Controlling the delicate balance of tetrapyrrole biosynthesis

Liang Yin and Carl E. Bauer

Department of Molecular and Cellular Biochemistry, Indiana University, Bloomington, IN 47405, USA

Tetrapyrroles are a family of compounds that contain four pyrrole rings. They are involved in many fundamental biological processes such as photoreception, electron transport, gas transport and also as cofactors for enzymatic reactions. As regulators of protein activity, tetrapyrroles mediate cellular response to light, oxygen and nutrient levels in the surrounding environment. Biosynthesis of haem tetrapyrroles shares conserved pathways and enzymes among all three domains of life. This is contrasted by chlorophyll biosynthesis that is only present in eubacteria and chloroplasts, or cobalamin biosynthesis that is only present in eubacteria and archaea. This implicates haem as the most ancient, and chlorophyll as the most recent, of the common tetrapyrroles that are currently synthesized by existing organisms. Haem and chlorophyll are both toxic when synthesized in excess over apo-proteins that bind these tetrapyrroles. Accordingly, the synthesis of these tetrapyrroles has to be tightly regulated and coordinated with apo-protein production. The mechanism of regulating haem and chlorophyll synthesis has been studied intensively in Rhodobacter species and will be discussed.

1. Branches of tetrapyrrole synthesis leading to haem, cobalamin and (bacterio)chlorophyll

There are many excellent reviews and books providing mechanistic details of individual steps in tetrapyrrole synthesis [1–5]. Consequently, we will not dwell on enzymatic details of the pathway here. In a generalized view, the ‘modern’ tetrapyrrole pathway can be divided into three branches that synthesize the end-products haem, cobalamin and chlorophyll (figure 1). The synthesis of all of these tetrapyrrole molecules starts with the common precursor 5-aminolevulinic acid (ALA). As shown in figure 1, tetrapyrrole synthesis is a pathway where ALA flows into all three different branches. Eight ALA molecules are assembled in the first three enzymatic steps to form uroporphyrinogen III (Urogen III), the last common intermediate for these three tetrapyrroles. The synthesis of cobalamin from Urogen III is a long and intricate metabolic process with more than 15 steps involving the insertion of cobalt, modification of the tetrapyrrole ring and assembly of a nucleotide tail (dimethylbenzimidazole) [3]. Beyond the split to the cobalt branch, the haem and chlorophyll branches both use three common reactions from Urogen III to protoporphyrin IX (Proto IX) [6]. The next branch point is at Proto IX, where insertion of iron by ferrochelatase results in the synthesis of haem, while the insertion of magnesium into Proto IX by Mg-chelatase produces Mg-Proto IX, which is the first committed step in the chlorophyll branch. After the insertion of Mg, the chlorophyll branch contains numerous enzymes that further modify the tetrapyrrole ring leading to the synthesis of chlorophyll, as in the case of oxygenic photosynthesis (cyanobacteria and chloroplasts), or bacteriochlorophyll in anoxygenic photosynthesis (green, purple and helio bacteria) [4].

2. The haem pathway is widely spread and probably the oldest among the three branches

Of the three tetrapyrrole branches, the haem branch easily stands out as the simplest one regarding its biosynthesis complexity. Haem synthesis involves...
that use haem, coupled with the diversity of haem-proteins, is consistent with haem being an ancient molecule.

In contrast to the presence of haem in all forms of life, the chlorophyll and cobalamin branches only exist in one or two kingdoms, and their functions are rather specialized [1,3–5]. For example, the chlorophyll-branch is only present in eubacteria and in chloroplasts. This branch uses numerous enzymatic steps beyond Proto IX to produce a heavily modified tetrapyrrole ring product, which together with light-harvesting and reaction centre antenna proteins, is dedicated to light absorption and photochemistry (figure 2b) [6]. Tetrapyrrole-based photosynthesis is the major way that life converts solar energy into cellular energy. As such, photosynthetic organisms form the base of the food chain that feeds energy into most life forms on Earth. Furthermore, the ability of oxygenic photosynthesis to obtain electrons from water, thereby releasing dioxygen as a byproduct, has resulted in a large change in the redox state of our atmosphere (nearly all of the atmospheric oxygen is a direct result of photosynthesis) [22]. This ability of photosynthetic organisms to use chlorophyll to convert solar energy into cellular energy, coupled with the evolution of oxygen, has had profound effects on the evolution of life on Earth.

The synthesis of cobalamin is only documented in eubacteria and in archaean species. However, the use of cobalamin as a cofactor for enzymes occurs in all three kingdoms. The functions of cobalamin predominantly fall into two categories: methyl group transfer and the generation of deoxyadenosyl radicals [3]. The structures of the cobalamin-binding domains identified to-date are relatively similar. In methyltransferases and most isomerases, the lower face of the corrin ring rests on a common Rossman (α/β) fold, while the other side of the ring is covered by a helix bundle in the former and by a triosephosphate isomerase (TIM) barrel in the latter cases [23–25]. An exception was found in ribonucleotide reductase in which a novel motif is used to bind cobalamin [26] (figure 2c). Interestingly, our laboratory has recently solved the structure of a haem-binding domain in the light-sensing protein AppA. This domain does not bind cobalamin even though its overall folding is almost identical to the cobalamin binding domain used by methyltransferase (L. Yin, V. Dragnea, G. Feldman, L. A. Hammad, J. A. Karty, C. Dann, C. E. Bauer 2013, unpublished data). One intriguing possibility is that this newly discovered haem-binding domain represents a turning point in the origination of cobalamin-binding protein from a related haem-binding precursor.

3. The origin and evolution of three tetrapyrrole branches

Even though we favour the haem pathway as the oldest among the three modern branches, haem may not be the first functional tetrapyrrole molecule to have evolved. The modern haem pathway, as shown in figure 1, might have evolved from a prototype resembling the siro-haem synthesis, which uses only three enzymatic steps from Urogen III via the intermediate precorrin-2 [5]. This is supported by the fact that archaea synthesize haem via the precorrin-2 pathway [27]. Furthermore, precorrin-haem could have evolved from an even earlier corrin-based tetrapyrrole molecule (figure 3a). The corrin ring, which is the structural base of cobalamin, is thought to be synthesized by all bacteria
cyanobacterium, chlorophyll biosynthesis is in all respects a product of an endosymbiotic event between a eukaryote and a redox poise of the atmosphere. Well after oxygenic photosynthesis began changing the rise to mitochondria would not have occurred until a biotic event involving a respiratory prokaryote that gave chlorophyll based photosynthesis. Thus, the endosymbiosis would not have been possible until significant levels of oxygen built up as a byproduct of aerobic respiration. Indeed, mitochondria converted to water. ATPase subsequently uses this proton gradient during respiration, where dioxygen is a proton donor.

Interestingly, mitochondrial cytochromes are responsible for cellular energy production by promoting a proton gradient during respiration, where dioxygen is converted to water. ATPase subsequently uses this proton gradient to synthesize ATP. Interestingly, mitochondrial aerobic respiration would not be necessary or even usable until significant levels of oxygen built up as a byproduct of chlorophyll based photosynthesis. Thus, the endosymbiotic event involving a respiratory prokaryote that gave rise to mitochondria would not have occurred until well after oxygenic photosynthesis began changing the redox poise of the atmosphere.

Besides prokaryotes, chlorophyll synthesis is undertaken in eukaryotic chloroplasts (figure 3b). Since chloroplasts are the product of an endosymbiotic event between a eukaryote and a cyanobacterium, chlorophyll biosynthesis is in all respects a prokaryotic process. This also indicates that the eukaryotic branch lost the ability to synthesize Mg-tetrapyrroles, or more probably, that the split between prokaryotes and eukaryotes occurred after the evolution of Mg-tetrapyrroles (figure 3a). Most genes that code for enzymes involved in the Mg-branch, including the first enzyme in this branch Mg-chelatase, have been transported from the chloroplast genome into the nuclear genome. This is despite the fact that chlorophyll biosynthesis occurs predominately in the chloroplast, necessitating the transport of numerous Mg-tetrapyrrole enzymes need into the chloroplast. It has been argued that nuclear localization gives photosynthetic eukaryotic cells the ability to regulate chlorophyll biosynthesis in response to changes in cytosolic redox poise. Consistent with this model, a symbiont sensor kinase was identified in Arabidopsis thaliana that couples photosynthesis to gene expression.

Photosynthetic eukaryotes with both chloroplast and mitochondria organelles have an interesting interdependency between haem and chlorophyll biosynthesis. In higher plant cells, all the enzymatic steps are processed in the chloroplast to produce chlorophyll as well as haem, with just a portion of haem thought to be produced in mitochondria. Finally, a rationale for retention of these organelles is that toxic tetrapyrrole products can be sequestered to these compartments, where they can be quickly assembled into their target proteins.

4. Balancing end-product synthesis in the branched tetrapyrrole pathway-I: feedback regulation

One interesting feature of the branched tetrapyrrole pathway is that the product of one branch influences the synthesis of products from the other two branches. Since they share the same ALA pool and many early common intermediates, feedback regulation must take all three branches into consideration in order to keep them at appropriate levels. Details of how different cell organelle synthesis of all three end products are sketchy but recent advances are beginning to show that...
there is a complex set of overlapping regulatory switches and interdependence of these branches in the pathway both at the transcriptional and post-transcriptional stages. In this section we will first consider post-transcriptional regulation.

The first enzyme of the tetrapyrrole pathway, 5-aminolevulinic acid synthase (ALAS), is known to be repressed by classic feedback control. Early work with *Rhodobacter sphaeroides* demonstrated that haem inhibits ALAS enzymatic activity [38,39]. Since all three branches share common early intermediates, each of the three branches will be affected when excess haem feedback inhibits ALA synthase activity.

Early on, the stability of ALAS was shown to be affected by haem in chick embryo liver [40] and in mammals [41]. More recently, it has been shown that haem binds to a Cys-Pro haem regulatory motif (HRM) in ALAS1, which is a non-specific isoform of ALAS. Haem binding to this site regulates mitochondrial import and stability [42,43]. This is contrasted by erythroid cells, where the isoform ALAS2 is mainly regulated translationally by iron but not by haem [43,44].

In the case of cobalamin biosynthesis, several *cob* genes are known to be controlled at the level of translation by a highly conserved riboswitch [45,46]. The riboswitch is located in the 5' untranslated mRNA of many *cob* genes where adenosine-cobalamin is known to bind. The binding of cobalamin results in a fold of the leader region that masks the ribosome-binding site, causing a decrease in translation. To our knowledge, similar riboswitches involving haem or chlorophyll genes have not previously been described.

The interdependence of these three pathways is also observed when analyzing enzymes in the chlorophyll-branch. For example, the cyclase enzyme encoded by *bchE* catalyzes the conversion of Mg-protoporphyrin monomethyl ester to protochlorophyllide in the Mg-tetrapyrrole branch. An *in vivo* assay has shown that its activity is cobalamin-dependent in many species of purple bacteria [47]. Cobalamin dependence is supported by sequence similarity between the BchE cyclase and the cobalamin-requiring F-methylase from *Streptomyces hygroscopicus* [47]. Likewise, haem is also an essential cofactor for cobalamin biosynthesis in some species. Specifically, cobalamin synthesis requires a monoxygenase CobG to generate a hydroxylactone intermediate. In *Rhodobacter capsulatus* CobZ is isofunctional with CobG and contains multiple cofactors,

![Figure 3.](http://rstb.royalsocietypublishing.org/)

**Figure 3.** Evolution of tetrapyrrole biosynthesis. (a) The coupled relationship between tetrapyrrole synthesis and endosymbiosis. (b) The distribution of haem, cobalamin and chlorophyll among different organisms. Note that the haem biosynthetic pathway is different in some archaea. (Online version in colour.)
including one haem, one flavin and two Fe-S centres [48]. Haem synthesis also requires S-adenosylmethionine (SAM) as a methyl group donor for oxygen-independent coproporphyrinogen III oxidase [49]. The production of SAM is dependent on the enzyme methionine-synthase (MetH) that, which uses cobalamin as a cofactor [23]. Thus, haem synthesis is not only feedback-regulated by the level of haem but also dependent on the availability of cobalamin. Collectively, the requirement of haem and cobalamin for synthesis of the three end-products shows that there is an interdependence of all three pathways on the presence of these two products.

5. Balancing end-product synthesis in the branched tetrapyrrole pathway-II: gene regulation using tetrapyrroles as cofactors

Regulation of tetrapyrrole gene expression in photosynthetic purple bacteria R. sphaeroides and R. capsulatus will be discussed from here on, unless otherwise noted. All three tetrapyrrole pathways are present in these species, with most research on the control of tetrapyrrole gene expression in eubacteria performed in Rhodobacter species. The section below focuses on the involvement of haem and cobalamin in the regulation of all three branches of the tetrapyrrole pathway. To our knowledge, there has been no previous report of the involvement of chlorophyll or any tetrapyrrole intermediates in regulating this pathway; however, their involvement cannot be ruled out.

The regulation of tetrapyrrole gene expression is relatively complex presumably owing to the interdependence among all three branches in the pathway, coupled with the fact that excess haem and bacteriochlorophyll synthesis are toxic to cells. Figure 4 depicts the large number of transcription factors that are known to control tetrapyrrole gene expression in R. capsulatus and R. sphaeroides. Several transcription factors such as RegB-RegA and FNR are global regulators that respond to changes in cellular redox poise [50–52]. In Rhodobacter species, the synthesis of bacteriochlorophyll occurs only under anaerobic growth conditions where these transcription factors are active. RegB and RegA mutations are unable to ramp up Mg-tetrapyrrole biosynthesis under anaerobic growth conditions, indicating that their main role is to increase the flow of tetrapyrroles through this pathway when the Mg-tetrapyrrole branch is de-repressed [50,51,53,54]. In addition to these global regulators, there are several additional transcription factors that have specialized roles in controlling tetrapyrrole biosynthesis. Emerging evidence indicates that this latter class of regulators uses haem and cobalamin as cofactors. This class of regulators will be the remaining focus of this discussion.

(a) CrtJ/PpsR

PpsR, or CrtJ depending on the species, is a redox-regulated repressor of tetrapyrrole gene expression [55–58]. It was shown in R. capsulatus that CrtJ aerobically inhibits hem and bch genes in the haem and bacteriochlorophyll branches, respectively [55–58]. This aerobic repressor also regulates the crt genes that code for enzymes in the carotenoid pathway and puc genes that encode the light-harvesting II peptides, which bind both chlorophyll and carotenoids [55–58]. Analysis of DNA binding by CrtJ has indicated that hem, bch, crt and puc promoter regions all contain two copies of a conserved palindromic sequence TGT-N12-ACA [57]. This palindrome is also present in a similar set of genes in R. sphaeroides [59]. The DNA-binding activity of CrtJ/PpsR is highly regulated by redox (oxygen), light, haem and cobalamin. In regard to
redox control, a conserved cysteine in the DNA-binding domain of CrtJ/PpsR is known to undergo oxidation when cells are grown in the presence of oxygen, leading to an increase in DNA binding and subsequent repression [60–62]. In addition to redox control, PpsR has also been shown to bind haem [13]. The presence of haem alters the DNA-binding pattern of PpsR and inhibits its ability to form a higher-ordered PpsR–DNA complex [13]. As the result, the expression level of a subset of PpsR-regulated genes increases. Among the various genes analysed, only puc and bchC genes increase in expression when PpsR binds to haem [13]. The reason why this set of genes are derepressed by haem binding to PpsR might be rooted in the observation that the puc and bchC promoters contain two PpsR-binding sites that are in close proximity (7 and 8 bp apart, respectively) [57]. This is contrasted by PpsR-binding sites that are relatively far from each other, as in the crtA and haem promoters (55 and 252 bp apart, respectively) [55]. Additional regulation of the DNA-binding activity of CrtJ/PpsR by light and cobalamin using a set of related antirepressors is discussed below.

(b) AppA
In R. sphaeroides, there is a light-sensing antirepressor of PpsR called AppA that can form a PpsR2–AppA complex, which releases PpsR from its target DNA [60]. Interestingly, the interaction of AppA with PpsR is highly controlled in response to light, redox and haem [60]. In regard to redox control, AppA has a cysteine-rich region (6 Cys at the carboxyl terminus) that is capable of reducing oxidized PpsR [60,63]. This event would inhibit DNA binding as PpsR needs to be in an oxidized state for optimal binding [60]. In regard to light control, the amino terminus of AppA was shown by Masuda & Bauer [60] to be a flavin-binding domain that represented a new class of photoreceptor called the blue-light using flavin (BLUF) photoreceptor. Currently, there are over 300 known examples of BLUF proteins present in a diverse number of prokaryotes, and in eukaryotic algae, with AppA representing the only transcription factor that uses blue light, redox and haem [60]. In vivo and in vitro studies indicate that AppA functions as a cobalamin-dependent antirepressor of CrtJ, not unlike that of AppA functioning as a haem-regulated antirepressor of PpsR (Z. Cheng, K. Li, L. A. Hammad, J. A. Karty, C. E. Bauer, 2013, unpublished data). In vivo and in vitro studies indicate that AerR functions as a cobalamin-dependent antirepressor of CrtJ, not unlike that of AppA functioning as a haem-regulated antirepressor of PpsR (Z. Cheng, K. Li, L. A. Hammad, J. A. Karty, C. E. Bauer, 2013, unpublished data). Thus, purple photosynthetic bacteria that contain the PpsR/CrtJ repressor family also have one or more antirepressors that sense the presence of haem or cobalamin.

(d) HbrL
A high-copy suppressor screen for genes that repress haem biosynthesis in R. capsulatus led to the identification of a haem-binding regulator of the LysR family (HbrL) [77]. Cells deleted for HbrL are unable to ramp up expression of hemA, hemB and hemZ in response to a reduction in haem levels [77]. Isolated HbrL was shown to bind haem with a Soret band at approximately 413 nm [77]. Recently, we have observed that an HbrL deletion strain is unable to grow on iron-depleted growth medium owing to the failure to transport iron (S. Zappa, C. E. Bauer 2013, unpublished data). Gene expression studies indicate that HbrL is responsible for regulating expression of a number of iron influx genes as well as genes that code for cytochrome apoproteins (S. Zappa, C. E. Bauer 2013, unpublished data). Thus, HbrL appears to be a central player in controlling the stoichiometry of all of the organic and inorganic components of cytochromes.

(e) Irr
Irr is a transcription factor intensively studied in Bradyrhizobium japonicum that is related to the ferric-uptake regulator (Fur) family of metalloregulators [16,78,79]. In B. japonicum, Irr is responsible for regulating iron uptake as well as the expression of the hemB gene [80], similar to that of HbrL from R. capsulatus. Since iron is the central atom in haem, it is not surprising that cells have evolved several mechanisms of coordinating iron uptake with the synthesis of haem. The importance of this coordination is emphasized by the fact that Irr responds to haem rather than to iron, as is the case for other members of the Fur family of transcription factors [16]. It has been shown that Irr forms a complex with ferrochelatase, the enzyme for the last step in haem pathway [81]. The formation of this complex enables Irr to sense haem directly at the point of haem synthesis. The binding of haem triggers Irr degradation depending on the AppA would be decreased PpsR repression of bch gene expression [13]. This will lead to an increase in Mg–tetrapyrrole synthesis, which would redirect the flow of Proto IX away from haem towards the bacteriochlorophyll branch.
Figure 5. The system of PprR/CrtJ-AppA/AerR regulated by oxygen, haem, cobalamin and light + haem. (Online version in colour.)

haem redox states and protein oxidation states [16,78,81]. Compared to the well-characterized B. japonicum Irr, R. capsulatus Irr lacks the HRM that is critical for haem-sensing in B. japonicum Irr [6], suggesting that it may function in a fundamentally different way.

6. Future challenges

As discussed above, there are several levels of feedback regulation used to ensure interdependence of all three tetrapyrrole branches. This complex set of transcriptional and post-transcriptional feedback controls is needed to keep appropriate balance of the tetrapyrrole end-products. Among the many challenges for further exploration will be to probe mechanistic details of regulatory proteins that bind either haem or cobalamin. Although it has been shown many times that haem can have a regulatory role targeting haem-sensing proteins in species ranging from bacteria to humans, the idea of “free haem pool” in vivo is still controversial because free haem is known to generate reactive oxygen species (ROS) and thus be highly toxic. This might be the reason why higher organisms confine in whole, or in part, tetrapyrrole synthesis in mitochondria and/or chloroplasts. A similar strategy of targeting haem-sensing at the location of haem synthesis is found in Irr, which can form a complex with ferrochelatase that catalyses the last step of haem-synthesis. Nonetheless, it would be important to understand the quantity and localization of free haem as many of the characterized transcription factors covered in the review bind haem with low affinity in vivo. The complex PpsR/CrtJ, AppA and AerR system also creates a unique challenge in understanding tetrapyrrole regulation, as these three proteins regulate tetrapyrrole biosynthesis in response to light, oxygen, haem and cobalamin (figure 5).

Clearly, additional work is needed to obtain a better understanding of how these proteins coordinate the synthesis of these important tetrapyrroles in vivo.

Research on tetrapyrrole biosynthesis in the C. Bauer laboratory is supported by the National Institutes of Health grant no. GM40941.


78. Yang J, Ishimori K, O'Brian MR. 2005 Two heme binding sites are involved in the regulated degradation of the bacterial iron response regulator (Irr) protein. J. Biol. Chem. 280, 7671–7676. (doi:10.1074/jbc.M411664200)


In the first line of §5a, the sentence ‘PpsR, or CrtJ depending on the species, is a redox-regulated repressor of tetrapyrrole gene expression [55–58]’ failed to cite that this repressor was initially discovered by Penfold & Pemberton [1,2].

References
