

# Kent Academic Repository

## Full text document (pdf)

### Citation for published version

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

### DOI

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

### Link to record in KAR

<http://kar.kent.ac.uk/58529/>

### Document Version

Author's Accepted Manuscript

#### Copyright & reuse

Content in the Kent Academic Repository is made available for research purposes. Unless otherwise stated all content is protected by copyright and in the absence of an open licence (eg Creative Commons), permissions for further reuse of content should be sought from the publisher, author or other copyright holder.

#### Versions of research

The version in the Kent Academic Repository may differ from the final published version.

Users are advised to check <http://kar.kent.ac.uk> for the status of the paper. **Users should always cite the published version of record.**

#### Enquiries

For any further enquiries regarding the licence status of this document, please contact:

[researchsupport@kent.ac.uk](mailto:researchsupport@kent.ac.uk)

If you believe this document infringes copyright then please contact the KAR admin team with the take-down information provided at <http://kar.kent.ac.uk/contact.html>

1 ***Desulfovibrio vulgaris* CbiK<sup>P</sup> cobaltochelatase: evolution of a haem binding protein**  
2 **orchestrated by the incorporation of two histidine residues**

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

5

6 <sup>1</sup>*Instituto de Tecnologia Química e Biológica NOVA, Avenida da República (EAN), 2780-157*  
7 *Oeiras, Portugal*

8 <sup>2</sup>*School of Biosciences, University of Kent, Giles Lane, Canterbury, Kent, CT2 7NJ, United*  
9 *Kingdom*

10 <sup>3</sup>*iBET, Instituto de Biologia Experimental e Tecnológica, Apartado 12, 2781-901 Oeiras,*  
11 *Portugal*

12 *\*These authors contribute equally to this work*

13

14 **<sup>+</sup>Corresponding author:**

15 Instituto de Tecnologia Química e Biológica NOVA

16 Av. da República (EAN) 2780-157 Oeiras, Portugal

17 Tel: +351214469328; Fax: +351214433644; e-mail: lst@itqb.unl.pt

18

19 **Running title:** *D. vulgaris* cobaltochelatase CbiK<sup>P</sup> haem binding protein

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

22

23 **Originality-Significance**

24 We show for the first time that a cobaltochelatase of an ecological important organism, such  
25 as the sulphate-reducer *Desulfovibrio vulgaris* cobaltochelatase CbiK<sup>P</sup>, evolved to a haem

26 binding protein by the presence of two histidine residues, which are absent from all other  
27 known bacterial cobaltochelatasases, permitting it to act as a potential haem chaperone or  
28 transporter.

29

### 30 **Abstract**

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

48

49

50

51 **Introduction**

52 Metallated tetrapyrroles support life by acting as prosthetic groups, cofactors and coenzymes  
53 in a wide range of biological processes such as photosynthesis, oxygen transport, respiration  
54 and metabolism. This family of compounds are synthesized via a pathway that is initiated from  
55 5-aminolevulinic acid, which is subsequently transformed into the first macrocyclic  
56 intermediate, uroporphyrinogen III. All the major metal-containing tetrapyrroles, including  
57 cobalamin,, coenzyme F<sub>430</sub>, sirohaem, chlorophyll and haem *b*, are based upon this molecular  
58 template. The different modified tetrapyrroles require specific metal ion chelatases to ensure  
59 the appropriate insertion of the correct metal into the centre of the macrocycle. Sulphate-  
60 reducing bacteria generally make three modified tetrapyrroles, cobamides, sirohaem and  
61 haem *b*. In this class of bacteria, all three of these compounds are derived from  
62 sirohydrochlorin that acts as the substrate for metal insertion (Lobo *et al.*, 2009; Bali *et al.*,  
63 2011).

64 Sulphate-reducing bacteria operate an anaerobic route for cobalamin biosynthesis, where  
65 cobalt is inserted into sirohydrochlorin by a class II ATP-independent chelatase called CbiK  
66 (Brindley *et al.*, 2003; Frank *et al.*, 2005). Amongst the cobaltochelateses, CbiX<sup>S</sup> is the  
67 structurally simplest enzyme consisting of a homodimer with approximately 130 amino acid  
68 residues per subunit. These enzymes are generally found in Archaea, and have been  
69 considered a primordial form of cobaltochelateses that by gene duplication and fusion events  
70 evolved to the bilobal class II ATP-independent cobaltochelateses (Brindley *et al.*, 2003).  
71 *Desulfovibrio vulgaris* Hildenborough is one of the best studied sulphate-reducing organisms  
72 that synthesises haem *b* via an alternative route that uses sirohaem as an intermediate (Bali *et*  
73 *al.*, 2011). *D. vulgaris* expresses two versions of CbiK which were named CbiK<sup>C</sup> and CbiK<sup>P</sup>,  
74 reflecting their predicted localisation in the cytoplasm and periplasm cellular compartments,  
75 respectively (Lobo *et al.*, 2008). Although both *D. vulgaris* CbiK<sup>C</sup> and CbiK<sup>P</sup> are active as

76 cobaltochelatasas, CbiK<sup>P</sup> is unique among the chelatasas in that it purifies with a bound haem  
77 group, with a stoichiometry of two molecules per tetramer (Lobo *et al.*, 2008; Romão *et al.*,  
78 2011).

79 A truncated protein lacking the signal peptide that was named  $\Delta 28\text{CbiK}^{\text{P}}$  contained no heme  
80 but kept the cobalt and iron sirohydrochlorin chelatase activity (Lobo *et al.*, 2008).

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

88

89 **Results**

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

98

99 **CbiK<sup>P</sup> chelatase activity**

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

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

110 Unexpectedly, substitution of Glu184 by a leucine gave rise to a protein with a  
111 cobaltochelate activity approximately three times higher than that of the wild type.  
112 However, the combination of the E184L mutation with H216L, in the E184L/H216L mutant

113 protein, strongly lowered activity, while the E184L/H154L and H154L/E184L/H216L mutant  
114 proteins had no cobaltochelatase activity.

115 To analyse the iron chelatase activity of *D. vulgaris* E184L-CbiK<sup>P</sup>, the same mutations were  
116 introduced in the cytoplasmic (truncated) version of *D. vulgaris* CbiK<sup>P</sup> ( $\Delta$ 28CbiK<sup>P</sup>), which was  
117 previously shown in complementation assays to insert iron into sirohydrochlorin *in vivo* (Lobo  
118 *et al.*, 2008). Similar assays were carried out by transforming plasmids expressing the wild type  
119 and mutated proteins into *E. coli* 302 $\Delta$ a pCIQ-SirCCobA, a cysteine auxotrophic strain that  
120 produces sirohydrochlorin but not sirohaem (Brindley *et al.*, 2003; Raux *et al.*, 2003). As  
121 expected, wild type and all mutated proteins grew well in minimal medium supplemented with  
122 cysteine (positive controls, Table 1 and Fig. S1 of Supplemental Material). Like  $\Delta$ 28CbiK<sup>P</sup>, the  
123 E184L mutant protein supported growth in cysteine-deficient minimal medium, as did all other  
124 single mutated proteins (Table 1; Fig. S1 of Supplemental Material). In contrast, the double  
125 H154L/H216L and H154L/E184L and triple H154L/H216L/E184L  $\Delta$ 28CbiK<sup>P</sup> mutant proteins were  
126 unable to rescue the *E. coli* 302 $\Delta$ a growth deficient phenotype.

127

### 128 **CbiK<sup>P</sup> is a periplasmic protein**

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

137 Expression of GFP- $\Delta$ 28CbiK<sup>P</sup> yielded cells whose cytoplasm exhibited a bright homogeneous  
138 fluorescence (Fig. 3A). Cells expressing the GFP-CbiK<sup>P</sup> fusion presented a peripheral  
139 fluorescence, with a polar protein accumulation that increased with time (Fig. 3B). These  
140 results show that CbiK<sup>P</sup> is exported to the periplasm by means of its 28 amino acid N-terminal  
141 signal peptide.

142

#### 143 **Haem binding properties of *D. vulgaris* CbiK<sup>P</sup>**

144 Biochemical characterisation of the *D. vulgaris* CbiK<sup>P</sup>, produced in *E. coli* grown in the presence  
145 of added haem precursor, showed that the protein contains two haem *b* groups per tetramer  
146 (Lobo *et al.*, 2008). From the crystallographic structure it was concluded that His96 is located  
147 within a coordination distance of *ca.* 2.0 Å from the haem iron atom, indicating that it is the  
148 haem axial ligand (Romão *et al.*, 2011). Replacement of His96 by leucine, indeed, does appear  
149 to prevent haem binding as H96L does not purify with any bound haem (Table 1). Furthermore,  
150 the cobaltochelatase activity of H96L-CbiK<sup>P</sup> increased by ~2-fold when compared with the wild  
151 type (Fig. 2), and H96L- $\Delta$ 28CbiK<sup>P</sup> retained the ability to complement the *E. coli* 302 $\Delta$ a pCIQ-  
152 SirCCobA strain (Table 1). These results confirm that haem binding and chelation are  
153 independent properties of the protein.

154 The *D. vulgaris* CbiK<sup>P</sup> wild type structure shows that each haem is located between two  
155 monomers, consistent with the haem/dimer ratio of 1:1 (Lobo *et al.*, 2008; Romão *et al.*,  
156 2011). Therefore, we examined the role of the haem ligand His96 in the oligomerisation state  
157 of the protein. The results in Table 1 show that mutation of this residue does not interfere with  
158 the quaternary structure of the enzyme as H96L-CbiK<sup>P</sup> retained its tetrameric form.

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



162 Glu76 by leucine led to the dismantling of the tetrameric form of CbiK<sup>P</sup> but did not affect the  
163 1:1 haem/dimer ratio (Table 1). CbiK<sup>P</sup> harbouring either the Arg54L or Glu76L substitutions  
164 exhibited cobaltochelatase activities in the same order of magnitude as wild type enzyme, with  
165 Arg54L having a slightly lower activity (Fig. 2); nonetheless, both mutated proteins retained the  
166 ability to complement the iron chelatase deficient *E. coli* 302Δa pCIQ-SirCCobA strain (Table 1).

167

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

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

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

188

189 How widespread are CbiK<sup>P</sup> proteins in bacterial systems

190 *D. gigas* contains a CbiK<sup>P</sup> homologue and a related CbiK protein is also found as an outer  
191 membrane haemin-binding protein potentially involved in iron assimilation (Dashper et al J

192 *Bacteriol.* 2000 Nov;182(22):6456-62.). As a number of CbiK proteins are clearly targeted to  
193 either the periplasm or the outer membrane, we have analysed how widespread extracellular

194 CbiK proteins are within biological systems, by searching protein databases for CbiK proteins  
195 with an N-terminal extension that could represent a signal peptide. Such an analysis revealed

196 the presence of a broad range of related CbiK proteins, from a diverse range of bacteria, that  
197 appear to be targeted outside of the bacterium. These included species within bacteriodes,

198 *Clostridia*, *Desulfo-*, *Eubacterium*, *Parabacteriodes*, *Porphyromonas*, *Roseburia* and *Veillonella*. It  
199 is not possible to predict whether these proteins are able to bind haem but the presence of

200 duplicate copies of CbiK, with and without a signaling peptide all suggests that the roles of CbiK  
201 potentially as a haem transporter is much more widespread than previously thought.

202

203

#### 204 **Revisiting *D. vulgaris* CbiK<sup>P</sup> structure**

205 In spite of several attempts to solve the structure of the *D. vulgaris* CbiK<sup>P</sup> loaded with 4 haems  
206 per tetramer, electron density for the two additional haems could not be observed, which is

207 most probably due to the high lability of the haems that may bind in the central cavity of the  
208 tetramer. Crystals, however, were obtained by incubation of CbiK<sup>P</sup> with its substrates cobalt

209 and sirohdrochlorin (herein named *D. vulgaris* CbiK<sup>P</sup> (CoSHC); PDB ID code 2xvy). This  
210 structure was determined to a higher resolution (1.7 Å) (Table S3 of Supplemental Material)

211 than the previous. The crystallographic data reveals new features in the orientation of His103  
212 and the haem propionate side chains, details that are discussed below.

213 The monomer organisation of CbiK<sup>P</sup> (CoSHC) is similar to that observed with the as-isolated  
214 CbiK<sup>P</sup> (PDB ID code 2xvx) and that of CbiK<sup>P</sup> upon incubation with cobalt, which was previously  
215 designated CbiK<sup>P</sup>(Co) (PDB ID code 2xvz) (Romão *et al.*, 2011). The three structures display a  
216 root mean square (r.m.s.) deviation between equivalent C<sup>α</sup> atoms of only 0.20 Å. In all  
217 structures, the monomer presents a  $\alpha/\beta$  topology consisting of two domains, domain A  
218 (residues 6-117) and domain B (residues 125-269), which are connected by a loop (residues  
219 118-124) (Fig. 5). Domain A is composed of one four-stranded parallel  $\beta$ -sheet and four helices  
220 ( $\alpha$ 1- $\alpha$ 4), with the haem ligand residue His96 located in the  $\alpha$ 4 helix. The strands of the  $\beta$ -sheet  
221 are organized as  $\beta$ 2 -  $\beta$ 1 -  $\beta$ 3 -  $\beta$ 4 with a twist of  $\alpha$ . 65 °, and the four  $\alpha$ -helices surround the  
222  $\beta$ -sheet core. Domain B is formed by a four-stranded  $\beta$ -sheet, arranged in a way similar to that  
223 of domain A ( $\beta$ 6 -  $\beta$ 5 -  $\beta$ 7 -  $\beta$ 8), that is surrounded by five helices ( $\alpha$ 5,  $\alpha$ 6,  $\alpha$ 7,  $\alpha$ 8 and  $\alpha$ 9) and  
224 harbours the residues His154, Glu184 and His216 that coordinate the cobalt ion. The  $\alpha$ 10 helix  
225 is positioned at the C-terminal, overlapping the two domains and is located near helices  $\alpha$ 1  
226 and  $\alpha$ 4 of domain A (Fig. 5A-C).

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

231 In the tetramer of all available *D. vulgaris* CbiK<sup>P</sup> structures, the monomeric porphyrin binding  
232 clefts face outwards, and are therefore accessible to the solvent (Fig. 1). Each monomer  
233 interacts with two others through hydrogen bonds that likely contribute to the stabilisation of  
234 the tetramer. There are only 28 inter-subunit hydrogen bonds out of a total of 1064 found  
235 within the tetramer. These hydrogen bonds can be further divided into groups involving

236 subunits AB and CD, and subunits AD and BC (Table 2). Additionally, the *D. vulgaris* CbiK<sup>P</sup>  
237 tetramer harbours an open cavity at the centre of the tetramer, with dimensions *ca.* 16 × 15 ×  
238 17 Å, where the haems are located (Fig. 6).

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

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

258

259 **Discussion**

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

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

283 The *D. vulgaris* CbiK<sup>P</sup> quaternary organisation is unique with the binding of two haem  
284 molecules in-between two monomers (Fig. 1). The tetrameric form of *D. vulgaris* CbiK<sup>P</sup> is

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

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

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

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

317 The titration of CbiK<sup>P</sup> with haemin showed that the protein is able to bind two extra haems,  
318 and the site-directed mutagenesis studies suggest that these haems interact with His103. The  
319 two extra haem groups could fit into the central cavity of the tetramer if the two existing haem  
320 groups take up the hypothetical conformations depicted in Fig. 8. The presence in  
321 CbiK<sup>P</sup>(CoSHC) of a sulphate molecule close to residue His103, and to the haem propionates,  
322 suggests that this position would be where the propionates of the extra haems could be  
323 stabilised. Based on this, and on the different conformations adopted by His103 in the  
324 presence of metal or porphyrin, a structural model in which the two extra haems would be  
325 coordinated by His103 has been devised, in agreement with the site-directed mutagenesis  
326 data. In this model, the proximity of Glu100 raises the possibility that this residue acts as the  
327 second axial ligand (Fig. 8A). However, carboxylates (Asp, Glu) are uncommon haem ligands  
328 with the *D. vulgaris* cytochrome *c* nitrite reductase NrfHA (Pdb 2j7a) exemplifying a rare case  
329 of an aspartate being located close to the haem iron (Rodrigues *et al.*, 2006). Moreover, a His-  
330 Glu axial coordination would most probably lead to a high-spin ferric haem, which does not  
331 agree with the observed spectroscopic data. Alternatively, the haems could be coordinated by  
332 His103 from two different subunits (Fig. 8B), which would generate a low-spin ferric haem and  
333 be consistent with the experimental results. Of course, the possibility exists that the binding of  
334 the extra haems causes structural rearrangements within the central cavity. [In the available](#)

335 structures, the haem binding His103 residue exhibits different conformations suggesting a high  
336 flexibility, which may facilitate the release of the haem upon interaction with other proteins  
337 (e.g. the haem transporters encoded in the operon in which the CbiK gene is located).

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

347

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

356 In *Desulfovibrio* and sulphate-reducing bacteria in general, sirohydrochlorin plays an important  
357 role not only as a protein cofactor but also as a key intermediate in the haem and vitamin B<sub>12</sub>  
358 biosynthetic pathways (Bali *et al.*, 2011; Lobo *et al.*, 2012). As mentioned in the introduction,  
359 *D. vulgaris* expresses cytoplasmic and periplasmic versions of CbiK that have very similar amino



360 acid sequences. In particular, CbiK<sup>C</sup> contains the two histidine residues equivalent to His154  
361 and His216 in CbiK<sup>P</sup>, which were shown here to be essential for the sirohydrochlorin chelatase  
362 activity. The two CbiKs therefore exhibit cobaltochelatase activity; however, as tetrapyrrole  
363 biosynthesis occurs in the cytosol, CbiK<sup>C</sup> is the most plausible candidate for the  
364 sirohydrochlorin cobaltochelatase in *Desulfovibrio* spp. (Lobo *et al.*, 2008). Reinforcing this  
365 hypothesis, our results have shown that the presence of His96 and His103 in CbiK<sup>P</sup> are  
366 associated with an additional haem-binding biological function, as site-directed mutagenesis of  
367 these two residues yielded proteins that no longer have the ability to bind haem. Therefore,  
368 although still a functional chelatase, the presence of His96 and His103 in CbiK<sup>P</sup>, which are  
369 absent from the sequences of *D. vulgaris* CbiK<sup>C</sup> and all other bacterial cobaltochelatases (Table  
370 S1 of the Supplemental Material), have prompted the evolution of a gain of function relating to  
371 haem transport. To the best of our knowledge, this is the first example of how a common  
372 ancestral protein has evolved a new functional role through acquisition of two histidine  
373 residues. Although CbiK<sup>P</sup> constitutes one of the first recognized haem chaperons in *D. vulgaris*,  
374 questions remain as to its physiological partners and how it participates in haem homeostasis  
375 in this sulphate-reducing bacterium.

376

## 377 **Experimental Procedures**

### 378 *Site-directed mutagenesis*

379 Twelve *D. vulgaris* CbiK<sup>P</sup> site-directed mutated proteins were constructed (seven single, three  
380 double and one triple mutated) with the following substitutions: H154L, H216L, E184L, R54L,  
381 E76L, H103L, H96L, H216L/H154L, H216L/E184L, H154L/E184L and H216L/H154L/E184L. Site-  
382 directed mutagenesis was carried out with the QuickChange II Site-Directed Mutagenesis Kit  
383 (Agilent Technologies). For this purpose, primers listed in Table S2 (Supplemental Material)  
384 and the following plasmids were used: pET28a(+)-*cbiK<sup>P</sup>*, which contains the DNA sequence

385 encoding the wild type CbiK<sup>P</sup>, and pET28a(+)-Δ28-*cbiK*<sup>P</sup> expressing a truncated form of CbiK<sup>P</sup>  
386 that lacks the periplasmic targeting signal sequence (Lobo *et al.*, 2008). Double and triple  
387 mutants were obtained as follows: plasmid pET28a(+)-H216L-*cbiK*<sup>P</sup> served as template to  
388 construct pET28a(+)-H216LH154L-*cbiK*<sup>P</sup> and pET28a(+)-H126LE184L, and pET28a(+)-H154L to  
389 produce pET28a(+)-H154LE184L-*cbiK*<sup>P</sup>. Plasmid pET28a(+)-H216LH154LE184L-*cbiK*<sup>P</sup> that  
390 expresses the triple mutated CbiK<sup>P</sup> was obtained using pET28a(+)-H216LH154L as template. In  
391 all cases, DNA sequencing confirmed that the correct amino acid changes were introduced and  
392 no undesirable mutations were present.

393

#### 394 *Protein expression, purification and characterisation*

395 Expression and purification of *D. vulgaris* CbiK<sup>P</sup> wild type and mutant variants was done as  
396 follows: each recombinant plasmid was transformed, separately, into *Escherichia coli*  
397 BL21Gold(DE3) (Stratagene) and cells were grown, at 30 °C, in Luria-Bertani (LB) medium  
398 containing kanamycin (30 µg.ml<sup>-1</sup>) until reaching an optical density at 600 nm (OD<sub>600</sub>) of 0.3.  
399 Expression of the proteins was achieved by addition of 200 µM isopropyl-D-  
400 thiogalactopyranoside (IPTG), 50 µM 5-aminolevulinic acid (ALA) and 100 µM FeSO<sub>4</sub>. After  
401 induction with IPTG, cells were grown overnight, at 20 °C, harvested by centrifugation (e.g. 11k  
402 ×g, 10 min, 4 °C), resuspended in 20 mM Tris-HCl pH 7.5 (buffer A), and disrupted in a French  
403 Press (ThermoScientific). The soluble fraction was separated by ultracentrifugation at 160000  
404 ×g, for 2 h. All centrifugation and purification steps were performed at 4 °C. The resultant  
405 soluble fraction was applied onto a Chelating Sepharose fast flow column (25 mL) (GE,  
406 Healthcare), previously charged with NiSO<sub>4</sub> in buffer A and containing 400 mM NaCl. A linear  
407 gradient up to 250 mM imidazole was applied and the protein was eluted approximately at 80  
408 mM imidazole. Protein fractions containing CbiK<sup>P</sup> were dialysed against buffer A and loaded  
409 onto a Q-Sepharose High-Performance column (25 mL) (GE, Healthcare), previously

410 equilibrated in buffer A and purified by means of a linear gradient up to 400 mM NaCl; the  
411 protein was collected at ~200 mM NaCl. Fractions containing CbiK<sup>P</sup> wild type and mutated  
412 proteins were dialyzed against 50 mM Tris-HCl pH 8 (buffer B).

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

419

#### 420 *Activity assays*

421 Sirohydrochlorin was produced as previously described (Lobo *et al.*, 2008). Plasmid pETcoco-  
422 2ABCDC harbouring the genes encoding *Methanosarcina barkeri* uroporphyrinogen III  
423 methyltransferase (CobA), *Methanothermobacter thermautotrophicus* porphobilinogen  
424 synthase (HemB) and precorrin-2 dehydrogenase (SirC), *B. megaterium* porphobilinogen  
425 deaminase (HemC) and uroporphyrinogen III synthase (HemD) (Frank *et al.*, 2007), was  
426 introduced into *E. coli* BL21Star(DE3)pLysS (Invitrogen). Cells were grown in LB medium  
427 containing ampicillin (50 µg.ml<sup>-1</sup>), chloramphenicol (34 µg.ml<sup>-1</sup>) and 0.2 % (w/v) glucose.  
428 Overnight grown cell cultures were used to inoculate fresh antibiotic supplemented LB  
429 medium and further cultured, at 37 °C, to an OD<sub>600</sub> of ~0.5. At this stage, the medium was  
430 supplemented with 0.02 % (w/v) L-arabinose and after two hours 400 µM IPTG was added.  
431 Following an overnight growth, at 24 °C, cells were collected by centrifugation and  
432 resuspended in buffer B containing 100 mM NaCl. Lysates were obtained by passage of cells  
433 through a French Press and debris was removed by centrifugation (~39000 xg, 30 min, 4 °C).  
434 Lysates were exposed to an argon atmosphere for 20-30 min, transferred to the anaerobic

435 chamber (Coy model A-2463 filled with a gas mixture of 95 % argon and 5 % hydrogen) and  
436 used in a reaction mixture to produce sirohydrochlorin. The reaction mixture was composed of  
437 2 ml cell lysate, 2 mg S-adenosyl-L-methionine, 1 mg aminolaevulinic acid and 1 mg NAD<sup>+</sup>, in a  
438 total volume of 6 ml buffer B with 100 mM NaCl, and the reaction mixture pH was adjusted to  
439 8 by addition of 1 M KOH. After overnight incubation at room temperature, the reaction  
440 mixture was filtered (0.22 µm PALL syringe filter) and passed through a DEAE-resin (Sigma)  
441 previously equilibrated in buffer B containing 100 mM NaCl. The column was washed with  
442 buffer B with 250 mM NaCl and sirohydrochlorin was eluted with buffer B with 1 M NaCl.  
443 The chelatase activity was measured under anaerobic conditions in a Shimadzu UV-1800  
444 spectrophotometer. The assays were performed in buffer B supplemented with 100 mM NaCl,  
445 using 4.2 µM sirohydrochlorin, 20 µM Co<sup>2+</sup> and several concentrations of *D. vulgaris* CbiK<sup>P</sup> wild  
446 type and mutated proteins. Reactions were monitored following the absorbance decrease of  
447 the 376 nm band (extinction coefficient  $2.4 \times 10^5 \text{ M}^{-1} \text{ cm}^{-1}$ ), which corresponds to the  
448 consumption of sirohydrochlorin, and the build-up of the 414 nm band associated with the  
449 cobalt-sirohydrochlorin formation (Schubert *et al.*, 2002).

450

#### 451 *Complementation assays*

452 The genes encoding *D. vulgaris* truncated CbiK<sup>P</sup> ( $\Delta 28\text{cbiK}^P$ ) wild type and site-directed mutants  
453 were cloned in pETac plasmid (Raux *et al.*, 1997). The resulting plasmids were transformed into  
454 *E. coli* 302Δa strain that lacks the multifunctional sirohaem synthase CysG and contains  
455 plasmid pCIQ-*sirCcobA* for expression of *M. thermoautotrophicus* *sirC* and *Pseudomonas*  
456 *denitrificans* *cobA* genes (Brindley *et al.*, 2003; Raux *et al.*, 2003). Selection of the recombinant  
457 plasmids harbouring CbiK<sup>P</sup> mutant in *E. coli* 302Δa strain was done on LB plates supplemented  
458 with ampicillin (100 µg.ml<sup>-1</sup>) and chloramphenicol (35 µg.ml<sup>-1</sup>). *E. coli* 302Δa was also  
459 transformed with plasmid pKK223.2-*cysG*, which expresses the *E. coli* CysG, and with the

460 empty pETac plasmid, to serve as positive and negative controls, respectively. All strains were  
461 cultured on minimal medium plates and their growth was evaluated in the absence and  
462 presence of cysteine ( $0.05 \text{ mg.ml}^{-1}$ ).

463

#### 464 *Haem titrations*

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

475

#### 476 *D. vulgaris* CbiK<sup>P</sup> cellular localisation

477 For the determination of the cellular localisation of *D. vulgaris* CbiK<sup>P</sup> in *E. coli*, two plasmids  
478 were constructed: pET23b(+)- $\Delta 28\text{cbiK}^P$ -GFP and pET23b(+)-*cbiK*<sup>P</sup>-sfGFP that express a green  
479 fluorescent protein (GFP) fusion of  $\Delta 28$  CbiK<sup>P</sup> and a superfold green fluorescent protein (sfGFP)  
480 fusion of CbiK<sup>P</sup>, respectively. The sfGFP probe is a GFP variant that has been shown to retain its  
481 activity when exported to the periplasm by the Sec pathway (Pedelacq *et al.*, 2006, Aronson *et*  
482 *al.*, 2011). pET23b(+)- $\Delta 28\text{cbiK}^P$ -*gfp* was obtained by sub-cloning the  $\Delta 28\text{cbiK}^P$  gene into  
483 pET23b(+)-*gfp*. The *sfgfp* gene was amplified by PCR from plasmid BBa\_I1746908  
484 ([http://partsregistry.org/Part:BBa\\_I746908](http://partsregistry.org/Part:BBa_I746908)) and using primers sfGFPfw (5'-

485 AAGCTAGCAAAGAGCTCATGCGTAAAGGC-3') and sfGFP<sub>Prev</sub> (5'-  
486 CCTGGCTCTCGAGTCATCATTTGTAC-3') containing SacI and XhoI restriction sites, respectively,  
487 and ligated to yield plasmid pET23b(+)-*cbiK<sup>P</sup>*-*sfgfp*. Plasmids containing *cbiK<sup>P</sup>*-*sfgfp* and  
488  $\Delta 28*cbiK<sup>P</sup>*$ -*sfgfp* gene fusions were introduced into *E. coli* BL21Gold(DE3), and cells were further  
489 grown in LB medium supplemented with ampicillin (50  $\mu\text{g}\cdot\text{ml}^{-1}$ ), at 30°C, to an OD<sub>600</sub> of ~0.3-  
490 0.5. At this stage, 100  $\mu\text{M}$  IPTG was added and cells after grown for 2 h and 3 h were collected  
491 by centrifugation (9700  $\times g$ , 5 min, room temperature), and resuspended in 100  $\mu\text{l}$  phosphate-  
492 buffered saline (PBS). Cell suspension (10  $\mu\text{l}$ ) was smeared onto a thin layer of 1.7 % agarose  
493 mounted in a microscope glass slide. Phase contrast and fluorescent visualisation were done in  
494 a Leica DM6000B microscope attached to a Leica EL6000 fluorescence light source and a  
495 MicroPoint. Images were obtained using an Andor Inox+ camera and Metamorph microscopy  
496 automation and image analysis software (Molecular Devices).

497

#### 498 *Protein crystallisation, structure determination and refinement*

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

507 An X-ray diffraction dataset to 1.7 Å resolution was collected at the BM14 beamline of the  
508 European Synchrotron Radiation Facility. The diffraction images were integrated and scaled  
509 with XDS (Kabsch, 1993) (Table S3). Merging of diffraction intensity data and conversion to

510 structure factors was carried out with SCALA and TRUNCATE in the CCP4 program package  
511 (Winn *et al.*, 2011).

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

520 Coordinates and structure factors of the *D. vulgaris* CbiK<sup>P</sup>(CoSHC) structure were deposited in  
521 the Protein Data Bank in Europe (Velankar *et al.*, 2011) with the accession code 2xvy.

522

### 523 **Acknowledgements**

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

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

536

537

538

539

540 **References**

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

543 Bali, S., Lawrence, A., Lobo, S.L., M., S.L., Golding, B., Palmer, D. *et al.* (2011) Molecular  
544 hijacking of siroheme for the synthesis of heme and  $d_1$  heme. *Proc Natl Acad Sci U S A* **108**:  
545 18260-18265.

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

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

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

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

555 Frank, S., Brindley, A.A., Deery, E., Heathcote, P., Lawrence, A.D., Leech, H.K. *et al.* (2005)  
556 Anaerobic synthesis of vitamin B<sub>12</sub>: characterisation of the early steps in the pathway. *Biochem*  
557 *Soc Trans* **33**: 811-814.



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

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

564 Krissinel, E., and Henrick, K. (2004) Secondary-structure matching (SSM), a new tool for fast  
565 protein structure alignment in three dimensions. *Acta Crystallogr D Biol Crystallogr* **60**: 2256-  
566 2268.

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

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

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

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

578 McCoy, A.J., Grosse-Kunstleve, R.W., Adams, P.D., Winn, M.D., Storoni, L.C., and Read, R.J.  
579 (2007) Phaser crystallographic software. *J Appl Crystallogr* **40**: 658-674.

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

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

584 Raux, E., Thermes, C., Heathcote, P., Rambach, A., and Warren, M.J. (1997) A role for  
585 *Salmonella typhimurium* cbiK in cobalamin (vitamin B<sub>12</sub>) and siroheme biosynthesis. *J Bacteriol*  
586 **179**: 3202-3212.

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

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

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

597 Schubert, H.L., Raux, E., Wilson, K.S., and Warren, M.J. (1999) Common chelatase design in the  
598 branched tetrapyrrole pathways of heme and anaerobic cobalamin synthesis. *Biochemistry* **38**:  
599 10660-10669.

600 Schubert, H.L., Raux, E., Brindley, A.A., Leech, H.K., Wilson, K.S., Hill, C.P., and Warren, M.J.  
601 (2002) The structure of *Saccharomyces cerevisiae* Met8p, a bifunctional dehydrogenase and  
602 ferrochelatase. *Embo J* **21**: 2068-2075.

603 Smith, P.K., Krohn, R.I., Hermanson, G.T., Mallia, A.K., Gartner, F.H., Provenzano, M.D. *et al.*  
604 (1985) Measurement of protein using bicinchoninic acid. *Anal Biochem* **150**: 76-85.

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

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

610 Winn, M.D., Ballard, C.C., Cowtan, K.D., Dodson, E.J., Emsley, P., Evans, P.R. *et al.* (2011)  
611 Overview of the CCP4 suite and current developments. *Acta Crystallogr D Biol Crystallogr* **67**:  
612 235-242.

613

614

615 **Table 1.** Properties of *D. vulgaris* CbiK<sup>P</sup> mutants

616

CbiK <sup>P</sup>	Biochemical characterisation		Complementation assays	
	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	+	+

617

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

623

624

625

626 **Table 2.** Hydrogen bonds between monomers pairs of *D. vulgaris* CbiK<sup>P</sup>(Co)

627

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

628

629 Hydrogen bonds between the CbiK<sup>P</sup> monomers A and B (equivalent to monomers D and C) and  
 630 monomers A and D (equivalent to monomers B and C), as depicted in Fig. 1.

631

632

633

634

635

636

637

638 **Figure Legends**

639

640 **Figure 1.** *D. vulgaris* CbiK<sup>P</sup>(Co) wild type.

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

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

649

650 **Figure 2.** Cobaltochelataase activity of CbiK<sup>P</sup> wild type and mutant variants.

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

654

655 **Figure 3.** Localisation of *D. vulgaris* CbiK<sup>P</sup>.

656 Phase contrast (PC) and fluorescent (GFP) visualisation of *E. coli* expressing  $\Delta 28$ CbiK<sup>P</sup>-GFP (A)  
657 and CbiK<sup>P</sup>-sfGFP (B). Cells expressing the fusion proteins were visualized after 2 h (A and B1)  
658 and 3 h (B2) post ITPG induction.

659

660 **Figure 4.** Haem binding assays.

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

665

666 **Figure 5.** Structure of the *D. vulgaris* CbiK<sup>P</sup>(CoSHC).

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

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

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

673 D. Cobalt site coordinated by three amino-acid residues: His154 (distances Co – NE2 is 2.05 Å  
674 and Co – CE1 is 2.04 Å), Glu184 (distance Co – NE2 is 2.45 Å) and His216 (distance Co – NE2 is  
675 2.07 Å), and by further two water molecules (distances Co – W1 is 2.19 Å and Co – W2 is 2.07  
676 Å) and a peroxide molecule (distance Co – O1 is 2.05 Å).

677

678 **Figure 6.** Central tetramer cavity of *D. vulgaris* CbiK<sup>P</sup>.

679 Detailed view of the central tetramer cavity of the as-isolated CbiK<sup>P</sup> wild type (CbiK<sup>P</sup>; PDB code  
680 2xvx), containing cobalt (CbiK<sup>P</sup>(Co); PDB code 2xvz) and after incubation with cobalt and  
681 sirohydrochlorin (CbiK<sup>P</sup>(CoSHC); PDB code 2xvy, this work). Each monomer is coloured as  
682 described in Fig. 1, with haems *b*, and axial iron ligand His96, His103 and His158 represented  
683 as sticks. For all structures, the right panel is rotated 90° showing a close-up of the haem  
684 between subunits A and D, indicating the distances between His158 and the propionate C, and  
685 His103 and propionate D. Close-up of the side-chain of His103 and haem propionate groups  
686 show that there are significant conformational differences among the three structures.

687

688 **Figure 7.** Superimposition of the *D. vulgaris* CbiK<sup>P</sup>(CoSHC) with *Salmonella enterica*  
689 sirohydrochlorin containing chelatase.

690 *D. vulgaris* CbiK<sup>P</sup>(CoSHC) was superimposed with the *Salmonella enterica* in complex with  
691 metalated sirohydrochlorin chelatase (pdb code 2xwp). The sulphate and glycerol molecules

692 observed in the *D. vulgaris* CbiK<sup>P</sup>(CoSHC) structure are in similar positions to the propionates  
693 and acetates groups from the metallated sirohydrochlorin observed in the *S. enterica* CbiK. The  
694 atoms are represented with the same colour code as in Fig. 1. The metallated sirohydrochlorin  
695 from *S. enterica* CbiK is represented in pink sticks.

696

697 **Figure 8.** Models for *D. vulgaris* CbiK<sup>P</sup> with the two extra haems in the central cavity of the  
698 tetramer.

699 A. Model 1, the extra haems are coordinated by His103, and Glu100 would be close to the  
700 other axial position. B. Model 2 where the extra haems would be coordinated by the two  
701 symmetrically related His103. The position 1 of His103 corresponds to that observed in the *D.*  
702 *vulgaris* CbiK<sup>P</sup>(CoSHC) structure, whereas the position 2 corresponds to that in the *D. vulgaris*  
703 CbiK<sup>P</sup>(Co) structure.