

**REVIEW
ARTICLE****A whole genome view of prokaryotic haem biosynthesis**

Heather Panek and Mark R. O'Brian

Author for correspondence: Mark R. O'Brian. Tel: +1 716 829 3200. Fax: +1 716 829 2725.
e-mail: mrobrian@buffalo.edu

Department of Biochemistry and Witebsky Center for Microbial Pathogenesis, State University of New York at Buffalo, Buffalo, NY 14214, USA

Keywords: tetrapyrroles, porphyrins, *Archaea*, degenerate genomes

Overview

Haem is the prosthetic group of proteins involved in many processes, and is also a regulatory molecule that mediates cellular responses to oxygen, iron and haem itself. Haem is synthesized in a multistep biosynthetic pathway with well-defined intermediates. Accordingly, the enzymes of the pathway are, to a large extent, conserved among prokaryotes and eukaryotes. Although individual haem biosynthesis genes and enzymes have been studied in many prokaryotic systems, only in few organisms has the entire pathway been examined. The availability of whole genome sequences allows an initial snapshot of the entire pathway at the genetic level in many diverse organisms. Unlike model systems, the genome represents a starting point for the detailed characterization of some unusual prokaryotes, and therefore examining haem synthesis genes can give important insight into core metabolism on which to lay the foundation for future experiments.

Haem function

Haems are involved in many facets of oxidative metabolism, including O₂ transport, electron-transport-dependent oxidative phosphorylation, oxidative stress responses, O₂-sensing, oxygenation reactions and detoxification. They are also involved in the synthesis or sensing of other diatomic gases such as carbon monoxide and nitric oxide (Rodgers, 1999). As a regulatory molecule, haem mediates gene expression at the level of transcription (Schmitt, 1999; Hach *et al.*, 1999; Ogawa *et al.*, 2001; Reddy *et al.*, 1996), translation (Chen & London, 1995), protein targeting (Lathrop & Timko, 1993), protein stability (Qi *et al.*, 1999; Wang *et al.*, 1999) and differentiation (Nakajima *et al.*, 1999).

Cytochromes are quantitatively the predominant haem proteins in most prokaryotes, and are essential for respiration. Prokaryotes can carry out oxidative respiration in the absence of oxygen using nitrate, sulfate or other oxidants. Cytochromes mediate electron transfer

and the final reduction of the terminal electron acceptor as occurs with O₂-based respiration. Also, some prokaryotes respond to hypoxia by expressing cytochrome oxidases with high affinity for oxygen, and therefore oxygen limitation actually increases the demand for haem synthesis. Thus the need for haem is not restricted to aerobic metabolism at ambient O₂ levels as it is in eukaryotes; rather, haem contributes to metabolic adaptation. Indeed, haems are found throughout the *Bacteria* and *Archaea*, and in most cases the demand is met by biosynthesis rather than by acquisition from the environment.

Haem biosynthetic pathway

Haems are tetrapyrroles, a family of compounds that includes chlorophylls, chlorins, corrins and bilins. Intermediates of the haem pathway may be incorporated into other tetrapyrroles depending on the organism (Fig. 1). The first universal tetrapyrrole precursor, δ -aminolevulinic acid (ALA), is synthesized either from glutamate by the C₅ pathway or from glycine and succinyl coenzyme A via ALA synthase (Fig. 2). No prokaryote is known to possess both routes of ALA formation. The synthesis of uroporphyrinogen III from ALA requires three successive enzymic steps catalysed by ALA dehydratase, porphobilinogen deaminase and uroporphyrinogen III synthase with conserved intermediate products (Fig. 1). Uroporphyrinogen III is the final common precursor for all tetrapyrroles and thus presents the major branch point in the haem pathway. Thus organisms that synthesize only non-haem tetrapyrroles contain genes necessary for uroporphyrinogen III synthesis, but lack the genes that catalyse the subsequent steps of the haem pathway.

For haem and chlorophyll synthesis, uroporphyrinogen III is metabolized by three successive enzymic steps that modify the side groups of the macrocycle to yield protoporphyrin (Fig. 1).

The two oxidative steps catalysed by coproporphyrino-

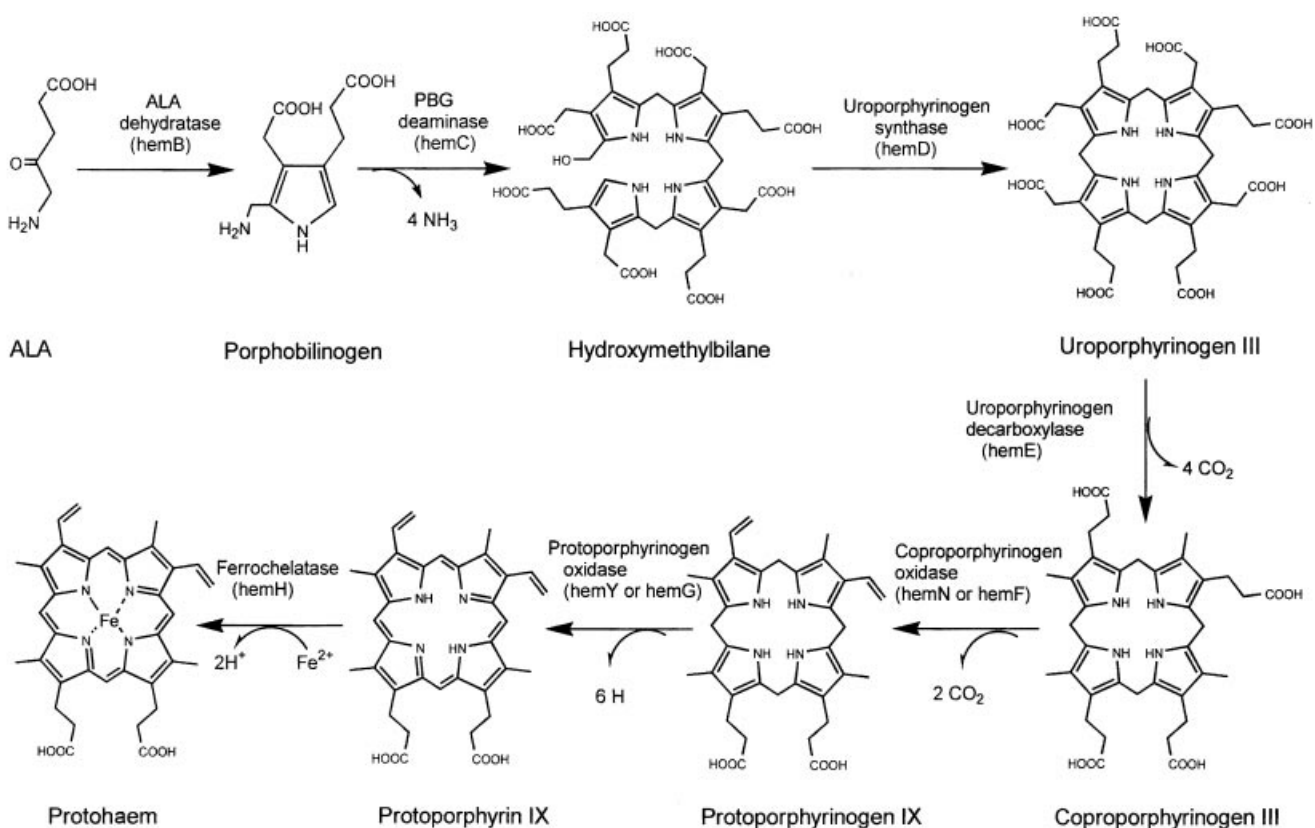


Fig. 1. Haem biosynthetic pathway from δ -aminolevulinic acid (ALA) to protohaem. The enzymes catalysing each step are noted, and the genes encoding them are denoted in parentheses.

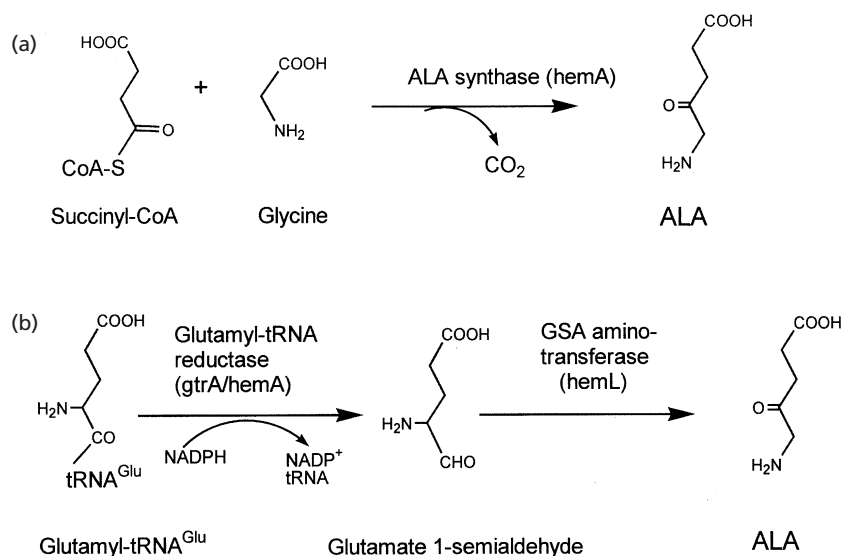


Fig. 2. Two routes of ALA synthesis by ALA synthase (a) and the C₅ pathway (b). The enzymes and genes are noted. Glutamyl-tRNA formation is catalysed by glutamyl-tRNA synthetase (not shown), which is also involved in protein synthesis and therefore not a committed step in ALA synthesis.

gen III oxidase and protoporphyrinogen IX oxidase use oxygen as a substrate in eukaryotes. Orthologues of those eukaryotic genes are found only in some prokaryotes, which in part reflects the fact that haem synthesis can occur in the absence of oxygen. Protoporphyrin

chelates ferrous iron to form protohaem, which is catalysed by ferrochelatase. Protohaem is the prosthetic group of numerous haem proteins, and is modified further to synthesize other types of haems found in cytochrome *c* and terminal oxidases.

Table 1. Prokaryotic enzymes and genes in haem biosynthesis

Enzyme	Gene
δ -Aminolevulinic acid synthase	<i>hemA</i>
Glutamyl-tRNA reductase	<i>gtrA</i> (<i>hemA</i>)*
Glutamate-1-semialdehyde aminotransferase†	<i>hemL</i>
δ -Aminolevulinic acid dehydratase (porphobilinogen synthase)	<i>hemB</i>
Porphobilinogen deaminase (hydroxymethylbilane synthase)	<i>hemC</i>
Uroporphyrinogen III synthase	<i>hemD</i>
Uroporphyrinogen III decarboxylase	<i>hemE</i>
Coproporphyrinogen III oxidase, O ₂ -dependent	<i>hemF</i>
Coproporphyrinogen III oxidase, O ₂ -independent	<i>hemN</i>
Protoporphyrinogen IX oxidase	<i>hemG</i>
Protoporphyrinogen IX oxidase	<i>hemY</i>
Ferrochelatase	<i>hemH</i>

* Both glutamyl-tRNA reductase and ALA synthase have been designated *hemA* but are unrelated enzymes. We recommend *hemA* and *gtrA* as designated.

† This enzyme is often annotated as glutamate semialdehyde 2,1-aminomutase. The enzyme catalyses an intermolecular amino transfer reaction between two glutamate 1-semialdehyde (GSA) molecules and therefore it is not a mutase reaction.

Gene designations

Table 1 lists the names of the genes encoding the enzymes of the haem biosynthetic pathway. The *hemA* designation has been ascribed to both the gene encoding ALA synthase and to that encoding the C₅ pathway glutamyl-tRNA reductase. Although both enzymes are involved in ALA formation, they are structurally dissimilar proteins that have different substrates and products, and therefore should not have the same gene name. We recommend that the ALA synthase gene be designated *hemA* and the gene encoding glutamyl-tRNA reductase as *gtrA*. *hemG* and *hemY* encode distinct protoporphyrinogen oxidase proteins, but the porphyrin substrate and product are the same for both enzymes. The oxygen-dependent and oxygen-independent coproporphyrinogen oxidases are dissimilar proteins encoded by *hemF* and *hemN*, respectively. A second O₂-independent coproporphyrinogen oxidase gene homologue has been identified in numerous bacteria and is called *hemZ* in *Rhodobacter sphaeroides* (Zeilstra-Ryalls & Kaplan, 1995; Yeliseev & Kaplan, 1999) and *Bacillus subtilis* (Homuth *et al.*, 1999). There is no evidence to suggest that the *hemZ* genes are functionally similar to each other, and thus the *hemZ* designation may be confusing.

Approaches used in evaluating haem synthesis gene homologues

During the initial perusal of haem synthesis genes, it became readily apparent that some assignments in the databases are incorrect, resulting in both a failure to identify homologues, and the erroneous assignment of

genes as encoding haem pathway enzymes. The Kyoto Encyclopedia of Genes and Genomes (KEGG, www.genome.ad.jp/kegg) allows a quick view of metabolic pathways in organisms based on genome sequences and is useful, but it relies on author annotation and therefore is incomplete. The approach taken here was to perform BLAST (Altschul *et al.*, 1990) alignments against the entire National Center for Biotechnology Information (NCBI, www.ncbi.nlm.nih.gov) database and against specific genomes within that database using proteins shown experimentally to be haem biosynthesis enzymes. In many cases, the enzymes from *Escherichia coli* or *B. subtilis* were used. At this writing, 69 prokaryotic genomes have been sequenced and are listed in Table 2.

Prokaryotic ALA synthase is found only in the α -proteobacteria

ALA synthase is the sole source of ALA formation in animals, fungi and non-photosynthetic eukaryotes as a whole. In prokaryotes, however, ALA synthase has only been found in the α -proteobacteria, and the C₅ pathway is found in all other groups that make ALA. The α -proteobacteria are a diverse taxonomic group that includes photosynthetic bacteria, obligate intracellular pathogens of eukaryotes, and plant symbionts. The bacterial ancestor of mitochondria also belongs to this taxonomic group, and eukaryotic ALA synthases were likely acquired by the ancient endosymbiotic event leading to mitochondria. Examination of the sequenced genomes in the dataset clearly confirms the restriction of ALA synthase to the α -proteobacteria. All seven α -proteobacterial genomes sequenced (six separate genera)

Table 2. Occurrence of haem biosynthesis gene homologues in prokaryotes in which the complete genome sequence has been determined

The organisms are grouped taxonomically. The letter designations A, B, C, etc., refer to *hemA*, *hemB*, *hemC*, and so on.

	Genes present	Genes absent but expected
Archaea, Crenarchaeota		
<i>Aeropyrum pernix</i> K1	<i>gtrA</i> , L, B, C	D, E, N, G or Y, H
<i>Sulfolobus solfataricus</i>	<i>gtrA</i> , L, B, C, D	E, N, G or Y, H
<i>Sulfolobus tokodaii</i>	<i>gtrA</i> , L, B, C, D	E, N, G or Y, H
<i>Pyrobaculum aerophilum</i>	<i>gtrA</i> , L, B, C	D, E, N, G or Y, H
Archaea, Euryarchaeota		
<i>Archaeoglobus fulgidus</i>	<i>gtrA</i> , L, B, C, D	
<i>Halobacterium</i> sp. NRC-1	<i>gtrA</i> , L, B, C, D	E, N, G or Y, H
<i>Methanobacterium thermoautotrophicum</i>	<i>gtrA</i> , L, B, C, D	
<i>Methanococcus jannaschii</i>	<i>gtrA</i> , L, B, C, D	
<i>Pyrococcus abyssi</i>	None	
<i>Pyrococcus horikoshii</i>	None	
<i>Thermoplasma acidophilum</i>	<i>gtrA</i> , L, B, C, E	D, N, G or Y, H
<i>Thermoplasma volcanium</i>	<i>gtrA</i> , L, B, C, D, E, N, H	G or Y
Bacteria, Gram-negative, α-proteobacteria		
<i>Agrobacterium tumefaciens</i>	A, B, C, D, E, F, N, H	G or Y
<i>Brucella melitensis</i>	A, B, C, D, E, F, N, H	G or Y
<i>Caulobacter crescentus</i>	A, B, C, E, F, N, H	D, G or Y
<i>Mesorhizobium loti</i>	A, B, C, D, E, F, N, G, H	
<i>Rickettsia conorii</i>	A, B, C, E, F, N, H	D, G or Y
<i>Rickettsia prowazekii</i>	A, B, C, E, F, N, H	D, G or Y
<i>Sinorhizobium meliloti</i>	A, B, C, D, E, F, N, H	G or Y
Bacteria, Gram-negative, β-proteobacteria		
<i>Neisseria meningitidis</i> (MC58 and Z2491)	<i>gtrA</i> , L, B, C, D, E, N, H	G or Y
<i>Ralstonia solanacearum</i>	<i>gtrA</i> , L, B, C, D, E, F, N, H	G or Y
Bacteria, Gram-negative, δ/ϵ-proteobacteria		
<i>Campylobacter jejuni</i>	<i>gtrA</i> , L, B, C, D, E, N, H	G or Y
<i>Helicobacter pylori</i> (26695 and J99)	<i>gtrA</i> , L, B, C, D, E, N, H	G or Y
Bacteria, Gram-negative, γ-proteobacteria		
<i>Buchnera</i> sp. APS	C	
<i>Escherichia coli</i> (K-12, O157:H7 EDL933 and O157:H7)	<i>gtrA</i> , L, B, C, D, E, F, N, G, H	
<i>Haemophilus influenzae</i>	H	
<i>Pasteurella multocida</i>	<i>gtrA</i> , L, B, C, D, E, N, G, H	
<i>Pseudomonas aeruginosa</i>	<i>gtrA</i> , L, B, C, D, E, F, N, H	G or Y
<i>Salmonella typhimurium</i>	<i>gtrA</i> , L, B, C, D, E, F, N, G, H	
<i>Salmonella typhi</i>	<i>gtrA</i> , L, B, C, D, E, F, N, G, H	
<i>Vibrio cholerae</i>	<i>gtrA</i> , L, B, C, D, E, F, N, G, H	
<i>Xylella fastidiosa</i>	<i>gtrA</i> , L, B, C, D, E, F, N, H	G or Y
<i>Yersinia pestis</i>	<i>gtrA</i> , L, B, C, D, E, F, N, G, H	
Bacteria, Gram-positive,		
Bacillus/Clostridium		
<i>Bacillus halodurans</i>	<i>gtrA</i> , L, B, C, D, E, N, Y, H	
<i>Bacillus subtilis</i>	<i>gtrA</i> , L, B, C, D, E, N, Y, H	
<i>Clostridium acetobutylicum</i>	<i>gtrA</i> , L, B, C, D	
<i>Clostridium perfringens</i>	<i>gtrA</i> , L, B, C, D	
<i>Lactococcus lactis</i> subsp. <i>lactis</i>	None	
<i>Listeria innocua</i>	<i>gtrA</i> , L, B, C, D, E, N, Y, H	

Table 2 (cont.)

	Genes present	Genes absent but expected
<i>Listeria monocytogenes</i>	<i>gtrA</i> , L, B, C, D, E, N, Y, H	
<i>Mycoplasma genitalium</i>	None	
<i>Mycoplasma pneumoniae</i>	None	
<i>Mycoplasma pulmonis</i>	None	
<i>Staphylococcus aureus</i> (N315 and Mu50)	<i>gtrA</i> , L, B, C, D, E, N, Y, H	
<i>Streptococcus pneumoniae</i> (TIGR4 and R6)	None	
<i>Streptococcus pyogenes</i>	None	
<i>Ureaplasma urealyticum</i>	None	
Bacteria, Gram-positive, Actinobacteria		
<i>Mycobacterium tuberculosis</i> (H37Rv and CDC1551)	<i>gtrA</i> , L, B, C, D, E, N, Y, H	
<i>Mycobacterium leprae</i>	<i>gtrA</i> , L, B, C, D, E, N, Y, H	
Bacteria, Chlamydia		
<i>Chlamydophila pneumoniae</i> (CWL029, AR39, J138)	<i>gtrA</i> , L, B, C, E, N, Y, H	D
<i>Chlamydia trachomatis</i>	<i>gtrA</i> , L, B, C, E, N, Y, H	D
<i>Chlamydia muridarum</i>	<i>gtrA</i> , L, B, C, E, N, Y, H	D
Bacteria, Aquificales		
<i>Aquifex aeolicus</i>	<i>gtrA</i> , L, B, C, D, E, N, Y, H	
Bacteria, Cyanobacteria		
<i>Nostoc</i> sp. PCC 7120	<i>gtrA</i> , L, B, C, D, E, F, N, H	G or Y
<i>Synechocystis</i> PCC 6803	<i>gtrA</i> , L, B, C, D, E, F, N, H	G or Y
Bacteria, Spirochaetes		
<i>Borrelia burgdorferi</i>	None	
<i>Treponema pallidum</i>	None	
Bacteria, Thermus/Deinococcus		
<i>Deinococcus radiodurans</i> R1	<i>gtrA</i> , L, B, C, D, E, N, Y, H	
Bacteria, Thermotogales		
<i>Thermotoga maritima</i>	None	

contain an ALA synthase gene, whereas none outside that taxon contain a homologue.

In addition to ALA synthase, the α -proteobacteria *Agrobacterium tumefaciens*, *Caulobacter crescentus*, *Mesorhizobium loti* and *Sinorhizobium meliloti* also contain a gene encoding a product with high similarity to the C_5 pathway enzyme glutamate-1-semialdehyde aminotransferase (GSAT), and the *C. crescentus* and *S. meliloti* gene products are annotated as GSAT in the databases. However, these proteins are highly homologous to aminotransferases in general. For example, *E. coli* GSAT shares 48% similarity with the *S. meliloti* protein annotated as GSAT (ID no. 1560716), but it also shares 49% similarity with an *S. meliloti* 4-aminobutyrate aminotransferase and high homology with other aminotransferases as well. More importantly, none of the α -proteobacteria have the other C_5 pathway gene encoding glutamyl-tRNA reductase. Finally, C_5 pathway enzyme activity could not be found in the α -proteobacteria *Bradyrhizobium japonicum* (Sangwan & O'Brian, 1991) or *Azorhizobium caulinodans* (Avissar et

al., 1989). Therefore, it is very unlikely that the C_5 pathway operates in those organisms.

Structural heterogeneity of uroporphyrinogen III synthases

Uroporphyrinogen III synthase, the product of the *hemD* gene in bacteria, synthesizes uroporphyrinogen III from the unstable linear tetrapyrrole hydroxymethylbilane. In the absence of uroporphyrinogen III synthase, hydroxymethylbilane readily cyclizes non-enzymically to the I isomer of uroporphyrinogen, which is not physiologically useful. The *hemD* gene is best characterized in *E. coli* and *B. subtilis*. Examination of *hemD* genes in whole genomes indicates that prokaryotic uroporphyrinogen synthases are more disparate from each other at the primary sequence level than are other haem synthesis enzymes, and that misannotation of some genes masks the fact that *hemD* homologues cannot be found in some genomes.

An intriguing *hemD*-like gene is found in the strict

anaerobe *Clostridium josui* that is an apparent gene fusion of the *cobA* gene with *hemD* predicted to encode a hybrid protein (Fujino *et al.*, 1995). The *cobA* gene encodes uroporphyrinogen III methyltransferase, an activity involved in cobalamin and sirohaem synthesis, but not in haem biosynthesis. The *cobA/hemD* fusion can complement an *E. coli hemD* mutant, showing that it has uroporphyrinogen III synthase activity. This unusual *cobA/hemD* fusion is also observed in the other clostridial species *Clostridium acetobutylicum* and *Clostridium perfringens*, as well as in *Synechocystis* sp. PCC 6803, *Nostoc* sp. PCC 7120 and in the *Mycobacterium* species.

Some organisms do not have a distinct CobA protein, but rather have a multifunctional protein called sirohaem synthase (CysG), which carries out the methyltransferase reaction as well as two additional reactions for sirohaem formation from uroporphyrinogen III. CysG has a C-terminal domain very homologous to CobA in accordance with its activity. The relevance of this to haem synthesis is noted below.

The *cobA/hemD* gene fusion originally described in *C. josui* has led to the incorrect annotation of other genes identified in genome sequencing projects. Some homologues of *cysG* or *cobA* have been annotated as *hemD* or as encoding uroporphyrinogen III synthase when there is no similarity to that protein. Some *hemD* genes have been annotated as encoding a synthase/methyltransferase as well. When these errors are taken into account, it is clear that some organisms predicted to synthesize tetrapyrroles do not contain a recognizable *hemD* gene. This issue was addressed systematically by searching the genomes for those that contain both the *hemB* and *hemC* gene, but not a *hemD* gene. The rationale for this is that the enzymes necessary for uroporphyrinogen III synthesis, including uroporphyrinogen III synthase, should be found in organisms that make tetrapyrroles since uroporphyrinogen is the branch point for all of them. Thus *hemB*, *hemC* and *hemD* genes should be present in all tetrapyrrole-synthesizing prokaryotes, even those that do not make haem (there are two ALA synthesis routes, thus those genes will not be universal).

BLAST searches of each genome containing *hemB* and *hemC* genes were carried out using bona fide uroporphyrinogen III synthases from *E. coli*, *B. subtilis* and mouse (protein id nos 41676, 399875 and 1363253, respectively). This search identified nine taxonomically diverse species from seven genera that contain no identifiable uroporphyrinogen III synthase (*hemD*) gene. *Rickettsia prowazekii*, *Rickettsia conorii* and *Caulobacter crescentus* are α -proteobacteria. *Aeropyrum pernix*, *Pyrobaculum aerophilum* and *Thermoplasma acidophilum* are archaea. *Chlamydophila pneumoniae* (strains AR39, J138, CWL029), *Chlamydia muridarum* and *Chlamydia trachomatis* are chlamydiaceae.

The inability to identify *hemD* homologues in these prokaryotes indicates an enzyme or enzymes that differ substantially in primary structure from known uroporphyrinogen III synthases that carry out that reaction. We

note that prokaryotic *hemD* gene products have been shown to be bona fide uroporphyrinogen III synthases only in a few species, and the model enzymes from *E. coli* and *B. subtilis* share only 25% identity to each other. Moreover, a uroporphyrinogen III synthase gene has not been identified in any plant, nor is a recognizable homologue found in the *Arabidopsis thaliana* genome. This is surprising considering the large chlorophyll requirement in plants and the successful purification of the protein from wheat (Higuchi & Bogorad, 1975).

Oxygen-dependent coproporphyrinogen oxidase is restricted to few eubacterial groups

Coproporphyrinogen oxidase catalyses the oxidative decarboxylation of two of the four propionate side groups of coproporphyrinogen to form protoporphyrinogen. Eukaryotic coproporphyrinogen oxidases are structurally conserved and require O₂ for the oxidative reaction. Several bacterial species also contain an O₂-dependent coproporphyrinogen oxidase, a protein approximately 300 amino acids in length that shares 60–70% similarity with the plant and animal oxidases, and about 45% similarity with the yeast enzyme. This protein is encoded by the *hemF* gene in bacteria. The prevalence of the O₂-dependent oxidase in eukaryotes and its early discovery in *E. coli* and *Salmonella typhimurium* may give the impression that this is a common haem synthesis enzyme. However, the *hemF* gene is relatively restricted, and is only found in several subclasses of the *Proteobacteria* and in cyanobacteria. However, an O₂-independent coproporphyrinogen oxidase encoded by the *hemN* gene is prevalent in prokaryotes that synthesize haem, even among those that contain a *hemF* gene. The *hemN* gene product is dissimilar from the O₂-dependent enzyme in its primary sequence even though the porphyrinogen substrates and products are identical in both cases. Early work indicated that NADP⁺ is the oxidant used by the O₂-independent enzyme (Keithly & Nadler, 1983; Seehra *et al.*, 1983), but Xu & Elliott (1994) could not demonstrate that activity from the *S. typhimurium hemN* product, thus identity of the physiological oxidizing substrate remains uncertain.

Absence of an identifiable protoporphyrinogen oxidase gene homologue in many prokaryotes

Protoporphyrinogen oxidase catalyses the six electron oxidation of protoporphyrinogen to protoporphyrin in the penultimate step of the haem pathway. The molecular characterization of bacterial protoporphyrinogen oxidase has focused primarily on *E. coli* and *B. subtilis*, which express entirely different oxidases. The 51 kDa protoporphyrinogen oxidase protein encoded by *hemY* in *B. subtilis* (Hansson & Hederstedt, 1992) is homologous to the oxidase in eukaryotes, and uses O₂ as the terminal electron acceptor. The *E. coli hemG* gene locus

was identified by analysis of a mutant that accumulates protoporphyrinogen and is defective in protoporphyrinogen oxidase activity (Sasarman *et al.*, 1979). The *hemG* gene expressed *in trans* from a plasmid complements a *hemG* mutant with respect to activity in crude extract and membrane preparations (Sasarman *et al.*, 1993). Enzyme activity of the purified 21 kDa product has not been demonstrated, therefore *hemG* may not be sufficient for protoporphyrinogen oxidation, and *hemG* could encode a subunit of a larger complex. We note that a gene named *hemY* is found in *E. coli* and related organisms and is often annotated as a protoporphyrinogen oxidase gene. However, this gene is dissimilar to bona fide protoporphyrinogen-oxidase-encoding *hemY* genes, and its role in haem synthesis has not been demonstrated.

The light sensitivity of an *E. coli hemH* (ferrochelatase) gene was reversed by a second mutation in a gene named *hemK* (Nakayashiki *et al.*, 1995), found in the putative *hemA-prfA-hemK* operon. The *hemK* mutant accumulates coproporphyrin and protoporphyrin when growth medium was supplemented with ALA, suggesting to the authors that *hemK* may be deficient in protoporphyrinogen oxidase activity. However, a mutation in a *hemK* homologue in yeast is not deficient in protoporphyrinogen oxidase, nor does overexpression of the *hemK* homologue confer activity (Le Guen *et al.*, 1999). Furthermore, HemK-related proteins have a consensus motif for S-adenosyl-methionine-dependent methyltransferases, and may be involved in that function. Finally, *hemK* homologues are found in many genomes, including those of *Borrelia burgdorferi* and *Treponema pallidum*, organisms that do not make haem or contain bona fide haem synthesis genes. From this, it is unlikely that *hemK* encodes a protoporphyrinogen oxidase, hence there must be another cause for the phenotype of the *E. coli hemK* strain. Nevertheless, many *hemK* homologues are annotated as encoding protoporphyrinogen oxidase.

We searched for *hemG* or *hemY* homologues in the genomes of organisms that contain the *hemE*, *hemN* and *hemH* genes, as would be expected in organisms that synthesize haem. The genomes of 34 species and 28 genera contain *hemE*, *hemN* and *hemH* and are either known to express haem proteins, or are predicted to based on identified genes (Table 2). *hemG* is found in six genera within the γ -Proteobacteria, and in *Mesorhizobium loti*, an α -proteobacterium. The *hemY* gene is found in eight genera that are taxonomically more distributed than is observed for *hemG*.

Interestingly, of the 28 genera examined, 14 of them contain no recognizable protoporphyrinogen oxidase gene (Table 2). This suggests a major gap in our understanding of haem biosynthetic pathways, even in organisms that are well studied such as *Pseudomonas aeruginosa*, *Caulobacter crescentus*, *Campylobacter jejuni* and the rhizobia. In addition to numerous Gram-negative bacteria, organisms that contain haem biosynthesis genes but not protoporphyrinogen oxidase also include the cyanobacteria *Synechocystis* PCC 6803

and *Nostoc* sp. PCC 7120, and the archaeon *Thermoplasma volcanium*.

Unlike uroporphyrinogen III synthase, the reaction catalysed by protoporphyrinogen oxidase can proceed non-enzymically *in vitro*, formally raising the possibility that this enzyme is not required in organisms where the gene has not been found. However, protoporphyrinogen oxidase mutants of *E. coli* (Sasarman *et al.*, 1993) and *B. subtilis* (Homuth *et al.*, 1999) are haem-defective. It seems more likely that a heretofore unidentified protoporphyrinogen oxidase exists in prokaryotes, or a known protein has an additional activity. In support of the latter idea, a secondary site suppressor mutant of a *hemG* strain has elevated promoter activity of *hemF*, the gene encoding the O₂-dependent coproporphyrinogen oxidase (Narita *et al.*, 1999). Furthermore, overexpression of *hemF* rescues a *hemG* mutant. However, it remains to be established whether the *hemF* product normally participates in protoporphyrinogen oxidation in wild-type cells.

Haem pathway genes corresponding to the branch from uroporphyrinogen III to protohaem are missing in some archaea that contain haem proteins

The absence of identifiable haem synthesis genes in an organism known to contain haem raises the possibility of an alternative synthetic pathway. In the cases where a single enzyme cannot be identified by homology (see above), it is likely that a similar reaction is carried out by a structurally disparate enzyme since both the substrate and product of the missing step can be accounted for. However, genome sequencing projects reveal cases where numerous enzymes catalysing successive steps of haem formation cannot be found in organisms that express haem enzymes.

The archaea *Sulfolobus acidocaldarius* and *Sulfolobus solfataricus* respire aerobically, contain haem proteins (Becker & Schafer, 1991; Lübben & Morand, 1994; Lübben *et al.*, 1994; Wright *et al.*, 1996; Yano *et al.*, 2000), and the genome of *S. solfataricus* indicates a plethora of haem-protein-encoding genes. However, our own homology searches as well as the author annotations find genes corresponding to the haem pathway only to uroporphyrinogen III formation. The final four genes, *hemE*, *hemN*, *hemY* (or *hemG*) and *hemH*, cannot be found. Similarly, the latter four enzymes of the pathway are unaccounted for in *Thermoplasma acidophilum* and in the extreme halophilic archaeon *Halobacterium* sp. NRC-1 despite the presence of haem proteins and haem protein gene homologues in those genera (Denda *et al.*, 1991, 1995; Fujiwara *et al.*, 1993; Gartner, 1991; Hallberg Gradin & Colmsjo, 1989). Although we could find no literature citations that identify haem proteins from *Aeropyrum pernix* or *Pyrobaculum aerophilum*, they are aerobic thermophiles with many haem-protein-encoding genes, but they also lack identifiable genes that metabolize uroporphyrinogen III to haem.

The discrepancy between a haem protein requirement and the apparent lack of enzymes that metabolize uroporphyrinogen III to haem is thus far confined to aerobic extremophilic archaea. However, the haem genes are observed in the thermophilic proteobacterium *Aquifex aeolicus* and in the thermophilic archaeon *Thermoplasma volcanium* (except for protoporphyrinogen oxidase, but see above) and thus neither thermophily nor taxonomy places an intrinsic restriction upon the expression of a conventional haem biosynthetic pathway.

The ecological niches of these extremophiles do not provide an obvious source of exogenous haem for uptake, and therefore it is very probable that haem is synthesized *de novo*. Even if they possess the alternative route for coproporphyrinogen III synthesis from uroporphyrinogen III as described for *Desulfovibrio vulgaris* (Akutsu *et al.*, 1993; Ishida *et al.*, 1998), the subsequent oxidations and iron chelation steps cannot be deduced based on homology. It seems likely that these organisms contain heretofore uncharacterized genes involved in haem biosynthesis. Further work is needed to determine whether the enzymology and intermediates differ from the conventional pathway.

Degenerate genomes

Some bacteria that form obligate associations with eukaryotes, either pathogenic or symbiotic, have very small genomes. Genome degeneracy is likely caused by the lack of selection pressure to maintain genes that the host renders non-essential. Thus the bacterial genome may reflect the organism's metabolic relationship with its eukaryotic host. *Haemophilus influenzae*, the organism from which the first complete genome sequence was learned, derives its genus name from its reliance on exogenous haem for growth ('haem loving'). *H. influenzae* can also use protoporphyrin and, accordingly, it contains only the ferrochelatase-encoding gene.

Buchnera

Buchnera is an obligate endosymbiont of aphids (Baumann *et al.*, 1995) with a genome of only 590 genes. The *Buchnera* sp. APS genome encodes the four subunits for the cytochrome *o* complex, as well as protohaem farnesyltransferase, the enzyme that places a farnesylethyl group on protohaem to form the cytochrome *o* prosthetic group. In addition, the genome encodes sulfite reductase, which has sirohaem as the prosthetic group, and also the enzymes that synthesize sirohaem from uroporphyrinogen. However, *Buchnera* contains only a homologue of *hemC*, the gene encoding porphobilinogen deaminase, and a truncated *hemN* homologue encoding O₂-independent coproporphyrinogen oxidase. The exchange of nutrients between *Buchnera* and its aphid host is documented (Baumann *et al.*, 1995), and analysis of the bacterial genome raises the possibility that the aphid may provide *Buchnera* with haem or haem intermediates. Similarly, deletion of the *hemA* gene in *Bradyrhizobium japonicum* does not abolish

haem synthesis when the organism is in association with soybean because the eukaryotic host can provide the endosymbiont with ALA (Sangwan & O'Brian, 1991).

Rickettsia

The *Rickettsia* are obligate intracellular bacteria responsible for epidemic typhus in humans. They belong to the α -proteobacteria, and taxonomically are the closest known relatives of mitochondria. The genomes of *R. prowazekii* and *R. conorii* indicate the presence of a haem biosynthetic pathway, as well as respiratory haem proteins, superoxide dismutase, and TCA cycle enzymes, which are all consistent with aerobic metabolism and similar to the other known α -proteobacterial genomes. However, other aspects of iron metabolism are absent, most notably genes encoding proteins for iron transport and regulation.

The Fur protein is a global transcriptional regulator of iron metabolism that is prevalent in eubacteria, but is absent in *Rickettsia*. Fur controls many facets of iron homeostasis, including the *hemA* gene in *B. japonicum* (Hamza *et al.*, 2000). Irr and PerR are structural homologues of Fur but are functionally distinct. Nevertheless, they also mediate iron control of haem synthesis in *B. japonicum* (Hamza *et al.*, 1998; Qi & O'Brian, 2002) and *B. subtilis* (Bsat *et al.*, 1998), respectively. The absence of Fur homologues in *Rickettsia* indicates the abandonment of the near universal bacterial-type iron regulation in that genus. However, this does not rule out other types of control by iron.

Most eubacteria have iron transport systems, utilizing either a permease system that can transport the metal directly, or siderophore-mediated transport. Siderophores are iron chelators which bind to a cognate cellular receptor for uptake. The *Rickettsia* genome appears to lack genes encoding these iron transport proteins despite the plethora of iron-containing proteins in that organism. However, *Rickettsia* does contain a highly conserved gene homologue of *atm1*, a mitochondrial iron transport protein prevalent in eukaryotes, and which is associated with X-linked sideroblastic anaemia and ataxia in humans (Allikmets *et al.*, 1999). *Atm1* is also found in other α -proteobacteria, and therefore it is likely a retained gene rather than one acquired from its eukaryotic host by horizontal gene transfer. The concomitant retention of *atm1* and loss of the bacterial-type iron transporter genes in both *Rickettsia* and mitochondria are likely evolutionary consequences of an obligatory intracellular milieu.

Borrelia burgdorferi

For the most part genome reduction is not an adaptive response, but reflects a lack of selective pressure to maintain genes rendered non-essential by the obligate association with a eukaryotic host (Ochman & Moran, 2001). An exception to this is observed in *B. burgdorferi*, the causative agent of Lyme disease. Most microbial pathogens overcome iron limitation, or so-called nutritional immunity, created by the host by sequestering

iron. *B. burgdorferi* has apparently circumvented this problem by eliminating the need for iron altogether (Posey & Gherardini, 2000), which is reflected by a genome that does not encode most iron proteins. Accordingly, the haem pathway genes, as well as those for haem proteins, are absent. A *fur* gene homologue was identified (Posey & Gherardini, 2000), but its product has greater similarity to PerR, which responds directly to manganese in *B. subtilis*, and only indirectly to iron (Bsat *et al.*, 1998). It is possible that *B. burgdorferi* represents the extreme case of a strategy employed by other pathogens with reduced genomes. *Mycoplasma pneumoniae*, a respiratory pathogen, and *Treponema pallidum*, the syphilis spirochaete, lack haem pathway enzymes, cytochromes, iron transport systems and most iron proteins. Exceptions to this are genes for rubredoxin, an iron-sulfur protein, found in *T. pallidum*, and NifS, a protein required for iron-sulfur cofactor synthesis. Both *M. pneumoniae* and *T. pallidum* have genes encoding dinucleotide reductase, an iron protein involved in nucleotide synthesis, which is missing in *B. burgdorferi*.

Concluding remarks

It is often said that an exception to the rule validates the rule, and this can be applied to a genomic assessment of prokaryotic haem biosynthesis. When one considers the overall diversity of prokaryotes, the general conservation of haem pathway intermediates and enzyme activities is noteworthy. Nonetheless, exceptions to the rule often define the new frontiers and challenge the established paradigms based on a relatively narrow set of species. The inability to account for protoporphyrinogen oxidation in the penultimate step of the pathway in so many organisms is striking, especially since it is not a taxonomically restricted phenomenon. One of the most astonishing features of archaeal genomes revealed by sequencing projects is the vast numbers of genes with no known homologues. Understanding the novel enzymes or intermediates in haem biosynthesis implicated by at least some of these organisms represents an important question in prokaryotic metabolism. Finally, the rapid development of technologies to assess global control of gene expression such as microarray and proteomic analyses should yield a whole new set of rules and exceptions for the regulation of metabolic pathways, including haem biosynthesis.

Acknowledgements

Work from M.R.O'B.'s laboratory is supported by the National Science Foundation (MCB-0089928 and MCB-0077628), and from the US Department of Agriculture (99-35305-8062). H.P. is supported by a postdoctoral fellowship from the National Institutes of Health (AI07614).

References

- Akutsu, H., Park, J.-S. & Sano, S. (1993). L-Methionine methyl is specifically incorporated into the C-2 and C-7 positions of the porphyrin of cytochrome c_3 in a strictly anaerobic bacterium, *Desulfovibrio vulgaris*. *J Am Chem Soc* **115**, 12185–12186.
- Allikmets, R., Raskind, W. H., Hutchinson, A., Schueck, N. D., Dean, M. & Koeller, D. M. (1999). Mutation of a putative mitochondrial iron transporter gene (ABC7) in X-linked sideroblastic anemia and ataxia (XLSA/A). *Hum Mol Genet* **8**, 743–749.
- Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. (1990). Basic local alignment tool. *J Mol Biol* **215**, 403–410.
- Avisar, Y. J., Ormerod, J. G. & Beale, S. I. (1989). Distribution of delta-aminolevulinic acid biosynthetic pathways among phototrophic bacterial groups. *Arch Microbiol* **151**, 513–519.
- Baumann, P., Baumann, L., Lai, C.-Y. & Rouhbakhsh, D. (1995). Genetics, physiology, and evolutionary relationships of the genus *Buchnera*: intracellular symbionts of aphids. *Annu Rev Microbiol* **49**, 55–94.
- Becker, M. & Schafer, G. (1991). Purification and spectral characterization of a b-type cytochrome from the plasma membrane of the archaeobacterium *Sulfolobus acidocaldarius*. *FEBS Lett* **291**, 331–335.
- Bsat, N., Herbig, A., Casillas-Martinez, L., Setlow, P. & Helmann, J. D. (1998). *Bacillus subtilis* contains multiple Fur homologs: identification of the iron uptake (Fur) and peroxide regulon (PerR) repressors. *Mol Microbiol* **29**, 189–198.
- Chen, J. J. & London, I. M. (1995). Regulation of protein synthesis by heme-regulated eIF-2 alpha kinase. *Trends Biochem Sci* **20**, 105–108.
- Denda, K., Fujiwara, T., Seki, M., Yoshida, M., Fukumori, Y. & Yamanaka, T. (1991). Molecular cloning of the cytochrome aa_3 gene from the archaeon (Archaeobacterium) *Halobacterium halobium*. *Biochem Biophys Res Commun* **181**, 316–322.
- Denda, K., Mogi, T., Anraku, Y., Yamanaka, T. & Fukumori, Y. (1995). Characterization of chimeric heme-copper respiratory oxidases using subunits I of *Escherichia coli* cytochrome b_o and *Halobacterium salinarium* cytochrome aa_3 . *Biochem Biophys Res Commun* **217**, 428–436.
- Fujino, E., Fujino, T., Karita, S., Sakka, K. & Ohmiya, K. (1995). Cloning and sequencing of some genes responsible for porphyrin biosynthesis from the anaerobic bacterium *Clostridium josui*. *J Bacteriol* **177**, 5169–5175.
- Fujiwara, T., Fukumori, Y. & Yamanaka, T. (1993). *Halobacterium halobium* cytochrome b -558 and cytochrome b -562: purification and some properties. *J Biochem (Tokyo)* **113**, 48–54.
- Gartner, P. (1991). Characterization of a quinole-oxidase activity in crude extracts of *Thermoplasma acidophilum* and isolation of an 18-kDa cytochrome. *Eur J Biochem* **200**, 215–222.
- Hach, A., Hon, T. & Zhang, L. (1999). A new class of repression modules is critical for heme regulation of the yeast transcriptional activator Hap1. *Mol Cell Biol* **19**, 4324–4333.
- Hallberg Gradin, C. & Colmsjo, A. (1989). Four different b-type cytochromes in the halophilic archaeobacterium, *Halobacterium halobium*. *Arch Biochem Biophys* **272**, 130–136.
- Hamza, I., Chauhan, S., Hassett, R. & O'Brian, M. R. (1998). The bacterial Irr protein is required for coordination of heme biosynthesis with iron availability. *J Biol Chem* **273**, 21669–21674.
- Hamza, I., Qi, Z., King, N. D. & O'Brian, M. R. (2000). Fur-independent regulation of iron metabolism by Irr in *Bradyrhizobium japonicum*. *Microbiology* **146**, 669–676.
- Hansson, M. & Hederstedt, L. (1992). Cloning and characterization of the *Bacillus subtilis* hemEHY gene cluster, which encodes protoheme IX biosynthetic enzymes. *J Bacteriol* **174**, 8081–8093.
- Higuchi, M. & Bogorad, L. (1975). The purification and properties of uroporphyrinogen I synthases and uroporphyrinogen III cosynthase. *Ann N Y Acad Sci* **244**, 401–418.

- Homuth, G., Rompf, A., Schumann, W. & Jahn, D. (1999).** Transcriptional control of *Bacillus subtilis* *hemN* and *hemZ*. *J Bacteriol* **181**, 5922–5929.
- Ishida, T., Yu, L., Akutsu, H., Ozawa, K., Kawanishi, S., Seto, A., Inubushi, T. & Sano, S. (1998).** A primitive pathway of porphyrin biosynthesis and enzymology in *Desulfovibrio vulgaris*. *Proc Natl Acad Sci USA* **95**, 4853–4858.
- Keithly, J. H. & Nadler, K. D. (1983).** Protoporphyrin formation in *Rhizobium japonicum*. *J Bacteriol* **154**, 838–845.
- Lathrop, J. T. & Timko, M. P. (1993).** Regulation by heme of mitochondrial protein transport through a conserved amino acid motif. *Science* **259**, 522–525.
- Le Guen, L., Santos, R. & Camadro, J. M. (1999).** Functional analysis of the *hemK* gene product involvement in protoporphyrinogen oxidase activity in yeast. *FEMS Microbiol Lett* **173**, 175–182.
- Lübben, M. & Morand, K. (1994).** Novel prenylated hemes as cofactors of cytochrome oxidases. Archaea have modified hemes A and O. *J Biol Chem* **269**, 21473–21479.
- Lübben, M., Warne, A., Albracht, S. P. & Saraste, M. (1994).** The purified SoxABCD quinol oxidase complex of *Sulfolobus acidocaldarius* contains a novel haem. *Mol Microbiol* **13**, 327–335.
- Nakajima, O., Takahashi, S., Harigae, H., Furuyama, K., Hayashi, N., Sassa, S. & Yamamoto, M. (1999).** Heme deficiency in erythroid lineage causes differentiation arrest and cytoplasmic iron overload. *EMBO J* **18**, 6282–6289.
- Nakayashiki, T., Nishimura, K. & Inokuchi, H. (1995).** Cloning and sequencing of a previously unidentified gene that is involved in the biosynthesis of heme in *Escherichia coli*. *Gene* **153**, 67–70.
- Narita, S., Taketani, S. & Inokuchi, H. (1999).** Oxidation of protoporphyrinogen IX in *Escherichia coli* is mediated by the aerobic coproporphyrinogen oxidase. *Mol Gen Genet* **261**, 1012–1020.
- Ochman, H. & Moran, N. A. (2001).** Genes lost and genes found: evolution of bacterial pathogenesis and symbiosis. *Science* **292**, 1096–1098.
- Ogawa, K., Sun, J., Taketani, S. & 8 other authors (2001).** Heme mediates derepression of Maf recognition element through direct binding to transcription repressor Bach1. *EMBO J* **20**, 2835–2843.
- Posey, J. E. & Gherardini, F. C. (2000).** Lack of a role for iron in the Lyme disease pathogen. *Science* **288**, 1651–1653.
- Qi, Z. & O'Brian, M. R. (2002).** Interaction between the bacterial iron response regulator and ferrochelatase mediates genetic control of heme biosynthesis. *Mol Cell* **9**, 155–162.
- Qi, Z., Hamza, I. & O'Brian, M. R. (1999).** Heme is an effector molecule for iron-dependent degradation of the bacterial iron response regulator (Irr) protein. *Proc Natl Acad Sci USA* **96**, 13056–13061.
- Reddy, S. V., Alcantara, G., Roodman, D. & Boldt, D. H. (1996).** Inhibition of tartrate-resistant acid phosphatase gene expression by hemin and protoporphyrin IX. Identification of a heme-responsive inhibitor of transcription. *Blood* **88**, 2288–2297.
- Rodgers, K. R. (1999).** Heme-based sensors in biological systems. *Curr Opin Chem Biol* **3**, 158–167.
- Sangwan, I. & O'Brian, M. R. (1991).** Evidence for an inter-organismic heme biosynthetic pathway in symbiotic soybean root nodules. *Science* **251**, 1220–1222.
- Sasarman, A., Chartrand, P., Lavoie, M., Tardif, D., Proschek, R. & Lapointe, C. (1979).** Mapping of a new *hem* gene in *Escherichia coli* K12. *J Gen Microbiol* **113**, 297–303.
- Sasarman, A., Letowski, J., Czaika, G., Ramirez, V., Nead, M. A., Jacobs, J. M. & Morais, R. (1993).** Nucleotide sequence of the *hemG* gene involved in the protoporphyrinogen oxidase activity of *Escherichia coli* K12. *Can J Microbiol* **39**, 1155–1161.
- Schmitt, M. P. (1999).** Identification of a two-component signal transduction system from *Corynebacterium diphtheriae* that activates gene expression in response to the presence of heme and hemoglobin. *J Bacteriol* **181**, 5330–5340.
- Seehra, J. S., Jordan, P. M. & Akhtar, M. (1983).** Anaerobic and aerobic coproporphyrinogen III oxidases of *Rhodospseudomonas spheroides*. Mechanism and stereochemistry of vinyl group formation. *Biochem J* **209**, 709–718.
- Wang, L., Elliott, M. & Elliott, T. (1999).** Conditional stability of the HemA protein (glutamyl-tRNA reductase) regulates heme biosynthesis in *Salmonella typhimurium*. *J Bacteriol* **181**, 1211–1219.
- Wright, R. L., Harris, K., Solow, B., White, R. H. & Kennelly, P. J. (1996).** Cloning of a potential cytochrome P450 from the archaeon *Sulfolobus solfataricus*. *FEBS Lett* **384**, 235–239.
- Xu, K. & Elliott, T. (1994).** Cloning, DNA sequence, and complementation analysis of the *Salmonella typhimurium hemN* gene encoding a putative oxygen-independent coproporphyrinogen III oxidase. *J Bacteriol* **176**, 3196–3203.
- Yano, J. K., Koo, L. S., Schuller, D. J., Li, H., Ortiz de Montellano, P. R. & Poulos, T. L. (2000).** Crystal structure of a thermophilic cytochrome P450 from the archaeon *Sulfolobus solfataricus*. *J Biol Chem* **275**, 31086–31092.
- Yeliseev, A. A. & Kaplan, S. (1999).** A novel mechanism for the regulation of photosynthesis gene expression by the TspO outer membrane protein of *Rhodobacter sphaeroides* 2.4.1. *J Biol Chem* **274**, 21234–21243.
- Zeilstra-Ryalls, J. H. & Kaplan, S. (1995).** Aerobic and anaerobic regulation in *Rhodobacter sphaeroides* 2.4.1: the role of the *fnrL* gene. *J Bacteriol* **177**, 6422–6431.