Multiple haem lyase genes indicate substrate specificity in cytochrome c biogenesis

S. Hartshorne*, D.J. Richardson* and J. Simon*†

*School of Biological Sciences, University of East Anglia, Norwich NR4 7JL, U.K., and †Institut für Molekulare Biowissenschaften, Johann Wolfgang Goethe-Universität, Marie-Curie-Strasse 9, D-60439 Frankfurt am Main, Germany

Abstract
c-Type cytochromes are a widespread class of proteins that play a vital role in the energy-conserving metabolism of prokaryotic and eukaryotic organisms. The key event in cytochrome c biogenesis is the covalent attachment of the haem cofactor to two (or rarely one) cysteine residues arranged in a haem c-binding motif such as CX2,4CH, CXXCK or X4CH. This reaction is catalysed by the membrane-bound enzyme CCHL (cytochrome c haem lyase). Different CCHLs have been described and some of them are dedicated to distinct haem c-binding motifs of cytochromes that are encoded in the vicinity of the respective CCHL gene. Various bacterial genomes contain multiple copies of CCHL-encoding genes, suggesting the presence of non-conventional or even as yet unrecognized haem c-binding motifs. This assumption is exemplified in the present study using the proteobacterium Wolinella succinogenes as a model organism whose genome encodes three CCHL isoenzymes. The discovery of a novel conserved multihæm cytochrome c (MccA) is described.

Introduction
c-Type cytochromes are a widespread class of proteins essential for the life of many different organisms [1,2]. They are characterized by the covalent attachment of haem (Fe-protoporphyrin IX) to a polypeptide chain via two (or rarely one) thioether bonds generated as a result of the reaction of thiol groups of reduced cysteine residues with haem vinyl groups. The two cysteine residues usually occur in the amino acid sequence CX2CH (or exceptionally CX3CH or CX4CH). Cytochromes c typically function in electron transfer, but c-type cytochrome centres are also present in the active site of many enzymes. Bacterial mono- or multihæm c-type cytochromes are located at the outside of the cytoplasmic membrane, in the periplasm or within the outer membrane of Gram-negative bacteria.

Remarkably, two bacterial cytochrome c biogenesis systems have been identified that completely differ in their enzymic components. These systems are named Ccm (cytochrome c maturation; also known as System I) and Ccs (cytochrome c synthesis; System II) (for reviews, see [3–6]). The Ccm system usually consists of eight proteins (CcmA–CcmH) and is found in α- and γ-proteobacteria, in Desmococcus and in plant and red algal mitochondria. The Ccs system involves at least four proteins (CcsA, CcsB, CcdA and ResA) and occurs in β-, δ- and ε-proteobacteria, Gram-positive bacteria, cytophagales, aquaphales, plant and algal chloroplasts and cyanobacteria. While a lot of experimental evidence is available for System I of Escherichia coli, knowledge of System II is limited [7–9].

The last steps of cytochrome c biogenesis comprise recognition of haem and of the reduced cysteine residue(s) of the haem c-binding motif followed by stereo-specific covalent haem attachment. These tasks are carried out by the enzyme CCHL (cytochrome c haem lyase). No CCHL has ever been purified, and enzymically catalysed haem attachment to apo-cytochrome has never been shown in vitro. The CCHL in System II is most likely CcsA, a membrane-bound protein of approx. 250–350 amino acid residues with five or six transmembrane domains [10,11]. The ccsA gene is often located adjacent to ccsB encoding another essential membrane protein that possibly forms a complex with CcsA [9,12]. In genomes of ε-proteobacteria, ccsB and ccsA genes appear to be fused, resulting in genes coding for membrane-bound CCHLs of approx. 900 residues. The System I protein that most likely corresponds functionally to CcsA is CcmF.

Unconventional haem c-binding motifs require dedicated haem lyases

In addition to the CX2,4CH motifs mentioned above, the sequences CXXCK and A/FX2CH have been described as haem c-binding motifs in bacterial and eukaryotic proteins respectively [13,14]. These motifs are apparently not processed by the components of the regular Ccm or Ccs systems but require a dedicated (private) CCHL, thus raising the question of how such enzymes are able to discriminate between distinct haem c-binding motifs. The unique CXXCK motif is found in most primary structures of pentahæm...
cytochrome c nitrite reductase (NrfA) where it binds the active site haem c group [13]. This high-spin haem group is axially ligated by the lysine residue of the CXXCK motif, whereas the other four low-spin haem c groups are conventionally bound by CX2CH motifs. The NrfE, NrfF and NrfG proteins that are similar to Ccm proteins have been shown to be involved specifically in covalent haem c attachment to the CXXCK motif of E. coli NrfA [15]. The NrfE, NrfF and NrfG proteins are therefore thought to make up a dedicated CCHL for NrfA that evolved from the Ccm system and that serves specifically in the recognition of the CXXCK motif in an unknown manner. A functionally similar system is present in the System II ε-proteobacterium Wolinella succinogenes. Here, the NrfI protein encoded by the gene situated immediately downstream of nrfA is predicted to be a CcsB–CcsA fusion-type CCHL belonging to the cytochrome c biogenesis System II (Figure 1A) [16]. The best evidence that NrfI is indeed a CCHL that recognizes the active site lysine residue of NrfA comes from the fact that the nrfI gene was found to be dispensable for maturation of a modified NrfA protein with five CX2CH motifs [17]. Likewise, it has to be assumed that conventional CCHLs are able to recognize the above-mentioned CX2–4CH motifs. It is intriguing that CCHLs dedicated to the CXXCK motifs of similar NrfA proteins (W. succinogenes and E. coli NrfA share 43% identical residues) have evolved on the basis of the respective cytochrome c biogenesis system.

The cytochromes c and c1 from mitochondria of Euglena species and trypanosomatids (various Trypanosoma and Leishmania species) contain only a single cysteine residue in the respective haem c-binding motif (AX2CH or FX2CH). The available genomes of several trypanosomatids do not encode proteins of the cytochrome c biogenesis Systems I and II, nor do they comprise the commonly found mitochondrial CCHL which is referred to as ‘System III’ [14]. Haem attachment to monocysteine motifs therefore also requires a dedicated maturation system that has not been identified to date.

Three haem lyase genes in W. succinogenes and the discovery of a novel Mcc (multihaem c-type cytochrome)

The W. succinogenes genome [18] predicts all components of a typical Ccs system (System II) of cytochrome c biogenesis including three different CCHLs, an as yet unprecedented feature in ε-proteobacteria (Figure 1). All three enzymes are CcsB–CcsA fusion proteins; they comprise 902 (NrfI, see above), 897 (CcsA1) and 935 (CcsA2) residues and their primary structures show 40–43% pairwise identity. The C-terminal third of each protein corresponds to the shorter CcsA proteins found in organisms outside the ε-proteobacteria and is predicted to contain at least seven
The context of mccA genes in different bacterial genomes

See text for details. The mccA gene clusters from S. putrefaciens CN-32 and Shewanella sp. PV-4 are similar to that of S. oneidensis MR-1 except that nrfC and nrfD homologues are absent from S. putrefaciens CN-32. The C. lari RM2100 gene designated as cyt/fdh encodes a putative electron transfer protein. Its N-terminal domain is predicted to be a hexa-haem cytochrome c (six CX2CH motifs) and its C-terminal domain resembles the di-haem cytochrome b membrane anchor of E. coli formate dehydrogenase N. See Figure 1 for further explanations; dsbC, gene encoding a putative thiol-disulphide interchange protein.

transmembrane domains. The N-terminal two-thirds of each CCHL are similar to CcsB proteins and contain three putative transmembrane domains close to their N-termini.

The ccsA1 gene is organized as part of a gene cluster that encodes MccA, a predicted Mcc of unknown function (Figure 1B). The mccA and ccsA1 genes are likely to belong to the same transcriptional unit. The close proximity of the two genes raises the possibility that CcsA1 is specifically involved in MccA maturation. The deduced primary structure of MccA contains seven CX2CH motifs as well as one monocysteine signature (AKGCH). The latter is conserved in all known primary structures similar to W. succinogenes MccA (consensus: A/NXGCH). The latter is conserved in all known primary structures similar to W. succinogenes MccA (consensus: A/NXGCH). These are CLA1177 from Campylobacter lari RM2100, Soo0479 from Shewanella oneidensis MR-1, two further amino acid sequences derived from incomplete genome sequences of Shewanella putrefaciens CN-32 and Shewanella sp. PV-4 (data obtained from the Oak Ridge National Laboratory website: http://genome.ornl.gov/microbial/), and protein AdehDRAFT_1113 predicted by the genome of Anaeromyxobacter dehalogenans 2CP-C. The mccA genes from C. lari and Shewanella spp. are also surrounded by respective putative CCHLs of either the CcsA-type (C. lari) or the NrfE-, NrfF-, NrfG-type (Shewanella spp.; Figure 2). This fact supports the assumption that a dedicated CCHL is required for MccA maturation, possibly recognizing and binding haem to the monocysteine motif.

To date, no report on the purification and characterization of an MccA protein has been published. Our preliminary MS analysis of the purified W. succinogenes MccA, however, suggests the covalent attachment of eight, rather than seven, haem groups (S. Hartshorne, R. Gross, D.J. Richardson and J. Simon, unpublished work). Such evidence might also explain the fact that W. succinogenes has previously been shown to be able to covalently attach haem to a modified SX2CH motif of the tetra-haem c-type cytochrome NrfH [19]. In contrast, E. coli is apparently unable to attach haem covalently to a monocysteine motif, which is in line with the fact that its genome encodes only two CCHLs that are apparently dedicated to the respective CXnCH and CXXCH motifs [15,20]. The W. succinogenes ccsA2 gene is located downstream of the bemH gene encoding ferrochelatase (Figure 1C). This enzyme catalyses the incorporation of iron into protoporphyrin IX, which is the last step of haem biosynthesis. A cytochrome c-encoding gene is absent from the vicinity of the ccsA2 gene. Hence, the CcsA2 protein is a legitimate candidate for the CCHL that attaches haem to conventional CXnCH motifs.

Conclusions and implications

An increasing number of bacterial genomes have been found to encode multiple copies of putative CCHLs, e.g.
W. succinogenes (see above) and Geobacter sulfurreducens (six CcsA homologues). CCHL-encoding genes located near putative c-type cytochrome genes suggest covalent haem attachment to non-conventional or even as yet unrecognized haem c-binding motifs. If true, such a finding will have far-reaching consequences for computer-based assignments of c-type cytochrome genes and the number of predicted haem c groups in multihaem cytochromes. W. succinogenes appears to be an ideal model organism concerning future studies on the capability of haem attachment to different haem c-binding motifs by using appropriate CCHL gene deletion mutants as host organisms. Finally, it will be of particular interest to examine the biophysical consequences of haem attachment to non-conventional haem c-binding motifs in order to explain the possible function of such haem groups in electron transfer or substrate turnover.

We are grateful for the opportunity to present our data at the 11th N-cycle meeting in Granada and for support by the Biochemical Society and by the European Science Foundation COST Action 856. J.S. is a Heisenberg fellow of the DFG (Deutsche Forschungsgemeinschaft). The work was supported by the Biotechnology and Biological Sciences Research Council (U.K.), Schlumberger (U.K.) and the DFG (Germany).

References

Received 22 September 2005