FEBS Letters 580 (2006) 4827–4834

A variant System I for cytochrome *c* biogenesis in archaea and some bacteria has a novel CcmE and no CcmH

James W.A. Allen[1], Edgar M. Harvat[1], Julie M. Stevens, Stuart J. Ferguson*

Department of Biochemistry, University of Oxford, South Parks Road, Oxford OX1 3QU, United Kingdom

Received 21 June 2006; revised 14 July 2006; accepted 20 July 2006

Available online 4 August 2006

Edited by Peter Brzezinski

Abstract ***C*-type cytochromes are characterized by post-translational covalent attachment of heme to thiols that occur in a Cys-Xxx-Xxx-Cys-His motif. Three distinct biogenesis systems are known for this heme attachment. Archaea are now shown to contain a significantly modified form of cytochrome *c* maturation System I (the Ccm system). The most notable adaptation relative to the well-studied apparatus from proteobacteria and plants is a novel form of the heme chaperone CcmE, lacking the highly conserved histidine that covalently binds heme and is essential for function in *Escherichia coli*. In most archaeal CcmEs this histidine, normally found in a His-Xxx-Xxx-Xxx-Tyr motif, is replaced by a cysteine residue that occurs in a Cys-Xxx-Xxx-Xxx-Tyr motif. The CcmEs from two halobacteria contain yet another form of CcmE, having HxxxHxxxH approximately corresponding in alignment to the H/CxxxY motif. The CxxxY-type of CcmE is, surprisingly, also found in some bacterial genomes (including *Desulfovibrio* species). All of the modified CcmEs cluster together in a phylogenetic tree, as do other Ccm proteins from the same organisms. Significantly, CcmH is absent from all of the complete archaeal genomes we have studied, and also from most of the bacterial genomes that have CxxxY-type CcmE.**

Keywords: Archaea; *C*-type cytochrome; Ccm; Cytochrome *c* maturation; Post-translational modification; Heme

1. Introduction

C-type cytochromes are widespread proteins characterized by covalent attachment of the iron cofactor heme to protein through thioether bonds formed between the heme vinyl groups and the thiols of two cysteines in a CxxCH motif [1]. The histidine is usually an axial ligand to the heme iron. Mitochondrial cytochrome *c* is the best known such protein, but there are many other distinct *c*-type cytochrome centres, *e.g.* in bacteria, that function in electron transfer or occur at the catalytic sites of enzymes.

Remarkably, evolution has produced multiple distinct biogenesis systems to achieve the post-translational covalent heme attachment to apocytochromes *c*; three have been characterized [2,3]. System I, also called the cytochrome *c* maturation (Ccm) system [4], consists of up to nine dedicated proteins (CcmABCDEFGHI) and several accessory proteins (mainly involved in disulfide bond oxidation/reduction). This apparatus is found in α-, some β- and most γ-proteobacteria, deinococci and the mitochondria of various eukaryotes including land plants, red algae and ciliates. System II [5] occurs in δ-, ε- and some β-proteobacteria, at least one γ-proteobacterium, most Gram-positive bacteria, plant and algal chloroplasts and cyanobacteria. Significant recent advances from the laboratory of Kranz [6,7] have shown that System II consists of just two essential proteins, ResB and ResC (sometimes found as a single fusion protein), and two accessory proteins for disulfide bond oxidation and reduction (ResA and either CcdA or DsbD). Biogenesis System III, the enzyme heme lyase, is found in the mitochondria of fungi, metazoans and some protozoa.

The Domain Archaea represents the major group of organisms in which cytochrome *c* biogenesis has not been studied significantly. Very recently, Bertini and co-workers undertook a detailed bioinformatic study of the occurrence of *c*-type cytochromes in all the domains of life [8]. Searching completely sequenced genomes, they identified four archaea that contain certain types of cytochrome(s) *c*. Following that work, and because a significant number of archaeal genomes are now complete, we have undertaken an *in silico* investigation of the cytochrome *c* biogenesis apparatus of the archaea.

2. Results and discussion

Our investigation commenced with the complete genomes of the four archaea identified by Bertini et al. [8] as containing at least one *c*-type cytochrome. These organisms are *Aeropyrum pernix* K1, *Archaeoglobus fulgidus* DSM4304, *Methanosarcina acetivorans* C2A and *Pyrobaculum aerophilum* IM2. We queried the archaeal genomes using tools at either Pedant (http://pedant.gsf.de/) or NCBI (http://www.ncbi.nih.gov/Genomes/) and protein sequences representative of cytochrome *c* biogenesis System I (from *Escherichia coli*, *Rhodobacter capsulatus* and *Deinococcus radiodurans*), System II (from *Arabidopsis thaliana*, *Geobacter metallireducens* and *Bacillus subtilis*) and System III (*Saccharomyces cerevisiae* and *Homo sapiens*). For System I, we searched using the characteristic marker

*Corresponding author. Fax: +44 1865 275259.
E-mail address: stuart.ferguson@bioch.ox.ac.uk (S.J. Ferguson).

[1] These authors contributed equally to this work.

0014-5793/$32.00
doi:10.1016/j.febslet.2006.07.073

proteins CcmB, CcmC, CcmE, CcmF and CcmH. [2] For System II, we chose ResB and ResC. For System III we searched using the two heme lyases from *S. cerevisiae* (cytochrome *c* heme lyase and cytochrome c_1 heme lyase), and the single human heme lyase. Searches were finalised as of 31 May 2006.

It was clear from these initial searches that the complete archaeal genomes that contain *c*-type cytochromes all contain biogenesis System I (Table 1). We found no evidence for the presence of Systems II or III in any of these organisms. This concurs with an early observation by Goldman and Kranz reporting on emerging genes from *A. fulgidus* [10], but is much more extensive both in the range of biogenesis proteins and the range of organisms investigated. Notably, however, the version of System I present in the archaea is significantly modified compared with that characterized in proteobacteria and plant mitochondria.

2.1. A novel form of CcmE highlighted by the archaea

CcmE is a heme chaperone and functionally by far the best-studied of the Ccm proteins [11]. The protein from proteobacteria and from plant mitochondria has been shown to bind heme covalently through a histidine residue (His130 using *E. coli* residue numbering [3]) [12,13]; the covalent adduct is believed to be an intermediate in cytochrome *c* maturation [12,14]. Thus, residue H130 has been thought to be universally conserved and functionally essential. However, inspection and sequence alignment of the CcmE proteins from *A. pernix*, *A. fulgidus*, *M. acetivorans* and *P. aerophilum* (Fig. 1A) indicates that this histidine is replaced by a cysteine. Residue 134 (*E. coli*) is a tyrosine that is thought to act as a ligand to the iron of the covalently bound heme of CcmE [15]; this tyrosine is conserved in these archaeal sequences (Fig. 1A), as well as in all the proteobacterial/plant (H130)-type CcmEs.

Using the sequences of archaeal CcmEs containing the novel "CxxxY motif" as BLAST queries, we have identified a total of 15 CcmEs with this motif (Fig. 1A and Table 1). These include six from archaea (*A. pernix*, *A. fulgidus*, *M. acetivorans*, *P. aerophilum*, *Methanosarcina mazei* and *Methanococcoides burtonii*) and, surprisingly, nine from diverse bacterial classes and species; *Chloroflexus aurantiacus* J-10-fl (Chloroflexi), *Cytophaga hutchinsonii* and *Salinibacter ruber* DSM13855 (Bacteroidetes), *Carboxydothermus hydrogenoformans* Z-2091 and *Halothermothrix orenii* H168 (Firmicutes), *Leptospira interrogans* serovar Copenhagen str. Fiocruz L1-130 (Spirochaetes), *Solibacter usitatus* Ellin6076 (Acidobacteria) and the δ-proteobacteria *Desulfovibrio vulgaris* Hildenborough and *Desulfovibrio desulfuricans* G20.

Supporting evidence for our assignment of these archaeal and bacterial proteins as a novel variant of CcmE (Table 1) comes initially from sequence data; for example, they share 16–26% identity and 26–43% similarity with *E. coli* CcmE, and this similarity including several very highly conserved residues (>95%) occurs throughout the polypeptide chain (Fig. 1A). The putative CcmEs are all around the expected (typical) size for CcmE. Further very strong evidence is found in the co-organisation of the *ccm* genes in some of the archaea and bacteria we have considered. In the *A. fulgidus* genome, the gene for the putative CcmE is contiguous with a very strong homologue of *E. coli* CcmF. In *P. aerophilum*, the gene we assign as *ccmE* is between genes encoding homologues of CcmC and CcmF. The genes coding for the readily recognisable proteins CcmB, CcmC, CcmE and CcmF (see footnote 2) all occur together in the bacteria *Cytophaga hutchinsonii*, *D. vulgaris*, *D. desulfuricans*, *Chloroflexus aurantiacus*, *S. usitatus* and *H. orenii*. Note that in Table 1, consecutive gi (gene identification) numbers indicate adjacent genes in the genome sequence. In some cases (*e.g.* in *M. acetivorans*, *A. pernix*, *S. ruber* and *C. hydrogenoformans*), the genes we have assigned as *ccmE* are not arranged with other *Ccm* genes in the genome. However, we are confident about our assignment of CcmE in these organisms because of the significant sequence similarity of their putative CcmEs with the CcmEs from organisms where *ccm* genes *are* co-organised (*cf.* Figs. 1A and 2A). Note that the overall genetic organisation of the Ccm system varies considerably in the organisms we discuss here. Sometimes, as described above, the *Ccm* genes are all found together. In other cases a few are together (*e.g.* CcmB and CcmC in *M. acetivorans*, *M. burtonii*, *M. mazei* and *L. interrogans*, or CcmE and CcmF in *A. fulgidus* and *L. interrogans*), with the remainder elsewhere in the genome. Finally, in a few of the organisms (*A. pernix*, *S. ruber* and *C. hydrogenoformans*), none of the Ccm genes we have identified are located together in the genome. It should be noted that *ccm* genes are frequently non-contiguous in other organisms, for example in many α-proteobacteria.

The substantial group of CcmEs with cysteine in place of the key histidine residue (Table 1) all cluster together in a phylogenetic tree (Fig. 2A). Thus, they seem likely to represent a distinct sub-group of CcmEs, which may be functionally different from those in other organisms. The effect of mutation of residue His130 to cysteine has been studied in *E. coli* CcmE and does not abolish covalent binding of heme to the protein [16,17], but renders the Ccm system incapable of maturing *c*-type cytochromes [17]. Thus, conceivably, the novel group of cysteine-containing CcmEs, described here in the archaea and some bacteria, could be non-functional in *c*-type cytochrome biogenesis. However, it seems much more likely that the CxxxY-type CcmE *will* be functional. Several lines of evidence support the latter hypothesis:

(i) Each archaeon containing this unusual cysteine form of CcmE contains the gene for at least one probable *c*-type cytochrome (see below). Some cytochromes *c* from these archaea have been purified or studied biochemically [18–21], so must have been handled by a functional biogenesis apparatus.

(ii) No other known cytochrome *c* maturation proteins are present in these archaeal genomes. Thus, Occam's Razor suggests that archaeal cytochrome *c* maturation will be mediated by the Ccm proteins, including the novel CcmE. CcmE is essential for cytochrome *c* maturation in the proteobacterial Ccm apparatus [12,22].

(iii) There are further modifications to the archaeal-type Ccm system compared with the well-studied apparatus from proteobacteria/plants, which presumably compensate

[2] Note that CcmA is a poor search choice because it contains the Walker motifs characteristic of ATPases and generates a large number of non-specific BLAST hits. CcmD is a very small protein (~60 amino acids) and generally very poorly conserved. CcmG has the thioredoxin fold [9] and again generates a large number of non-specific hits. CcmI contains TPR repeat motifs and also generates many non-specific BLAST hits. The remaining Ccm proteins can be considered as distinct markers for the Ccm system.

[3] Position 190 in the sequence alignment in Fig. 1A.