Available online at www.sciencedirect.com
ScienceDirect
Biochimica et Biophysica Acta 1768 (2007) 2164–2181
Biochimica et Biophysica Acta BBA
www.elsevier.com/locate/bbamem

Evolutionary origins of members of a superfamily of integral membrane cytochrome *c* biogenesis proteins

Jong-Hoon Lee [a,c,1], Edgar M. Harvat [a,b,1], Julie M. Stevens [b], Stuart J. Ferguson [b], Milton H. Saier Jr. [a,*]

[a] *Division of Biological Sciences, University of California at San Diego, La Jolla, CA 92093-0116, USA*
[b] *Department of Biochemistry, University of Oxford, South Parks Rd., Oxford OX1 3QU, UK*
[c] *Department of Food Science and Biotechnology, Kyonggi University, Suwon 443-760, South Korea*

Received 9 December 2006; received in revised form 22 March 2007; accepted 24 April 2007
Available online 3 May 2007

Abstract

We have analyzed the relationships of homologues of the *Escherichia coli* CcmC protein for probable topological features and evolutionary relationships. We present bioinformatic evidence suggesting that the integral membrane proteins CcmC (*E. coli*; cytochrome *c* biogenesis System I), CcmF (*E. coli*; cytochrome *c* biogenesis System I) and ResC (*Bacillus subtilis*; cytochrome *c* biogenesis System II) are all related. Though the molecular functions of these proteins have not been fully described, they appear to be involved in the provision of heme to *c*-type cytochromes, and so we have named them the putative Heme Handling Protein (HHP) family (TC #9.B.14). Members of this family exhibit 6, 8, 10, 11, 13 or 15 putative transmembrane segments (TMSs). We show that intragenic triplication of a 2 TMS element gave rise to a protein with a 6 TMS topology, exemplified by CcmC. This basic 6 TMS unit then gave rise to two distinct types of proteins with 8 TMSs, exemplified by ResC and the archaeal CcmC, and these further underwent fusional or insertional events yielding proteins with 10, 11 and 13 TMSs (ResC homologues) as well as 15 TMSs (CcmF homologues). Specific evolutionary pathways taken are proposed. This work provides the first evidence for the pathway of appearance of distantly related proteins required for post-translational maturation of *c*-type cytochromes in bacteria, plants, protozoans and archaea.

Keywords: Transport; Protein evolution; Transporter topology; Gene duplication; Gene fusion; Gene deletion

1. Introduction

c-type cytochromes are essential constituents of numerous multicomponent electron transfer chains present in virtually all major groups of living organisms [1,2]. These proteins contain heme that is covalently linked via heme vinyl groups to two cysteine residues in the apocytochromes. Thioether bonds interconnect the heme and the protein [3]. The conserved signature motif, Cys-X-X-Cys-His, provides the site of covalent binding. Three distinct systems are believed to have evolved independently of each other for the insertion of the chromophore and assembly of the holocytochrome *c* [4]. Prokaryotes, plant mitochondria and chloroplasts use Systems I (>10 components) and II (4 recognized components) [5]. Archaea seem to possess a significantly modified version of System I [6]. Fungal and animal mitochondria use System III, where a single enzyme, CCHL (*C*ytochrome *c H*eme *L*yase), is sufficient for the covalent attachment of heme to the apocytochrome (cytochrome polypeptide). A membrane bound flavoprotein, Cyc2p, has recently been suggested to play a role in the reduction of the heme iron in this system [7]. In each case, the overall process requires heme transport and delivery, apocytochrome ushering, reductant provision and thioreduction [8].

Members of what we define here as the Heme Handling Protein (HHP) families are present in Systems I and II [4]. They display a well-conserved tryptophan-rich motif flanked by conserved histidines [9]. They are believed to function in heme delivery, possibly presenting the cofactor to the target apocytochrome in the periplasm [9,10].

Abbreviations: TMS, Transmembrane Segments; HHP, Heme Handling Protein; Ccm, *C*ytochrome *c m*aturation; CCHL, *C*ytochrome *c H*eme *L*yase; MC, Mitochondrial Carrier

* Corresponding author. Tel.: +1 858 534 4084; fax: +1 858 534 7108.
E-mail address: msaier@ucsd.edu (M.H. Saier).
[1] These authors contributed equally to this work.

0005-2736/$ - see front matter © 2007 Elsevier B.V. All rights reserved.
doi:10.1016/j.bbamem.2007.04.022

CcmC, a component of System I, is required for heme provision and covalent attachment to the heme chaperone CcmE [11]. Subsequently, heme is transferred stereospecifically from CcmE to the apocytochrome, a process that depends on a distant CcmC homologue, CcmF [12]. The biochemical functions of CcmC and CcmF are not established. CcmC has been hypothesized to be a subunit of an ABC transporter, along with CcmAB [13]. However, there is evidence that CcmC functions independently of CcmAB [11,14], and it has been postulated to be an independently functioning heme exporter [10,11]. CcmF is proposed to facilitate heme transfer from CcmE to the apocytochrome, in cooperation with CcmH, a thioredoxin-like protein that may reduce the two thiols of the CXXCH motif of the cytochrome with electrons provided by the cytoplasmic thioredoxin via DsbD and CcmG [12].

Another distant homologue of CcmC is ResC of *Bacillus subtilis*. ResC has been shown to be essential for cytochrome *c* production in organisms that use System II and is thought to be involved, together with ResB, in heme delivery to apocytochromes [15]. Similarly to System I, reduction of the heme binding thiols of the cytochrome is required prior to covalent heme attachment. This role is fulfilled by two thioredoxin-like proteins, ResA and CcdA [15,16].

We have previously shown [17] that the 6 TMS heme binding protein, YedZ of *Escherichia coli*, contains three recognizable repeat units of two TMSs. From this observation, we inferred that the current 6 TMS YedZ protein arose by triplication of a 2 TMS-encoding genetic element. A 6 TMS topology was established experimentally for YedZ with the N- and C-termini inside [18]. von Rozycki et al. [17] also noted weak sequence similarity of this protein with CcmC.

In this work, we show, using established bioinformatic methodology, that cytochrome *c* biogenesis proteins CcmC, CcmF and ResC are related, having been derived from a primordial basic unit of 6 putative TMSs. We further show that, like YedZ, these proteins arose by triplication of a 2 TMS element. This primordial 6 TMS protein then underwent a variety of intragenic and extragenic duplication, deletion and fusion events to give homologous proteins of various topologies with 6, 8, 10, 11, 13 and 15 probable TMSs.

2. Methods

In this study, the CcmC homologue involved in cytochrome *c* biogenesis in *Desulfovibrio desulfuricans* G20 (gi 23475583) was used as the query sequence for BLAST searches [19]. Hundreds of homologous proteins were retrieved from the NCBI database (e-values=10^{-4} on 9/29/2005). Redundant sequences were eliminated, and 324 proteins were retained for topological analyses. Since the CcmF subfamily was less well represented than the CcmC and ResC subfamilies in this initial BLAST search, a subsequent search was performed to retrieve all CcmF homologues using the CcmF protein from *Pasteurella multocida* str. Pm70 (gi 12720209) as query. The BLAST search yielded about 380 sequences. Using the protein clustering program CD-hit [20,21], redundant sequences (defined as proteins of greater than 90% identity) were eliminated, and representatives of each cluster were randomly chosen. This left 180 CcmF sequences.

The CLUSTAL X program [22] and the TreeView program [23] were used, respectively, for multiple alignment of homologous sequences and construction of the phylogenetic trees. The default parameters of the CLUSTAL X program were used for generating the phylogenetic trees (number of bootstrap trials: 1000). Topological analyses of individual proteins were performed using the WHAT [24] and HMMTOP [25] programs.

Statistical sequence similarity comparisons between proteins, and between internal regions of these proteins, were conducted using the IC [26] and GAP [27] programs. These programs randomly shuffle the primary sequences of the proteins or protein segments under scrutiny, and compare these shuffled sequences with the native sequences. They therefore correct for abnormal protein compositions that can occur in integral membrane proteins. Five hundred random shuffles and default settings have proven to be satisfactory for obtaining statistically significant values. A value of 9 standard deviations (S.D.) for comparable regions of the two proteins of at least 60 amino acyl residues (aas), corresponding to a probability of 10^{-27} that the observed degree of sequence similarity arose by chance, is considered sufficient to establish homology as described previously [28–32].

After establishing the evolutionary relationships between the various proteins, and between the internal segments of these proteins, closely related proteins were further eliminated in order to simplify phylogenetic tree construction. Two hundred and fifty-three proteins remained, and these were analyzed topologically and phylogenetically. Although there is general agreement about the topology of CcmC proteins, which have 6 TMSs [5,33], topological studies with members of the CcmF and ResC subfamilies within the HHP family have given conflicting predicted topologies [5,9,34]. Reference to TMSs therefore refers throughout to putative TMSs, based on hydropathy analyses described in Results.

3. Results

Tables 1 and 2 list the examples of CcmC, CcmF and ResC analyzed in this study. A multiple alignment of these proteins (Fig. S1a), together with a simplified alignment displaying only representatives of each major cluster (Fig. S1b), can be found on our website (www.biology.ucsd.edu/~msaier/supmat/CcmC/). The topological and statistical sequence studies carried out in this work demonstrate sufficient sequence similarity to allow establishment of homology between CcmC, CcmF and ResC (see below). These proteins are therefore designated members of a single protein family, here named the Heme Handling Protein (HHP) family (TC #9.B.14), in recognition of their incompletely understood roles in insertion of heme into *c*-type cytochromes and their relationship to the heme protein YedZ.

3.1. Numbers of paralogues of the HHP family in individual organisms

Numerous organisms have homologues of the HHP family; organisms from all three domains of life, bacteria, archaea and eukaryotes, have genes that encode these proteins. Almost all of the major recognized bacterial classes, as well as two of the archaeal phyla, the crenarchaeote and the euryarchaeote, are represented. However, in eukaryotes, only plants, algae, flagellates and other photosynthetic organisms possess these homologues. Thus, the non-photosynthetic fungal, protozoan, and animal kingdoms are not represented. These organisms have the cytochrome *c* biogenesis System III. Eukaryotic HHP family proteins function specifically for the assembly of the *c*-type cytochrome components (including cytochrome *f* of chloroplasts) in plant mitochondrial and thylakoid electron transfer chains [4,35].

There seem to be at least three distinct functional subfamilies of the HHP family, and each of them includes several distinct phylogenetic clusters. For optimal clarity, we present individual phylogenetic trees for each subfamily (Figs. 1A, B and 2A). Subfamily 1 (CcmC) and subfamily 2 (CcmF) (Fig. 1 and Table 1) consist exclusively of CcmC and CcmF homologues, respectively. They are both constituents of cytochrome *c* biogenesis System I.