Kent Academic Repository

Lobo, Susana A.L., Videira, Marco A.M., Pacheco, Isabel, Wass, Mark N., Warren, Martin J., Teixeira, Miguel, Matias, Pedro M., Romão, Célia V. and Saraiva, Lígia M. (2016) ***Desulfovibrio vulgarisCbiKPcobaltochelatase: evolution of a haem binding protein orchestrated by the incorporation of two histidine residues.*** **Environmental Microbiology, 19 (1). pp. 106-118. ISSN 1462-2912.**

Downloaded from

https://kar.kent.ac.uk/58529/ The University of Kent's Academic Repository KAR

The version of record is available from

https://doi.org/10.1111/1462-2920.13479

This document version

Author's Accepted Manuscript

DOI for this version

Licence for this version

UNSPECIFIED

Additional information

Versions of research works

Versions of Record

If this version is the version of record, it is the same as the published version available on the publisher's web site. Cite as the published version.

Author Accepted Manuscripts

If this document is identified as the Author Accepted Manuscript it is the version after peer review but before type setting, copy editing or publisher branding. Cite as Surname, Initial. (Year) 'Title of article'. To be published in *Title of Journal* , Volume and issue numbers [peer-reviewed accepted version]. Available at: DOI or URL (Accessed: date).

Enquiries

If you have questions about this document contact ResearchSupport@kent.ac.uk. Please include the URL of the record in KAR. If you believe that your, or a third party's rights have been compromised through this document please see our Take Down policy (available from https://www.kent.ac.uk/guides/kar-the-kent-academic-repository#policies).

Desulfovibrio vulgaris CbiK^P cobaltochelatase: evolution of a haem binding protein orchestrated by the incorporation of two histidine residues

Susana A.L. Lobo[1*], Marco A.M. Videira[1*], Isabel Pacheco[1], Mark N. Wass[2], Martin J. Warren[2], Miguel Teixeira[1], Pedro M. Matias[1,3], Célia V. Romão[1], and Lígia M. Saraiva[1]+

[1]*Instituto de Tecnologia Química e Biológica NOVA, Avenida da República (EAN), 2780-157 Oeiras, Portugal*

[2]*School of Biosciences, University of Kent, Giles Lane, Canterbury, Kent, CT2 7NJ, United Kingdom*

[3]*iBET, Instituto de Biologia Experimental e Tecnológica, Apartado 12, 2781-901 Oeiras, Portugal*

**These authors contribute equally to this work*

[+]***Corresponding author:***

Instituto de Tecnologia Química e Biológica NOVA
Av. da República (EAN) 2780-157 Oeiras, Portugal
Tel: +351214469328; Fax: +351214433644; e-mail: lst@itqb.unl.pt

Running title: *D. vulgaris* cobaltochelatase CbiK^P haem binding protein

Keywords: cobaltochelatase, haem binding protein, *Desulfovibrio vulgaris*, sulphate reducing bacteria

Originality-Significance

We show for the first time that a cobaltochelatase of an ecological important organism, such as the sulphate-reducer *Desulfovibrio vulgaris* cobaltochelatase CbiKP, evolved to a haem binding protein by the presence of two histidine residues, which are absent from all other known bacterial cobaltochelatases, permitting it to act as a potential haem chaperone or transporter.

Abstract

The sulphate-reducing bacteria of the *Desulfovibrio* genus make three distinct modified tetrapyrroles, haem, sirohaem and adenosycobamide, where sirohydrochlorin acts as the last common biosynthetic intermediate along the branched tetrapyrrole pathway. Intriguingly, *D. vulgaris* encodes two sirohydrochlorin chelatases, $\mathrm{CbiK}^{\mathrm{P}}$ and $\mathrm{CbiK}^{\mathrm{C}}$, that insert cobalt/iron into the tetrapyrrole macrocycle but are thought to be distinctly located in the periplasm and cytoplasm, respectively. Fusing GFP onto the C-terminus of $\mathrm{CbiK}^{\mathrm{P}}$ confirmed that the protein is transported to the periplasm. The structure-function relationship of $\mathrm{CbiK}^{\mathrm{P}}$ was studied by constructing twelve site-directed mutants and determining their chelatase activities, oligomeric status and haem binding abilities. Residues His154 and His216 were identified as essential for metal-chelation of sirohydrochlorin. The tetrameric form of the protein is stabilized by ArgR54 and Glu76, which form hydrogen bonds between two subunits. His96 is responsible for the binding of two haem groups within the main central cavity of the tetramer. Unexpectedly, $\mathrm{CbiK}^{\mathrm{P}}$ is shown to bind two additional haem groups through interaction with H103. Thus, although still retaining cobaltochelatase activity, the presence of His96 and His103 in $\mathrm{CbiK}^{\mathrm{P}}$, which are absent from all other known bacterial cobaltochelatases, has evolved $\mathrm{CbiK}^{\mathrm{P}}$ a new function as a haem binding protein permitting it to act as a potential haem chaperone or transporter.

Introduction

Metallated tetrapyrroles support life by acting as prosthetic groups, cofactors and coenzymes in a wide range of biological processes such as photosynthesis, oxygen transport, respiration and metabolism. This family of compounds are synthesized via a pathway that is initiated from 5-aminolevulinic acid, which is subsequently transformed into the first macrocyclic intermediate, uroporphyrinogen III. All the major metal-containing tetrapyrroles, including cobalamin,, coenzyme F_{430}, sirohaem, chlorophyll and haem *b*, are based upon this molecular template. The different modified tetrapyrroles require specific metal ion chelatases to ensure the appropriate insertion of the correct metal into the centre of the macrocyle. Sulphate-reducing bacteria generally make three modified tetrapyrroles, cobamides, sirohaem and haem *b*. In this class of bacteria, all three of these compounds are derived from sirohydrochlorin that acts as the substrate for metal insertion (Lobo *et al*., 2009; Bali *et al*., 2011).

Sulphate-reducing bacteria operate an anaerobic route for cobalamin biosynthesis, where cobalt is inserted into sirohydrochlorin by a class II ATP-independent chelatase called CbiK (Brindley *et al*., 2003; Frank *et al*., 2005). Amongst the cobaltochelatases, $\mathrm{CbiX^S}$ is the structurally simplest enzyme consisting of a homodimer with approximately 130 amino acid residues per subunit. These enzymes are generally found in Archaea, and have been considered a primordial form of cobaltochelatases that by gene duplication and fusion events evolved to the bilobal class II ATP-independent cobaltochelatases (Brindley *et al*., 2003). *Desulfovibrio vulgaris* Hildenborough is one of the best studied sulphate-reducing organisms that synthesises haem *b* via an alternative route that uses sirohaem as an intermediate (Bali *et al*., 2011). *D. vulgaris* expresses two versions of CbiK which were named $\mathrm{CbiK^C}$ and $\mathrm{CbiK^P}$, reflecting their predicted localisation in the cytoplasm and periplasm cellular compartments, respectively (Lobo *et al*., 2008). Although both *D. vulgaris* $\mathrm{CbiK^C}$ and $\mathrm{CbiK^P}$ are active as cobaltochelatases, CbiK^P is unique among the chelatases in that it purifies with a bound haem group, with a stoichiometry of two molecules per tetramer (Lobo *et al*., 2008; Romão *et al*., 2011).

A truncated protein lacking the signal peptide that was named Δ28CbiK^P contained no heme but kept the cobalt and iron sirohydrochlorin chelatase activity (Lobo *et al*., 2008).

The unusual structural features of *D. vulgaris* CbiK^P led us to investigate, by site-directed mutagenesis, the role of several residues associated with catalytic activity, oligomerisation and prosthetic group binding. Additionally, the structure of the *D. vulgaris* CbiK^P was revisited through the analysis of a crystallographic structure of CbiK^P obtained after incubation with cobalt and sirohydrochlorin. A study of the haem binding properties of wild type and mutant *D. vulgaris* CbiK^P has revealed that two specific histidine residues are associated with the evolution of a new function for this protein.

Results

In this research we have constructed 12 site-directed mutated *D. vulgaris* CbiK^P variants, namely seven single, three double and one triple amino acid substitutions. The residues that were selected for mutagenesis had been proposed to modulate either chelatase activity or haem binding to *D. vulgaris* CbiK^P (Fig. 1). The selected amino acid residues were replaced by leucine residues, and the mutated and wild type proteins were produced and purified. All proteins were shown to be stable and were analysed for their chelatase activity, oligomeric form and their ability to bind haem. The crystallographic structure of *D. vulgaris* CbiK^P after incubation with cobalt and sirohydrochlorin was also analysed.

CbiK^P chelatase activity

Our previous studies had shown that *D. vulgaris* CbiK^P inserts cobalt into sirohydrochlorin, and the structural analysis suggested that the amino acid residues His154, Glu184 and His216 are involved in metal ligation (Lobo *et al.*, 2008; Romão *et al.*, 2011). Three individual site-directed mutants of CbiK^P, namely H154L, E184L and H216L were constructed and their cobaltochelatase activity was evaluated. Additionally, proteins carrying double mutations such as H154L/E184L, H154L/H216L and E184L/H216L and the triple mutation H154L/E184L/H216L were also produced and analysed.

The wild type CbiK^P exhibited a cobaltochelatase activity of 47 nmol.min^{-1}.mg^{-1} (Fig. 2), and the individual replacement of histidine residues 154 and 216 generated proteins with negligible activity and, not surprisingly, the double mutated H154L/H216L protein was also inactive.

Unexpectedly, substitution of Glu184 by a leucine gave rise to a protein with a cobaltochelatase activity approximately three times higher than that of the wild type. However, the combination of the E184L mutation with H216L, in the E184L/H216L mutant protein, strongly lowered activity, while the E184L/H154L and H154L/E184L/H216L mutant proteins had no cobaltochelatase activity.

To analyse the iron chelatase activity of *D. vulgaris* E184L-CbiK^{P}, the same mutations were introduced in the cytoplasmic (truncated) version of *D. vulgaris* CbiK^{P} (Δ28CbiK^{P}), which was previously shown in complementation assays to insert iron into sirohydrochlorin *in vivo* (Lobo *et al*., 2008). Similar assays were carried out by transforming plasmids expressing the wild type and mutated proteins into *E. coli* 302Δa pCIQ-SirCCobA, a cysteine auxotrophic strain that produces sirohydrochlorin but not sirohaem (Brindley *et al*., 2003; Raux *et al*., 2003). As expected, wild type and all mutated proteins grew well in minimal medium supplemented with cysteine (positive controls, Table 1 and Fig. S1 of Supplemental Material). Like Δ28CbiK^{P}, the E184L mutant protein supported growth in cysteine-deficient minimal medium, as did all other single mutated proteins (Table 1; Fig. S1 of Supplemental Material). In contrast, the double H154L/H216L and H154L/E184L and triple H154L/H216L/E184L Δ28CbiK^{P} mutant proteins were unable to rescue the *E. coli* 302Δa growth deficient phenotype.

CbiK^{P} is a periplasmic protein

Cobaltochelatase enzymes are associated with cobalamin biosynthesis which occurs in the cytoplasm. The presence of a transit peptide on the N-terminal sequence of CbiK^{P} suggests that this enzyme locates outside the cytoplasm. To assess whether the 28 amino acid signal sequence exports CbiK^{P}, the cellular localisation of the *D. vulgaris* CbiK^{P} was investigated and compared with that of the truncated Δ28CbiK^{P} form. For this purpose, CbiK^{P} and Δ28CbiK^{P} were at the C-terminal end fused to GFP and sfGFP, respectively and their localisation was visualized by fluorescence microscopy. These studies were performed using *E. coli* as the host since no fluorescent probes are available for *D. vulgaris*.

Expression of GFP-Δ28CbiK^P yielded cells whose cytoplasm exhibited a bright homogeneous fluorescence (Fig. 3A). Cells expressing the GFP-CbiK^P fusion presented a peripheral fluorescence, with a polar protein accumulation that increased with time (Fig. 3B). These results show that CbiK^P is exported to the periplasm by means of its 28 amino acid N-terminal signal peptide.

Haem binding properties of *D. vulgaris* CbiK^P

Biochemical characterisation of the *D. vulgaris* CbiK^P, produced in E. coli grown in the presence of added haem precursor, showed that the protein contains two haem *b* groups per tetramer (Lobo *et al*., 2008). From the crystallographic structure it was concluded that His96 is located within a coordination distance of *ca.* 2.0 Å from the haem iron atom, indicating that it is the haem axial ligand (Romão *et al*., 2011). Replacement of His96 by leucine, indeed, does appear to prevent haem binding as H96L does not purify with any bound haem (Table 1). Furthermore, the cobaltochelatase activity of H96L-CbiK^P increased by ~2-fold when compared with the wild type (Fig. 2), and H96L-Δ28CbiK^P retained the ability to complement the *E. coli* 302Δa pCIQ-SirCCobA strain (Table 1). These results confirm that haem binding and chelation are independent properties of the protein.

The *D. vulgaris* CbiK^P wild type structure shows that each haem is located between two monomers, consistent with the haem⁄dimer ratio of 1:1 (Lobo *et al*., 2008; Romão *et al*., 2011). Therefore, we examined the role of the haem ligand His96 in the oligomerisation state of the protein. The results in Table 1 show that mutation of this residue does not interfere with the quaternary structure of the enzyme as H96L-CbiK^P retained its tetrameric form.

Since amino acid residues Arg54 and Glu76 are also located at the dimer interface and within hydrogen bonding distance with respect to each other (Fig. 1, Table 2), their contribution to the oligomerisation state and enzyme activity of CbiK^P was analysed. Substitution of Arg54 and

Glu76 by leucine led to the dismantling of the tetrameric form of CbiK^P but did not affect the 1:1 haem/dimer ratio (Table 1). CbiK^P harbouring either the Arg54L or Glu76L substitutions exhibited cobaltochelatase activities in the same order of magnitude as wild type enzyme, with Arg54L having a slightly lower activity (Fig. 2); nonetheless, both mutated proteins retained the ability to complement the iron chelatase deficient *E. coli* 302Δa pCIQ-SirCCobA strain (Table 1).

The possibility that *D. vulgaris* CbiK^P could bind more haem was also investigated. For this purpose, haem binding to isolated CbiK^P was followed by UV/vis spectroscopy by adding exogenous haemin and following changes in the spectra (Fig. 4). Up to about the addition of 2 equivalents of haemin per CbiK^P the spectra showed an increase in absorbance at 412 and 540 nm, indicating the binding of haem, possibly, in a low-spin hexa-coordinated form. Indeed, no bands around 640 nm, typical of high-spin haems, were observed after subtracting the spectral contribution from free haemin. Spectral features of free haem become more evident as more haemin is added to the protein.A plot of the change in absorbance at 412 nm, after subtracting the spectrum of the native protein (which already harbours two haems *b*) from the remaining spectra, demonstrates that CbiK^P accommodates two extra haem groups (Fig. 4), as the changes in absorbance level out after the addition of approximately two equivalents of haemin. Hence, CbiK^P is able to bind a total of 4 haems per tetramer, all of which remain bound after the passage of CbiK^P through a desalting column. A global haem association constant was determined to be 1 µM.

The structure of *D. vulgaris* CbiK^P suggests that His103 could be involved in the ligation of the other two haem moieties (Fig. 1). Thus, a protein in which His103 was replaced by leucine was constructed and analysed. H103L-CbiK^P was purified as a mixture of dimers and tetramers (Table 1), and exhibited slightly lower cobalt chelatase activity (Fig. 2). More importantly, the

H103L-CbiK^P tetramer variant was found to bind less haem than the wild type protein, apart from the two haems that are ligated to the protein through His96.

How widespread are CbiK^P proteins in bacterial systems

D. gigas contains a CbiK^P homologue and a related CbiK protein is also found as an outer membrane haemin-binding protein potentially involved in iron assimilation (Dashper et al J Bacteriol. 2000 Nov;182(22):6456-62.). As a number of CbiK proteins are clearly targeted to either the periplasm or the outer membrane, we have analysed how widespread extracellular CbiK proteins are within biological systems, by searching protein databases for CbiK proteins with an N-terminal extension that could represent a signal peptide. Such an analysis revealed the presence of a broad range of related CbiK proteins, from a diverse range of bacteria, that appear to be targeted outside of the bacterium. These included species within bacteriodes, *Clostridia*, *Desulfo-, Eubacterium, Parabacteriodes, Porphyromoas, Roseburia* and *Veillonella*. It is not possible to predict whether these proteins are able to bind haem but the presence of duplicate copies of CbiK, with and without a signaling peptide all suggests that the roles of CbiK potentially as a haem transporter is much more widespread than previously thought.

Revisiting *D. vulgaris* CbiK^P structure

In spite of several attempts to solve the structure of the *D. vulgaris* CbiK^P loaded with 4 haems per tetramer, electron density for the two additional haems could not be observed, which is most probably due to the high lability of the haems that may bind in the central cavity of the tetramer. Crystals, however, were obtained by incubation of CbiK^P with its substrates cobalt and sirohydrochlorin (herein named *D. vulgaris* CbiK^P (CoSHC); PDB ID code 2xvy). This structure was determined to a higher resolution (1.7 Å) (Table S3 of Supplemental Material) than the previous. The crystallographic data reveals new features in the orientation of His103 and the haem propionate side chains, details that are discussed below.

The monomer organisation of CbiK^P (CoSHC) is similar to that observed with the as-isolated CbiK^P (PDB ID code 2xvx) and that of CbiK^P upon incubation with cobalt, which was previously designated CbiK^P(Co) (PDB ID code 2xvz) (Romão *et al.*, 2011). The three structures display a root mean square (r.m.s.) deviation between equivalent C^α atoms of only 0.20 Å. In all structures, the monomer presents a α/β topology consisting of two domains, domain A (residues 6-117) and domain B (residues 125-269), which are connected by a loop (residues 118-124) (Fig. 5). Domain A is composed of one four-stranded parallel β-sheet and four helices (α1-α4), with the haem ligand residue His96 located in the α4 helix. The strands of the β-sheet are organized as β2 - β1 - β3 - β4 with a twist of *ca.* 65 º, and the four α-helices surround the β-sheet core. Domain B is formed by a four-stranded β-sheet, arranged in a way similar to that of domain A (β6 - β5 - β7 - β8), that is surrounded by five helices (α5, α6, α7, α8 and α9) and harbours the residues His154, Glu184 and His216 that coordinate the cobalt ion. The α10 helix is positioned at the C-terminal, overlapping the two domains and is located near helices α1 and α4 of domain A (Fig. 5A-C).

The cobalt in domain B is coordinated by His154, Glu184, His216, two water molecules and a peroxide molecule (Fig. 5D). His154 does not coordinate to cobalt by its NE2 atom but instead by the NE2-CE2 edge, an unusual orientation for a metal binding histidine, also previously observed for CbiK^P(Co) (Romão *et al.*, 2011).

In the tetramer of all available *D. vulgaris* CbiK^P structures, the monomeric porphyrin binding clefts face outwards, and are therefore accessible to the solvent (Fig. 1). Each monomer interacts with two others through hydrogen bonds that likely contribute to the stabilisation of the tetramer. There are only 28 inter-subunit hydrogen bonds out of a total of 1064 found within the tetramer. These hydrogen bonds can be further divided into groups involving subunits AB and CD, and subunits AD and BC (Table 2). Additionally, the *D. vulgaris* CbiK^P tetramer harbours an open cavity at the centre of the tetramer, with dimensions *ca.* 16 × 15 × 17 Å, where the haems are located (Fig. 6).

Although no electron density for sirohydrochlorin could be modelled in the porphyrin binding cleft in the *D. vulgaris* CbiK^P(CoSHC), as was also observed for the structures of *Salmonella enterica* and *Archaeoglobus fulgidus* proteins (PDB ID codes 2xwp and 2xwq, respectively) (Romão *et al*., 2011), the structure did nonetheless show significant changes in the central region of the tetramer. These changes were most prominent in the conformation of His103 and of the haem propionate side chains (Fig. 6C). In the as-isolated CbiK^P, the side-chain of His103 is hydrogen bonded with one of the haem propionates (D), at a distance of 2.5 Å from His103(ND1), while the other haem propionate (C) is hydrogen bonded at a distance of 2.9 Å to His158 (NE2) from the symmetry related monomer (Fig. 6A). In CbiK^P(Co), His103 faces the centre of the tetramer, partially blocking it, and the distance of His158 (NE2) to the haem propionate (C) is 4.4 Å (Fig. 6A-C). In the CbiK^P(CoSHC) structure, His158 (NE2) is 3.8 Å from the haem propionate (C), while His103 retains a similar position as that in the as-isolated protein structure. Hence, in all structures of CbiK^P the ring C propionate is in an identical position which may hydrogen bond to the NE2 atom of His158, while the ring D propionate undergoes a conformational change, assuming a bent geometry (Fig. 6C).

Furthermore, in the crystal structure CbiK^P(CoSHC) the central region of the tetramer, including His103, exhibited a $2|F_o|-|F_c|$ and $|F_o|-|F_c|$ electron density, which although could not be modelled, may be due to the binding of a tetrapyrrole molecule, suggesting that this region may accommodate the tetrapyrrole substrate.

Discussion

In this work, His154 and His216 are shown to mediate the metal chelatase activity of *D. vulgaris* CbiK^{P}. Moreover, complementation studies indicate that His154 is the most important residue for iron chelation into sirohydrochlorin, since the variants H154L/H216L, H154L/E184L and H154L/E184L/H216L-CbiK^{P} did not overcome the cysteine auxotrophy of the *E. coli* 302Δa strain while the E184L/H216L-Δ28CbiK^{P} mutant did. A reason why His154 is so crucial for the activity of *D. vulgaris* CbiK^{P} may be due to the fact that it is the only residue that appears to change conformation upon the binding of cobalt; furthermore, it is interesting to note that His154 coordination to cobalt is through the NE2-CE2 edge, instead of the NE2 lone pair as occurs with His216. This result agrees with previous data of *S. enterica* CbiK in which the mutation of His207 (His216 in *D. vulgaris* CbiK^{P}) led to a protein that is still able to complement the 302Δa strain, whereas the His145 mutation (His154 in *D. vulgaris* CbiK^{P}) exhibited a very poor ability to rescue the *E. coli* 302Δa phenotype (Schubert *et al*., 1999). The key role of these residues in CbiK^{P} is consistent with their sequence conservation in the ATP-independent type II chelatase family that includes *D. vulgaris* CbiK^{C} (Table S2 of the Supplemental Material) (Brindley *et al*., 2003).

Although the *D. vulgaris* CbiK^{P}(Co) and CbiK^{P}(CoSHC) structures showed that the negatively charged Glu184 residue binds the metal (Romão *et al*., 2011), in this work the mutation of Glu184 to leucine did not lower the enzyme activity but instead increased it. The higher activity of the Glu184L mutant may be due to a weakening of the Co(II) binding, facilitating its insertion into the sirohydrochlorin moiety. Interestingly, Glu184 is not conserved in *D. vulgaris* CbiK^{C} where it is replaced by a glycine (Table S1 of the Supplemental Material), and in *S. enterica* CbiK the binding of sirohydrochlorin causes a shift in the position of the equivalent glycine (Romão *et al*., 2011).

The *D. vulgaris* CbiK^{P} quaternary organisation is unique with the binding of two haem molecules in-between two monomers (Fig. 1). The tetrameric form of *D. vulgaris* CbiK^{P} is shown to be maintained by at least two residues, namely Arg54 and Glu76 that form hydrogen bonds between subunits AB and CD (Fig. 1 and Table 2). Moreover, the conversion of the tetrameric to a dimeric form caused by mutation of Arg54 and Glu76 does not modify the haem binding ability of the protein. While the Glu76L mutation does not alter the cobaltochelatase activity of *D. vulgaris* CbiK, the Arg54 mutation lowered this activity. Superimposition of the structures of *Salmonella enterica* CbiK-metallated SHC complex and *D. vulgaris* CbiK^{P} (CoSHC) shows that in this last structure a glycerol and two sulphate molecules occupy positions similar to that of the propionates and acetates of sirohydrochlorin in *S. enterica* CbiK: propionate A and acetate (C), and propionate (D), respectively (Fig. 7). In particular, Arg54 (atom NH1) is 3.1 Å from the carbonyl oxygen of Thr49, and the latter is 3.3 Å from the glycerol molecule (O1), which is about the same distance of the propionate A from the carbonyl of Thr49 (Fig. 7). Therefore, mutation of Arg54 to leucine could disrupt the hydrogen bonding network that may be required for either porphyrin stabilisation or distortion.

Mutation of His96 generated a CbiK^{P} protein that is unable to bind two haems per tetramer, although it retains its quaternary structure. Moreover, the cobaltochelatase activity of the CbiK^{P} H96L mutant was found to be higher, most probably due to the removal of steric hindrance caused by the presence of the haem groups. Furthermore, the ND1 of His96 is approximately 3 Å from the carbonyls of Pro91 and Gly92, which are themselves close to the acetate (D) group of the sirohydrochlorin, and could affect the binding of the tetrapyrrole substrate (Fig. 7).

D. vulgaris CbiK^{P}(CoSHC) exhibited structural differences in the central region of the tetramer when compared to the metal free and cobalt loaded *D. vulgaris* CbiK^{P} structures, mainly the conformation of the His103 side chain and haem propionate groups (Fig. 6). In CbiK^{P}(CoSHC), the side chain of His103 points towards the haem groups as also observed with the as-isolated

CbiK^P, but the haem propionate attached to ring D is found in a bent conformation similar to that recorded in CbiK^P(Co) (Fig. 6A-C). Moreover, the bent conformation of the propionates in CbiK^P(CoSHC) move them away from His103 (Fig. 6A and C), thereby hindering the formation of the hydrogen bonds between the His103 side-chain and one of the carboxylate oxygens from the haem propionate group. Hence, both His103 and the haem propionate side chains appear to respond to the presence of the substrate. This new conformation may underpin a regulatory function of the central cavity, to allow it to control the entry or exiting of haem.

The titration of CbiK^P with haemin showed that the protein is able to bind two extra haems, and the site-directed mutagenesis studies suggest that these haems interact with His103. The two extra haem groups could fit into the central cavity of the tetramer if the two existing haem groups take up the hypothetical conformations depicted in Fig. 8. The presence in CbiK^P(CoSHC) of a sulphate molecule close to residue His103, and to the haem propionates, suggests that this position would be where the propionates of the extra haems could be stabilised. Based on this, and on the different conformations adopted by His103 in the presence of metal or porphyrin, a structural model in which the two extra haems would be coordinated by His103 has been devised, in agreement with the site-directed mutagenesis data. In this model, the proximity of Glu100 raises the possibility that this residue acts as the second axial ligand (Fig. 8A). However, carboxylates (Asp, Glu) are uncommon haem ligands with the *D. vulgaris* cytochrome *c* nitrite reductase NrfHA (Pdb 2j7a) exemplifying a rare case of an aspartate being located close to the haem iron (Rodrigues *et al.*, 2006). Moreover, a His-Glu axial coordination would most probably lead to a high-spin ferric haem, which does not agree with the observed spectroscopic data. Alternatively, the haems could be coordinated by His103 from two different subunits (Fig. 8B), which would generate a low-spin ferric haem and be consistent with the experimental results. Of course, the possibility exists that the binding of the extra haems causes structural rearrangements within the central cavity. In the available structures, the haem binding His103 residue exhibits different conformations suggesting a high flexibility, which may facilitate the release of the haem upon interaction with other proteins (e.g. the haem transporters encoded in the operon in which the CbiK gene is located).

A bioinformatics analysis of the UniProt database after a BLAST search for CbiK and a signal peptide has revealed the presence of CbiK proteins with N-terminal signal peptides in several bacteria, demonstrating that CbiK^P in *D. gigas* is far from an isolated example. The proteins identified within the database search include a range of different sized CbiK proteins. Although the localisation and association of *cbiK* with other genes within the various genomes was not investigated, it is tempting to suggest that these other CbiK proteins are also associated with haem transport. Further research is required to understand the exact role played by CbiK$_P$ and its evolution from a cytoplasmic enzyme associated with cobalamin biosynthesis into a component of an extracellular transport system.

In conclusion, we have shown that the *D. vulgaris* CbiK^P is located in the periplasm, which *a priori* precludes it from playing a role in the cytoplasmic synthesis of tetrapyrroles. As a tetramer, the protein is able to bind haems most likely in the central cavity, suggesting that CbiK^P could act as a haem transporter, in support of the observation that *cbiK*^{P} is located within a cluster that encodes proteins involved in transport processes (lobo et al 2009). Interestingly, in *Yersinia pestis*, the periplasmic haem-containing HmuT protein was also shown to bind two stacked haems in a central binding cleft as part of a haem uptake and delivery system (Mattle *et al*., 2010).

In *Desulfovibrio* and sulphate-reducing bacteria in general, sirohydrochlorin plays an important role not only as a protein cofactor but also as a key intermediate in the haem and vitamin B_{12} biosynthetic pathways (Bali *et al*., 2011; Lobo *et al*., 2012). As mentioned in the introduction, *D. vulgaris* expresses cytoplasmic and periplasmic versions of CbiK that have very similar amino acid sequences. In particular, CbiK^{C} contains the two histidine residues equivalent to His154 and His216 in CbiK^{P}, which were shown here to be essential for the sirohydrochlorin chelatase activity. The two CbiKs therefore exhibit cobaltochelatase activity; however, as tetrapyrrole biosynthesis occurs in the cytosol, CbiK^{C} is the most plausible candidate for the sirohydrochlorin cobaltochelatase in *Desulfovibrio* spp. (Lobo *et al.*, 2008). Reinforcing this hypothesis, our results have shown that the presence of His96 and His103 in CbiK^{P} are associated with an additional haem-binding biological function, as site-directed mutagenesis of these two residues yielded proteins that no longer have the ability to bind haem. Therefore, although still a functional chelatase, the presence of His96 and His103 in CbiK^{P}, which are absent from the sequences of *D. vulgaris* CbiK^{C} and all other bacterial cobaltochelatases (Table S1 of the Supplemental Material), have prompted the evolution of a gain of function relating to haem transport. To the best of our knowledge, this is the first example of how a common ancestral protein has evolved a new functional role through acquisition of two histidine residues. Although CbiK^{P} constitutes one of the first recognized haem chaperons in *D. vulgaris*, questions remain as to its physiological partners and how it participates in haem homeostasis in this sulphate-reducing bacterium.

Experimental Procedures

Site-directed mutagenesis

Twelve *D. vulgaris* CbiK^{P} site-directed mutated proteins were constructed (seven single, three double and one triple mutated) with the following substitutions: H154L, H216L, E184L, R54L, E76L, H103L, H96L, H216L/H154L, H216L/E184L, H154L/E184L and H216L/H154L/E184L. Site-directed mutagenesis was carried out with the QuickChange II Site-Directed Mutagenesis Kit (Agilent Technologies). For this purpose, primers listed in Table S2 (Supplemental Material) and the following plasmids were used: pET28a(+)-*cbiK^{P}*, which contains the DNA sequence encoding the wild type CbiK^{P}, and pET28a(+)-Δ28-*cbiK*^{P} expressing a truncated form of CbiK^{P} that lacks the periplasmic targeting signal sequence (Lobo *et al.*, 2008). Double and triple mutants were obtained as follows: plasmid pET28a(+)-H216L-*cbiK*^{P} served as template to construct pET28a(+)-H216LH154L-*cbiK*^{P} and pET28a(+)-H126LE184L, and pET28a(+)-H154L to produce pET28a(+)-H154LE184L-*cbiK*^{P}*.* Plasmid pET28a(+)-H216LH154LE184L-*cbiK*^{P} that expresses the triple mutated CbiK^{P} was obtained using pET28a(+)-H216LH154L as template. In all cases*,* DNA sequencing confirmed that the correct amino acid changes were introduced and no undesirable mutations were present.

Protein expression, purification and characterisation

Expression and purification of *D. vulgaris* CbiK^{P} wild type and mutant variants was done as follows: each recombinant plasmid was transformed, separately, into *Escherichia coli* BL21Gold(DE3) (Stratagene) and cells were grown, at 30 °C, in Luria-Bertani (LB) medium containing kanamycin (30 µg.ml^{-1}) until reaching an optical density at 600 nm (OD$_{600}$) of 0.3. Expression of the proteins was achieved by addition of 200 µM isopropyl-D-thiogalactopyranoside (IPTG), 50 µM 5-aminolevulinic acid (ALA) and 100 µM $FeSO_4$. After induction with IPTG, cells were grown overnight, at 20 °C, harvested by centrifugation (e.g. 11k *xg*, 10 min, 4 °C), resuspended in 20 mM Tris-HCl pH 7.5 (buffer A), and disrupted in a French Press (ThermoScientific). The soluble fraction was separated by ultracentrifugation at 160000 *×g*, for 2 h. All centrifugation and purification steps were performed at 4 °C. The resultant soluble fraction was applied onto a Chelating Sepharose fast flow column (25 mL) (GE, Healthcare), previously charged with $NiSO_4$ in buffer A and containing 400 mM NaCl. A linear gradient up to 250 mM imidazole was applied and the protein was eluted approximately at 80 mM imidazole. Protein fractions containing CbiK^{P} were dialysed against buffer A and loaded onto a Q-Sepharose High-Performance column (25 mL) (GE, Healthcare), previously equilibrated in buffer A and purified by means of a linear gradient up to 400 mM NaCl; the protein was collected at ~200 mM NaCl. Fractions containing CbiK^P wild type and mutated proteins were dialyzed against 50 mM Tris-HCl pH 8 (buffer B).

The purity of the proteins was confirmed by SDS-PAGE, protein concentration was evaluated by the bicinchoninic acid method (Smith *et al.*, 1985) using Sigma protein standards, and haem content was determined by the haemochromopyridine method (Berry and Trumpower, 1987). The protein molecular mass was determined by gel filtration in a Superdex 200 column (GE, Healthcare) using 50 mM sodium-phosphate pH 7, 150 mM NaCl buffer and molecular mass standards.

Activity assays

Sirohydrochlorin was produced as previously described (Lobo *et al.*, 2008). Plasmid pETcoco-2ABCDC harbouring the genes encoding *Methanosarcina barkeri* uroporphyrinogen III methyltransferase (CobA), *Methanothermobacter thermautotrophicus* porphobilinogen synthase (HemB) and precorrin-2 dehydrogenase (SirC)*, B. megaterium* porphobilinogen deaminase (HemC) and uroporphyrinogen III synthase (HemD) (Frank *et al.*, 2007), was introduced into *E. coli* BL21Star(DE3)pLysS (Invitrogen). Cells were grown in LB medium containing ampicillin (50 µg.ml^{-1}), chloramphenicol (34 µg.ml^{-1}) and 0.2 % (w/v) glucose. Overnight grown cell cultures were used to inoculate fresh antibiotic supplemented LB medium and further cultured, at 37 °C, to an OD_{600} of ~0.5. At this stage, the medium was supplemented with 0.02 % (w/v) L-arabinose and after two hours 400 µM IPTG was added. Following an overnight growth, at 24 °C, cells were collected by centrifugation and resuspended in buffer B containing 100 mM NaCl. Lysates were obtained by passage of cells through a French Press and debris was removed by centrifugation (~39000 *xg*, 30 min, 4 °C). Lysates were exposed to an argon atmosphere for 20-30 min, transferred to the anaerobic chamber (Coy model A-2463 filled with a gas mixture of 95 % argon and 5 % hydrogen) and used in a reaction mixture to produce sirohydrochlorin. The reaction mixture was composed of 2 ml cell lysate, 2 mg S-adenosyl-L-methionine, 1 mg aminolaevulinic acid and 1 mg NAD^+, in a total volume of 6 ml buffer B with 100 mM NaCl, and the reaction mixture pH was adjusted to 8 by addition of 1 M KOH. After overnight incubation at room temperature, the reaction mixture was filtered (0.22 µm PALL syringe filter) and passed through a DEAE-resin (Sigma) previously equilibrated in buffer B containing 100 mM NaCl. The column was washed with buffer B with 250 mM NaCl and sirohydrochlorin was eluted with buffer B with 1 M NaCl.

The chelatase activity was measured under anaerobic conditions in a Shimadzu UV-1800 spectrophotometer. The assays were performed in buffer B supplemented with 100 mM NaCl, using 4.2 µM sirohydrochlorin, 20 µM Co^{2+} and several concentrations of *D. vulgaris* $CbiK^P$ wild type and mutated proteins. Reactions were monitored following the absorbance decrease of the 376 nm band (extinction coefficient 2.4×10^5 $M^{-1}cm^{-1}$), which corresponds to the consumption of sirohydrochlorin, and the build-up of the 414 nm band associated with the cobalt-sirohydrochlorin formation (Schubert *et al.*, 2002).

Complementation assays

The genes encoding *D. vulgaris* truncated $CbiK^P$ (Δ28*cbiK*^{P}) wild type and site-directed mutants were cloned in pETac plasmid (Raux *et al.*, 1997). The resulting plasmids were transformed into *E. coli* 302Δa strain that lacks the multifunctional sirohaem synthase CysG and contains plasmid pCIQ-*sirCcobA* for expression of *M. thermoautotrophicus sirC* and *Pseudomonas denitrificans cobA* genes (Brindley *et al.*, 2003; Raux *et al.*, 2003). Selection of the recombinant plasmids harbouring $CbiK^P$ mutant in *E. coli* 302Δa strain was done on LB plates supplemented with ampicillin (100 µg.ml^{-1}) and chloramphenicol (35 µg.ml^{-1}). *E. coli* 302Δa was also transformed with plasmid pKK223.2-*cysG*, which expresses the *E. coli* CysG, and with the empty pETac plasmid, to serve as positive and negative controls, respectively. All strains were cultured on minimal medium plates and their growth was evaluated in the absence and presence of cysteine (0.05 mg.ml^{-1}).

Haem titrations

D. vulgaris CbiK^{P} wild type and mutant H103L-CbiK^{P} proteins were diluted in buffer B to a final concentration of 1 µM in a final volume of 500 µl. A stock solution of haemin (Frontier Scientific) was prepared in 0.1 M NaOH and added at increasing concentrations (0.2 - 4 µM) to the protein solution, at room temperature, or to a buffer solution. The UV-visible spectra after each addition were measured in a Shimadzu UV-1700 spectrophotometer, at room temperature. The binding stoichiometry of haemin was determined by plotting the absorbance values at 412 nm of the protein solution, after subtraction of the initial CbiK^{P} spectrum and of the corresponding free haemin amountin the buffer, as a function of the haemin/protein concentration ratio. The data was fitted to an equation with a single binding constant, as there was not sufficient data to discriminate between two distinct binding constants.

D. vulgaris CbiK^{P} cellular localisation

For the determination of the cellular localisation of *D. vulgaris* CbiK^{P} in *E. coli*, two plasmids were constructed: pET23b(+)-Δ28*cbiK^{P}*-GFP and pET23b(+)-*cbiK^{P}*-sfGFP that express a green fluorescent protein (GFP) fusion of Δ28 CbiK^{P} and a superfold green fluorescent protein (sfGFP) fusion of CbiK^{P}, respectively. The sfGFP probe is a GFP variant that has been shown to retain its activity when exported to the periplasm by the Sec pathway (Pedelacq *et al.*, 2006, Aronson *et al.*, 2011). pET23b(+)-Δ28*cbiK^{P}-gfp* was obtained by sub-cloning the Δ28*cbiK^{P}* gene into pET23b(+)-*gfp*. The *sfgfp* gene was amplified by PCR from plasmid BBa_I1746908 (http://partsregistry.org/Part:BBa_I746908) and using primers sfGFPfw (5´-

AAGCTAGCAAAGAGCTCATGCGTAAAGGC-3´) and sfGFPrev (5´-CCTGGCTCTCGAGTCATCATTTGTAC-3´) containing SacI and XhoI restriction sites, respectively, and ligated to yield plasmid pET23b(+)-*cbiK^P-sfgfp*. Plasmids containing *cbiK^P-sfgfp* and Δ28*cbiK^P-sfgfp* gene fusions were introduced into *E. coli* BL21Gold(DE3), and cells were further grown in LB medium supplemented with ampicillin (50 µg.ml^{-1}), at 30°C, to an OD_{600} of ~0.3-0.5. At this stage, 100 µM ITPG was added and cells after grown for 2 h and 3 h were collected by centrifugation (9700 *x g*, 5 min, room temperature), and resuspended in 100 µl phosphate-buffered saline (PBS). Cell suspension (10 µl) was smeared onto a thin layer of 1.7 % agarose mounted in a microscope glass slide. Phase contrast and fluorescent visualisation were done in a Leica DM6000B microscope attached to a Leica EL6000 fluorescence light source and a MicroPoint. Images were obtained using an Andor Inox+ camera and Metamorph microscopy automation and image analysis software (Molecular Devices).

Protein crystallisation, structure determination and refinement

Co-crystallisation of CbiK^P with cobalt and sirohydrochlorin was performed by the hanging drop vapour diffusion method using crystallisation conditions similar to those previously reported (Romão *et al.*, 2011): 1 µl *D. vulgaris* CbiK^P (10 mg.ml^{-1}) in buffer A was mixed with 1.8 µl reservoir solution (100 mM Tris-HCl pH 8.5 and 2 M ammonium sulphate), 0.2 µl cobalt(II) chloride (1 M) and 0.2 µl sirohydrochlorin (2 mM). The crystallisation drops were equilibrated against a 500 µl reservoir. Crystals appeared after 1 day and grew to typical dimensions of 200 x 75 x 75 µm^3. One crystal was harvested and cryo-protected with the reservoir solution supplemented with 20 % glycerol prior to flash-cooling in liquid nitrogen.

An X-ray diffraction dataset to 1.7 Å resolution was collected at the BM14 beamline of the European Synchrotron Radiation Facility. The diffraction images were integrated and scaled with XDS (Kabsch, 1993) (Table S3). Merging of diffraction intensity data and conversion to structure factors was carried out with SCALA and TRUNCATE in the CCP4 program package (Winn *et al.*, 2011).

The structure was solved by molecular replacement with PHASER (McCoy *et al.*, 2007) using as search model the native structure of *D. vulgaris* CbiK^{P} (PDB 2xvx) (Romão *et al.*, 2011), and refined with REFMAC (Murshudov *et al.*, 1997). Model rebuilding was done with COOT (Emsley *et al.*, 2010). The refinement statistics are presented in Table S3 (Supplemental Material). The structure was analysed with PROCHECK and SFCHECK (Vaguine *et al.*, 1999) and all stereochemical parameters were similar to or better than the average values from structures determined at similar resolution. Anisotropic atomic displacement parameters were refined for the cobalt ion.

Coordinates and structure factors of the *D. vulgaris* CbiK^{P}(CoSHC) structure were deposited in the Protein Data Bank in Europe (Velankar *et al.*, 2011) with the accession code 2xvy.

Acknowledgements

We thank the EMBL Grenoble Outstation for providing support for measurements at the ESRF EMBL-CRG BM14 beamline under the European Community – Access to Research Infrastructure Action FP6 Programme "Structuring the European Research Area Specific Program", with contract number RII3-CT-2004-506008. The help and advice given by Martin A. Walsh and Hassan Belrhali (EMBL, Grenoble) during the BM14 data collections is also gratefully acknowledged. We thank Luís Sobral, Mafalda Figueiredo and Sara Sousa for technical contributions. This work was financially supported by: Project LISBOA-01-0145-FEDER-007660 (Microbiologia Molecular, Estrutural e Celular) funded by FEDER funds through COMPETE2020 - Programa Operacional Competitividade e Internacionalização (POCI) and by national funds through Fundação para a Ciência e a Tecnologia, and PTDC/BBB-BQB/5069/2014 project.

C.V. Romão, S.A. Lobo and M.A.M. Videira are recipients of FCT grants SFRH/BPD/94050/2013, SFRH/BPD/63944/2009 and SFRH/BD/19813/2014, respectively.

References

Aronson, D.E., Costantini, L.M., and Snapp, E.L. (2011) Superfolder GFP is fluorescent in oxidizing environments when targeted via the Sec translocon. Traffic **12**: 543-548.

Bali, S., Lawrence, A., Lobo, S.L., M., S.L., Golding, B., Palmer, D. *et al*. (2011) Molecular hijacking of siroheme for the synthesis of heme and d_1 heme. Proc Natl Acad Sci U S A **108**: 18260-18265.

Berry, E.A., and Trumpower, B.L. (1987) Simultaneous determination of hemes *a*, *b*, and *c* from pyridine hemochrome spectra. Anal Biochem **161**: 1-15.

Brindley, A.A., Raux, E., Leech, H.K., Schubert, H.L., and Warren, M.J. (2003) A story of chelatase evolution: identification and characterisation of a small 13-15-kDa "ancestral" cobaltochelatase (CbiX^S) in the archaea. J Biol Chem **278**: 22388-22395.

Diederichs, K., and Karplus, P.A. (1997) Improved R-factors for diffraction data analysis in macromolecular crystallography. Nat Struct Biol **4**: 269-275.

Emsley, P., Lohkamp, B., Scott, W.G., and Cowtan, K. (2010) Features and development of Coot. Acta Crystallogr D Biol Crystallogr **66**: 486-501.

Frank, S., Brindley, A.A., Deery, E., Heathcote, P., Lawrence, A.D., Leech, H.K. *et al*. (2005) Anaerobic synthesis of vitamin B_{12}: characterisation of the early steps in the pathway. Biochem Soc Trans **33**: 811-814.

Frank, S., Deery, E., Brindley, A.A., Leech, H.K., Lawrence, A.D., Heathcote, P. *et al*. (2007) Elucidation of substrate specificity in the cobalamin (vitamin B_{12}) biosynthetic methyltransferases; structure and function of the C20 methyltransferase (CbiL) from *Methanothermobacter thermautotrophicus*. J Biol Chem **282**: 23957-23969.

Kabsch, W. (1993) Automatic processing of rotation diffraction data from crystals of initially unknown symmetry and cell constants. Journal of Applied Crystallography **26**: 795-800.

Krissinel, E., and Henrick, K. (2004) Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. Acta Crystallogr D Biol Crystallogr **60**: 2256-2268.

Lobo, S.A., Warren, M.J., and Saraiva, L.M. (2012) Sulfate-reducing bacteria reveal a new branch of tetrapyrrole metabolism. Adv Microb Physiol **61**: 267-295.

Lobo, S.A., Brindley, A., Warren, M.J., and Saraiva, L.M. (2009) Functional characterization of the early steps of tetrapyrrole biosynthesis and modification in *Desulfovibrio vulgaris* Hildenborough. Biochem J **420**: 317-325.

Lobo, S.A., Brindley, A.A., Romão, C.V., Leech, H.K., Warren, M.J., and Saraiva, L.M. (2008) Two distinct roles for two functional cobaltochelatases (CbiK) in D*esulfovibrio vulgaris* Hildenborough. Biochemistry **47**: 5851-5857.

Mattle, D., Zeltina, A., Woo, J.S., Goetz, B.A., and Locher, K.P. (2010) Two stacked heme molecules in the binding pocket of the periplasmic heme-binding protein HmuT from *Yersinia pestis*. *J Mol Biol* **404**: 220-231.

McCoy, A.J., Grosse-Kunstleve, R.W., Adams, P.D., Winn, M.D., Storoni, L.C., and Read, R.J. (2007) Phaser crystallographic software. J Appl Crystallogr **40**: 658-674.

Murshudov, G.N., Vagin, A.A., and Dodson, E.J. (1997) Refinement of macromolecular structures by the maximum-likelihood method. Acta Crystallogr D Biol Crystallogr **53**: 240-255.

Pedelacq, J.D., Cabantous, S., Tran, T., Terwilliger, T.C., and Waldo, G.S. (2006) Engineering and characterization of a superfolder green fluorescent protein. Nat Biotechnol **24**: 79-88.

Raux, E., Thermes, C., Heathcote, P., Rambach, A., and Warren, M.J. (1997) A role for *Salmonella typhimurium* cbiK in cobalamin (vitamin B_{12}) and siroheme biosynthesis. J Bacteriol **179**: 3202-3212.

Raux, E., Leech, H.K., Beck, R., Schubert, H.L., Santander, P.J., Roessner, C.A. *et al*. (2003) Identification and functional analysis of enzymes required for precorrin-2 dehydrogenation and metal ion insertion in the biosynthesis of sirohaem and cobalamin in *Bacillus megaterium*. Biochem J **370**: 505-516.

Rodrigues, M.L., Oliveira, T.F., Pereira, I.A., and Archer, M. (2006) X-ray structure of the membrane-bound cytochrome c quinol dehydrogenase NrfH reveals novel haem coordination. EMBO J **25**: 5951-5960.

Romão, C.V., Ladakis, D., Lobo, S.A., Carrondo, M.A., Brindley, A.A., Deery, E. *et al*. (2011) Evolution in a family of chelatases facilitated by the introduction of active site asymmetry and protein oligomerization. Proc Natl Acad Sci U S A **108**: 97-102.

Schubert, H.L., Raux, E., Wilson, K.S., and Warren, M.J. (1999) Common chelatase design in the branched tetrapyrrole pathways of heme and anaerobic cobalamin synthesis. Biochemistry **38**: 10660-10669.

Schubert, H.L., Raux, E., Brindley, A.A., Leech, H.K., Wilson, K.S., Hill, C.P., and Warren, M.J. (2002) The structure of *Saccharomyces cerevisiae* Met8p, a bifunctional dehydrogenase and ferrochelatase. Embo J **21**: 2068-2075.

Smith, P.K., Krohn, R.I., Hermanson, G.T., Mallia, A.K., Gartner, F.H., Provenzano, M.D. *et al*. (1985) Measurement of protein using bicinchoninic acid. Anal Biochem **150**: 76-85.

Vaguine, A.A., Richelle, J., and Wodak, S.J. (1999) SFCHECK: a unified set of procedures for evaluating the quality of macromolecular structure-factor data and their agreement with the atomic model. Acta Crystallogr D Biol Crystallogr **55**: 191-205.

Velankar, S., Alhroub, Y., Alili, A., Best, C., Boutselakis, H.C., Caboche, S. *et al*. (2011) PDBe: Protein Data Bank in Europe. Nucleic Acids Res **39**: D402-410.

Winn, M.D., Ballard, C.C., Cowtan, K.D., Dodson, E.J., Emsley, P., Evans, P.R. *et al*. (2011) Overview of the CCP4 suite and current developments. Acta Crystallogr D Biol Crystallogr **67**: 235-242.

Table 1. Properties of *D. vulgaris* CbiK^{P} mutants

	Biochemical characterisation		Complementation assays	
CbiK^{P}	**Oligomeric form**	**Haem/Oligomer**	**Growth MM**	**Growth MM+Cys**
Wild type	Tetramer	2	+	+
H154L	Tetramer	2	+	+
E184L	Tetramer	2	+	+
H216L	Tetramer	2	+	+
E184L/H216L	Tetramer	2	+	+
H154L/H216L	Tetramer	2	-	+
H154L/E184L	Tetramer	2	-	+
H154L/E184/H216L	Tetramer	2	-	+
H96L	Tetramer	0	+	+
R54L	Dimer	1	+	+
E76L	Dimer	1	+	+
H103L	Dimer/Tetramer	1/2	+	+

Proteins were analysed regarding oligomeric form, number of haem groups, cobalt and iron chelatase activity. Complementation assays of *E. coli* 302Δa-pCIQ-*sirCcobA* sirohaem ferrochelatase deficient strain on minimal media (MM), in the absence and presence of cysteine (Cys), were done in triplicate. Full growth, reduced growth, and no growth are indicated as (++), (+) and (-), respectively.

Table 2. Hydrogen bonds between monomers pairs of *D. vulgaris* CbiK^P(Co)

Monomer	Residue1	Distance (Å)	Residue2	Monomer
A	Arg54 NH2	3.05	Glu76 OE1	**B**
	Arg54 NE	2.67	Glu76 OE2	
	Arg58 NH1	2.91	Glu76 OE2	
A	Tyr125 OG	2.75	Asp161 OD2	**D**
	Tyr170 OH	2.51	Ser186 OG	
	Ser186 OG	2.51	Tyr170 OH	

Hydrogen bonds between the CbiK^P monomers A and B (equivalent to monomers D and C) and monomers A and D (equivalent to monomers B and C), as depicted in Fig. 1.

Figure Legends

Figure 1. *D. vulgaris* CbiK^P(Co) wild type.

Tetrameric assembly and detail view of the tetramer central region highlighting the amino acids that have been mutated in this work.

Each monomer (A-D) is coloured differently: A - green; B - blue; C - yellow; D - pink. Amino acid residues are represented in sticks and colour coded: carbon in the corresponding colour of the subunit, nitrogen in blue and oxygen in red. Haems *b,* located in-between subunits are represented as sticks, with carbon in white and the remaining atoms depicted as mentioned above. Haem irons and cobalt atoms are shown as black and grey spheres, respectively. Sirohydrochlorin (SHC) binding sites are indicated by dashed arrows.

Figure 2. Cobaltochelatase activity of CbiK^{P} wild type and mutant variants.

Specific activity was measured by following the decrease of the sirohydrochlorin absorbance band at 376 nm of the UV-visible spectrum, as described in Experimental Procedures. Activities were measured in triplicate for two independent batches of purified protein.

Figure 3. Localisation of *D. vulgaris* CbiK^{P}.

Phase contrast (PC) and fluorescent (GFP) visualisation of *E. coli* expressing Δ28CbiK^{P}-GFP (A) and CbiK^{P}-sfGFP (B). Cells expressing the fusion proteins were visualized after 2 h (A and B1) and 3 h (B2) post ITPG induction.

Figure 4. Haem binding assays.

Haemin binding curve of isolated *D. vulgaris* CbiK^{P} wild type (■) and H103L-CbiK^{P} (▲).The binding stoichiometry of haemin was determined by plotting the absorbance values at 412 nm, after subtraction of the initial CbiK^{P} spectrum and the equivalent amount of haemin in the buffer (see Materials and Methods), as a function of the haemin/protein concentration ratio.

Figure 5. Structure of the *D. vulgaris* CbiK^{P}(CoSHC).

A. Cartoon diagram of the *D. vulgaris* CbiK^{P}(CoSHC) monomer coloured in rainbow from blue (N-terminal) to red (C-terminal). Cobalt ligands His154, Glu184 and His216, haem *b* and haem ligand His96 are represented as sticks. Colour code with carbon in white, nitrogen in blue and oxygen in red. Cobalt and iron are drawn as grey and black spheres, respectively.

B. Domain from residues 6 to 117 with His96 and haem *b* represented and coloured as in A.

C. Domain from residues 125 to 269 depicting the cobalt site.

D. Cobalt site coordinated by three amino-acid residues: His154 (distances Co – NE2 is 2.05 Å and Co – CE1 is 2.04 Å), Glu184 (distance Co – NE2 is 2.45 Å) and His216 (distance Co – NE2 is 2.07 Å), and by further two water molecules (distances Co – W1 is 2.19 Å and Co – W2 is 2.07 Å) and a peroxide molecule (distance Co – O1 is 2.05 Å).

Figure 6. Central tetramer cavity of *D. vulgaris* CbiK^{P}.

Detailed view of the central tetramer cavity of the as-isolated CbiK^{P} wild type (CbiK^{P}; PDB code 2xvx), containing cobalt (CbiK^{P}(Co); PDB code 2xvz) and after incubation with cobalt and sirohydrochlorin (CbiK^{P}(CoSHC); PDB code 2xvy, this work). Each monomer is coloured as described in Fig. 1, with haems *b*, and axial iron ligand His96, His103 and His158 represented as sticks. For all structures, the right panel is rotated 90° showing a close-up of the haem between subunits A and D, indicatingthe distances between His158 and the propionate C, and His103 and propionate D. Close-up of the side-chain of His103 and haem propionate groups show that there are significant conformational differences among the three structures.

Figure 7. Superimposition of the *D. vulgaris* CbiK^{P}(CoSHC) with *Salmonella enterica* sirohydrochlorin containing chelatase.

D. vulgaris CbiK^{P}(CoSHC) was superimposed with the *Salmonella enterica* in complex with metalled sirohydrochlorin chelatase (pdb code 2xwp). The sulphate and glycerol molecules observed in the *D. vulgaris* CbiK^{P}(CoSHC) structure are in similar positions to the propionates and acetates groups from the metallated sirohydrochlorin observed in the *S. enterica* CbiK. The atoms are represented with the same colour code as in Fig. 1. The metallated sirohydrochlorin from *S. enterica* CbiK is represented in pink sticks.

Figure 8. Models for *D. vulgaris* CbiK^{P} with the two extra haems in the central cavity of the tetramer.

A. Model 1, the extra haems are coordinated by His103, and Glu100 would be close to the other axial positon. B. Model 2 where the extra haems would be coordinated by the two symmetrically related His103. The position 1 of His103 corresponds to that observed in the *D. vulgaris* CbiK^{P}(CoSHC) structure, whereas the position 2 corresponds to that in the *D. vulgaris* CbiK^{P}(Co) structure.