(9)C)	CX9C
NDUFS5	CI15k	LECA		Yes (Cx(9)C)	CX9C
NDUFB10	CI-PDSW	LECA		Cx(11–13)C (?)	
NDUFA9	NDUFS2L	LECA	NSDHL, SDR42E1,...	Partially (S/TY**K**)	GDP-mannose 4,6 dehydratase
NDUFA10	CI-42k	Metazoa	DGUOK, TK2,...	Yes	Deoxynucleoside kinase
NDUFAB1	ACP	LECA	AASDH	Yes (DSL)	Phosphopantetheine attachment site (ACP)
NDUFS6	CI-13kA	α-prot	COX5B	Yes (CHCC)	Zinc-finger domain (CHCC)
NDUFAF1	CIA30	LECA			Carbohydrate binding domain (family 11)
NDUFAF2	NDUFA12L	LECA	NDUFA12		NADH ubiquinone oxidoreductase subunit NDUFA12
NDUFAF3	C3orf60	Bacteria	AAMDC		DUF498
NDUFAF4	C6orf66	LECA	NGRN		Neugrin
NDUFAF5	C20orf7	Bacteria	SAM-dependent methyltransferases	Yes (GxGxG)	SAM-dependent methyltranferase (family 23)
NDUFAF7	C2orf56	Bacteria	SAM-dependent methyltransferases	Yes (GxGxG)	SAM-dependent methyltranferase (family 28)
NDUFAF6	C8orf38	Bacteria	FDFT1	No (2x DXXXD)	Squalene/phytoene synthase
ACAD9		Metazoa	VLCAD, GCDH,...	Yes (426E)	Acyl-CoA dehydrogenase
ECSIT		Filozoa	PTCD3		PPR repeat family
FOXRED1	FP364	Metazoa	SARDH, DMGDH,...	Yes (FAD binding)	FAD dependent oxidoreductase
NUBPL	C14orf127	Bacteria	NUBP1, NUBP2,...	Yes (Walker A&B, CXXC)	NUBPL iron-transfer P-loop NTPase
TIMMDC1	C3orf1	Metazoa	TIMM22, TIMM17B,s...		Tim17/Tim22/Tim23/Pmp24 family
TMEM126B		Mammalia	TMEM126A		DUF1370