Title: Identification of the sirohaem biosynthesis pathway in *Staphylococcus aureus*

Running Title: Sirohaem biosynthesis in *Staphylococcus aureus*

Marco A.M. Videira[1], Susana A.L. Lobo[1,2], Filipa L. Sousa[3], and Lígia M. Saraiva[1,*]

[1] Instituto de Tecnologia Química e Biológica António Xavier, Universidade Nova de Lisboa, Avenida da República, 2780-157 Oeiras, Portugal

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

[3] Department of Ecogenomics and Systems Biology, University of Vienna, 1090 Vienna, Austria.

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

Instituto de Tecnologia Química e Biológica António Xavier, Universidade Nova de Lisboa

Av. da República, 2780-157 Oeiras, Portugal

Phone: +351 21 4469328. Fax: +351 21 4433 644. E-mail: lst@itqb.unl.pt

Keywords: sirohaem, *Staphylococcu*s, haem, ferrochelatase, tetrapyrrole biosynthesis

Abbreviations:

HmbS -hydroxymethylbilane synthase

IPTG- isopropyl β-D-1-thiogalactopyranoside

$NaNO_2$ - sodium nitrite

P2D- precorrin-2 dehydrogenase

SAM- S-adenosyl-L-methionine

ShfC - sirohydrochlorin ferrochelatase

SUMT- S-adenosyl-L-methionine-dependent uroporphyrinogen III methyltransferase

Uro´gen III -uroporphyrinogen III

UroM - S-adenosyl-L-methionine uroporphyrinogen III methyltransferase

UroS - uro´gen III synthase

Abstract

Sirohaem is a modified tetrapyrrole and a key prosthetic group of several enzymes involved in nitrogen and sulfur metabolisms. This work shows that *Staphylococcus aureus* produces sirohaem through a pathway formed by three independent enzymes. Of the two putative sirohaem synthases encoded in the *S. aureus* genome and annotated as *cysG*, one is herein shown to be a uroporphyrinogen III methyltransferase that converts uroporphyrinogen III to precorrin-2, and was renamed as UroM. The second *cysG* gene encodes a precorrin-2 dehydrogenase that converts precorrin-2 to sirohydrochlorin, and was designated as P2D. The last step was found to be performed by the gene *nirR* that, in fact, codes for a protein with sirohydrochlorin ferrochelatase activity, labeled as ShfC. Additionally, site-directed mutagenesis studies of *S. aureus* ShfC revealed that residues H22 and H87, which are predicted by homology modelling to be located at the active site, control the ferrochelatase activity. Within bacteria, sirohaem synthesis may occur via one, two or three enzymes, and we propose to name the correspondent pathways as Type 1, 2 and 3, respectively. A phylogenetic analysis revealed that Type 1 is the most used pathway in Gammaproteobacteria and Streptomycetales, Type 2 predominates in Fibriobacteres and Vibrionales, and Type 3 in Firmicutes of the Baciliales order. Altogether, we concluded that the current distribution of sirohaem pathways within bacteria, that changes at the genus or species level and within taxa, seems to be the result of evolutionary multiple fusion/fission events.

Introduction

Staphylococcus aureus is a facultative anaerobic bacterium responsible for a large number of infections of difficult treatment due to the emergence of antibiotic resistant strains. *S. aureus* colonizes niches with low oxygen levels, such as abscesses, intestinal lumen, skin, kidneys and liver, where it grows fermentatively or uses nitrate or nitrite as terminal electron acceptors [1–6]. Therefore, *S. aureus* encodes enzymes that reduce nitrate and nitrite, such as nitrate reductase (NarGHI) and nitrite reductase (NirBD) [7,8]. In the *S. aureus* genome, the nitrite reductase *nir* cluster contains four genes, namely *nirR*, *nirB*, *nirD*, and *cysG*. These genes are predicted to encode a nitrite reductase regulator (NirR), the two subunits of a NADH-dependent nitrite reductase (NirBD), and an uroporphyrinogen III methyltransferase CysG, as they share, respectively, 34%, 78%, 67% and 54% amino acid sequence identity with their homologues in *S. carnosus* [4,9]. In *S. carnosus*, this cluster contains a *sirA* gene, predicted to encode a putative precorrin-2 dehydrogenase/sirohydrochlorin ferrochelatase enzyme, which is not present in the *S. aureus nir* cluster.

The function of the nitrite reductases NirBD depends on the presence of a sirohaem cofactor, which transfers six electrons from reduced nicotinamide adenine dinucleotide (NADH) to nitrite to form ammonia [10]. In general, sirohaem is synthesized in bacteria from uroporphyrinogen III in a three step reaction that includes methylation of uroporphyrinogen III, NAD^+-dependent oxidation of precorrin-2, and insertion of iron into sirohydrochlorin [11,12] (Figure 1). Sirohaem is also an important intermediate of the alternative haem biosynthesis pathway, that is mainly utilized by Archaea and sulfate reducing bacteria, and in which sirohaem is converted to protohaem via an oxygen-independent four-enzyme-step process [13–16].

In *Escherichia coli* and *Salmonella enterica*, the three reactions are accomplished by a single multifunctional enzyme called sirohaem synthase usually referred as CysG [11]. The protein (of 457 amino acid residues) is composed by a N-terminal region ($CysG^B$) with precorrin-2 dehydrogenase and sirohydrochlorin ferrochelatase activities, and a C-terminal region ($CysG^A$), with S-adenosyl-L-methionine (SAM)-dependent uroporphyrinogen III (uro´gen III) methyltransferase (SUMT) activity that converts uro´gen III to precorrin-2 [12].

In *Saccharomyces cerevisiae*, two enzymes are required for the synthesis of sirohaem, encoded by *MET1* and *MET8* genes [17]. The product of *MET1* (Met1p) is a protein that shares amino acid sequence similarity to the C-terminal region of the *E. coli* CysG, and like CysG, it has SUMT activity [18]. Met8p protein is a bifunctional enzyme that shares amino acid sequence similarity to the N-terminal region of *E. coli* CysG, and exhibits two activities: an NAD^+-dependent precorrin-2 dehydrogenase activity and sirohydrochlorin chelatase activity [18].

In the Gram-positive bacterium *Bacillus megaterium*, the synthesis of sirohaem occurs via three proteins, namely SirA, SirB and SirC that are organized in the gene cluster *sirABC*. These proteins

exhibit SUMT activity (SirA), precorrin-2 dehydrogenase activity (SirC) and sirohydrochlorin ferrochelatase activity (SirB) [19,20].

In this work, we investigated the sirohaem biosynthetic pathway of *S. aureus* and show that three gene products are required for the conversion of uro´gen III to sirohaem. We also report that the gene annotated as *nirR* in the genome of *S. aureus*, and proposed to encode a nitrite reductase transcriptional regulator, is in fact a sirohydrochlorin ferrochelatase that performs the last step of the sirohaem biosynthesis pathway. A phylogenetic analysis was done to infer the occurrence of sirohaem pathways in bacteria.

Results

The amino acid sequences of enzymes known to be involved in the synthesis of sirohaem were used to search homolog proteins encoded in the *S. aureus* genome, which allowed identifying two CysG-like proteins. One is encoded by the *cysG* gene included in *nir* cluster (herein designated as *cysG1*), and the second is codified by a *cysG* gene (*cysG*2) located downstream of the sulfite reductase *cysJ* gene.

To test whether these proteins participate in sirohaem formation in *S. aureus*, the genes *cysG1* and *cysG2* were cloned, and the correspondent proteins were recombinantly produced and biochemically characterized.

The uroporphyrinogen III methyltransferase of *S. aureus*

Staphylococcus aureus cysG1 gene encodes 325 amino acids and the BLASTp search showed that the protein shares a significant degree of amino acid sequence identity and similarity to *E. coli* CysG C-terminal region and SirA/CobA proteins of *B. megaterium*, *Pseudomonas denitrificans and Paracoccus denitrificans* (Table 1, Figure 2A). In the sirohaem pathway, these proteins perform the first step that converts uro´gen III to precorrin-2, in a SAM-dependent way. To investigate its activity, *S. aureus* CysG1 was produced and the purified protein exhibited a molecular mass of approximately 40 kDa (Figure 3A). Incubation of CysG1 with uro´gen III and SAM generated a yellow product that exhibited an UV-Visible spectrum typical of precorrin-2 with broad absorption features from 500–400 nm and 400–350 nm (Figure 3B) [21], which indicated that the enzyme is an uro´gen III methyltransferase. Next, we determined the kinetic parameters of the S-adenosyl-L-methionine-dependent uroporphyrinogen III (SUMT) activity of CysG1, which was done by coupling the CysG1 and the *D. vulgaris* SirC activities, as described in Materials and methods. The data fitted well to an allosteric sigmoidal kinetics indicating that the substrates bind cooperatively. *S. aureus* CysG1 converted uro´gen

III to precorrin-2 with a specific activity of 0.72 ± 0.02 nmol.mg^{-1}min^{-1} and a K_M for uro´gen III of 1.24 ± 0.29 μM (Figure 3C). These values are in the same order of magnitude of those reported for *D. vulgaris* CobA (3 nmol.mg^{-1}min^{-1} and K_M of 0.4 μM for uro´gen III) [22]. Since *S. aureus* CysG1 functions as an uro´gen III methyltransferase, the enzyme was renamed as UroM.

S. aureus CysG2 is a precorrin-2 dehydrogenase

Comparison of the *S. aureus cysG*2 encoded sequence against the protein database showed that it shares similarity to the precorrin-2 dehydrogenase of *E. coli* CysG N-terminal region, the SirC proteins of *B. megaterium* and *D. vulgaris*, and *S. cerevisiae* Met8p (Table 1, Figure 2B). The *cysG2* gene was cloned and the protein was produced in *E. coli* for further purification and evaluation of its activity. *S. aureus* CysG2 is a 201 amino acid protein with the expected molecular mass of approximately 25 kDa (Figure 3A). The activity of CysG2 was tested by incubation of the protein with NAD^+ and precorrin-2, which was previously generated by incubating SAM, δ-Ala and the enzymes expressed from plasmid pET-coco2-ABCD (Materials and methods). Formation of sirohydrochlorin was deduced by a colour change of the reaction mixture from yellow to purple, and further confirmed by the presence of absorbance maxima at 376 nm and 590 nm (Figure 3D), which is consistent with what was reported in previous studies [19]. Additionally, when the reaction was done in the presence of iron, no sirohydrochlorin ferrochelatase activity was observed for CysG2 (data not shown).

The catalytic parameters of the NAD^+-dependent precorrin-2 dehydrogenase activity of *S. aureus* CysG2 were also determined. Measurements were done using as substrates NAD^+ and precorrin-2, the latter being generated in a separated reaction containing the enzymes expressed from plasmid pET-coco2-ABCD. The reaction was monitored by UV-visible spectroscopy following the increase of sirohydrochlorin absorbance at 376 nm. *S. aureus* CysG2 exhibited a specific activity of 62 ± 2 nmol.mg^{-1}min^{-1} and a K_M for NAD^+ of 340 ± 40 μM (Figure 3E). Therefore, CysG2 of *S. aureus* proved to function as a precorrin-2 dehydrogenase *in vitro*, and was renamed as P2D.

S. aureus NirR is a sirohydrochlorin ferrochelatase and is required for nitrite reduction

S. aureus UroM and P2D proteins showed the highest amino acid sequence identity and similarity to enzymes of *B. megaterium*. Thus, we used *B. megaterium* sirohydrochlorin ferrochelatase to search in *S. aureus* for the enzyme that could perform the last step of the sirohaem pathway (Table 1). Unexpectedly, the search retrieved the gene product of *nirR* located upstream of *nirBD,* which encode the large subunit (NirB) and the small subunit (NirD) of the nitrite reductase enzyme, and which was annotated as its regulator. For this reason, we first tested whether NirR was indeed the nitrite reductase

regulator by evaluating the expression of *nirB* and *cysG*1 in the *S. aureus* Δ*nirR* mutant. However, the expression of these genes had no differences between wild type and the mutant strain (fold changes: *nirB*=0.87; *cysG*1=0.77), which rule out the regulatory role of NirR. Consequently, the hypothesis of NirR being a sirohydrochlorin ferrochelatase was next addressed.

The *nirR* gene has 732 bp and its initiation codon is annotated as GTG, which encodes a valine amino acid instead of methionine. After several attempts, this gene product could not be expressed successfully. A more detailed examination of the *nirR* sequence revealed that the most probable ATG start codon was located 39 bp upstream of the one annotated in the genome giving rise to a gene with 771 bp. This DNA segment was cloned into pET-23b, and a protein of 255 amino acids and a molecular mass of approximately 29 kDa was produced.

Due to the low solubility of the protein, the sirohydrochlorin ferrochelatase activity of NirR was determined in complementation experiments using the *E. coli* Δ302a mutant. This strain lacks the *E. coli cysG* gene and, because it is unable to synthesize sirohaem, it only grows in medium supplemented with cysteine or sulfide, or when the strain expresses other proteins capable of producing sirohaem [23]. Thus, the *E. coli* Δ302a expressing *M. thermoautotrophicus sirC* and *P. denitrificans cobA* from plasmid pCIQ was transformed with plasmid pET-23 expressing *S. aureus nirR* gene and the growth in minimal medium with no cysteine was evaluated. Figure 4A shows that while the strain containing the empty pET-23 vector did not grow, the strain complemented with pET-23-wild type-nirR grew as well as that complemented with *E. coli cysG.*

To further infer about the role of the protein in sirohaem synthesis, a *S. aureus* sirohydrochlorin ferrochelatase mutant strain (*ΔnirR)* was tested on its capacity to reduce nitrite. We observed that the *S. aureus* wild type strain grown anaerobically in nitrite-containing medium reduces the nitrite within approximately 5 h. On the contrary, under similar conditions, the sirohydrochlorin ferrochelatase mutant strain showed no nitrite consumption, which is consistent with the lack of sirohaem formation (Figure 4B).

Altogether, these experiments demonstrated that *S. aureus* NirR is a sirohydrochlorin ferrochelatase and, therefore, it was renamed as ShfC.

Histidines 22 and 87 control the enzyme activity of *S. aureus* ShfC

The active site of sirohydrochlorin ferrochelatase-like proteins has been proposed to include histidine residues. Comparison analysis of the *S. aureus* ShfC amino acid sequence with previously studied sirohydrochlorin ferrochelatases led us to select histidines 22, 87 and 146 as potential amino acid residues involved in controlling its function (Figure 5A). Hence, we have generated, by site-directed mutagenesis, three independent genes in which the codons for these histidines were replaced by leucines. The plasmids containing the mutated genes of *S. aureus* ShfC were introduced into an *E. coli* Δ302a strain expressing *M. thermoautotrophicus sirC* and *P. denitrificans cobA*, and the growth was evaluated

in minimal medium. *E. coli* Δ302a expressing either genes with H22L or the H87L mutations did not revert the cysteine auxotrophy of the strain (Figure 4A), which is consistent with the need of these residues are necessary for the function of the enzyme. Although H146 is highly conserved among sirohydrochlorin ferrochelatase-like proteins, replacement by leucine did not cause alterations on the phenotype (Figure 4A).

To predict the localization of the histidine residues in the structure of *S. aureus* ShfC, a homology-based modelling was used considering as template *S. enterica* CbiK (PDB code: 2XWP), which contains a metallated sirohydrochlorin at the active site [24]. An overlay between the two proteins indicates that *S. aureus* ShfC adopts the typical overall structure of type II class of metal chelatase enzymes with two domains of a mixed α/β architecture (Figure 5B). Analysis of the central cavity of *S. aureus* ShfC where the metallated substrate would be inserted also predicts that H22 and H87 modulate the enzyme activity. Although H146 is located in the vicinity of the active centre it is not expected to be involved in metal binding, which agrees with experimental data indicating that it is not a functionally important residue (Figure 5C).

Distribution of sirohaem biosynthesis pathways among bacteria

The enzymatic reactions forming the three step sirohaem pathway are performed by one, two or three independent enzymes, depending on the organism. CysG denomination has been used for several sirohaem-related proteins that exhibit quite different activities. For this reason, we proposed a new set of abbreviations that correlates with the enzymes´ activity. In this nomenclature, CysG is used for the multifunctional enzymes that combine three activities in a modular architecture and different combination of the core domains. UroM designates proteins with S-adenosyl-L-methionine (SAM)-dependent uroporphyrinogen III (uro´gen-III) methyltransferase activity, P2D stands for precorrin-2 dehydrogenase enzymes, and ShfC describes sirohydrochlorin chelatases (Table 2).

We also searched for homologous of the sirohaem biosynthesis proteins in prokaryotic complete and annotated genomes. A comprehensive analysis of the amino acid sequences of 4856 complete bacterial genomes was done using BlastP tool [25] and several characterized or bona-fide enzymes as queries (see Materials and methods). The large dataset obtained was further refined using as selection criteria the presence of the core domains, sequence length and alignment coverage with the different bona-fide queries.

The presence of two independent enzymes, one with an UroM activity and another with a P2D-ShfC fusion, which until now was only observed for eukaryotic yeast, also occurs in bacteria. Therefore, considering that the synthesis of sirohaem can take place via one, two or three enzymes, we propose to name the correspondent pathways as Type 1, 2 and 3, respectively (Figure 6A). Type 1 refers to a pathway performed by the trifunctional enzyme CysG, which contains the SUMT, dehydrogenation and

ferrochelation activities in a single polypeptide. Type 2 refers to a pathway formed by 2 independent enzymes, such as that occurring in organisms that contain an enzyme with the UroM domain and another enzyme with a P2D-ShfC fusion motif. Finally, the Type 3 pathway involves three proteins, having each one a single activity (Figure 6A). Within the organisms that perform sirohaem biosynthesis via the Type 3 pathway, the chelatase operative in the last step may be performed by four different enzymes, namely SirB, $CbiX^L$, $CbiX^S$ and CbiK. The last three proteins have been described as sirohydrochlorin cobaltochelatases but they also exhibit ferrochelatase activity [26,27].

The organization of the genes encoding sirohaem related proteins is quite diverse even within the genus level (Figure 6B). The sirohaem genes of *Staphylococcus* spp. are located in the vicinity of the nitrite reductase operon. However, while in *S. carnosus* these three genes are organized in a single cluster, in *S. aureus* and *S. epidermidis* this cluster only contains two of these genes and the third gene is located elsewhere in the genome. Interestingly, many bacteria apparently contain more than one *uroM*-like gene, such as in *S. carnosus*, *S. epidermidis* and *Bacillus* spp. Sirohaem-related genes are also located in gene clusters that encode sulfite reductase enzymes, which are sirohaem containing proteins.

We next studied the distribution of the 3 types of sirohaem biosynthesis pathways in bacterial organisms considering 4856 genomic records (Figure 7). A complete sirohaem biosynthetic pathway was found in organisms belonging to 15 out of the 31 bacterial phyla surveyed. Among the organisms apparently lacking any sirohaem biosynthesis are the phyla of Chlamydiae, Cyanobacteria, Elusimicrobia, Tenericute and Verrucomicrobia.

The synthesis of sirohaem via Type 1 is the preferable route utilized by bacteria, as it accounts for 33% of the sirohaem synthesizing organisms, while only almost 4% synthesize sirohaem via the Type 2, and 9% use Type 3. In general, the Type 1 pathway is mainly present in Corynebacteriales (29%), Betaproteobacteria (31%), Alphaproteobacteria (52%), Gammaproteobacteria (80%) and Streptomycetales (91%) order of the Actinobacteria phyla. Noteworthy, some Gamma and Alphaproteobacteria harbours two or even 3 *cysG* like genes, as it is the case of some members of the Aeromonadales order. Exceptions in these phyla occur in Vibrionales and Caulobacterales order that contain more organisms with enzymes of the Type 2 pathway, and in *Shewanella* genus with all genomes housing enzymes for a Type 2 pathway.

A bacterial version of the Type 2 route, similar to that present in yeast, occurs in Fibrobacteres phylum (100%), Vibrionales (98%), almost all organisms from the actinobacterial Propionibacteriaceae family (86%), 80% genomes from the Thermaceae family (Deinococcus-Thermus), Negativicutes (80%), Gammaproteobacteria from the Alteromodales order (35%), and Clostridia (10%, mainly in Desulfitobacterium and Desulfosporosinus genus). Interestingly, the P2D-ShfC of *Propionibacteriaceae* have an extended N-terminal region that shares sequence similarity to the type II chelatase superfamily.

The synthesis of sirohaem via the Type 3 pathway is predicted to occur mainly in Firmicutes of the Baciliales order (77% of 407 organisms such as *Bacillus, Geobacillus, Lynsinibacillus, Staphylococcus*),

Nitrosospira (75%), Spirochaetes (19%), and Planctomycetes (17%). The variant (UroM+P2D+CbiK), *i.e.,* in which the ShfC enzyme is replaced by CbiK, occurs mainly in genomes of the Fusobacteria (79%), Deltaproteobacteria (59%), Chlorobia (36%) and Clostridia (27%) classes, and is present in all of the 76 Listeria genomes surveyed. Moreover, the Type 3 pathway appears sporadically in other phyla.

Discussion

In this work, we have identified and characterized the proteins that perform the biosynthesis of sirohaem in *S. aureus*. The genome contains two *cysG*-like genes whose proteins share amino acid sequence similarity to SirA and SirC of *B. megaterium*. The recombinant *S. aureus* CysG proteins exhibit uroporphyrinogen III methyl-transferase and precorrin-2 dehydrogenase activities that are in the range of values determined for their bacterial homologues. For this reason, the proteins were renamed UroM and P2D, respectively. The initial analysis of the *S. aureus* genome did not retrieve a homolog of a sirohaem ferrochelatase. This led us to hypothesize that *S. aureus* P2D could have a dehydrogenase/chelatase activity similar to what is observed for *E. coli* $CysG^B$ and *S. cerevisiae* Met8p [12,18]. But this was not the case, as *S. aureus* P2D showed no ferrochelatase activity. However, a suitable candidate was found in the *nirR* gene, which according to the genome annotation was predicted to encode a nitrite regulator, but that shares some degree of sequence similarity to the *B. megaterium* SirB. We show that *S. aureus* NirR is not a transcription factor involved in the regulation of the gene cluster where it is inserted, which contains genes encoding two subunits of the nitrite reductase enzyme (NirB and NirD). Moreover, we proved that *S. aureus nirR* encodes a functional sirohydrochlorin ferrochelatase, and therefore it was designated as ShfC. Additionally, ShfC is shown to be required for the nitrite reductase activity of *S. aureus* cells as a mutant strain in this gene is unable of nitrite consumption.

The comparison of the amino acid sequences of several sirohydrochlorin chelatases (Figure 5A) indicates that *S. aureus* ShfC contains histidine residues that are proposed to modulate the chelatase activity and/or be involved in metal binding. For example, two of the residues that in *A. fulgidus* $CbiX^s$ (H10 and H75) bind the metal [24] are also present in *S. aureus* ShfC (H22 and H87) (Figure 5A). The amino acid sequence alignment also indicates that H146 of *S. aureus* ShfC is a conserved residue, which in *D. vulgaris* $Cbik^p$ (H182), *P. pantotrophus* CbiX (H127), and *S. enterica* CbiK modulate the iron chelatase activity (H145) [24, 27-28]. In the absence of structural data for *S. aureus* ShfC, we generated a homology model and performed site-directed mutagenesis studies to investigate in complementation assays the role of the histidines predicted to be functionally important. According to the model, which was generated using the structure of *S. enterica* CbiK, the *S. aureus* ShfC adopts a typical type II chelatase structure with residues H22, H87 and H146 located near the active site. Furthermore, our site-directed mutagenesis results showed that H22 and H87 are essential for the function of the protein.

Although in the model residue H87 is pointing out to the center of the molecule (Figure 5C), it was previously reported for *S. enterica* CbiK that H207, which occupies the same structural position as *S. aureus* ShfC H87, can undergo a structural change, alternating between free metal binding or moving away from the tetrapyrrole, upon sirohydrochlorin binding, to facilitate the insertion of the metal ion into the macrocycle [24]. This would explain why the mutation of H87 by leucine in *S. aureus* ShfC impaired the ferrochelatase activity. The lack of function is most probably related to modification of the catalytic properties of the protein resultant from site-directed mutagenesis. However, it cannot be definitively excluded that H22 and H87 mutations could have impaired the protein's ability to bind to sirohydrochlorin or iron substrates, a study that requires a stable purified protein.

CysG type enzymes are present in enteric bacteria such as *S. enterica* and *E. coli*. Until now, the sirohaem synthesis done by two enzymes was only described for yeast, which uses the bifunctional precorrin-2 dehydrogenase and chelatase enzyme Met8p and uro´gen-III methyltransferase Met1p.

This work revelead that *S. aureus* uses three individual enzymes to transform uro´gen III into sirohaem, as occurs in other bacilli. In higher plants, separated enzymes also exist but, so far, no precorrin-2 dehydrogenase was reported.

Our analysis of the genome organization of the sirohaem genes allow us to conclude that there is a large diversity in the cluster organization among organisms. Despite the amino acid sequence similarities with the *B. megaterium* sirohaem related proteins, the *S. aureus* genes encoding enzymes involved in sirohaem synthesis are not organized in a single operon [19], but instead are widespread in the genome, with *uroM* and *shfC* being only relatively closely located: *uroM* is part of the *nirBD* operon which is located downstream of *nirD*, while *shfC* is located upstream of the *nirBD* operon. The *p2D* gene is found in a different region of the genome. This organization significantly differs from that of *S. carnosus* that contains a gene cluster with three genes (Figure 6B), which were previously referred as being involved in sirohaem synthesis [9].

The distribution of the three types of pathways (Figure 7) shows that Type 1 is the most used route in bacteria. Interestingly, although Cyanobacteria depend on sirohaem for key functions that involve the assimilatory nitrite and sulfite reductase enzymes, no complete sirohaem synthase pathway seems to be present, containing only disperse gene products that share meaningful sequence similarity to ShfC and UroM proteins. It is possible that these cyanobacteria synthesize sirohaem using only these two proteins, similarly to what takes place in higher plants that apparently lack P2D homologues, in which a completely distinct enzyme may be involved in this step. However, in many organisms, *shfC* gene is also not found. It has been proposed that instead of using the long form of the $CbiX^L$ chelatase, these organisms recruited either the short version (*e.g.*, Clostridia – Thermoanaerobacteraceae family) or the CbiK chelatase such as Deltaproteobacteria, Chlorobia, and Clostridium that express sirohaem containing proteins**.**

Organisms that could utilize multiple types of sirohaem pathways are rare and can only be found in *Fimbriimonas ginsengisoli* (Type 1 and Type 2), *Magnetococcus marinus* (Type 2 and Type 3), in the

Desulfuromonadales *Pelobacter carbinolicus* and *Geobacter lovleyi* (Type 1 and Type 3), and in *Photobacterium gaetbulicola* species (Type 1 and Type 2). In many cases, Type 2 or Type 3 pathways change at genus or species level, within a taxa, which suggest that their distribution arose from multiple fusion/fission events. Additionally, *Shewanella* organisms seem to be an example of a fission event at the ancestral of the Gammaproteobacteria as although it uses only the Type 2 pathway, its P2D-ShfC enzyme shares a high amino acid sequence similarity to *E. coli* CysG (Type 1 pathway).

Also, the larger type II chelatases, such as SirB, $CbiX^L$, CbiK and HemH, with the residues forming the active site either located in the N-terminal or C-terminal regions, may result from $CbiX^S$ by a gene duplication and fusion event [24,29]. Noteworthy, although organisms belonging to the Corynebacteriales and Streptomycetales order (Type 1) have some residues previously shown to be important for P2D/ShfC activity, the amino acid sequence region of domain III^B that is normally present in CysG proteins, is not present in the enzymes of these organisms. Thus, more studies will have to be performed to understand if these organisms can really perform the three reaction steps in a single enzyme.

Materials and methods

Bacterial strains, growth conditions and nitrite consumption assays

Strains and plasmids used in this work are described in Table 3. All *E. coli* and *S. aureus* strains were grown aerobically in LB and in Tryptic Soy Broth (TSB) at 37 °C and 150 rpm, respectively. Exception was made for the *E. coli* 302Δa strain, which for the complementation experiments was grown on minimal medium composed of M9Salts (12.8 g/L Na_2HPO_4, 3 g/L KH_2PO_4, 0.5 g/L NaCl, 1 g/LNH_4Cl) and 20 mM glucose, 2 mM $MgSO_4$, 0.1 mM $CaCl_2$, and 1.5% agar. Antibiotic resistance was selected using chloramphenicol (35 μg.ml^{-1}) and ampicillin (100 μg.ml^{-1}). *S. aureus* NE1352 (Δ*nirR*) strain was selected for erythromycin resistance (10 μg.ml^{-1}). Competent cells were prepared essentially as described previously [30].

S. aureus JE2 and *S. aureus* Δ*nirR* were used for RNA extraction from cells grown in TSB, aerobically (OD_{600}=1) in Erlenmeyer flasks at 37°C. The harvested cells were treated with an ice-cold ethanol/phenol RNA protective solution (5%), centrifuged at 2000 *x g* for 5 min, and the pellets flash frozen in liquid nitrogen.

For nitrite consumption assays, overnight cultures of *S. aureus* JE2 and Δ*nirR* cells were inoculated in TSB supplemented with 2 mM of sodium nitrite ($NaNO_2$), to an OD_{600} of 0.1 and grown under anaerobic conditions at 37 ºC and 150 rpm. The growth of *S. aureus* JE2 and Δ*nirR* under was monitored hourly by measuring the optical density at 600 nm in a spectrophotometer (Multiskan™ GO, ThermoFisher Scientific). Every hour, cells of *S. aureus* JE2 and Δ*nirR* were collected, loaded in

in a 96-well plates spectrophotometer, and nitrite concentration was determined by addition of the Griess reagent (1 % sulfanilamide, 0.1% naphthylene diamine dihydrochloride and 2 % H3PO4) and measuring the optical density at 540 nm (Multiskan™ GO, ThermoFisher Scientific). The cellular nitrite levels were determined against a calibration curve

RNA isolation, RT-PCR and quantitative real-time RT-PCR assays

For RNA isolation, cell pellets were thawed on ice, resuspended in 10 mM Tris pH 8 and lysed with 2 mg ml^{-1} lysozyme and 30 µg ml^{-1} lysostaphin, at 37 °C for 30 min. The lysates were transferred to Aurum RNA Binding Mini Columns and total RNA was extracted with Aurum™ Total RNA Mini Kit (Bio-Rad) following the manufacturer´s instructions. Contaminating DNA was removed by treatment with Ambion® TURBO DNA-free™DNase kit (Life Technologies). The concentration, purity and integrity of the RNA were evaluated in a Nanodrop ND-1000 UV–visible spectrophotometer (Thermo Fisher Scientific) and by gel electrophoresis.

For cDNA synthesis, 1 µg of RNA was reverse transcribed with the Transcriptor High Fidelity cDNA Synthesis Kit (Roche) using the Anchored-oligo (dT)18 and Random Hexamer primers. Quantitative real-time RT-PCR reactions were prepared with LightCycler® 480 SYBR Green I Master kit, using the primers listed in Table 4 and done in LightCycler® 480 (Roche). The fold change was calculated using the comparative CT method and 16S rRNA as reference gene. Assays were done for two independent biological samples analysed in triplicate.

Cloning and site-directed mutagenesis experiments

The two *S. aureus cysG* genes, which were denominated *cysG1* (NWMN_2298) and *cysG2* (NWMN_2517), and *nirR* (NWMN_2301) were amplified by PCR using genomic DNA and oligonucleotides described in Table 4. The DNA fragments were digested and ligated to pET-23b (Novagen) plasmids which had been pre-treated with restriction enzymes selected so that His-tag encoding sequences were introduced into each gene (Table 3).

Site-directed mutagenesis was done for *S. aureus* NirR: three independent NirR recombinant proteins were generated each containing a single amino acid residue mutation by substitution of H22, H87 and H146 with hydrophobic and non-aromatic leucine residues. The mutagenesis was carried on pET-23b-SanirR using the primers listed in Table 4 and QuickChange II Site-Directed Mutagenesis Kit (Agilent Technologies) to generate plasmids pET-23b-SanirRH22L, pET-23b-SanirRH87L, and pET-23b-SanirRH146L.

For all wild type and mutated genes, the correct sequence was confirmed by DNA sequencing.

Protein expression and purification

In order to facilitate the purification procedure, *S. aureus* proteins were produced containing His-tag fusions, which were bound either to the N-terminus (CysG1) or to the C-terminus (CysG2 and NirR). *Desulfovibrio vulgaris* HmbS (hydroxymethylbilane synthase) UroS^{d} (an UroS domain that contains uro´gen III synthase activity) and precorrin-2 dehydrogenase SirC His-tagged proteins were expressed from plasmids pET-14b-DvhmbS, pET-14b-DvsirC and pET-14b-DvuroS^{d} which were constructed as described previously [31].

All plasmids, pET-23b-SacysG1, pET-23b-SanirR, pET-23b-SacysG2, pET-14b-DvhmbS, pET-14b-DvsirC and pET-14b-DvuroS^{d}, were transformed separately into *E. coli* BL21STAR(DE3) pLysS competent cells, which were cultured in LB medium to an OD_{600}=0.6. At this stage, expression was induced with 400 µM of isopropyl β-D-1-thiogalactopyranoside (IPTG) and cells grown for 16-20 h, at 20 °C. Cells were harvested by centrifugation (11000 *x g*, 10 min, 4°C), resuspended in 20 mM Tris-HCl buffer pH 8 and lysed in a French press at 1000 psi. Cells were centrifuged twice (48300 *x g*, 10 min, 4°C), and the soluble fraction was applied onto a Ni^{2+} Sepharose fast flow column (GE Healthcare), previously equilibrated with 20 mM Tris-HCl buffer pH 8, 500 mM NaCl and 10 mM imidazole. *S. aureus* CysG1, NirR and CysG2 were eluted in 20 mM Tris-HCl buffer pH 8 with 500 mM NaCl and 500 mM imidazole. The *D. vulgaris* HmbS, UroS^{d} and SirC proteins were eluted in the same buffer but containing only 400 mM imidazole. Protein fractions were concentrated in an Amicon Stirred Ultrafiltration Cell using a 10 kDa membrane (Millipore), and buffer exchanged against 50 mM Tris-HCl buffer pH 8 in PD-10 columns (GE Healthcare). This last step was done under anaerobic conditions (Coy model A-2463 and Belle Technology chambers) to avoid inactivation of the enzymes. The purity and molecular masses of the recombinant proteins were determined by SDS-PAGE, and their concentration was determined spectrophotometrically, at 280 nm, using the following theoretical extinction coefficient (ExPASy software) [32]: ε_{280nm}=22920 M^{-1} cm^{-1} for *S. aureus* CysG2, ε_{280nm}=31665 M^{-1} cm^{-1} for *S. aureus* CysG1 and ε_{280nm}=8480 M^{-1} cm^{-1} for *D. vulgaris* SirC.

Complementation experiments

Plasmid pCIQ-sirCcobA for expression of *Methanothermobacter thermoautotrophicus sirC* and *Pseudomonas denitrificans* uroporphyrinogen III methyltransferase *cobA* genes [26] was introduced into *E. coli* 302Δa, which is a strain that lacks sirohaem synthase *cysG* and, thus, is unable to synthesize

sirohaem. Therefore this strain, that does not reduce sulfite to sulfide due to the absence of sirohaem in the sulfite reductase, only grows in the presence of sulfide or cysteine [23].
Strains of *E. coli* 302Δa (pCIQ-sirCcobA) transformed with pET-23b (negative control) or pKK223.2-cysG [26] were used in the complementation experiments as negative and positive controls of sirohaem synthesis, respectively.
Strains of *E. coli* 302Δa (pCIQ-sirCcobA) co-transformed with pET-23b expressing *S. aureus* ferrochelatase SirB wild type and H22L, H87L and H146L variants were tested for ferrochelatase activity.
In all cases, cell colonies were diluted in PBS to an OD_{600} of 0.1 and platted on minimal medium with or without cysteine (0.05 mg ml^{-1}) and antibiotics, and grown at 37°C, for 20 h.

Enzyme assays

All reactions were performed in solutions that contained 50 mM Tris-HCl buffer pH 8, and enzyme activities were evaluated, at room temperature, in a Shimadzu UV-1800 spectrophotometer inside an anaerobic Coy chamber.

Substrate production

Due to commercial unavailability of uro´gen III and precorrin-2, the two substrates were prepared in the laboratory in reactions done under anaerobic conditions.
For uro´gen III, 5 mg *D. vulgaris* HmbS-His and $UroS^{d}$-His were incubated with 1 mg of porphobilinogen (PBG) (Frontier Scientific) in a final volume of 2 ml. Uro´gen III was separated from the *D. vulgaris* enzymes through passage of the reaction mixtures in a Ni^{2+} Sepharose fast flow columns. Uro´gen III was diluted in 1 M HCl and its concentration was determined spectrophotometrically ($\varepsilon_{405}=5.4\times10^{5}$ M^{-1} cm^{-1}).
Precorrin-2 was produced using the recombinant plasmid pETcoco-2*ABCD* (Table 3), which expresses a His-tagged form of the following enzymes required for its formation from δ-aminolevulinic acid (δ-Ala): *Methanosarcina barkeri* uro´gen III methyltransferase (CobA), *Methanothermobacter thermautotrophicus* porphobilinogen synthase (PbgS), *Bacillus megaterium* hydroxymethylbilane synthase (HmbS) and uro´gen III synthase (UroS). Plasmid pETcoco-2*ABCD* was transformed into *E. coli* BL21(DE3)Gold competent cells, which were grown in LB medium supplemented with 0.2% (w/v) glucose and ampicillin to an OD_{600} of 0.5. At this time, the cells were supplemented with L-arabinose (0.02% w / v), cultured for two more hours, after which 400 μM IPTG was added and cells were grown overnight at 20 ° C. Cells were harvested by centrifugation (11000 *x g*, 10 min, 4°C), the pellets were resuspended in 50 mM Tris-HCl buffer pH 8 with 100 mM NaCl, and disrupted in a French cell at 1000

psi. The soluble fraction collected after centrifugation was buffer exchanged, under anaerobic conditions, against 50 mM Tris-HCl pH 8 using a PD-10 column (GE Healthcare). Finally, precorrin-2 was prepared in a reaction mixture that contained 3 ml of soluble fraction, 2 mg SAM and 2 mg δ-Ala, and that was incubated overnight. Next, the reaction mixture was loaded into a Ni^{2+} Sepharose fast flow column to separate the His-tagged proteins from precorrin-2 formed in the reaction. Due to the lack of the precorrin-2 extinction coefficient value, the quantification of precorrin-2 was done indirectly through the formation of sirohydrochlorin in a reaction coupled to precorrin-2 dehydrogenase activity of *D. vulgaris* SirC [33].

Uroporphyrinogen III methyltransferase and precorrin-2 dehydrogenase activities

The uro´gen III methyltransferase activity of CysG1 was determined by coupling the activity to the precorrin-2 dehydrogenase activity of *D. vulgaris* SirC being the formation of precorrin-2 evaluated indirectly through the production of sirohydrochlorin. The uroporphyrinogen III methyltransferase reaction mixture consisted of 155 µg of *S. aureus* CysG1, 8 µg *D. vulgaris* SirC, 125 µM S-adenosyl-L-methionine (SAM), 100 µM NAD^+ and uro´gen III (0.5-5 µM) in a final volume of 1 ml.
Precorrin-2 dehydrogenase activity was measured by incubation of precorrin-2 (2.5 µM) with several concentrations of NAD^+ (0.05-2 mM) and *S. aureus* CysG2 (4 µg), in a final reaction volume of 1 ml. The assays were done in triplicate.
The two reactions yielded as final product sirohydrochlorin that was quantified by following the increase of the absorbance of the characteristic band at 376 nm (ε_{376}=2.4×10^5 M^{-1} cm^{-1}).
Kinetic parameters were obtained by fitting the activity and substrate concentration values to an allosteric sigmoidal or Michaelis-Menten equation for uroporphyrinogen III methyltransferase and precorrin-2 dehydrogenase activities, respectively, using GraphPad Prism (GraphPad Software, La Jolla, CA, USA).

Homology modelling of *S. aureus* ShfC

The protein structure prediction of *S. aureus* ShfC was done using the SWISS-MODEL [34]. *S. enterica* CbiK (PDB code: (2XWP)) that shares with the amino acid *S. aureus* protein (identity 14% and similarity 32%) was used as template for the construction of the model. The quality of the homology model was examined using SWISS-MODEL workspace and RMSD-derived PyMOL (PyMOL Molecular Graphics System, Version 2.0 Schrödinger, LLC). The sirohydrochlorin molecule was introduced into *S. aureus* ShfC with PyMOL by alignment of the *S. enterica* CbiK crystal structure with the homology model.

Genomic analysis

A dataset of 4856 bacterial complete genomes was downloaded from NCBI and parsed against NCBI taxonomy. The following selected sequences were used as queries: *Bacillus megaterium* WP_116077515.1; *Bacillus subtilis* NP_389445.1 and NP_388210; *Bacillus thuringiensis* HD-771 WP_000521277.1; *Desulfovibrio vulgaris* YP_009955, YP_010682, YP_009872 and YP_010584; *Escherichia coli* NP_417827, NP_418247 and NP_418248; *Methanosarcina barkeri* AAZ70732 and AAZ70418; *Pseudomonas fluorescens* F113 AEV65677; *Saccharomyces cerevisiae* NP_012995 and NP_009772; *Staphylococcus aureus* WP_001014063.1 and AAW36821.1; and *Xanthomonas oryzae* AAW76356. A BLAST [25] search was performed using as a cut of identity of at least 25% and an E-value lower than 10^{-10}. Due to the similarity between the domains present in the single and multifunctional enzymes as well as unrelated enzymes, each retrieved sequence was characterized in terms of PFAM domains (Version 31). The domain architecture, sequence similarity, alignment coverage and sequence length were used to classify each single potential hit into their corresponding protein family by using an in-house Perl script. The genomic protein dataset used in this analysis was retrieved from publicly available repositories (RefSeq) and is available under the accession number(s) given in Supplementary Table S1.

References

1 Simmen H-P & Blaser J (1993) Analysis of pH and pO_2 in abscesses, peritoneal fluid, and drainage fluid in the presence or absence of bacterial infection during and after abdominal surgery. *Am. J. Surg.* **166**, 24–27.

2 Balasubramanian D, Harper L, Shopsin B & Torres VJ (2017) *Staphylococcus aureus* pathogenesis in diverse host environments. *Pathog. Dis.* **75**, 1–13.

3 Burke KA & Lascelles J (1975) Nitrate reductase system in *Staphylococcus aureus* wild type and mutants. *J. Bacteriol.* **123**, 308–316.

4 Schlag S, Fuchs S, Nerz C, Gaupp R, Engelmann S, Liebeke M, Lalk M, Hecker M & Götz F (2008) Characterization of the oxygen-responsive NreABC regulon of *Staphylococcus aureus*. *J. Bacteriol.* **190**, 7847–7858.

5 Evans JB (1975) Uracil and pyruvate requirements for anaerobic growth of staphylococci. *J. Clin. Microbiol.* **2**, 14–17.

6 Fuchs S, Pané-Farré J, Kohler C, Hecker M & Engelmann S (2007) Anaerobic gene expression in

Staphylococcus aureus. *J. Bacteriol.* **189**, 4275–4289.

7 Neubauer H & Götz F (1996) Physiology and interaction of nitrate and nitrite reduction in *Staphylococcus carnosus*. *J. Bacteriol.* **178**, 2005–2009.

8 Pantel I, Lindgren PE, Neubauer H & Götz F (1998) Identification and characterization of the *Staphylococcus carnosus* nitrate reductase operon. *Mol. Gen. Genet.* **259**, 105–114.

9 Neubauer H, Pantel I & Götz F (1999) Molecular characterization of the nitrite-reducing system of *Staphylococcus carnosus*. *J. Bacteriol.* **181**, 1481–1488.

10 Castiglione N, Rinaldo S, Giardina G, Stelitano V & Cutruzzolà F (2012) Nitrite and nitrite reductases: from molecular mechanisms to significance in human health and disease. *Antioxid. Redox Signal.* **17**, 684–716.

11 Spencer JB, Stolowich NJ, Roessner CA & Scott AI (1993) The *Escherichia coli cysG* gene encodes the multifunctional protein, sirohaem synthase. *FEBS Lett.* **335**, 57–60.

12 Warren MJ, Bolt EL, Roessner CA, Scott AI, Spencert JB & Woodcock SC (1994) Gene dissection demonstrates that the *Escherichia coli cysG* gene encodes a multifunctional protein. *Biochem. J* **302**, 837–844.

13 Bali S, Lawrence AD, Lobo SA, Saraiva LM, Golding BT, Palmer DJ, Howard MJ, Ferguson SJ & Warren MJ (2011) Molecular hijacking of sirohaem for the synthesis of haem and d_1 haem. *Proc. Natl. Acad. Sci. U. S. A.* **108**, 18260–5.

14 Kühner M, Haufschildt K, Neumann A, Storbeck S, Streif J & Layer G (2014) The alternative route to haem in the methanogenic archaeon *Methanosarcina barkeri*. *Archaea* **2014**, 1-13.

15 Dailey HA, Dailey TA, Gerdes S, Jahn D, Jahn M, O'Brian MR & Warren MJ (2017) Prokaryotic haem biosynthesis: multiple pathways to a common essential product. *Microbiol. Mol. Biol. Rev.* **81,** e00048–16.

16 Lobo SAL, Lawrence AD, Romão CV., Warren MJ, Teixeira M & Saraiva LM (2014) Characterisation of *Desulfovibrio vulgaris* haem b synthase, a radical SAM family member. *Biochim. Biophys. Acta - Proteins Proteomics* **1844**, 1238–1247.

17 Hansen J, Muldbjerg M, Chérest H & Surdin-Kerjan Y (1997) Sirohaem biosynthesis in *Saccharomyces cerevisiae* requires the products of both the *MET1* and *MET8* genes. *FEBS Lett.* **401**, 20–24.

18 Raux E, McVeigh T, Peters SE, Leustek T & Warren MJ (1999) The role of *Saccharomyces cerevisiae* Met1p and Met8p in sirohaem and cobalamin biosynthesis. *Biochem. J.* **338 (Pt 3**, 701–708.

19 Raux E, Leech HK, Beck R, Schubert HL, Santander PJ, Roessner CA, Scott AI, Martens JH, Jahn D, Thermesr C, Rambach A & Warren MJ (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.

20 Leech HK, Raux-Deery E, Heathcote P & Warren MJ (2002) Production of cobalamin and

sirohaem in *Bacillus megaterium*: an investigation into the role of the branchpoint chelatases sirohydrochlorin ferrochelatase (SirB) and sirohydrochlorin cobalt chelatase (CbiX). *Biochem. Soc. Trans.* **30**, 610–613.

21 Storbeck S, Walther J, Müller J, Parmar V, Schiebel HM, Kemken D, Dülcks T, Warren MJ & Layer G (2009) The *Pseudomonas aeruginosa nirE* gene encodes the S-adenosyl-L-methionine-dependent uroporphyrinogen III methyltransferase required for haem d1biosynthesis. *FEBS J.* **276**, 5973–5982.

22 Lobo SAL, Brindley A, Warren MJ & Saraiva LM (2009) Functional characterization of the early steps of tetrapyrrole biosynthesis and modification in *Desulfovibrio vulgaris* Hildenborough. *Biochem. J.* **420**, 317-326.

23 Kolko MM, Kapetanovich LA & Lawrence JG (2001) Alternative pathways for sirohaem synthesis in *Klebsiella aerogenes*. *J. Bacteriol.* **183**, 328–335.

24 Romão CV, Ladakis D, Lobo SAL, Carrondo MA, Brindley AA, Deery E, Matias PM, Pickersgill RW, Saraiva LM & Warren MJ (2011) Evolution in a family of chelatases facilitated by the introduction of active site asymmetry and protein oligomerization. *Proc. Natl. Acad. Sci.* **108**, 97–102.

25 Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W & Lipman DJ (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25**, 3389–402.

26 Lobo SAL, Brindley AA, Romão C V., Leech HK, Warren MJ & Saraiva LM (2008) Two distinct roles for two functional cobaltochelatases (CbiK) in *Desulfovibrio vulgaris* Hildenborough. *Biochemistry* **47**, 5851–5857.

27 Lobo SAL, Videira MAM, Pacheco I, Wass MN, Warren MJ, Teixeira M, Matias PM, Romão C V. & Saraiva LM (2017) *Desulfovibrio vulgaris* CbiK P cobaltochelatase: evolution of a haem binding protein orchestrated by the incorporation of two histidine residues. *Environ. Microbiol.* **19**, 106–118.

28 Bali S, Rollauer S, Roversi P, Raux-Deery E, Lea SM, Warren MJ & Ferguson SJ (2014) Identification and characterization of the “missing” terminal enzyme for sirohaem biosynthesis in α-proteobacteria. *Mol. Microbiol.* **92**, 153–163.

29 Brindley AA, Raux E, Leech HK, Schubert HL & Warren MJ (2003) A story of chelatase evolution: Identification and characterization of a small 13-15-kDa “ancestral” cobaltochelatase (CbiXs) in the archaea. *J. Biol. Chem.* **278**, 22388–22395.

30 Inoue H, Nojima H & Okayama H (1990) High efficiency transformation of *Escherichia coli* with plasmids. *Gene* **96**, 23–28.

31 Lobo SAL, Scott A, Videira MAM, Winpenny D, Gardner M, Palmer MJ, Schroeder S, Lawrence AD, Parkinson T, Warren MJ & Saraiva LM (2015) *Staphylococcus aureus* haem biosynthesis: characterisation of the enzymes involved in final steps of the pathway. *Mol. Microbiol.* **97**, 472–

487.

32 Tools D, Mirrors S & Contact A (2010) ExPASy Proteomics Server. *Search*, 1–2.

33 Schubert HL, Raux E, Brindley AA, Leech HK, Wilson KS, Hill CP & Warren MJ (2002) The structure of *Saccharomyces cerevisiae* Met8p, a bifunctional dehydrogenase and ferrochelatase. *EMBO J.* **21**, 2068–2075.

34 Arnold K, Bordoli L, Kopp J & Schwede T (2006) The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling. *Bioinformatics* **22**, 195–201.

35 Griffiths L & Cole JA (1987) Lack of redox control of the anaerobically-induced nirB+ gene of *Escherichia coli* K-12. *Arch. Microbiol.* **147**, 364–369.

36 Lorenz LL & Duthie ES (1952) Staphylococcal coagulase:mode of action and antigenicity. *Microbiology* **6**, 95–107.

37 Frank S, Deery E, Brindley AA, Leech HK, Lawrence A, Heathcote P, Schubert HL, Brocklehurst K, Rigby SEJ, Warren MJ & Pickersgill RW (2007) Elucidation of substrate specificity in the cobalamin (vitamin B12) biosynthetic methyltransferases. Structure and function of the C20 methyltransferase (CbiL) from *Methanothermobacter thermautotrophicus*. *J. Biol. Chem.* **282**, 23957–23969.

38 Edgar RC (2004) MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32**, 1792–1797.

Tables

Table 1. Amino acid sequence identity and similarity of CysG1, CysG2 and NirR encoded in the *S. aureus* genome with previously studied microbial uro´gen III methyltransferase, precorrin-2 dehydrogenase, and sirohydrochlorin ferrochelatase enzymes, respectively. Accession numbers: *S. aureus* CysG1 (WP_000109968.1); *B. megaterium* SirA (AAA22317.1); *Pseudomonas (P.) denitrificans* CobA (AAA25773.1); *E. coli* CysG C-terminal 216-457 (WP_000349855.1); *Paracoccus (Pa.) denitrificans* CobA (WP_011748768.1); *S. aureus* CysG2 (A0A0H3KAM1); *B. megaterium* SirC (CAD48923); *E. coli* CysG N-terminal 1-223 (WP_000349855.1); *Desulfovibrio (D.) vulgaris* SirC (YP_010682); *Saccharomyces (Sac.) cerevisae* Met8p (NP_009772.1); *S. aureus* NirR (YP_001333335); *B. megaterium* SirB (CAD48922); *Synechocystis sp.* CbiX (BAA10794.1); *Archaeoglobus (A.) fulgidus* CbiX (WP_010878224.1); and *Paracoccus (Pa.) pantotrophus* CbiX (A0A023GPI5)

	Identity (%)	Similarity (%)
***S. aureus* CysG1**		
B. megaterium SirA	44	62
P. denitrificans CobA	37	54
E. coli CysG C-terminal	38	55
Pa. denitrificans CobA	37	52
***S. aureus* CysG2**		
B. megaterium SirC	35	61
E. coli CysG N-terminal	27	50
D. vulgaris SirC	23	42
Sach. cerevisae Met8p	22	37
***S. aureus* NirR**		
B. megaterium SirB	25	43
Synechocystis sp. CbiX	23	40
A. fulgidus CbiX	29	53
Pa. pantotrophus CbiX	25	42

Table 2. Names and abbreviations for bacterial sirohaem biosynthesis enzymes

Enzyme function	Old abbreviation	New abbreviation
Multifunctional sirohaem synthase	CysG	CysG
SAM-dependent uroporphyrinogen III methyltransferase	SirA/CobA/SUMT/CysG	UroM
Precorrin-2 dehydrogenase	CysG/SirC	P2D
Sirohydrochorin ferrochelatase	CysG/SirB/CbiX/CbiK	ShfC

Table 3. Strains and plasmids used in this work

Strain	Genotype	Application	Resistance	Reference
E. coli				
E. coli BL21(DE3)Gold	F⁻ *ompT hsdS$_B$*($r_B^-m_B^-$) *gal dcm*	Recombinant protein expression	NA	Novagen
E. coli BL21STAR(DE3)pLysS	F⁻ *ompT hsdS$_B$*($r_B^-m_B^-$) *gal dcm*	Recombinant protein expression	Cm^R	Novagen
E. coli 302Δa	Nir^S, Lac^+, Cys^-; *cysG*⁻	cysG⁻ strain (cysteine auxothrophic strain) used in complementation experiments	NA	[34]
E. coli XL1-Blue	*recA1 endA1 gyrA96 thi-1 hsdR17 supE44 relA1 lac [F′ proAB lacI^q ZΔM15 Tn10 (Tet^R)*	Host strain used for genetic manipulation	Tet^R	Laboratory strain
S. aureus				
S. aureus Newman	Wild type Laboratory strain	Used for gDNA PCR amplification	NA	[35]
S. aureus JE2	plasmid-cured derivative of strain LAC	Used as wild type strain in q/RT-PCR and nitrite consumption assays	NA	Nebraska Transposon Mutant Library (UNMC)
S. aureus NE1352 (Δ*nirR*)		Used in q/RT-PCR and nitrite consumption assays		Nebraska Transposon Mutant Library (UNMC)
Plasmid				
pETcoco-2ABCD	*M. barkeri cobA, M.thermautotrophicus pbgS, B. megaterium hmbS and uroS*	Plasmid for production of precorrin-2	Amp^R	M.J Warren´s lab
pETcoco-2ABCDC	Same as pETcoco-2*ABCD* plus *M. thermautotrophicus sirC*	Plasmid for production of sirohydrochlorin	Amp^R	[36]
pET-23b-SacysG1	*S. aureus cysG1*	*cysG1* fused to His-tag at N-terminus, for SUMT activity testing	Amp^R	This study
pET-23b-SanirR	*S. aureus nirR*	*nirR* fused to His-tag at C-terminus, for complementation experiments	Amp^R	This study
pET-23b-SacysG2	*S. aureus cysG2*	*cysG2* fused to His-tag at C-terminus, for precorrin-2 dehydrogenase activity determination	Amp^R	This study
pET-14b-DvhmbS	*D. vulgaris hmbS*	Production of uroporphyrinogen III	Amp^R	[21]
pET-14b-DvuroS^d	*D. vulgaris uroS^d*	Production of uroporphyrinogen III	Amp^R	[21]
pET-14b-DvsirC	*D. vulgaris sirC*	SUMT activities	Amp^R	[21]
pKK223.2-cysG	*E. coli cysG*	Positive control in complementation experiments	Amp^R	[25]
pCIQ-sirCcobA	*M.thermoautotrophicus sirC and P.denitrificans cobA*	Complementation experiments	Cm^R	[25]
pET-23b-SanirRH22L	*S. aureus nirR*	*nirR* H22 mutant for complementation experiments	Amp^R	This study
pET-23b-SanirRH87L	*S. aureus nirR*	*nirR* H87 mutant for for complementation experiments	Amp^R	This study

pET-23b-SanirRH146L	*S. aureus nirR*	*nirR* H146 mutant for for complementation experiments	Amp^R	This study

NA, not applicable

Table 4. Oligonucleotides used in this work

Primer	Oligonucleotides 5´-> 3´	Restriction Sites
Cloning		
cysG1_Fwd	GAGCTAGCCACCACCACCACCACCACATGTCTGTAGAGGAATATG	NheI
cysG1_Rev	CAAAGCTTTTTACTAGTTTAGTGACATAACACTGTATTAG	HindIII
nirR_Fwd	CACAAAGCTAGCATGATTTGGTATACAATAC	NheI
nirR_Rev	CGTTAGTGCTCGAGTATTTTCATTGGAATC	XhoI
cysG2_Fwd	AAGGAGCTAGCTCAACATGAATATGCCATTAATG	NheI
cysG2_Rev	CGCTACTCGAGTCTTACATCCAACCACGCTA	XhoI
Mutagenesis		
nirRH22L_Fwd	GGAATATCATTGTTGCACTGGGCATGAGGCACGGACGA	
nirRH22L_Rev	TCGTCCGTGCCTCATGCCCAGTGCAACAATGATATTCC	
nirRH87L_Fwd	CCATTGCTAATCTTTAGTGCAATGCTGTATCTCAAGGATATACCGAATATCG	
nirRH87L_Rev	CGATATTCGGTATATCCTTGAGATACAGCATTGCACTAAAGATTAGCAATGG	
nirRH146L_Fwd	AGTTGGAGTTATGGTTGTTGCACTGGGCAATATTAATGGAAAGTTTAC	
nirRH146L_Rev	GTAAACTTTCCATTAATATTGCCCAGTGCAACAACCATAACTCCAACT	
q/RT-PCR		
cysG1_RT_Fwd	AGCAGCGCGTCGATATAACAAG	
cysG1_RT_Rev	CGACCAAATATCGCTGGATCGC	
nirB_RT_Fwd	ACACGACACGACTTGGCATTCG	
nirB_RT_Rev	TTCTTGGGCAGCCTGATACACC	

Figure Legends

Figure 1. Outline of the three reactions necessary to synthesize sirohaem.

Figure 2. Analysis of the amino acid sequence of *S. aureus* CysG proteins.
(A) *S. aureus* CysG1 sequence alignment with the following bacterial SUMT proteins: *S. aureus* CysG1 (WP_000109968.1); *B. megaterium* SirA (AAA22317.1); *P. denitrificans* CobA (AAA25773.1.); *E. coli* CysG C-terminal 216-457 (WP_000349855.1); *Pa. denitrificans* CobA (WP_011748768.1). (B) *S. aureus* CysG2 sequence alignment with selected precorrin-2 dehydrogenases: *S. aureus* CysG2 (YP_001333335); *B. megaterium* SirC (CAD48923); *D. vulgaris* SirC (YP_010682); *E. coli* CysG N-terminal 1-223 (WP_000349855.1); and *Sac. cerevisae* Met8p (NP_009772.1). The NAD^+ binding motif is marked with asterisks.
Accession numbers are indicated within parenthesis, and proteins were aligned using MUSCLE [38]. Colours represent the degree of conservation among residues from dark blue (highest) to light blue (lowest).

Figure 3. Enzyme activity of *S. aureus* CysG1 and CysG2.
(A) SDS-PAGE of purified *S. aureus* CysG1 and CysG2. Proteins were analysed on a 12% gel. Lanes 1-3 depict the molecular marker, and proteins CysG1 and CysG2, respectively.
(B) UV-Visible spectrum of precorrin-2 generated by incubation of uro´gen III (20 μM) with SAM (40 μM) and CysG1 (0.7 mg) (green) and precorrin-2 generated by the enzymes expressed from plasmid pETcoco-2*ABCD* (red), as described in the Materials and Methods.
(C) *S. aureus* CysG1 activity was measured by coupling with the reaction for precorrin-2 dehydrogenase activity, and using *D. vulgaris* SirC, NAD^+, SAM and uro´gen III.
(D) UV-Visible spectrum of sirohydrochlorin obtained upon incubation of precorrin-2 (2.5 μM) with NAD^+ (160 μM) and CysG2 (6 μg) (green) and precorrin-2, which was generated by the enzymes expressed from plasmid pETcoco-2*ABCDC* (red).
(E) *S. aureus* CysG2 activity was evaluated in a reaction containing as substrates precorrin-2 and NAD^+ (0.05-2 mM).
In (C) and (E), data was fitted to an allosteric sigmoidal equation and a Michaelis-Menten equation, respectively. Error bars represent the standard deviation of the activity measured in triplicate.

Figure 4. Sirohydrochlorin ferrochelatase (ShfC) is essential for nitrite reduction.
(A) Complementation experiments were done in *E. coli* 302ΔapCIQ-SirCCobA cells transformed with plasmids harboring native and mutated versions of *S. aureus* ShfC.

E. coli 302ΔapCIQ-SirCCobA strain was transformed with *E. coli* CysG (pKK-*cysG*, positive control) pET-23b expressing ShfC in the wild type form (pET-*nirR*), and mutated in histidines 22, 87 and 146 (pET-SanirRH22L, pET-SanirRH87L, pET-SanirRH146L) and pET-23b alone (pET-23b). The same strain carrying pKK223.2-*cysG* and empty pET-23b plasmids was used as positive and negative control, respectively. (B1) Growth and (B2) nitrite consumption of *S. aureus* wild type and *ΔnirR* cells in medium supplemented with $NaNO_2$ (2 mM), measured every hour. In (B), the error bars represent the standard deviation of n value = 4.

Figure 5. Structural features of histidines required for *S. aureus* ShfC ferrochelatase activity.
(A) Alignment of *S. aureus* NirR amino acid sequence with the best studied bacterial homologs.
Alignment of *S. aureus* NirR with selected bacterial SirB/CbiX enzymes, namely: *S. aureus* NirR (YP_001333335); *B. megaterium* SirB (CAD48922); *S. enterica* CbiK (NP_460970.1); *D. vulgaris* CbiKP (WP_010937953.1); *A. fulgidus* CbiX (WP_010878224.1); and *P. pantotrophus* CbiX (A0A023GPI5). Proteins were aligned using MUSCLE [38]. Colors represent the degree of conservation among residues from dark blue (highest) to light blue (lowest). Asterisks represent histidine amino acids previously shown to be involved in the sirohydrochlorin ferrochelatase activity in selected SirB/CbiX homologs.
(B) Modelling of *S. aureus* ShfC using *S. enterica* CbiK (2XWP) and SWISS-Model.
(C) Inset depicts a close up of the active site where structurally homologous residues are highlighted. *S. aureus* ShfC is represented in green, *S. enterica* CbiK in turquoise, and metallated sirohydrochlorin in pink. (B) and (C) show structures obtained with PyMOL.

Figure 6. Classification of the sirohaem biosynthesis pathways, and genome organization of the related genes.
(A) Classification of the bacterial sirohaem pathways. Figure depicts the enzyme activity modules that form proteins acting in sirohaem pathways: UroM - S-adenosyl-L-methionine uroporphyrinogen III methyltransferase (blue); P2D- precorrin-2 dehydrogenase (green); ShfC - sirohydrochlorin ferrochelatase (orange). Type 1 pathway includes CysG-like enzymes that have three activities in a single polypeptide. UroM refers to SirA and CobA enzymes. ShfC designates SirB, CbiX and CbiK enzymes with sirohydrochlorin ferrochelatase activity. P2D-ShfC describes enzymes with precorrin-2 dehydrogenase and sirohydrochlorin ferrochelatase activities in a single polypeptide chain (Table 2).
(B) Organization of genes encoding enzymes putatively involved in sirohaem biosynthesis in selected bacteria. Colored genes are named according to nomenclature described in (A), while genes coloured grey are named as annotated in KEGG database or by gene locus position.

Figure 7. Distribution of the three main routes for sirohaem biosynthesis among bacteria. The 4856 organisms analysed were grouped into 146 orders that are represented in the central cladogram obtained using the NCBI common tree tool and the taxonomic information given in Supplementary Information Table S1. The outer rectangles in blue, orange and green represent the normalized occurrence of the sirohaem pathway Types 1, 2 and 3 respectively within a taxon (scale bar on the left). Only organisms with a complete pathway were considered. The complete table with genomic accession codes, retrieved from publicly available repositories (RefSeq), and presence and absence of proteins per genome is given in Supplementary Information (Table S1).

Acknowledgments

We are grateful to Cátia Família for the technical support. 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 grants PTDC/BBB-BQB/5069/2014 and PTDC/BIA-BQM/28642/2017 from Fundação para a Ciência e a Tecnologia. This work has also received funding from the European Union's Horizon 2020 research and innovation program under grant agreement number 810856.

Author contributions

MAMV and LMS designed the research and interpreted the data. SALL designed the research. MAMV performed the experimental part. Genomic data and analysis were carried out by FLS. MAMV and LMS wrote the paper with contributions from the other co-authors.

Conflicts of Interest

There are no conflicts of interest.

Supporting Information

Supplementary Table S1- Complete distribution of sirohaem biosynthesis enzymes among bacteria.

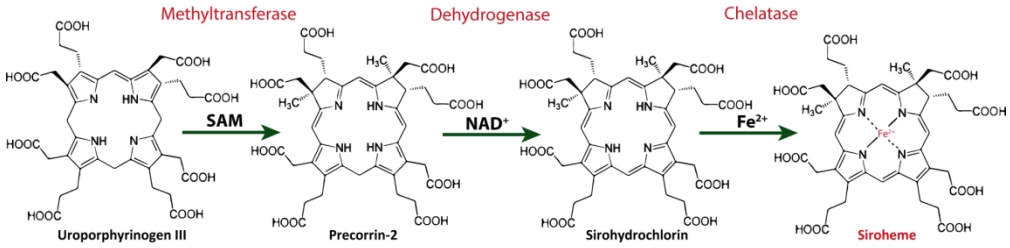

Figure 1. Outline of the three reactions necessary to synthesize sirohaem.

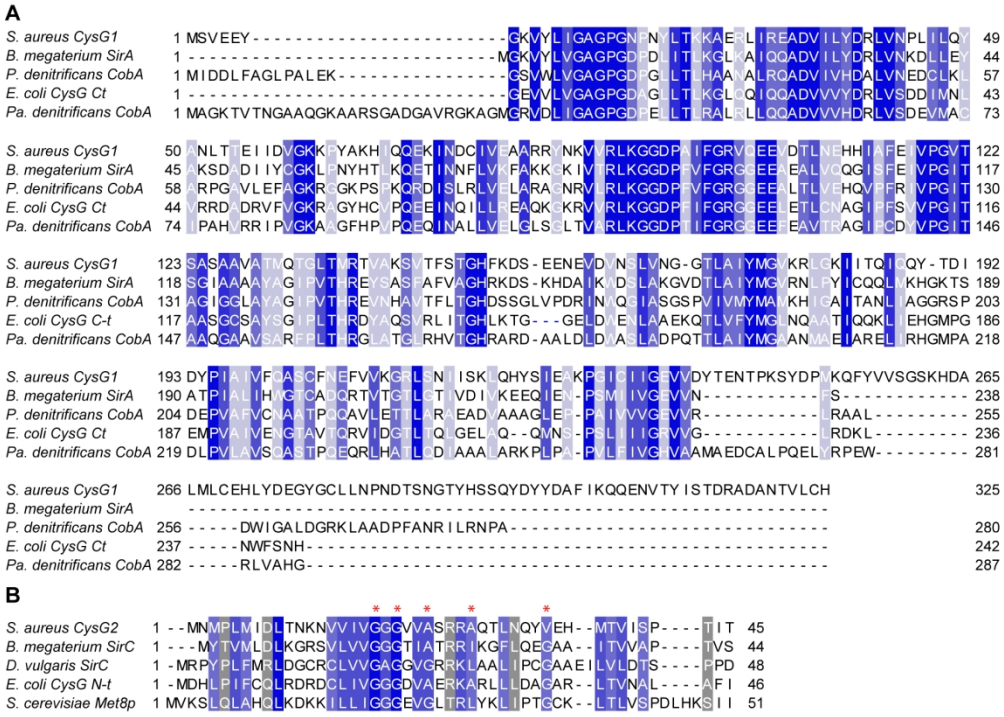

Figure 2. Analysis of the amino acid sequence of S. aureus CysG proteins.
(A) S. aureus CysG1 sequence alignment with the following bacterial SUMT proteins: S. aureus CysG1 (WP_000109968.1); B. megaterium SirA (AAA22317.1); P. denitrificans CobA (AAA25773.1.); E. coli CysG C-terminal 216-457 (WP_000349855.1); Pa. denitrificans CobA (WP_011748768.1). (B) S. aureus CysG2 sequence alignment with selected precorrin-2 dehydrogenases: S. aureus CysG2 (YP_001333335); B. megaterium SirC (CAD48923); D. vulgaris SirC (YP_010682); E. coli CysG N-terminal 1-223 (WP_000349855.1); and Sac. cerevisae Met8p (NP_009772.1). The NAD+ binding motif is marked with asterisks.
Accession numbers are indicated within parenthesis, and proteins were aligned using MUSCLE [38]. Colours represent the degree of conservation among residues from dark blue (highest) to light blue (lowest).

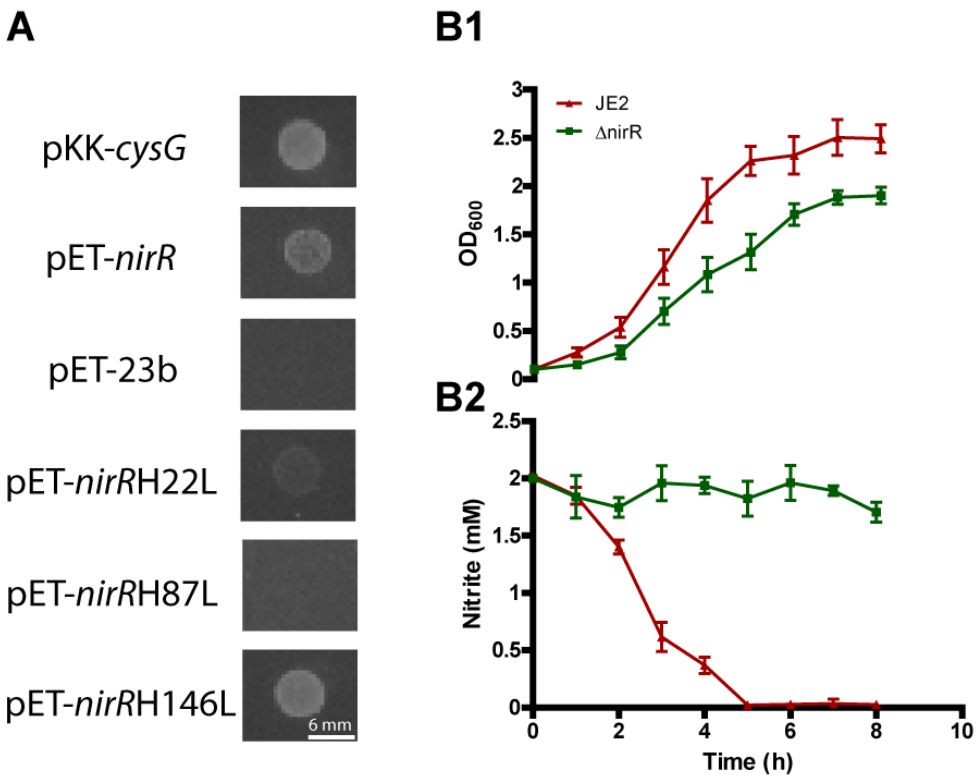

Figure 4. Sirohydrochlorin ferrochelatase (ShfC) is essential for nitrite reduction.
(A) Complementation experiments were done in E. coli 302ΔapCIQ-SirCCobA cells transformed with plasmids harboring native and mutated versions of S. aureus ShfC.
E. coli 302ΔapCIQ-SirCCobA strain was transformed with E. coli CysG (pKK-cysG, positive control) pET-23b expressing ShfC in the wild type form (pET-nirR), and mutated in histidines 22, 87 and 146 (pET-SanirRH22L, pET-SanirRH87L, pET-SanirRH146L) and pET-23b alone (pET-23b). The same strain carrying pKK223.2-cysG and empty pET-23b plasmids was used as positive and negative control, respectively. (B1) Growth and (B2) nitrite consumption of S. aureus wild type and ΔnirR cells in medium supplemented with NaNO2 (2 mM), measured every hour. In (B), the error bars represent the standard deviation of n value = 4.

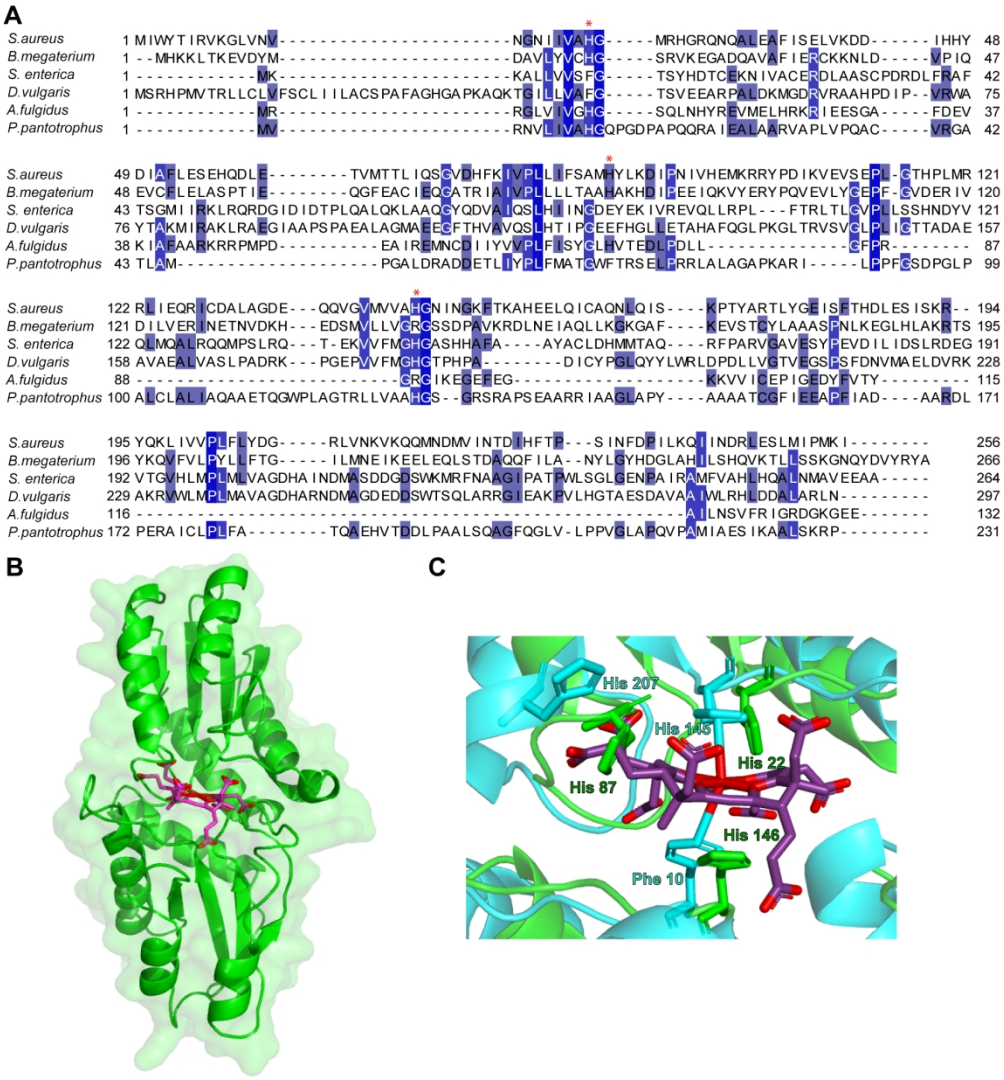

Figure 5. Structural features of histidines required for S. aureus ShfC ferrochelatase activity.
(A) Alignment of S. aureus NirR amino acid sequence with the best studied bacterial homologs. Alignment of S. aureus NirR with selected bacterial SirB/CbiX enzymes, namely: S. aureus NirR (YP_001333335); B. megaterium SirB (CAD48922); S. enterica CbiK (NP_460970.1); D. vulgaris CbiKP (WP_010937953.1); A. fulgidus CbiX (WP_010878224.1); and P. pantotrophus CbiX (A0A023GPI5). Proteins were aligned using MUSCLE [38]. Colors represent the degree of conservation among residues from dark blue (highest) to light blue (lowest). Asterisks represent histidine amino acids previously shown to be involved in the sirohydrochlorin ferrochelatase activity in selected SirB/CbiX homologs.
(B) Modelling of S. aureus ShfC using S. enterica CbiK (2XWP) and SWISS-Model.
(C) Inset depicts a close up of the active site where structurally homologous residues are highlighted. S. aureus ShfC is represented in green, S. enterica CbiK in turquoise, and metallated sirohydrochlorin in pink. (B) and (C) show structures obtained with PyMOL.

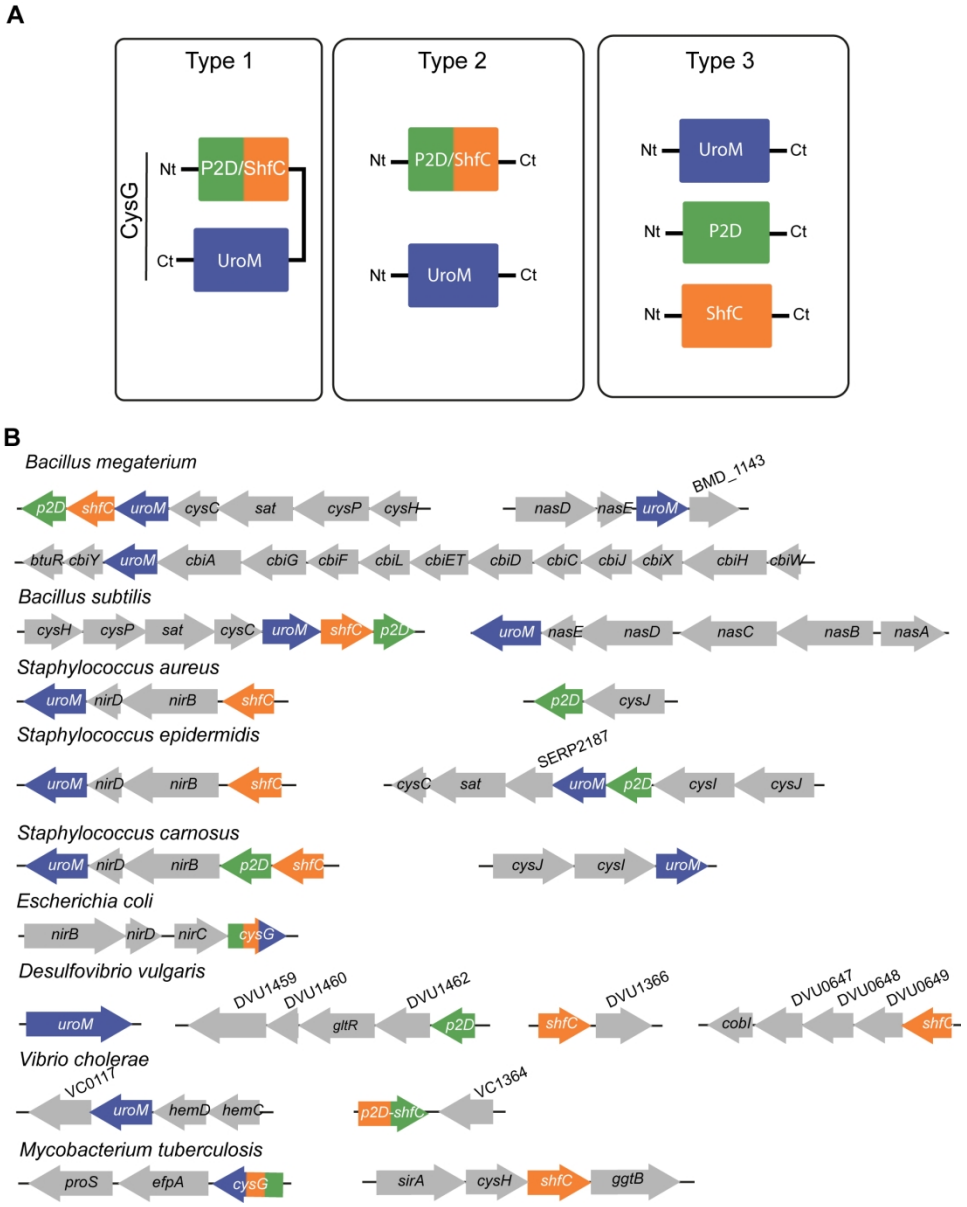

Figure 6. Classification of the sirohaem biosynthesis pathways, and genome organization of the related genes.

(A) Classification of the bacterial sirohaem pathways. Figure depicts the enzyme activity modules that form proteins acting in sirohaem pathways: UroM - S-adenosyl-L-methionine uroporphyrinogen III methyltransferase (blue); P2D- precorrin-2 dehydrogenase (green); ShfC - sirohydrochlorin ferrochelatase (orange). Type 1 pathway includes CysG-like enzymes that have three activities in a single polypeptide. UroM refers to SirA and CobA enzymes. ShfC designates SirB, CbiX and CbiK enzymes with sirohydrochlorin ferrochelatase activity. P2D-ShfC describes enzymes with precorrin-2 dehydrogenase and sirohydrochlorin ferrochelatase activities in a single polypeptide chain (Table 2).

(B) Organization of genes encoding enzymes putatively involved in sirohaem biosynthesis in selected bacteria. Colored genes are named according to nomenclature described in (A), while genes coloured grey are named as annotated in KEGG database or by gene locus position.

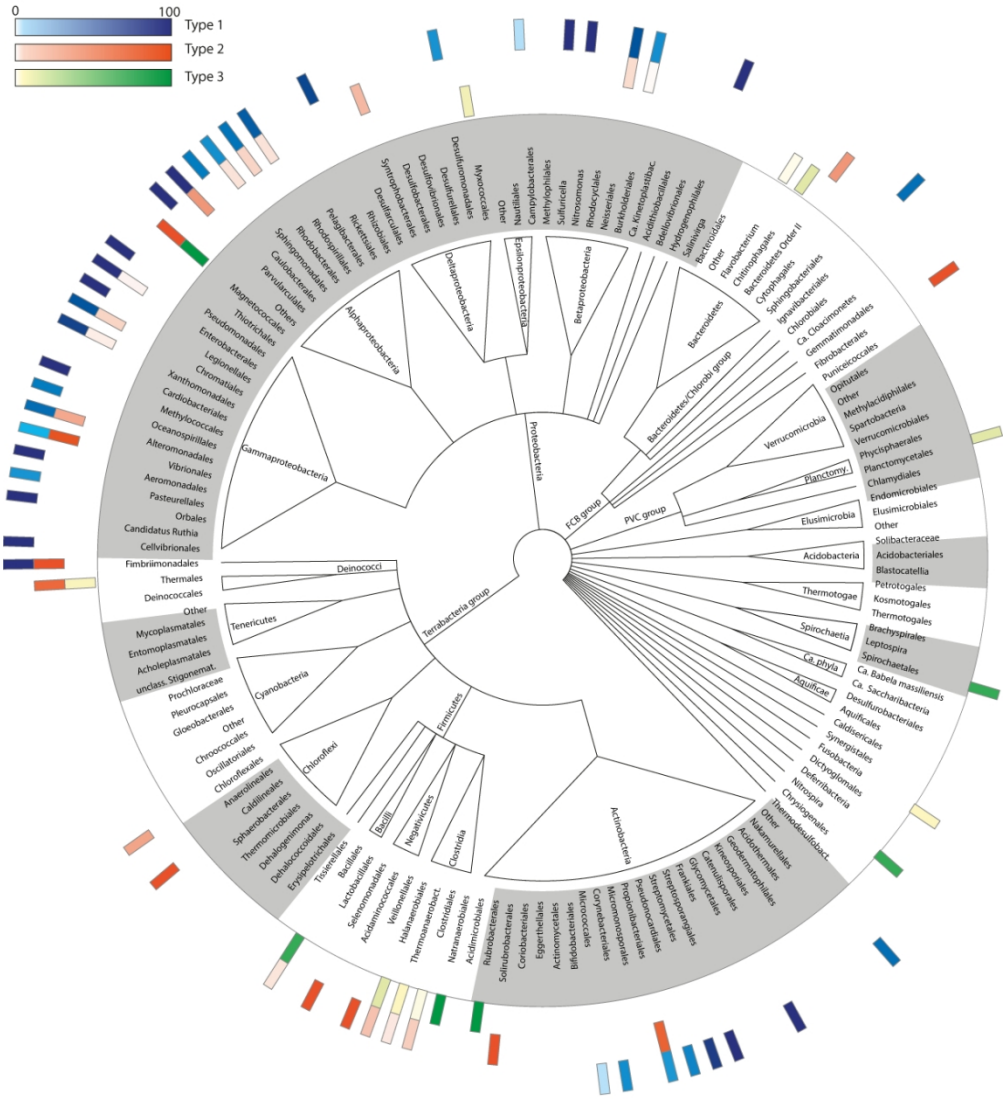

Figure 7. Distribution of the three main routes for sirohaem biosynthesis among bacteria. The 4856 organisms analysed were grouped into 146 orders that are represented in the central cladogram obtained using the NCBI common tree tool and the taxonomic information given in Supplementary Information Table S1. The outer rectangles in blue, orange and green represent the normalized occurrence of the sirohaem pathway Types 1, 2 and 3 respectively within a taxon (scale bar on the left). Only organisms with a complete pathway were considered. The complete table with genomic accession codes, retrieved from publicly available repositories (RefSeq), and presence and absence of proteins per genome is given in Supplementary Information (Table S1).