Acta Crystallographica Section D
**Biological Crystallography**
ISSN 0907-4449

# Insights into the mechanism of pyrrole polymerization catalysed by porphobilinogen deaminase: high-resolution X-ray studies of the *Arabidopsis thaliana* enzyme

**A. Roberts,[a] R. Gill,[b] R. J. Hussey,[a] H. Mikolajek,[a] P. T. Erskine,[b] J. B. Cooper,[b]* S. P. Wood,[b] E. J. T. Chrystal[c]‡ and P. M. Shoolingin-Jordan[a]**

[a]School of Biological Sciences, University of Southampton, Southampton SO16 1BJ, England, [b]Laboratory of Protein Crystallography, Centre for Amyloidosis and Acute Phase Proteins, UCL Division of Medicine (Royal Free Campus), Rowland Hill Street, London NW3 2PF, England, and [c]Syngenta, Jealott's Hill International Research Centre, Bracknell, Berkshire RG42 6EY, England

‡ Deceased.

Correspondence e-mail: jon.cooper@ucl.ac.uk

Received 3 December 2012
Accepted 31 December 2012

**PDB Reference:**
porphobilinogen deaminase, 4htg

This paper is dedicated to the memory of Ewan Chrystal.

The enzyme porphobilinogen deaminase (PBGD; hydroxymethylbilane synthase; EC 2.5.1.61) catalyses a key early step of the haem- and chlorophyll-biosynthesis pathways in which four molecules of the monopyrrole porphobilinogen are condensed to form a linear tetrapyrrole. The active site possesses an unusual dipyrromethane cofactor which is extended during the reaction by the sequential addition of the four substrate molecules. The cofactor is linked covalently to the enzyme through a thioether bridge to the invariant Cys254. Until recently, structural data have only been available for the *Escherichia coli* and human forms of the enzyme. The expression of a codon-optimized gene for PBGD from *Arabidopsis thaliana* (thale cress) has permitted for the first time the X-ray analysis of the enzyme from a higher plant species at 1.45 Å resolution. The *A. thaliana* structure differs appreciably from the *E. coli* and human forms of the enzyme in that the active site is shielded by an extensive well defined loop region (residues 60–70) formed by highly conserved residues. This loop is completely disordered and uncharacterized in the *E. coli* and human PBGD structures. The new structure establishes that the dipyrromethane cofactor of the enzyme has become oxidized to the dipyrromethenone form, with both pyrrole groups approximately coplanar. Modelling of an intermediate of the elongation process into the active site suggests that the interactions observed between the two pyrrole rings of the cofactor and the active-site residues are highly specific and are most likely to represent the catalytically relevant binding mode. During the elongation cycle, it is thought that domain movements cause the bound cofactor and polypyrrole intermediates to move past the catalytic machinery in a stepwise manner, thus permitting the binding of additional substrate moieties and completion of the tetrapyrrole product. Such a model would allow the condensation reactions to be driven by the extensive interactions that are observed between the enzyme and the dipyrromethane cofactor, coupled with acid–base catalysis provided by the invariant aspartate residue Asp95.

## 1. Introduction

The vital biological role of chlorophyll as the major energy-harvesting pigment in the biosphere coupled with the importance of haem in respiration emphasizes the key significance of tetrapyrrole biosynthesis to living systems (Warren & Smith, 2009). Whilst in all cells there is a constant need to generate haem for incorporation into cytochromes and other haemoproteins, in photosynthetic cells chlorophyll must be synthesized at much greater levels. During the initial stages of greening, chlorophyll levels increase dramatically in a matter of hours, while haem levels remain essentially constant

(Cornah *et al.*, 2003). In higher plants, the first committed steps of chlorophyll biosynthesis involve the formation of aminolevulinic acid (ALA) from glutamate, which is accomplished by glutamyl-tRNA synthase, glutamyl-tRNA reductase (GluTR) and glutamate-1-semialdehyde aminotransferase (GSAT). GluTR is one of the very few enzymes which utilize aminoacyl-tRNA in a process other than protein synthesis. Specifically, it catalyses the reduction of glutamyl-tRNA, yielding glutamate-1-semialdehyde (GSA), in the presence of NADPH. The next enzyme, GSAT, catalyses the pyridoxal-5′-phosphate (PLP)-dependent rearrangement of GSA to form 5-aminolaevulinic acid. GluTR appears to be a key site of regulation of the pathway in plants; it is inhibited by haem and expression of its gene (*hemA*) is subject to a wide range of regulatory signals including hormones, the circadian clock, plastid signals and light. In contrast, in animals ALA is synthesized by a different enzyme, 5-aminolaevulinic acid synthase (ALAS), which catalyses the PLP-dependent condensation of glycine and succinyl-CoA, and is also a key site of regulation. The remaining steps of haem biosynthesis, which are common to animals, plants and prokrayotes, initially involve the condensation of two ALA moieties to form the pyrrole porphobilinogen in a reaction which is catalysed by 5-aminolaevulinic acid dehydratase (ALAD).

The subsequent step in the pathway is catalysed by the enzyme porphobilinogen deaminase (PBGD), which is also known as hydroxymethylbilane synthase (EC 2.5.1.61). This reaction constitutes the fifth step of the haem-biosynthesis pathway in plants (the third step in animals) and involves the polymerization of four molecules of the monopyrrole porphobilinogen in a stepwise head-to-tail manner to form the linear tetrapyrrole preuroporphyrinogen or hydroxymethylbilane (Fig. 1; Jordan, 1991). In humans, genetic lesions in the gene for this enzyme give rise to the disease acute intermittent porphyria (AIP), which is one of the most common of the hereditary porphyrias (Wood *et al.*, 1995). PBGDs are monomeric enzymes with molecular weights in the range 34–44 kDa depending on the species. The enzymes in this family exhibit high thermal stability and have pH optima in the range 8.0–8.5, with isoelectric points in the range 4.0–4.5. Isotopic labelling and single-turnover studies showed that the pyrrole-forming ring *A* of the tetrapyrrole (Fig. 1) is the first to bind to the enzyme, followed by rings *B*, *C* and finally *D* (Jordan, 1991). NMR, isotopic labelling and mass-spectrometric studies of the *Escherichia coli* PBGD enzyme showed that it possesses a dipyrromethane cofactor (Fig. 2) which is covalently bound to the enzyme by a thioether linkage involving an invariant cysteine residue (Cys242 in *E. coli* numbering; Jordan & Warren, 1987; Warren & Jordan, 1988; Scott *et al.*, 1988). Whilst the apoenzyme possesses no catalytic activity, incubation with PBG for a period of several hours at pH 8.0 generates active holoenzyme (Scott *et al.*, 1989). During this time, the cofactor can be assembled from two molecules of the normal substrate porphobilinogen. *In vivo*, the cofactor is derived from cleavage of the product preuroporphyrinogen, which reacts rapidly with the apoenzyme (Awan *et al.*, 1997). During catalysis, the cofactor acts as a primer with which four porphobilinogen molecules react sequentially to give an enzyme-bound hexapyrrole; mass spectrometry has allowed the identification of all of the corresponding enzyme–substrate complexes ($ES_1$–$ES_4$; Aplin *et al.*, 1991). After the assembly of $ES_4$, cleavage of the link between the cofactor and the first substrate molecule (S1) completes the reaction. The latter editing function ensures that the cofactor remains permanently attached to the enzyme as the tetrapyrrole product is released.

The X-ray structure of the *E. coli* enzyme has been solved at high resolution (Louie *et al.*, 1992, 1996; Hädener *et al.*, 1999) and the closely related structure of the human enzyme is also available (Gill *et al.*, 2009; Song *et al.*, 2009). The polypeptide of the enzyme is folded into three domains (1–3), each of approximately the same size. The general architecture of domains 1 and 2 shows a strong resemblance to the type II periplasmic binding proteins (Louie *et al.*, 1992; Louie, 1993) whereas domain 3 has a distinct fold. The dipyrromethane cofactor is covalently attached to a cysteine residue in a loop of domain 3 so that it is positioned within a deep active-site cleft formed between domains 1 and 2. Here, the enzyme provides several crucial arginine residues which bind the side-chain carboxylates of the cofactor and/or substrate within the active site. Indeed, many of the disease-associated point mutations in AIP sufferers affect these conserved arginine residues (Wood *et al.*, 1995).

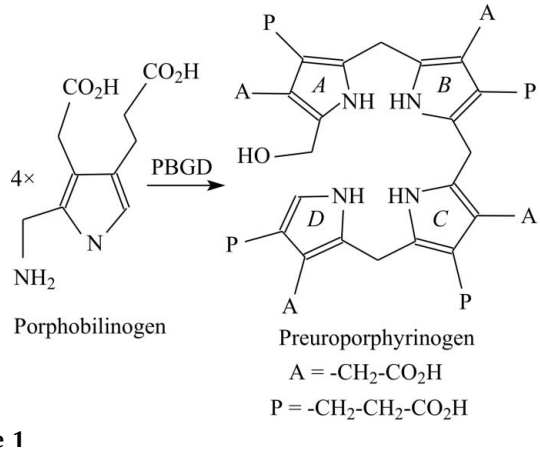


**Figure 1**
The reaction catalysed by porphobilinogen deaminase. Four molecules of the pyrrole porphobilinogen are condensed to form the linear tetrapyrrole preuroporphyrinogen (hydroxymethylbilane).

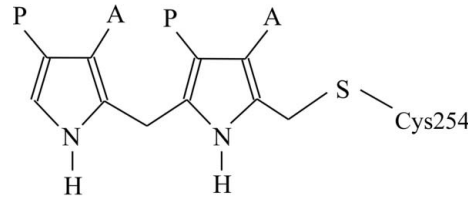


**Figure 2**
The dipyrromethane cofactor of porphobilinogen deaminase. The cofactor is covalently attached to the enzyme by a thioether bond to Cys254. Four substrate pyrroles are added linearly to the cofactor to give a hexapyrrole; finally, hydrolysis of the linkage between the first substrate moiety and the cofactor releases the tetrapyrrole product preuroporphyrinogen.

The tetrapyrrole-biosynthetic pathway is rather more elaborate in plants than it is in animals to allow the production of the crucial photosynthetic pigment chlorophyll, which is localized in the chloroplast thylakoid membranes. However, PBGD catalyses one of the early steps which is common to the biosynthesis of both haem and chlorophyll. PBGDs have been characterized at both the genetic and the protein levels in a number of plant species, most notably the pea *Pisum sativum* (Smith, 1988; Spano & Timko, 1991; Witty *et al.*, 1993). Like other tetrapyrrole-biosynthetic enzymes, PBGD is expressed at its highest level in leaves a few days after germination (He *et al.*, 1994) and the loss of PBGD activity in dark-incubated plants has been implicated as a key factor in leaf senescence (Hukmani & Tripathy, 1994). Immuno-gold electron microscopy and other studies have shown that the pro-enzyme is targeted to the chloroplast (Smith, 1988; Witty *et al.*, 1993, 1996), where it is confined in both leaves and roots. This contrasts with the situation in animals and yeast, where PBGD is a cytosolic enzyme. Mutations affecting the PBGD gene in higher plants have been implicated in the premature cell death that is associated with the camouflage leaf patterning of certain maize mutants (Huang *et al.*, 2009) and albinism in wheat (Chen *et al.*, 2006).

There is a single gene for PBGD in chromosome 5 of the *Arabidopsis thaliana* (thale cress) genome which encodes a precursor protein of 382 residues. In the chloroplast, the precursor is processed by the removal of an N-terminal transit peptide which is around 60 residues in length (Lim *et al.*, 1994). The *A. thaliana* enzyme has high sequence similarity to other known plant PBGD sequences (identities of >80%), suggesting that it is highly representative of the phytal enzyme. The genomic DNA possesses four introns within the coding region of the mature protein. Structural studies of a plant PBGD enzyme have potential for the discovery of novel inhibitory compounds that could act as herbicides. Indeed, a number of commonly used herbicides, such as acifluorfen, target this particular pathway. Here, we report the crystallization of *A. thaliana* PBGD in a form that diffracted X-rays to near-atomic resolution (1.45 Å) and allowed the first structure analysis of a PBGD from a higher plant species.

## 2. Methods

Native PBGD has previously been purified from *A. thaliana* by Jones & Jordan (1994), who obtained enough functional enzyme for activity studies, although insufficient enzyme was obtained for crystallization. For the current investigation, a codon-optimized version of the *A. thaliana* PBGD gene was designed for overexpression in *E. coli.* Cloning and expression of this gene and the subsequent purification of recombinant *A. thaliana* PBGD enzyme have been described in detail previously (Roberts *et al.*, 2012). Since the dipyrromethane cofactor of PBGD is light-sensitive, crystals were grown in the dark using the hanging-drop method with a stock protein solution of 5 mg ml$^{-1}$ PBGD in 20 m*M* Tris–HCl buffer pH 8.0, 5 m*M* DTT. Crystals were obtained using 25%(*w*/*v*) PEG 4000, 100 m*M* sodium citrate pH 5.6, 200 m*M* ammonium sulfate. These were cryoprotected by the addition of glycerol to a final concentration of 30%(*v*/*v*) and were mounted in loops before flash-cooling in liquid ethane in a liquid-nitrogen bath. Data collection at station ID29 at the ESRF, Grenoble, France using an ADSC Q315 CCD detector revealed that the crystals were monoclinic, belonging to space group *C*2 with unit-cell parameters $a = 141.6$, $b = 37.3$, $c = 55.1$ Å, $\beta = 105.0^\circ$. Data processing with *MOSFLM* (Leslie, 2006), *SCALA* (Evans, 2006) and other programs in the *CCP*4 suite (Winn *et al.*, 2011) yielded intensity data to a resolution of $d_{\rm min} = 1.45$ Å with an overall $R_{\rm merge} = 9.8\%$ (see Table 1 for details). The crystals were found to have one PBGD monomer per crystallographic asymmetric unit and a solvent content of 39%. Structure analysis using the molecular-replacement program *MOLREP* (Vagin & Teplyakov, 2010) with *E. coli* PBGD (42% identity; PBD entry 1pda; Louie *et al.*, 1992) as the search model was successful. Refinement of the *A. thaliana* PBGD structure was undertaken using *REFMAC* (Murshudov *et al.*, 2011) and *SHELX* (Sheldrick, 2008) with manual rebuilding using *Coot* (Emsley & Cowtan, 2004). The geometric restraints for refinement of the dipyrromethane cofactor were generated using *PRODRG* (Schüttelkopf & van Aalten, 2004). The final refined structure and reflection data set were analysed using the validation programs *PROCHECK* (Laskowski *et al.*, 1993), *SFCHECK* (Vaguine *et al.*, 1999) and *MolProbity* (Lovell *et al.*, 2003; Chen *et al.*, 2010), and have been deposited in the PDB (http://www.wwpdb.org) with accession code 4htg. Structure figures were prepared using *CueMol* (http://www.cuemol.org/en) and *POV-Ray* (http://

**Table 1**
X-ray statistics for the *A. thaliana* porphobilinogen deaminase structure.

Values in parentheses are for the outer resolution shell.

| | |
|---|---|
| Beamline | ID29, ESRF |
| Wavelength (Å) | 0.979 |
| Space group | *C*2 |
| Unit-cell parameters | |
| *a* (Å) | 141.6 |
| *b* (Å) | 37.3 |
| *c* (Å) | 55.1 |
| $\beta$ (°) | 105.0 |
| Mosaic spread (°) | 0.6 |
| Resolution (Å) | 32.89–1.45 (1.53–1.45) |
| $R_{\rm merge}$† (%) | 9.8 (64.1) |
| $R_{\rm meas}$‡ (%) | 11.4 (76.7) |
| Completeness (%) | 99.3 (99.4) |
| Average $I/\sigma(I)$ | 5.6 (2.2) |
| Multiplicity | 3.6 (3.5) |
| No. of observed reflections | 179303 (24674) |
| No. of unique reflections | 49235 (7129) |
| Wilson plot *B* factor (Å$^2$) | 15.1 |
| Solvent content (%) | 39.0 |
| Refinement | |
| *R* factor (%) | 14.5 |
| Free *R* factor (%) | 21.7 |
| R.m.s.d. bond lengths (Å) | 0.02 |
| R.m.s.d. bond angles (°) | 2.05 |
| No. of reflections in working set | 46662 |
| Mean holoenzyme *B* factor (Å$^2$) | 19.9 |

† $R_{\rm merge} = \sum_{hkl}\sum_i |I_i(hkl) - \langle I(hkl)\rangle| / \sum_{hkl}\sum_i I_i(hkl)$. ‡ $R_{\rm meas} = \sum_{hkl}\{N(hkl)/[N(hkl)-1]\}^{1/2}\sum_i |I_i(hkl) - \langle I(hkl)\rangle| / \sum_{hkl}\sum_i I_i(hkl)$, where $\langle I(hkl)\rangle$ is the mean intensity of the $N(hkl)$ observations $I_i(hkl)$ of each unique reflection *hkl* after scaling.

www.povray.org), in which the solvent-accessible surface and electrostatic potentials were calculated using the *MSMS* (Sanner *et al.*, 1996) and *APBS* (Baker *et al.*, 2001) software, respectively. Lattice contacts were analysed using the *PISA* server (Krissinel & Henrick, 2007). For molecular modelling of the tetrapyrrole intermediate, initial coordinates and restraints for the ligand were generated using *PRODRG*, and fitting to the active-site cleft was performed using on-the-fly regularization with *Coot*; final geometry idealization was completed using *REFMAC*. Bioinformatic analysis of known plant PBGD genes was conducted using Phytozome (http://www.phytozome.net; Goodstein *et al.*, 2012).

## 3. Results

### 3.1. Overall structure

Expression of a codon-optimized gene for the enzymatic moiety of *A. thaliana* PBGD in *E. coli* allowed the protein to be purified by ion-exchange and gel-filtration chromatography with sufficient yield for crystallization trials. The $M_r$ of the purified enzyme was determined to be 34 930 by electrospray mass-spectrometry, and kinetic analysis revealed that it had a $K_m$ of 7 $\mu M$ and a $V_{max}$ of 5000 nmol $h^{-1}$ per milligram of protein. Crystals which diffracted synchrotron radiation to 1.45 Å resolution were obtained in space group *C*2, with one molecule per asymmetric unit, and allowed the structure to be solved by molecular replacement and refined to an *R* factor of 14.5% and an $R_{free}$ of 21.7%. The resulting model of the enzyme (Fig. 3*a*) was found to have all of the amino acids within the allowed regions of the Ramachandran plot and 92.8% within the 'most favoured' areas by the *PROCHECK* criteria (Laskowski *et al.*, 1993). Validation by *MolProbity* (Chen *et al.*, 2010) indicated that 99.3% of all residues fall within favoured regions of the Ramachandran plot and that all residues are in allowed regions. The electron density for the refined model at 1.45 Å resolution is of very high quality and the estimated r.m.s. coordinate error (Read, 1986) of 0.04 Å suggests that the structure is defined with high accuracy. Indeed, this structure has the highest resolution of all PBGDs that have been analysed by X-ray diffraction to date.

The N-terminal sequence of the protein purified from *Arabidopsis* leaves was determined to be XVAVE... (Jones & Jordan, 1994), which corresponds to the sequence ...KACVAVE... deduced from gene sequencing. In this sequence, the Ala-Cys dipeptide is predicted to be the cleavage site for the chloroplast stromal processing peptidase (Lim *et al.*, 1994). This corresponds to cleavage between positions 62 and 63 of the precursor protein. The amino acids in the X-ray structure reported here have been numbered assuming that the above cysteine (residue 63 of the precursor) forms the N-terminal residue of the mature enzyme, although the first nine amino acids of the expressed protein are not visible in the electron-density map, presumably owing to disorder. The first amino acid that is defined by the electron density for the *A. thaliana* enzyme is therefore Thr10, which corresponds to residue 3 of the *E. coli* enzyme. The electron

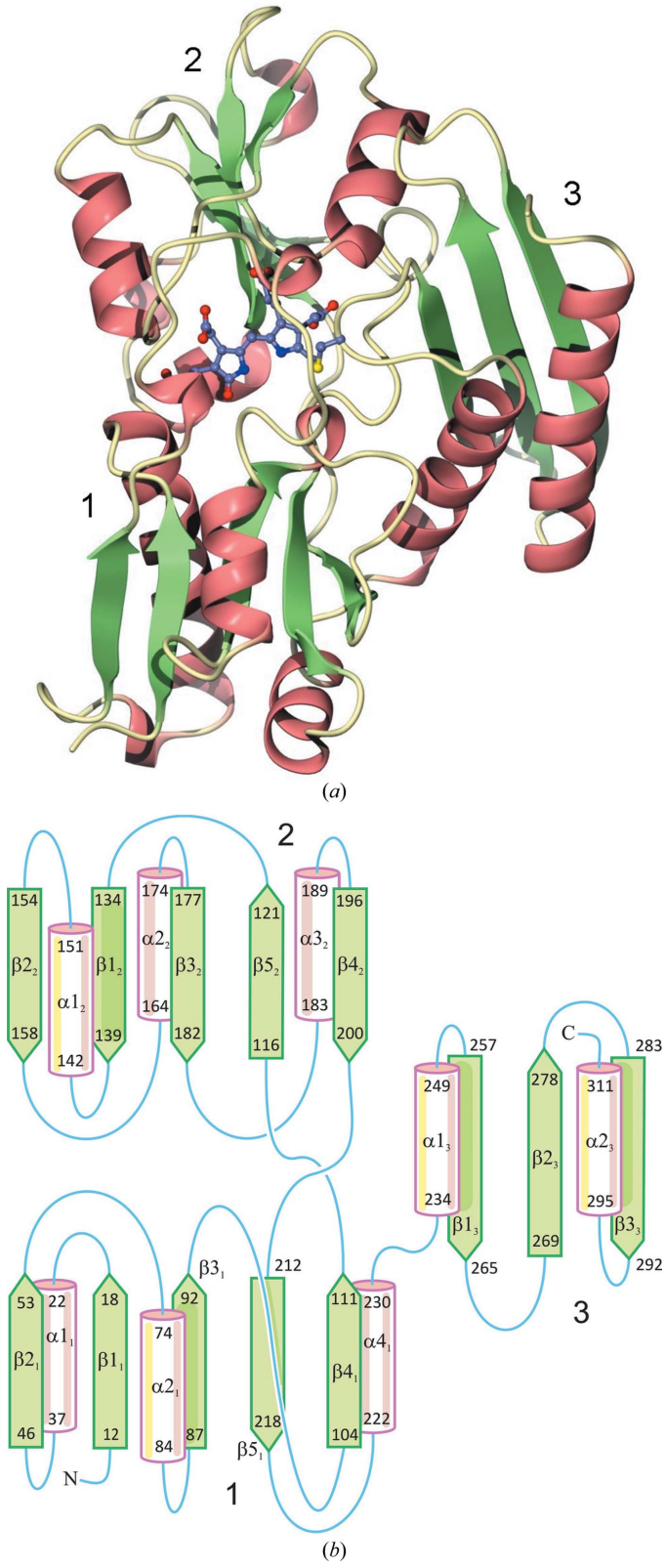


**Figure 3**
Structure of *A. thaliana* porphobilinogen deaminase. (*a*) The tertiary structure is shown with domains 1 and 2 on the left-hand side and domain 3 on the right. The dipyrromethane cofactor and the residue to which it is covalently bound (Cys254) are shown in ball-and-stick representation. (*b*) The overall fold of the protein, emphasizing the topological symmetry of domains 1 and 2. The secondary-structure elements are labelled using the nomenclature of the *E. coli* enzyme (Louie *et al.*, 1992).

density at the C-terminal end of the protein diminishes after Ser311 within the ...ELLSRAG... sequence motif which is strongly conserved in PBGDs from higher plants, suggesting that this marks the end of the ordered region of the protein in these species.

The backbone of the enzyme is folded into three domains (numbered 1–3), each of approximately the same size (Fig. 3*a*). Partway through the N-terminal domain (domain 1) the polypeptide leaves to form domain 2 (residues 111–211) and then returns to complete the last strand of domain 1; the first domain of the structure is thus formed by residues 10–110 and 212–229. The two short connecting regions between domains 1 and 2 are thought to act as hinge points which allow the protein to have appreciable flexibility. The general architecture of domains 1 and 2 shows a strong resemblance to the type II family of periplasmic binding proteins (Louie *et al.*, 1992; Louie, 1993), which are known to adopt open and closed states in response to ligand binding. Each of these domains has an $\alpha/\beta$ topology which consists of a doubly wound five-stranded $\beta$-sheet that is flanked on each face by $\alpha$-helical segments; the penultimate $\beta$-strand of each domain is antiparallel to the other four strands as a result of topological strand exchange between the two domains (Fig. 3*b*). There is a topological twofold axis relating these domains that passes along the substrate-binding channel. In this arrangement, the amino-terminal ends of the $\alpha$-helical segments within both domains 1 and 2 point towards the active-site cleft, suggesting that their helix dipoles are oriented to interact favourably with the negatively charged porphobilinogen moieties of the cofactor and substrate. Domain 3 (residues 230–316) has a completely different topology consisting of a three-stranded $\beta$-meander fold that is preceded and succeeded sequentially by $\alpha$-helices that spatially lie in parallel on one side of the sheet. The dipyrromethane cofactor is covalently attached to the loop connecting $\alpha1_3$ and $\beta1_3$ in domain 3 and is positioned in the deep active-site cleft formed between domains 1 and 2 (Fig. 3*a*).

There are many interactions between the carboxyl groups of the dipyrromethane cofactor and the enzyme, most of which involve basic residues from domain 2 (Figs. 4*a* and 4*b*). The two aromatic pyrrole rings themselves occupy a large cavity, in which they make some direct interactions with the protein and other indirect interactions mediated by ordered water molecules. In *E. coli* PBGD, the cofactor ring furthest from the attachment point (the outer or terminal pyrrole) can adopt two conformations, apparently depending on its oxidation state. In the *A. thaliana* enzyme the terminal pyrrole is extremely well defined in one position, corresponding to the oxidized form in the *E. coli* enzyme (Fig. 4*c*). In both conformations, the pyrrole N atoms form hydrogen bonds to a

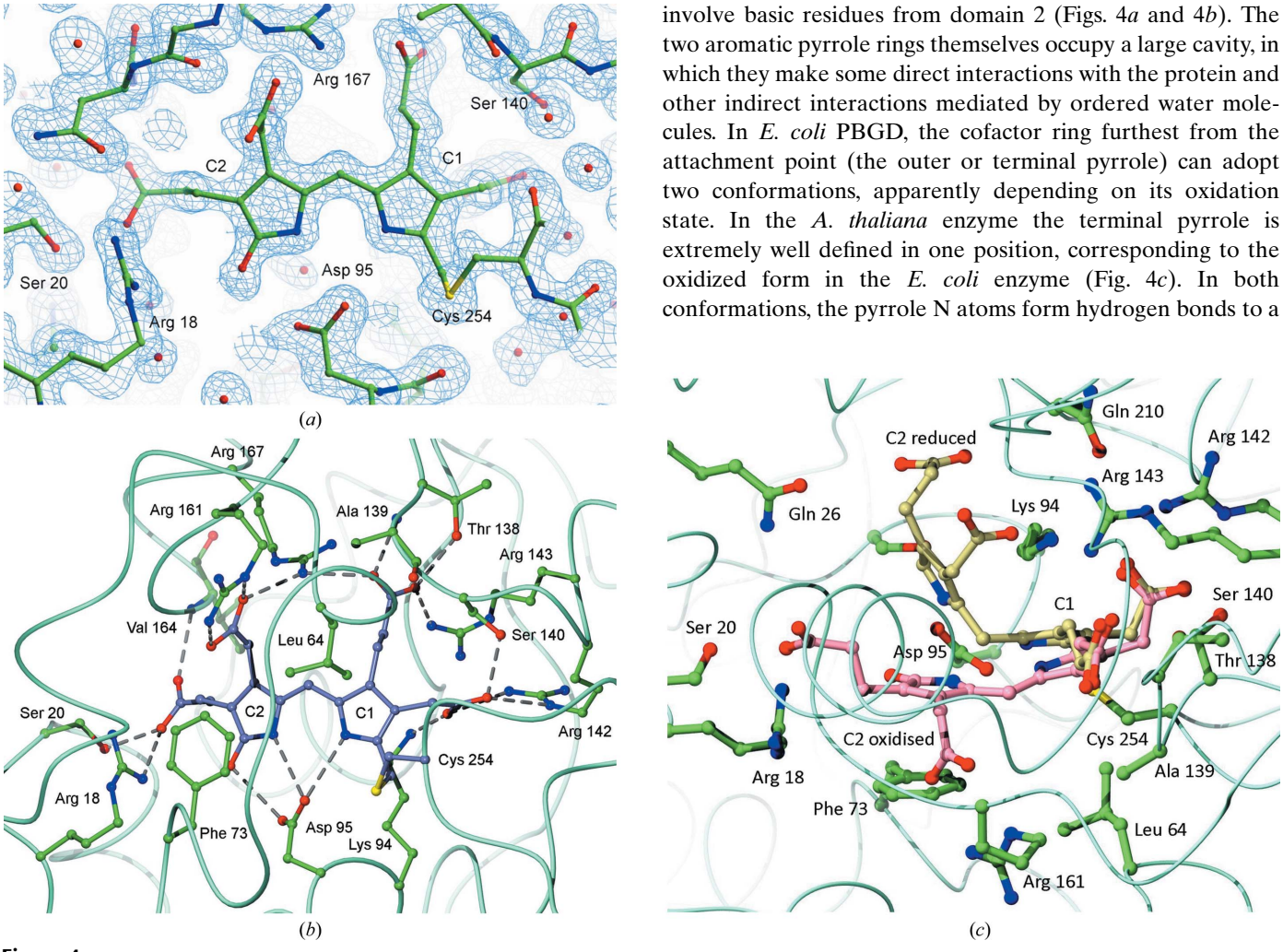


**Figure 4**
The dipyrromethane cofactor. (*a*) The electron density for the cofactor and a selection of the surrounding active-site residues contoured at 2 r.m.s. The oxidation at the $\alpha$-position of the C2 ring (lower left corner of the cofactor) is very clear. Both pyrrole N atoms and the extra O atom are close to the catalytic Asp95 shown at the bottom. (*b*) The residues forming hydrogen bonds and ion-pairs with the cofactor. The cofactor and the cysteine to which it is attached are coloured mauve and the hydrogen bonds are shown as dashed grey lines. The two conformations of the cofactor found in the 'reduced' *E. coli* structure (beige) and the 'oxidized' *A. thaliana* structure (pink) are shown edge-on in (*c*).

domain 1 residue, Asp95, which is invariant and is a very important residue in catalysis. Mutagenesis of the equivalent residue in *E. coli* PBGD (Asp84) to glutamate reduces the $k_{cat}$ by two orders of magnitude without affecting the $K_m$ (Woodcock & Jordan, 1994). Replacing the catalytic aspartate with alanine or asparagine inactivates the enzyme although, intriguingly, the mutant enzyme appears to form a stable $ES_2$ complex (*i.e.* the dipyrromethane is linked to two further pyrromethane rings) as a result of the mutant apoenzyme reacting with the tetrapyrrole preuroporphyrinogen (Shoolingin-Jordan *et al.*, 1996, 1997). Asp95 is thought to function in catalysis by deaminating an incoming porphobilinogen moiety, yielding a reactive azafulvene intermediate which is susceptible to nucleophilic attack by the terminal pyrrole of the enzyme-bound cofactor or intermediate. It should be noted that there are two conserved lysine residues in the vicinity of the catalytic aspartate (Lys70 and Lys94) which probably explains the partial inhibition of the *E. coli* enzyme in the presence of pyridoxal-5′-phosphate and sodium borohydride (Warren & Jordan, 1988). Such treatment would irreversibly modify whichever lysine is able to form a Schiff base with the pyridoxal-5′-phosphate and thereby disrupt the binding of the substrate.

There is evidence that the conformation of PBGD changes during the tetrapolymerization, which comes from the observation that the *E. coli* enzyme becomes increasingly susceptible to inactivation by *N*-ethylmaleimide (NEM) during the reaction (Warren *et al.*, 1995). The susceptible cysteine in *E. coli* PBGD (Cys134) is at some distance from the active site (~12 Å) and lies in a cleft between domains 2 and 3, adjacent to the loop that the cofactor is covalently attached to. Conformational changes during the catalytic cycle may result in the separation of domains 2 and 3 owing to the enzyme 'pulling' the growing chain of pyrroles through the active-site cleft such that each incoming substrate moiety has access to the catalytic aspartate. Alternatively, these conformational changes may be due to reorganization of the polypyrrole as it is assembled. In the final stages of the reaction, the polypyrrole must be repositioned so that the catalytic machinery of the enzyme can facilitate cleavage of the bond linking ring *A* to the cofactor, thereby releasing the tetrapyrrole product. The NEM-susceptible cysteine in *E. coli* PBGD is replaced by a serine residue (Ser145) in the *A. thaliana* enzyme and accordingly the latter is not susceptible to inhibition by NEM (Jones & Jordan, 1994).

### 3.2. Comparison with other PBGD structures

A superposition of *A. thaliana* PBGD with the oxidized *E. coli* enzyme (PDB entry 1pda; Fig. 5*a*) shows the close relationship between these two proteins, which is reflected in a $C^{\alpha}$ r.m.s.d. of 0.8 Å for 286 structurally equivalent residues. A few amino acids at the extreme N-terminal ends of both proteins are in different positions, presumably owing to disorder in this exposed region of the two structures. The two proteins superimpose closely from residue 12 of the *A. thaliana* enzyme up to residue 40, where there is an insertion of four residues with respect to the *E. coli* enzyme. This insertion forms a $\beta$-bulge at the N-terminal end of an exposed $\beta$-strand $\beta2_1$ in the nomenclature of Louie *et al.* (1992). It is some distance from the active site (~30 Å) and so is unlikely to have an effect on the catalytic machinery of the enzyme. In contrast, the other end of this $\beta$-strand leads into a large loop or flap (residues 60–70) over the active site that was completely disordered in the *E. coli* enzyme.

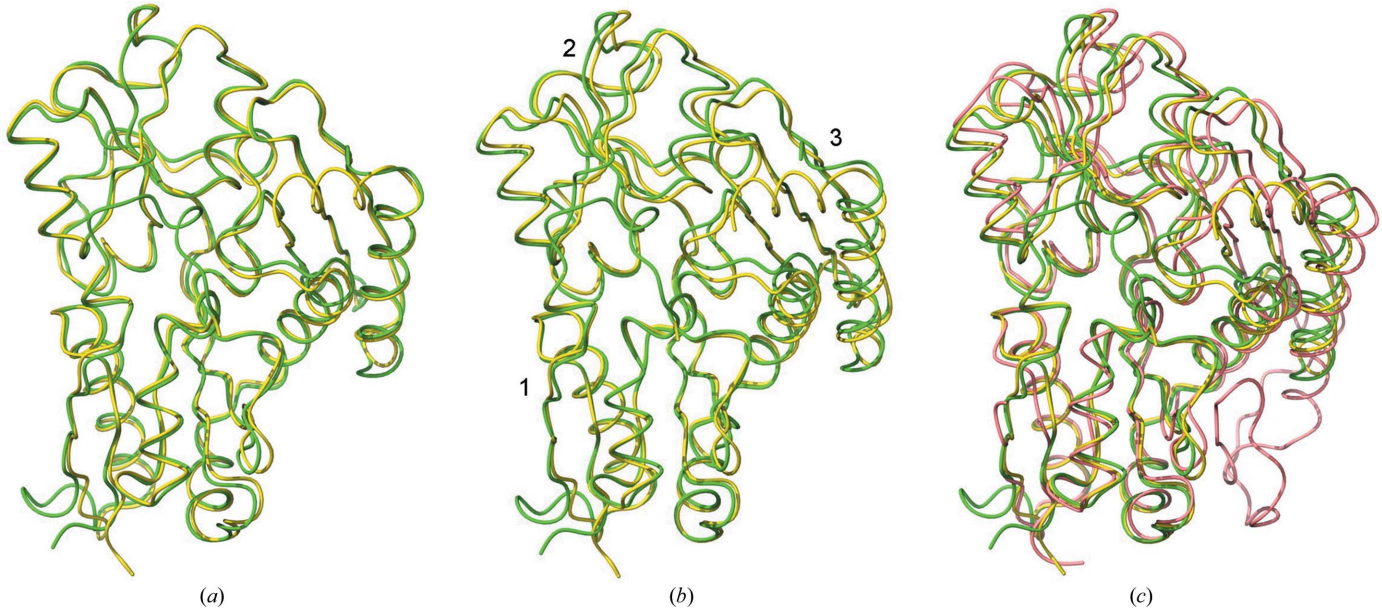


**Figure 5**
Superposition of *A. thaliana* PBGD with the *E. coli* and human enzymes. (*a*) Overall least-squares superposition of plant PBGD (green) with the *E. coli* enzyme (yellow). A superposition of the two enzymes based on domain 1 alone is shown in (*b*) to emphasize the concerted shifts of domains 2 and 3. An overlay of the plant, *E. coli* and human enzymes when superposed by using domain 1 only is shown in (*c*) with the human enzyme coloured pink.

This flap region is extraordinarily well defined in the *A. thaliana* structure and contributes a number of side chains to the active site that are likely to be important in binding the substrate and cofactor as well as in aiding catalysis (Fig. 4*b*). The most notable residue is Leu64, the side chain of which interacts with the first pyrrole ring of the dipyrromethane cofactor (C1) that is bonded directly to Cys254. In addition, the loop contains Lys70, which is situated at the mouth of the active-site cavity and, although the side chain of this residue is somewhat disordered and is at some distance from the cofactor (∼10 Å), its position at one of the openings to the active-site cleft and its degree of conservation are suggestive of an important role. It is likely that the improved definition of the flap region in the *A. thaliana* enzyme is partly due to a neighbouring protein molecule being appreciably closer in the crystal lattice than is the case in the *E. coli* PBGD crystal structure. Whilst the flap shows a marked reduction in disorder in the *A. thaliana* enzyme and thus provides a more effective cover over the active site, the pyrroles of the dipyrromethane remain in contact with several partly buried water molecules that appear to be trapped beneath the flap region.

The remainder of the *A. thaliana* and *E. coli* PBGD structures are highly similar, with the exception of an insertion at position 159 of the plant enzyme which introduces a $\beta$-bulge in strand $\beta2_2$, another single-residue insertion at residue 267 in the $\beta$-hairpin loop between strands $\beta1_3$ and $\beta2_3$, and a two-residue insertion at position 294 which extends the N-terminal end of helix $\alpha2_3$ by approximately half a turn. One final significant difference between the *A. thaliana* and *E. coli* PBGD structures involves the C-terminal $\alpha$-helix that is formed by residues 298–307 of *E. coli* PBGD (helix $\alpha3_3$). This helix apparently has no counterpart in *A. thaliana* PBGD owing to the last few residues of this enzyme being invisible in the electron-density map. Whilst this difference might seem to be of little significance, in *E. coli* PBGD helix $\alpha3_3$ comes quite close to the active site (∼15 Å) and interacts extensively with the cofactor-binding loop. The lack of this helix in *A. thaliana* PBGD causes the residues of the cofactor-binding loop and the preceding helix $\alpha4_1$ to move appreciably (∼1.5 Å in the extremity) into a more exposed position and this seems to be coupled to rigid-body domain movements within the enzyme (see below).

Superposition of *A. thaliana* PBGD with the human enzyme (Song *et al.*, 2009) yields a somewhat higher r.m.s.d. of 1.1 Å for 236 structurally equivalent $C^{\alpha}$ atoms. The fact that human PBGD differs structurally from the plant enzyme to a greater extent than the *E. coli* enzyme is in accordance with the amino-acid sequence identities for the structured regions of all three proteins. The *A. thaliana* enzyme has a sequence identity of 46% to the *E. coli* enzyme and an appreciably lower sequence identity of 38% to human PBGD. Like the *E. coli* enzyme, human PBGD exhibits major disorder in the region of the active-site flap, with some 20 residues being undefined by the electron-density map. In addition, the human enzyme has a large ordered insertion of approximately 30 amino acids in the loop connecting the last $\beta$-strand of domain 3 to the ensuing $\alpha$-helix (Gill *et al.*, 2009). The residues forming this insertion lie in a gap between domains 1 and 3 which in the other structures is simply a large solvent-filled channel.

### 3.3. Domain movements

Superposition of *A. thaliana* PBGD with the structure of the *E. coli* enzyme (PDB entry 1pda) suggests that there are small but significant differences in the relative orientations of the three domains of each enzyme. Accordingly, the individual domains may be superimposed with r.m.s.d. values (typically 0.6 Å) that are appreciably lower than that obtained for the structures as a whole (0.8 Å). These differences in the relative domain orientation are emphasized most clearly when the two enzymes are superimposed using only domain 1, as shown in Fig. 5(*b*). Least-squares superposition with this domain reveals that the other two domains of the *A. thaliana* enzyme (domains 2 and 3) are both tilted by 3.3° in a concerted manner with respect to their counterparts in the *E. coli* enzyme. Intriguingly, these domains provide the bulk of the ionic bonding and the covalent link to the cofactor. Whilst a rotation of this order is quite small, it equates to a shift of approximately 2 Å at the extremities of domains 2 and 3. The concerted nature of this domain tilt probably stems from the fact that the interactions between domains 2 and 3 are slightly more extensive in terms of interface area (∼850 $Å^2$) than the other domain–domain interfaces within the molecule (∼800 $Å^2$ each). The concerted movement of the second and third domains is redolent of the action of a ratchet handle and may allow the enzyme to 'pull' on the growing polypyrrole chain that is covalently attached to domain 3 during the elongation process. Domains 1 and 2 are covalently linked by two antiparallel segments of polypeptide in the vicinity of residues 111 and 211. The irregularity of the backbone structure of these segments suggests that in solution they are capable of acting together as a fairly flexible inter-domain hinge. The same rigid-body effect is found when comparing the *A. thaliana* enzyme with the structure of reduced *E. coli* PBGD (Hädener *et al.*, 1999), suggesting that the oxidation state of the cofactor does not greatly influence the domain movements.

The differences in domain orientation in the plant and *E. coli* PBGD crystal structures may, of course, stem from differences in crystal lattice contacts or from differences in crystallization conditions such as pH, although this particular parameter appears to be relatively constant for the structures being compared. There may also be a contribution from the marked ordering of the active-site loop (residues 60–70) in the *A. thaliana* enzyme. However, the domain movements are suggestive of an important flexibility in the structure which may well have a bearing on the mechanism of action. Whilst the lack of inter-domain contacts between domains 1 and 2 emphasizes their potential for independent movement, it must be remembered that there are numerous inter-domain contacts that are mediated by the dipyrromethane cofactor and therefore any movement of the cofactor is likely to affect the relative orientation of these two domains.

A superposition of the *A. thaliana* PBGD structure with that of the human enzyme (Gill *et al.*, 2009; Song *et al.*, 2009), shown in Fig. 5(*c*), reveals a similar concerted movement of domains 2 and 3 with respect to domain 1. Although the direction of the domain shift in the human enzyme is slightly different from that found for *E. coli* PBGD, it is clear that the last two domains of the enzyme appear to move, with respect to domain 1, as a relatively independent rigid group. The net rotation of domains 2 and 3 in the human enzyme is 4.0°, which corresponds to a displacement of approximately 4.1 Å at the extremity of the protein. The greater relative domain movements in human PBGD may stem partly from the substantial insertion of 30 residues in domain 3; these extra residues form a large loop that resides at the interface between domains 1 and 3 (Gill *et al.*, 2009). Whilst it is interesting to speculate on the biological role of the small rigid-body movements that we observe on comparing these PBGD structures, it should be noted that these are relatively small compared with the substantial domain movements that have been observed crystallographically in other members of the type II periplasmic binding protein family, which can be of the order of 50° (Louie, 1993). However, the domain shifts that we report here may give an indication of the larger movements that take place during the catalytic reaction and it is quite likely that all three domains can move with some degree of independence, as suggested by NEM-modification studies of *E. coli* PBGD (Warren *et al.*, 1995).

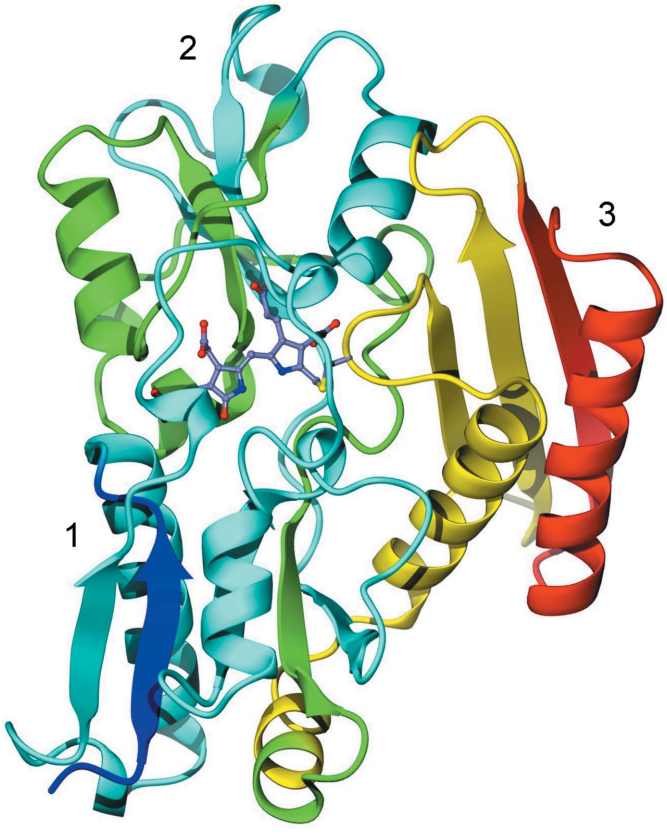


**Figure 6**
The intron–exon boundaries of *A. thaliana* PBGD. The exons are coloured from blue to red starting from the N-terminal end of the protein (the numbers refer to the protein domains). Exon 1 (dark blue) codes for the amino acids up to residue 19, exon 2 (cyan) encodes residues 20–155, exon 3 (green) encodes residues 156–224, exon 4 (yellow) encodes residues 225–283 and exon 5 (red) encodes the remainder of the protein.

### 3.4. Gene structure

The human PBGD gene has no less than 15 exons (Chretien *et al.*, 1988; Yoo *et al.*, 1993), in contrast to five exons for the *A. thaliana* enzyme (Lim *et al.*, 1994). The intron–exon boundaries of the plant gene are displayed on the tertiary structure of *A. thaliana* PBGD in Fig. 6. The first exon codes for the transit peptide and the first 19 residues of the mature protein, although only nine amino acids of this region are visible in the electron-density map. The second exon encodes residues 20–155 which form a large part of the first two domains of the protein, between which the cofactor is held by salt bridges and other noncovalent interactions. The third exon (residues 156–224) essentially codes for the remainder of these two domains, excluding the last helix, $\alpha 4_1$. This short helix and the third domain of the protein, to which the cofactor is covalently attached, are encoded by the fourth and fifth exons. These span residues 225–283 and 284–320, respectively, although the electron density for residues beyond Ser311 is of poor quality, suggesting a disordered C-terminus.

The exon–intron boundaries of *A. thaliana* PBGD all correspond to loop regions of the tertiary structure, as might be expected from evolutionary arguments (see, for example, Go, 1981). There are also some interesting similarities and differences between the intron–exon boundaries of the plant and human genes (Wood *et al.*, 1995). The first and last exons of the plant gene code for regions of the protein that correspond quite closely to those parts encoded by exons 3 and 15 of the human gene. The second plant exon corresponds to exons 4–9 of the human gene, the third plant exon corresponds to exons 10, 11 and part of exon 12 of the human gene and the fourth plant exon corresponds to the remainder of exon 12 in human PBGD along with exons 13 and 14. The results are broadly consistent with comparative genomic studies, which have shown that around 14% of the intron positions in animal genes match those of plant genes (Fedorov *et al.*, 2002). These studies provide strong evidence for ancestral introns that predate the animal–plant divergence of eukaryotes. In contrast to the human PBGD gene, which exhibits complex alternative splicing patterns for the erythroid and non-erythroid forms of the enzyme (Chretien *et al.*, 1988), the plant enzyme would appear to be coded for by a single housekeeping gene that is expressed constitutively and spliced identically in all tissues (Lim *et al.*, 1994). Finally, whilst the positions of the intron–exon boundaries are strongly conserved with respect to the protein fold in higher plants, there are appreciable differences in simpler plant organisms such as the *Volvox* and *Chlamydomonas* algae.

### 3.5. Cofactor-binding site

PBGD possesses a dipyrrole group or dipyrromethane which is covalently bound to an active-site cysteine, Cys254, in the *A. thaliana* enzyme. The electron density for the cofactor is

shown in Fig. 4(*a*) along with the interacting residues of the enzyme, which are predominantly polar in nature. The four carboxylate groups of the dipyrromethane form extensive ion-pair interactions with conserved or invariant basic residues of the enzyme, principally arginines. The resolution of the X-ray data in the current analysis shows unambiguously that the active-site dipyrromethane is oxidized to the dipyrromethenone form, in which an O atom is double-bonded to the $\alpha$-C atom that would normally act as the reaction site for an incoming porphobilinogen. When the cofactor is oxidized in this manner, the enzyme is inactive (Jordan, 1991). Needless to say, the extremely well defined electron density for the cofactor in the *A. thaliana* PBGD structure suggests that perhaps this oxidized state is fortuitous for the purposes of obtaining a highly detailed crystallographic 'snapshot' of the enzyme active site.

The porphobilinogen moiety that is covalently attached to Cys254 is denoted the C1 ring of the dipyrromethane and the adjacent PBG is denoted C2 (Fig. 4*b*). The acetate group of the C1 ring forms a hydrogen bond to the side chain of Ser140 as well as two charge-assisted hydrogen bonds to the guanidinium group of Arg142 and another to the side chain of Lys94. The propionate of this pyrrole forms hydrogen bonds to the main-chain N atom of Ala139 and the side-chain O atom of the preceding residue Thr138. In addition, this propionate group forms two charge-assisted hydrogen bonds to the guanidiniums of Arg143 and potentially Arg167. The acetate group of the C2 ring forms two charge-assisted hydrogen bonds to the guanidinium group of Arg161 and another to Arg167, which is also involved in binding the C1 propionate. The propionate group of the C2 pyrrole forms hydrogen bonds to the main-chain N atom of Val164 and the side-chain O atom of Ser20 as well as an ion-pair interaction with Arg18. Each of the C2 pyrrole carboxylates forms a hydrogen bond with well defined solvent molecules in the active-site cleft. The important roles of the arginine residues which interact with the cofactor have been confirmed by mutagenesis of the equivalent residues in the *E. coli* enzyme (Lander *et al.*, 1991; Jordan & Woodcock, 1991).

In addition to these polar interactions, the cofactor makes a limited number of contacts with nonpolar residues, with the two most notable being Leu64 and Phe73, which form a rather disjointed hydrophobic patch that interacts with the front face of the dipyrromethane, as viewed in Fig. 4(*b*). The interaction with Leu64 is particularly interesting because this residue lies within a region of the *A. thaliana* structure (residues 60–70) that was completely disordered in the *E. coli* and human enzymes. The flexibility of this region may stem partly from the fact that it possesses a conserved glycine-rich consensus sequence: $GGK^{70}G$. However, in the *A. thaliana* PBGD structure it appears that crystal contacts, predominantly with the symmetry-related molecule at $(\frac{1}{2} - x, \frac{1}{2} + y, -z)$, stabilize this region of the molecule such that it is extremely well defined by the electron-density map, in contrast to the *E. coli* and human structures, in which the lack of equivalent crystal contacts leads to disorder in this region. Whilst it might be argued that crystal contacts could perturb the conformation of this loop, the lattice contacts made by this region are not as extensive as, for example, those involving the C-terminal $\beta$-strand of the protein ($\beta 3_3$). The conformation of this latter region is highly conserved in the other PBGD crystal structures, despite its involvement in numerous different crystal contacts in the known structures. One point of interest is that the insertion of asparagine at position 159 of *A. thaliana* PBGD introduces a $\beta$-bulge in strand $\beta 2_2$ that is oriented towards the active-site flap. Both the asparagine and the preceding residue (Glu158) have their side chains oriented towards the tip of the flap and interact with it by water-mediated hydrogen bonds and van der Waals contacts, presumably stabilizing its conformation.

The remaining interactions with the dipyrromethane cofactor involve water molecules that occupy substantial cavities on each side of the cofactor. Indeed, the rearwards face of the cofactor, from the view shown in Figs. 4(*a*) and 4(*b*), points towards a substantial pocket with a volume that is larger than that occupied by the cofactor itself. The base of this pocket is isolated from the bulk solvent, principally by the residue Arg188 along with a number of other conserved ionic residues which form salt bridges and appear to form a tight molecular plug. Mutagenesis of the residue equivalent to Arg188 in *E. coli* PBGD significantly reduces the $k_{cat}/K_m$ of the enzyme, suggesting that it has an important role (Lander *et al.*, 1991; Jordan & Woodcock, 1991).

From a catalytic standpoint, the most important interactions made by the cofactor with the enzyme are likely to be those involving the pyrrole N atoms, both of which hydrogen-bond to one of the side-chain O atoms of Asp95. The other side-chain O atom of this aspartate is hydrogen-bonded to the oxygen substituent on the C2 ring which arises from oxidation of the cofactor. The very precise way in which the dipyrromethane appears to be held close to this invariant aspartate residue suggests that it plays a very important role in catalysing the reaction, as has been confirmed by mutagenesis (Woodcock & Jordan, 1994).

The two pyrrole rings of the cofactor in the *A. thaliana* enzyme appear to be almost coplanar, as they are in the predominant conformation of the cofactor present in the original structure of the *E. coli* enzyme (Louie *et al.*, 1992). In contrast, the structures of the human and *E. coli* enzymes crystallized under reducing conditions (Gill *et al.*, 2009; Hädener *et al.*, 1999; Louie *et al.*, 1996; Song *et al.*, 2009) show that the cofactor has a more folded conformation, with the C2 ring being largely buried within the active site (see Fig. 4*c*). Despite these conformational differences, the C1 pyrrole carboxylates make essentially the same salt-bridge interactions with the enzyme owing to compensating reorientations of these pyrrole side groups. However, the C2 rings make very different interactions with the enzyme in the inactive oxidized and active reduced states. Based on the reduced *E. coli* PBGD structure, the acetate group of the C2 ring would interact with the guanidinium group of Arg143 in *A. thaliana* PBGD along with the main-chain N atom of Ala182, whereas the propionate group would occupy a position in which it would make no direct hydrogen bonds or ionic interactions with the

enzyme, although it is oriented in the directions of Ser92 and Arg188 and could potentially interact with them *via* water molecules. Intriguingly, the existence of a well defined feature in the electron-density map of *A. thaliana* PBGD closely resembling the C2 acetate group of the reduced *E. coli* enzyme suggests either that it represents a minor conformation of the cofactor or that an acetate anion has been retained in this position by the enzyme during purification and crystallization. In addition, the conserved nature of amino acids forming the 'reduced' and 'oxidized' binding pockets for the C2 ring suggests that both may well be functionally important. Movement of the Arg188 side chain could in principle provide an opening to bulk solvent and allow PBG molecules to enter the active site directly into the 'reduced' C2 pocket. However, the Arg188 side chain appears to be held in place reasonably strongly by neighbouring amino acids and inspection of the solvent-accessible surface (shown in Fig. 7) establishes that it is relatively neutral close to this residue, suggesting that this is unlikely to be an attractive entry site for the negatively charged substrate. In contrast, the active site has three other substantial openings to bulk solvent, the largest of which is in the general vicinity of the conserved flap residue Lys70 and the other two of which lie close to Arg18, which contributes to the 'oxidized' C2 pocket. These openings have a substantial electropositive character, suggesting that they would be the predominant entry sites for incoming PBG moieties.

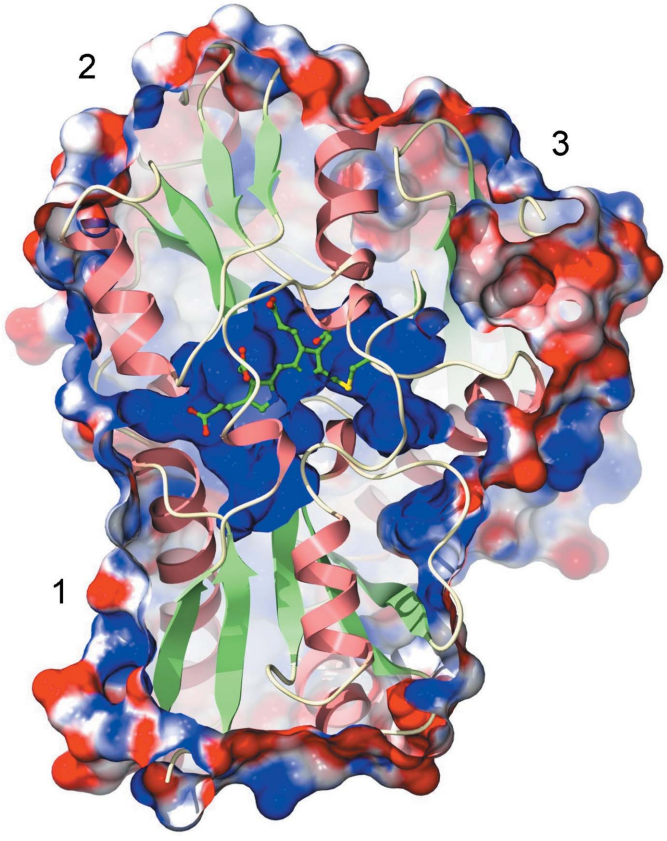


**Figure 7**
The electrostatic surface of *A. thaliana* PBGD. The solvent-accessible surface of the enzyme is shown coloured according to the electrostatic potential. The surface has been clipped to show the large and highly electropositive (blue) binding site for the cofactor, which is formed predominantly of conserved arginine residues. The dipyrromethane and the cysteine residue to which it is attached (Cys254) are displayed in ball-and-stick representation.

In the *A. thaliana* PBGD structure, the C2 ring clearly has an oxygen substituent at the $\alpha$-C atom which would normally act as the reaction site for an incoming porphobilinogen. If the cofactor remains in the same position during the catalytic cycle, our structure suggests that a pyrrole could be added to the C2 cofactor ring with relatively small conformational changes in either the enzyme or cofactor. In contrast, the conformation of the C2 ring adopted in the 'reduced' PBGD structures is such that an incoming pyrrole would not have sufficient space to bind close enough to the free $\alpha$-position of the C2 ring in order for a reaction to occur. If the cofactor maintains the interactions that we observe in the *A. thaliana* structure during the elongation reaction, it is possible that the active conformation of the cofactor is that found in the structure reported here, in which the $\alpha$-position C atom is oxidized. If this model of elongation is correct, it is possible that the cofactor would have to be in the planar conformation in order to react with the first incoming substrate moiety, but this conformation presumably also makes the cofactor susceptible to oxidation and inactivation. In contrast, if the cofactor moves appreciably during elongation and vacates the C2 site, this could allow an incoming pyrrole to bind in the now empty C2 pocket. Given that there are two conformations observed for the C2 ring in *E. coli* PBGD, depending on its oxidation state, the incoming pyrrole could in principle bind in either of the corresponding positions. Although binding in the 'reduced' position would cause the carboxylates of the newly bound PBG to have fewer hydrogen bonds and ionic interactions with the enzyme than in the 'oxidized' position, there are plenty of direct contacts between the C2 pyrrole ring and the enzyme in the 'reduced' conformer, most of which involve the invariant active-site residues Ser92, Lys94 and Asp95.

### 3.6. Solvent channels forming the active site

In the structure of *E. coli* PBGD, the loop that is predicted to cover the active site (residues 49–59) is substantially disordered and consequently the dipyrromethane cofactor has an appreciable solvent-accessible area of 47 Å$^2$. In contrast, the equivalent loop of the *A. thaliana* enzyme (residues 60–70) is much better defined and seems to shield the cofactor from bulk solvent to a greater extent. Accordingly, the solvent-accessible area of the cofactor in the plant enzyme has a lower value of 37 Å$^2$ (omitting the additional oxygen present in this structure for comparison purposes). However, the active-site dipyrromethane still has substantial solvent-accessibility owing to a deep dumbbell-shaped solvent channel that essentially surrounds the active-site cysteine residue and allows the C1 and C2 rings of the cofactor to have access to the bulk solvent *via* openings which are at opposite ends of the channel. The solvent-accessible surface of the substrate-binding channel is shown in Fig. 7, where it is colour-coded according to the electrostatic potential of the surrounding

enzyme moiety. The markedly electropositive nature of this channel stems from the abundance of arginine and lysine residues that form the dipyrromethane (and substrate) binding site. This figure illustrates how the active site of the enzyme possesses a positively charged surface to match the increase in negative charge of the polypyrrole intermediates formed during elongation and emphasizes how the enzyme could permit these intermediates to adopt a range of conformations. The surface representation also demonstrates the more tenuous nature of the interface between domains 1 and 3, which has a number of invaginations that presumably allow solvent access to residues that become exposed in different conformational states of the enzyme.

### 3.7. Modelling a polypyrrole intermediate in the active site

If the cofactor is omitted from the surface and electrostatic calculations (as shown in Fig. 7), the enzyme is seen to have an extraordinarily large and complex active-site cavity which is lined with basic residues. In addition to those described above for the C1 and C2 binding sites, the following basic enzyme residues clearly have the potential to bind pyrrole carboxylate groups: His91, Arg244 and Arg255. Interestingly, rotation of the dipyrromethane around the covalent bond linking the methylene C atom of the C1 pyrrole with the Cys254 S atom would cause the C1 and C2 rings to vacate their respective binding sites and potentially allow another pyrrole moiety to bind at the catalytic centre of the enzyme. The C1 ring can move into a substantial empty pocket in which its carboxylates could interact with two arginine residues. Whilst the accommodation of a pyrrole in this predicted pocket would require slight conformational change within the enzyme, it does allow quite a substantial movement of the cofactor such that the C2 ring moves into the pocket previously occupied by C1. This would 'free up' the binding site previously occupied by the C2 ring and facilitate the binding of an incoming pyrrole moiety in this pocket. Large movements of the cofactor would presumably be aided by the presence of two almost completely invariant glycine residues which precede the active-site cysteine in the consensus sequence ...LEGGC... within the cofactor attachment loop of domain 3.

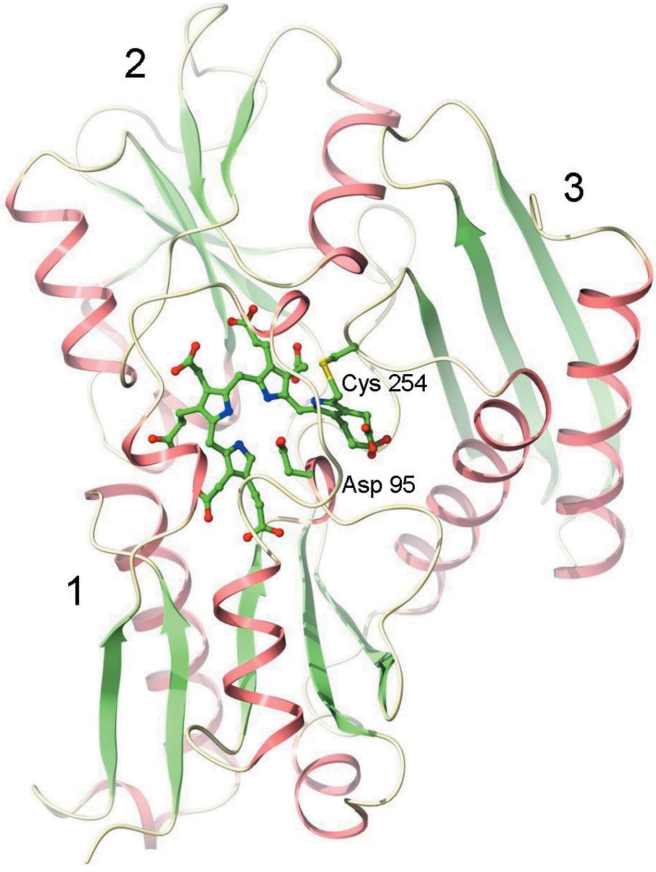


**Figure 8**
A putative model of a covalently bound linear tetrapyrrole. This model was constructed by extending the dipyrromethane cofactor by one PBG unit at both ends followed by molecular-geometry optimization. The additional pyrroles occupy vacant space in the substrate-binding cavity shown in Fig. 7, although small movements of domain 1 and the Cys254 loop were needed. The catalytic residue Asp95 (shown) interacts with three of the pyrrole N atoms.

Inspection suggests that at least four pyrroles could be accommodated without substantial movement of the protein domains or the active-site residues. To test this hypothesis, a linear tetrapyrrole was modelled into the active-site cleft with the two central pyrroles occupying roughly the same position as the dipyrromethane cofactor in the current structure. In attempting to predict additional pyrrole-binding subsites, it seemed most informative to try to model an additional porphobilinogen moiety at each end of the cofactor, as is found in the high-resolution crystal structure reported here. A model built in this way may mimic one of the conformational states of the $ES_2$ intermediate which is formed during elongation of the polypyrrole chain. Interestingly, this intermediate has been shown to be more stable than either of the other isolable intermediates $ES_1$ or $ES_3$ (Scott *et al.*, 1988; Warren & Jordan, 1988), suggesting that the enzyme active site can accommodate a covalently bound linear tetrapyrrole particularly favourably. This is expected since the cofactor-assembly process involves the binding of the tetrapyrrole preuroporphyrinogen ($S_4$) to the apoenzyme (Awan *et al.*, 1997). Indeed, $S_4$ has been shown to be the preferred substrate for the apoenzyme and covalent binding of $S_4$ to it will automatically generate $ES_2$.

In modelling the covalently bound tetrapyrrole, a movement of the N-terminal domain of approximately 1 Å was necessary to relieve steric hindrance and a local rearrangement of the cofactor attachment loop was needed to ensure satisfactory geometry of the thioether linkage with Cys254 after regularization, as described in §2. In this speculative model (Fig. 8), the second, third and fourth pyrroles are approximately coplanar, with their NH groups oriented towards and close to the catalytic aspartate residue Asp95. Importantly, this model suggests that the ring N atoms of these three pyrroles can simultaneously interact with the catalytic aspartate group. To accommodate four pyrroles in the active site, the ring which is covalently attached to Cys254 has to be oriented appreciably differently from the other three pyrroles. One important feature of this model is that the extra

carboxylate groups of the tetrapyrrole can interact with a number of conserved basic residues in addition to those which are already known to interact with the dipyrromethane cofactor. For example, the carboxylates of the first pyrrole are close to the guanidiniums of Arg244 and Arg255 such that only minor reorientations of these side chains allows them to interact favourably with the ligand. Of these two residues, Arg244 is strongly conserved or invariant. The interactions made by the second and third pyrroles in the model are essentially the same as those made by the dipyrromethane cofactor observed in the X-ray structure (see above). The fourth ring of the tetrapyrrole is putatively positioned such that its carboxylates can interact with the side chains of Arg18 and His91. Both of these are strongly conserved or invariant residues, and mutations of Arg18 in the human gene are associated with the hereditary disease AIP. It should be noted that since Arg18 is also involved in interactions with the cofactor at the C2 site, mutations of this residue potentially affect two pyrrole-binding sites.

### 3.8. Catalytic residues and implications for the mechanism of elongation

The extraordinarily tightly bound nature of the cofactor in the *A. thaliana* PBGD structure suggests that the positions of its two constituent pyrroles might possibly represent the two PBG moieties which are condensed in the elongation reaction. The fact that the carboxylate groups of the dipyrromethane cofactor are held in place by a large number of charge-assisted hydrogen bonds suggests that the conformation of the cofactor found in this crystal structure is highly relevant to the mechanism of catalysis and that any mechanistic proposals should encapsulate this state as the endpoint of the condensation reaction. However, one important qualifier to this statement is that the cofactor is observed to adopt two well defined conformations depending on its oxidation state in the different crystal structures of the PBGD enzyme (see §3.5). Since the 'oxidized' conformation of the cofactor allows it to form more hydrogen bonds and ionic interactions with the enzyme than the 'reduced' conformation, it is possible that the 'oxidized' conformation, as observed in the current structure, is more stable. Thus, intermediates formed during elongation of the polypyrrole chain are likely to bind in a manner which mimics either the 'reduced' or 'oxidized' conformation of the cofactor, or perhaps both, at different stages of the reaction.

One obvious prerequisite of this hypothesis is that elongation of the pyrrole chain would involve movement of the cofactor such that the C2 ring vacates its binding site and subsequently occupies the former position of the C1 ring in order for an incoming porphobilinogen to be accommodated. Movement of the cofactor is presumably linked to movement of the protein domains. The alternative hypothesis is that the cofactor can remain close to the position observed in the X-ray structure as an incoming pyrrole is condensed with it. If the newly added pyrrole can then be repositioned, this local movement may be sufficient to 'free up' the

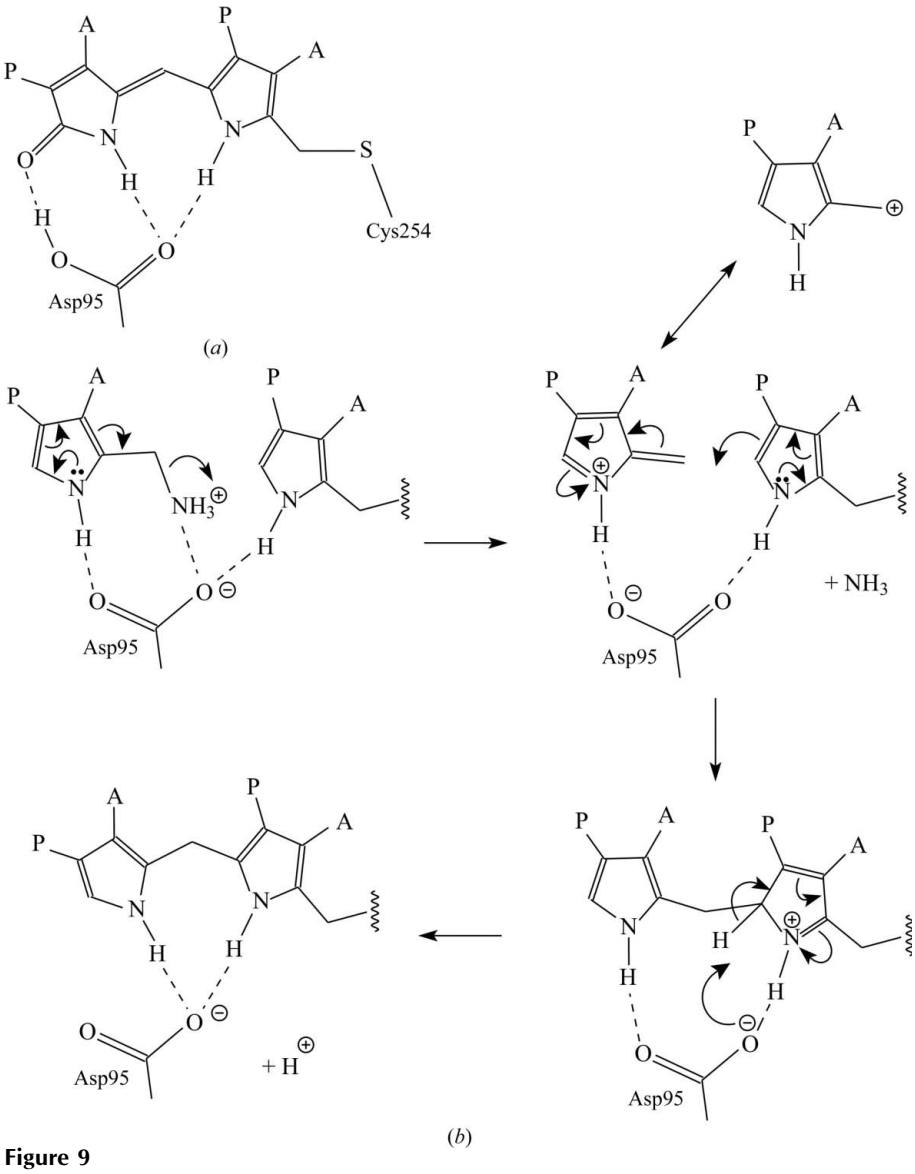


**Figure 9**
The catalytic mechanism of PBGD. (*a*) The interactions made by Asp95 with the oxidized cofactor. (*b*) A putative mechanism showing the likely involvement of the aspartate. Asp95 deaminates the incoming porphobilinogen to form a carbocation intermediate that is stabilized by resonance (as shown) and is susceptible to nucleophilic attack by the terminal enzyme-bound pyrrole, with this too potentially being activated by the catalytic aspartate.

catalytic apparatus for the next elongation cycle. In the structure of the reduced enzyme the C2 ring of the cofactor is buried further within the active-site cavity, which creates more space for an incoming pyrrole to bind close to the catalytic aspartate without displacement of the cofactor. However, this possibility does not seem very feasible given the tight geometric constraints of the active site. The main problem with this proposal is that steric hindrance between the C2 ring and the incoming pyrrole would be considerable and would probably force the C2 ring into a conformation in which its free $\alpha$-position would be even less accessible to the incoming substrate. From steric considerations, it would be very unlikely for an incoming pyrrole to approach sufficiently close to the C2 ring, as observed in the reduced enzyme structure, for condensation to occur. However, if we assume that the cofactor moves during the reaction, the C2 ring will vacate its pocket and an incoming PBG moiety can then bind in its position, mimicking either the 'reduced' or 'oxidized' conformation of the cofactor. Thus, the hypothesis that the polypyrrole chain is pulled through the active site has the advantage that each condensation reaction would involve pyrrole moieties binding tightly to the enzyme in the manner that is observed for the dipyrromethane cofactor in the holoenzyme structure. Our modelling has indicated that whilst additional pyrrole moieties can readily be accommodated in the active-site cavity, the interactions that they make with the enzyme are less extensive than those of the cofactor. One of the general themes of structural enzymology is that those parts of a substrate which bind close to the catalytic centre of an enzyme are bound the most tightly by the protein moiety either to induce strain or to stabilize the transition state. The numerous contacts made between the C1 and C2 carboxylates of the cofactor and conserved residues of the protein is indeed redolent of the sort of interactions observed in complexes of enzymes with transition-state analogues (Mobashery & Kotra, 2002), suggesting that the C1 and C2 rings are bound in the catalytically relevant mode. The tight-binding nature of these contacts is also consistent with the known stereospecificity of the enzyme (Jordan, 1991 and references therein). Hence, the hypothesis which we will consider further is that during elongation the enzyme-bound polypyrrole moves such that its terminal ring occupies the same position as the C1 ring of the cofactor in the current structure to allow the incoming pyrrole to bind tightly in the pocket that is occupied by C2 in the present structure.

We therefore now focus on the amino-acid residues which are observed to interact with the pyrrole-ring N atoms of the cofactor, as these are likely to be important in the chemical mechanism of catalysis. The invariant aspartate residue Asp95 forms a number of important hydrogen bonds, *e.g.* one of the carboxylate O atoms makes a hydrogen bond to each of the pyrrole N atoms. The other O atom of this carboxylate forms a hydrogen bond to the side chain of Ser92, which resides in the same stretch of conserved polypeptide as the aspartate that forms the catalytic floor of the active site ($VHS^{92}XKD^{95}$). Intriguingly, this O atom of the aspartate forms a well defined hydrogen bond to the keto O atom arising from oxidation of the C2 ring of the cofactor (Fig. 9*a*). Hence the carboxyl-group O atom making this interaction must be protonated and the other O atom, which interacts with the two pyrrole N atoms, must be the keto O atom of the carboxyl. These observations unambiguously define the protonation state of the catalytic aspartate in the structure. Since both O atoms of the catalytic aspartate also form a single hydrogen bond to a water molecule, the hydrogen-bonding capacity of the carboxylate appears to be fully satisfied. These interactions suggest that the side chain of Asp95 is uniquely placed close to both the pyrrole N atom and the amino side group of incoming substrate in order for it to function in acid–base catalysis of pyrrole condensation.

In chemical terms, each of the ring-coupling reactions involves two steps: (i) deamination of the incoming porphobilinogen side group and (ii) nucleophilic attack by the $\alpha$-C atom of the terminal ring of the enzyme-bound polypyrrole on the C atom from which the amino group was removed in step (i). Significant insight into the mechanism of catalysis has been provided by studies using the substrate analogue *N*-methylporphobilinogen (Pichon *et al.*, 1992). This compound, which has a methyl substituent at the pyrrole N atom, is almost as good a substrate as porphobilinogen, although only one moiety is incorporated, forming a stable chain-termination $ES_1$ complex that is not reactive with further pyrrole moieties. This suggested that the proton attached to the pyrrole N atom is not necessary for the initial deamination of incoming porphobilinogen and that this reaction must be an E1 elimination that generates a resonance-stabilized carbocation (Pichon *et al.*, 1992). A mechanism which encapsulates these findings and key information from the X-ray structure of the *A. thaliana* enzyme, notably the role of Asp95, is shown in Fig. 9(*b*). At the optimal pH of the enzyme and when the cofactor is in a reduced state, the catalytic aspartate carboxyl would most probably be deprotonated and could therefore bind the amino side group of incoming porphobilinogen ionically. Elimination of ammonia generates the resonance-stabilized intermediate that is shown at the top right in Fig. 9(*b*) along with its carbocation canonical form. This intermediate will be highly vulnerable to nucleophilic attack by the $\alpha$-position of the terminal ring of the enzyme-bound polypyrrole. The fact that *N*-methylporphobilinogen acts as a chain-termination inhibitor suggests that whilst the proton on the pyrrole N atom of the incoming PBG is not important in the deamination in step (i) above, it is however important for the terminal pyrrole which acts as the nucleophile in step (ii) or in the translocation mechanism. In this mechanism the aspartate side chain has an important role in stabilizing a positive charge on the ring N atom of the terminal pyrrole of the cofactor, or ES complex, *via* a hydrogen bond. Replacement of the corresponding hydrogen by an *N*-methyl group would therefore disrupt the nucleophilic attack of the terminal pyrrole on the carbocation form of the incoming PBG. Altough not essential, it is likely that the ability of the enzyme to stabilize a partial positive charge at the pyrrole N atom of incoming PBG will assist catalysis (top right-hand intermediate in Fig. 9*b*) and our structural studies suggest that a

charge-assisted hydrogen bond between the pyrrole N atom and the side chain of Asp95 will fulfil this role. The formation of a covalent bond between the two substrate pyrrole moieties gives rise to a tetrahedral carbon in what was previously the terminal enzyme-bound pyrrole group (the lower right-hand pyrrole in Fig. 9*b*). In the final step, Asp95 may also catalyse the deprotonation of the tetrahedral $\alpha$-C atom of the 'nucleophilic' pyrrole, which can then revert to the trigonal state. These reactions are known to occur with the retention of stereochemical configuration at the aminomethyl C atom (Jones *et al.*, 1984).

## 4. Discussion

The first X-ray structure of porphobilinogen deaminase from a higher plant reported here establishes that the enzyme, like its animal and prokaryotic counterparts, is folded into three domains, the first two of which topologically resemble the type II periplasmic binding proteins. The third domain has a rather different fold and contains a cysteine residue (Cys254) to which the dipyrromethane cofactor is covalently attached. The cofactor is additionally bound by many ionic interactions between its carboxyl groups and a number of highly conserved basic residues, notably arginines, that are located in a cavity between the first and second domains. Both pyrrole N atoms of the cofactor form hydrogen bonds to the domain 1 residue Asp95, which is likely to be an important functional group in catalysis. In essence, domains 2 and 3 provide the bulk of the interactions with the cofactor and domain 1 provides the catalytic group. Superimposing the structure of *A. thaliana* PBGD with the human and *E. coli* enzymes establishes that the domains are oriented slightly differently in each of the structures. Although the different crystalline environment of each enzyme may be a significant factor, the differences in domain orientation may also be due to unique structural features of each enzyme, such as the highly ordered loop covering the active-site cleft that we observe in *A. thaliana* PBGD. Additionally, the domain flexibility may be an important feature of the enzyme that is required for elongation of the pyrrole chain. The current structure indicates that domains 2 and 3 move in a concerted manner relative to domain 1 and may thus act in the manner of a 'ratchet handle' during elongation. The active-site dipyrromethane is buried in a deep solvent channel that has sufficient space to accommodate a number of additional pyrromethane moieties, although attempts to model the interactions that additional PBG units could make with the enzyme suggest, importantly, that they would not bind to the enzyme quite as strongly as those of the cofactor itself. This indicates that the bound conformation of the cofactor is highly relevant to the mechanism of condensation. The proximity of Asp95 to both pyrrole N atoms suggests that it catalyses the deprotonation of the side-chain amino group of an incoming PBG to generate a resonance-stabilized carbocation intermediate. This is susceptible to nucleophilic attack by the terminal ring of the enzyme-bound polypyrrole which is potentially activated by the same aspartate. Interestingly, these reactions occur rapidly with the nonphysiological substrate hydroxyporphobilinogen (Jordan, 1991).

It seems very likely that the active-site aspartate will be involved in the final hydrolysis reaction, in which the linear tetrapyrrole product hydroxymethylbilane or preuroporphyrinogen is released from the enzyme. Hence, the catalytic apparatus would appear to catalyse the condensation of pyrrole moieties in the elongation reaction, as well as a competing reaction involving hydrolysis of the bound polypyrrole chain. Ultimately, the latter process releases preuroporphyrinogen, which is the substrate of the next enzyme in the pathway. PBGD can also hydrolyse the bound polypyrrole chain, yielding individual pyrrole moieties of hydroxyporphobilinogen (Warren & Jordan, 1988). Indeed, any enzyme-bound intermediate can react with ammonia to give PBG and with hydroxylamine or water (at elevated temperature) to give hydroxyporphobilinogen, and both of these products can act as substrates in the condensation reaction. During elongation, random movements of the protein domains may cause the bound polypyrrole to undergo something of a random walk, perhaps shuffling backwards and forwards in the active-site cleft in a stepwise manner. This would allow each of the pyrrole moieties to spend some time sampling the tightest binding pockets of the enzyme in contact with the catalytic aspartate. These parts of the polypyrrole will then be susceptible to the hydrolysis reaction, although this must proceed at a slower rate than condensation. A stepwise random walk of the polypyrrole will eventually free up the site at which incoming porphobilinogen can bind sufficiently tightly to undergo the exergonic condensation reaction. Thus, it is easy to see how in the presence of a high substrate concentration the reaction would be driven in the net direction of elongation rather than hydrolysis of bound intermediates.

In these processes, the movement of the negatively charged polypyrrole will be facilitated by the large number of basic residues which form an extensive buried electropositive surface within the protein moiety. Similar considerations also suggest that the number of pyrrole moieties that can be assembled into a single chain would be restricted by any physical constraints limiting the domain movements within the enzyme. This is consistent with the known ability of the enzyme to assemble at most six pyrroles in the $ES_4$ state. Having assembled a linear chain of six pyrroles, it may be that the final specific cleavage of the tetrapyrrole preuroporphyrinogen from $ES_4$ would be triggered by a conformational change in the enzyme resulting from the steric or electrostatic effects of accommodating a polypyrrole chain of this length in the active-site cavity. The intriguing details of how this final step is orchestrated remain to be established. The near-atomic resolution structure that we report here provides a sound basis for the design of mutants to further explore all of these effects and to probe the role of the uniquely ordered active-site flap in this enzyme.

We gratefully acknowledge the BBSRC, UK for a PhD case studentship to AR (with EJTC, Syngenta) and the School

of Biological Sciences, University of Southampton for a studentship award to RJH. We acknowledge the ESRF (Grenoble, France) for beam time and travel support. We thank Dr Matthew Terry (School of Biological Sciences, University of Southampton) for many helpful discussions on chlorophyll bioynthesis.

## References


Aplin, R. T., Baldwin, J. E., Pichon, C., Roessner, C. A., Scott, A. I., Schofield, C. J., Stolowich, N. J. & Warren, M. J. (1991). *Bioorg. Med. Chem. Lett.* **1**, 503–506.

Awan, S. J., Siligardi, G., Shoolingin-Jordan, P. M. & Warren, M. J. (1997). *Biochemistry*, **36**, 9273–9282.

Baker, N. A., Sept, D., Joseph, S., Holst, M. J. & McCammon, J. A. (2001). *Proc. Natl Acad. Sci. USA*, **98**, 10037–10041.

Chen, D.-M., Fan, S.-H., Liu, X.-L., Chen, G.-Y., Deng, Z.-Y. & Guo, A.-G. (2006). *Chin. J. Biochem. Mol. Biol.* **22**, 973–978.

Chen, V. B., Arendall, W. B., Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., Murray, L. W., Richardson, J. S. & Richardson, D. C. (2010). *Acta Cryst.* D**66**, 12–21.

Chretien, S., Dubart, A., Beaupain, D., Raich, N., Grandchamp, B., Rosa, J., Goossens, M. & Romeo, P. H. (1988). *Proc. Natl Acad. Sci. USA*, **85**, 6–10.

Cornah, J. E., Terry, M. J. & Smith, A. G. (2003). *Trends Plant Sci.* **8**, 224–230.

Emsley, P. & Cowtan, K. (2004). *Acta Cryst.* D**60**, 2126–2132.

Evans, P. (2006). *Acta Cryst.* D**62**, 72–82.

Fedorov, A., Merican, A. F. & Gilbert, W. (2002). *Proc. Natl Acad. Sci. USA*, **99**, 16128–16133.

Gill, R., Kolstoe, S. E., Mohammed, F., Al d-Bass, A., Mosely, J. E., Sarwar, M., Cooper, J. B., Wood, S. P. & Shoolingin-Jordan, P. M. (2009). *Biochem. J.* **420**, 17–25.

Go, M. (1981). *Nature (London)*, **291**, 90–92.

Goodstein, D. M., Shu, S., Howson, R., Neupane, R., Hayes, R. D., Fazo, J., Mitros, T., Dirks, W., Hellsten, U., Putnam, N. & Rokhsar, D. S. (2012). *Nucleic Acids Res.* **40**, D1178–D1186.

Hädener, A., Matzinger, P. K., Battersby, A. R., McSweeney, S., Thompson, A. W., Hammersley, A. P., Harrop, S. J., Cassetta, A., Deacon, A., Hunter, W. N., Nieh, Y. P., Raftery, J., Hunter, N. & Helliwell, J. R. (1999). *Acta Cryst.* D**55**, 631–643.

He, Z.-H., Li, J., Sundqvist, C. & Timko, M. P. (1994). *Plant Physiol.* **106**, 537–546.

Huang, M., Slewinski, T. L., Baker, R. F., Janick-Buckner, D., Buckner, B., Johal, G. S. & Braun, D. M. (2009). *Mol. Plant*, **2**, 773–789.

Hukmani, P. & Tripathy, B. C. (1994). *Plant Physiol.* **105**, 1295–1300.

Jones, R. M. & Jordan, P. M. (1994). *Biochem. J.* **299**, 895–902.

Jones, C., Jordan, P. M. & Akhar, M. (1984). *J. Chem. Soc. Perkin Trans. 1*, pp. 2625–2633.

Jordan, P. M. (1991). *New Compr. Biochem.* **19**, 1–66.

Jordan, P. M. & Warren, M. J. (1987). *FEBS Lett.* **225**, 87–92.

Jordan, P. M. & Woodcock, S. C. (1991). *Biochem. J.* **280**, 445–449.

Krissinel, E. & Henrick, K. (2007). *J. Mol. Biol.* **372**, 774–797.

Lander, M., Pitt, A. R., Alefounder, P. R., Bardy, D., Abell, C. & Battersby, A. R. (1991). *Biochem. J.* **275**, 447–452.

Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. (1993). *J. Appl. Cryst.* **26**, 283–291.

Leslie, A. G. W. (2006). *Acta Cryst.* D**62**, 48–57.

Lim, S. H., Witty, M., Wallace-Cook, A. D., Ilag, L. I. & Smith, A. G. (1994). *Plant Mol. Biol.* **26**, 863–872.

Louie, G. V. (1993). *Curr. Opin. Struct. Biol.* **3**, 401–408.

Louie, G. V., Brownlie, P. D., Lambert, R., Cooper, J. B., Blundell, T. L., Wood, S. P., Malashkevich, V. N., Hädener, A., Warren, M. J. & Shoolingin-Jordan, P. M. (1996). *Proteins*, **25**, 48–78.

Louie, G. V., Brownlie, P. D., Lambert, R., Cooper, J. B., Blundell, T. L., Wood, S. P., Warren, M. J., Woodcock, S. C. & Jordan, P. M. (1992). *Nature (London)*, **359**, 33–39.

Lovell, S. C., Davis, I. W., Arendall, W. B., de Bakker, P. I. W., Word, J. M., Prisant, M. G., Richardson, J. S. & Richardson, D. C. (2003). *Proteins*, **50**, 437–450.

Mobashery, S. & Kotra, L. P. (2002). In *Encyclopedia of Life Sciences*. Chichester: John Wiley & Sons. doi:10.1038/npg.els.0000618.

Murshudov, G. N., Skubák, P., Lebedev, A. A., Pannu, N. S., Steiner, R. A., Nicholls, R. A., Winn, M. D., Long, F. & Vagin, A. A. (2011). *Acta Cryst.* D**67**, 355–367.

Pichon, C., Clemens, K. R., Jacobson, A. R. & Scott, A. I. (1992). *Tetrahedron*, **48**, 4687–4712.

Read, R. J. (1986). *Acta Cryst.* A**42**, 140–149.

Roberts, A., Gill, R., Hussey, R. J., Mikolajek, H., Erskine, P. T., Cooper, J. B., Wood, S. P., Chrystal, E. J. T. & Shoolingin-Jordan, P. M. (2012). *Acta Cryst.* F**68**, 1491–1493.

Sanner, M. F., Olson, A. J. & Spehner, J. C. (1996). *Biopolymers*, **38**, 305–320.

Schüttelkopf, A. W. & van Aalten, D. M. F. (2004). *Acta Cryst.* D**60**, 1355–1363.

Scott, A. I., Clemens, K. R., Stolowich, N. J., Santander, P. J., Gonzalez, M. D. & Roessner, C. A. (1989). *FEBS Lett.* **242**, 319–324.

Scott, A. I., Roessner, C. A., Stolowich, N. J., Karuso, P., Williams, H. J., Grant, S. K., Gonzalez, M. D. & Hoshino, T. (1988). *Biochemistry*, **27**, 7984–7990.

Sheldrick, G. M. (2008). *Acta Cryst.* A**64**, 112–122.

Shoolingin-Jordan, P. M., Warren, M. J. & Awan, S. J. (1996). *Biochem. J.* **316**, 373–376.

Shoolingin-Jordan, P. M., Warren, M. J. & Awan, S. J. (1997). *Methods Enzymol.* **281**, 317–327.

Smith, A. G. (1988). *Biochem. J.* **249**, 423–428.

Song, G., Li, Y., Cheng, C., Zhao, Y., Gao, A., Zhang, R., Joachimiak, A., Shaw, N. & Liu, Z.-J. (2009). *FASEB J.* **23**, 396–404.

Spano, A. J. & Timko, M. P. (1991). *Biochim. Biophys. Acta*, **1076**, 29–36.

Vagin, A. & Teplyakov, A. (2010). *Acta Cryst.* D**66**, 22–25.

Vaguine, A. A., Richelle, J. & Wodak, S. J. (1999). *Acta Cryst.* D**55**, 191–205.

Warren, M. J., Gul, S., Aplin, R. T., Scott, A. I., Roessner, C. A., O'Grady, P. & Shoolingin-Jordan, P. M. (1995). *Biochemistry*, **34**, 11288–11295.

Warren, M. J. & Jordan, P. M. (1988). *Biochemistry*, **27**, 9020–9030.

Warren, M. J. & Smith, A. G. (2009). *Tetrapyrroles: Birth, Life and Death.* Austin: Landes Biosciences.

Winn, M. D. *et al.* (2011). *Acta Cryst.* D**67**, 235–242.

Witty, M., Jones, R. M., Robb, M. S., Jordan, P. M. & Smith, A. G. (1996). *Planta*, **199**, 557–564.

Witty, M., Wallace-Cook, A. D., Albrecht, H., Spano, A. J., Michel, H., Shabanowitz, J., Hunt, D. F., Timko, M. P. & Smith, A. G. (1993). *Plant Physiol.* **103**, 139–147.

Wood, S., Lambert, R. & Jordan, P. M. (1995). *Mol. Med. Today*, **1**, 232–239.

Woodcock, S. C. & Jordan, P. M. (1994). *Biochemistry*, **33**, 2688–2695.

Yoo, H.-W., Warner, C. A., Chen, C.-H. & Desnick, R. J. (1993). *Genomics*, **15**, 21–29.