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

^aSchool of Biological Sciences, University of Southampton, Southampton SO16 7PX, England, ^bLaboratory 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 ^cSyngenta, Jealott's Hill International Research Centre, Bracknell, Berkshire RG42 6EY, England

† Deceased.

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

Received 13 August 2012

Accepted 8 October 2012

Crystallization and preliminary X-ray characterization of the tetrapyrrole-biosynthetic enzyme porphobilinogen deaminase from *Arabidopsis thaliana*

The enzyme porphobilinogen deaminase (PBGD; hydroxymethylbilane synthase; EC 2.5.1.61) catalyses a key early step of the haem-biosynthesis pathway in which four molecules of the monopyrrole porphobilinogen are condensed to form a linear tetrapyrrole. The enzyme possesses a dipyrromethane cofactor which is covalently linked by a thioether bridge to an invariant cysteine residue. Since PBGD catalyses a reaction which is common to the biosynthesis of both haem and chlorophyll, structural studies of a plant PBGD enzyme offer great potential for the discovery of novel herbicides. Until recently, structural data have only been available for the *Escherichia coli* and human forms of the enzyme. Expression in *E. coli* of a codon-optimized gene for *Arabidopsis thaliana* PBGD has permitted for the first time the crystallization and preliminary X-ray analysis of the enzyme from a plant species at high resolution.

1. Introduction

The enzyme porphobilinogen deaminase (PBGD), which is also known as hydroxymethylbilane synthase (EC 2.5.1.61), catalyses the third step of the haem-biosynthesis pathway in animals (the fifth step in plants), in which four molecules of the monopyrrole porphobilinogen are condensed to form a linear tetrapyrrole, preuroporphyrinogen or hydroxymethylbilane (Fig. 1). 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. PBGD is a monomeric protein with a molecular weight in the range 34–44 kDa depending on the species (Jordan, 1991). 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 (Fig. 1) is the first to bind to the enzyme, followed by rings B, C and finally D. 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*; 254 in *Arabidopsis thaliana* PBGD; Jordan & Warren, 1987). This cofactor can be assembled from two molecules of the normal substrate porphobilinogen or derived from cleavage of the product preuroporphyrinogen (Awan *et al.*, 1997). The cofactor acts as a primer to which four porphobilinogen molecules are attached sequentially prior to cleavage of the link between the cofactor and the first substrate molecule on completion of the reaction. Thus, the cofactor remains covalently attached to the enzyme when the product of the reaction is released.

X-ray structures of the *E. coli* enzyme have been solved at resolutions better than 2.0 Å (PDB entries 1pda, 1gtk, 1ah5, 1ypn and 2ypn; Louie *et al.*, 1992, 1996; Hädener *et al.*, 1999) and a number of structures of the human enzyme are also available (PDB entries 3eq1 and 3ecr; Gill *et al.*, 2009; Song *et al.*, 2009). The polypeptide 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 a number of periplasmic binding proteins. The dipyrromethane cofactor is attached to a loop on domain 3 and is positioned at the

mouth of a deep active-site cleft formed between domains 1 and 2. These structural studies have shown that many of the mutations occurring in AIP victims affect crucial arginine residues which bind the side-chain carboxylates of the cofactor and/or substrate (Wood *et al.*, 1995).

The tetrapyrrole-biosynthetic pathway is considerably more elaborate in plants than it is in animals to allow production of the crucial photosynthetic pigment chlorophyll. However, the PBGD enzyme catalyses one of the early steps which is common to the biosynthesis of both haem and chlorophyll. Structural studies of a plant PBGD enzyme have great potential for the discovery of novel inhibitory compounds that could act as herbicides. Indeed, a number of commonly used herbicides, such as acifluorfen, already target this particular pathway. Here, we report the expression of a codon-optimized gene for *A. thaliana* PBGD and the subsequent purification and crystallization of the enzyme in a form that diffracts synchrotron radiation to very high resolution.

2. Gene synthesis

Native PBGD was first purified from *A. thaliana* by Jones & Jordan (1994), who obtained enough functional enzyme for activity studies but not enough for crystallization. For the current investigation, two versions of the *A. thaliana* PBGD gene were designed for over-expression in *E. coli*. The first version (gene 1) had the same nucleotide sequence as the wild-type protein but lacked the N-terminal chloroplast-import sequence. The second version (gene 2) also lacked the chloroplast-import sequence but was codon-optimized for *E. coli* expression using the online *OptimumGene* server at <http://www.genscript.com>. The two genes were synthesized and cloned into

the pUC57 plasmid (GenScript Inc). For expression studies, both genes were excised from the cloning vector by digestion with the restriction endonucleases *NdeI* and *BamHI* and then ligated into the pT7-7 vector by standard molecular-biology methods (Roberts, 2008). The resulting plasmids were transformed into *E. coli* BL21(DE3) and expression of the cloned gene was induced using IPTG, which was added to 1 l cultures to give a final concentration of 0.1 mM. The cultures were grown initially at 310 K; following induction at the mid-log phase, they were grown for a further 6 h at a lower temperature of 303 K.

3. Purification

The harvested cells were resuspended in 20 mM Tris-HCl buffer pH 8.0 containing 5 mM DTT and 100 μ M phenylmethylsulfonyl fluoride (PMSF) prior to sonication on ice and ultracentrifugation at 40 000g for 40 min. The resulting supernatant was heat-treated by incubation at 333 K for 10 min with stirring. Impurities that were precipitated by this step were removed by further ultracentrifugation. The supernatant was then applied onto a DEAE-Sephacel column pre-equilibrated with 20 mM Tris-HCl buffer pH 8.0 (containing 5 mM DTT and 100 μ M PMSF) and the bound protein was eluted with a linear gradient of KCl (0–300 mM). Fractions with PBGD activity were pooled and concentrated by ultrafiltration before being applied onto a Superdex G75 (16/60) gel-filtration column that was equilibrated with the same buffer. Further purification involved the use of a Biogel hydroxyapatite column and the active fractions from this step were concentrated and desalted again by ultrafiltration. The codon-optimized gene was found to give significantly higher yields of purified protein and so was used in further studies. The M_r of the purified enzyme was determined by electrospray mass-spectrometry to be 34 930.

4. Crystallization

Crystals were grown at room temperature using the hanging-drop method with a stock protein solution of 5 mg ml⁻¹ PBGD in 20 mM Tris-HCl buffer pH 8.0, 5 mM DTT. All droplets were hand-pipetted and consisted of 1 μ l protein solution mixed with 1 μ l well solution on a siliconized glass coverslip that was subsequently suspended over each corresponding well of a greased 24-well plate. JBScreen Classic 3 condition C6 [25% (w/v) PEG 4000, 100 mM sodium citrate pH 5.6, 200 mM ammonium sulfate] gave yellow irregular-looking crystals, as

Figure 1

The reaction catalysed by porphobilinogen deaminase. Four molecules of the pyrrole porphobilinogen are condensed to form the linear tetrapyrrole preuroporphyrinogen (hydroxymethylbilane). The acetic and propionic acid side chains of each pyrrole are abbreviated as A and P, respectively, and the four rings of the tetrapyrrole product are indicated in italics as A, B, C and D.

Figure 2

The dipyrromethane cofactor of porphobilinogen deaminase is covalently attached to the enzyme by a thioether bond with a cysteine residue. Four substrate pyrroles are added linearly to the cofactor and, finally, hydrolysis of the linkage between the substrate and the cofactor releases the tetrapyrrole product hydroxymethylbilane.

Figure 3

Crystals of *A. thaliana* porphobilinogen deaminase obtained by the hanging-drop method. The marked yellow colouration probably arises from oxidation of the dipyrromethane cofactor. In spite of their irregular appearance, the best crystals diffracted synchrotron radiation to around 1.5 Å resolution. The largest crystal shown here is approximately 0.3 mm in its longest dimension.

Figure 4

A diffraction image of *A. thaliana* PBGD obtained on beamline ID29 at the ESRF using an ADSC Q315 detector. In spite of slight splitting of some spots and the relatively high background, the diffraction data obtained from this crystal were processed to a resolution of 1.45 Å. The inset figure shows a sample of the diffraction spots at the resolution limit of the data set (the dashed line indicates the 1.5 Å resolution shell).

shown in Fig. 3, after 2–3 weeks of growth. Since the dipyrromethane cofactor of PBGD is light-sensitive, the crystals were left to grow in the dark. The largest crystals obtained were mounted by transferring them into 10 µl well solution using a loop and stirring in four 1 µl droplets of glycerol stepwise to give a final glycerol concentration of approximately 30% (v/v). The crystals were then flash-cooled in loops by dipping them in liquid ethane in a liquid-nitrogen bath prior to storage in cryovials.

5. Preliminary X-ray analysis

X-ray data collection at station ID29 at the ESRF (Grenoble, France) revealed that, in spite of their unusual appearance, the *A. thaliana* PBGD crystals were of very high diffraction quality (Fig. 4) and hence no further optimization of the crystallization conditions was required. Using 1° oscillations, 190° of data were collected from a single crystal maintained at a temperature of 100 K using an ADSC Q315 CCD detector with an exposure time of 10 s per image (three passes) and a crystal-to-detector distance of 194.1 mm. A low-resolution pass was performed using 3 s exposures (one pass) and a crystal-to-detector distance of 366.6 mm to re-collect the spots that were overloaded in the initial high-resolution run. For both data collections the incident beam had a wavelength of 0.979 Å and was attenuated by 90% to reduce radiation damage to the sample. Data processing with *MOSFLM* (Leslie, 2006), *SCALA* (Evans, 2006) and other programs from the *CCP4* suite (Winn *et al.*, 2011) revealed that the crystals were monoclinic *C2*, with unit-cell parameters $a = 141.6$, $b = 37.3$, $c = 55.1$ Å, $\beta = 105.0^\circ$. Inspection of the correlation coefficient for half-data-set intensities, as recommended by Karplus & Diederichs (2012) and Evans (2012), suggested that the diffraction data extend to a resolution of $d_{\min} = 1.45$ Å with an overall $R_{\text{merge}} = 9.8\%$ and $R_{\text{meas}} = 11.4\%$ (for details see Table 1). Using the method of Matthews (1968), as implemented by Kantardjieff & Rupp (2003), it was estimated that the crystals have a single PBGD monomer per crystallographic asymmetric unit, with a solvent content of 39%. Structure analysis using the molecular-replacement program

Table 1

Data-collection and processing statistics for *A. thaliana* porphobilinogen deaminase.

Values in parentheses are for the outer resolution shell.

Beamline	ID29, ESRF
Wavelength (Å)	0.979
Space group	<i>C2</i>
Unit-cell parameters	
a (Å)	141.6
b (Å)	37.3
c (Å)	55.1
β (°)	105.0
Mosaic spread (°)	0.6
Resolution (Å)	32.89–1.45 (1.53–1.45)
$R_{\text{merge}}^\dagger$ (%)	9.8 (64.1)
$R_{\text{meas}}^\ddagger$ (%)	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
No. of molecules per asymmetric unit	1

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

MOLREP (Vagin & Teplyakov, 2010) with *E. coli* PBGD (42% identity; PDB entry 1pda; Louie *et al.*, 1992) as the search model was successful and refinement of the *A. thaliana* PBGD structure is currently in progress using this high-resolution data set.

We gratefully acknowledge the UK BBSRC for a studentship to AR 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 Neville Wright (University of Southampton) for the mass spectrometry.

References

- Awan, S. J., Siligardi, G., Shoolingin-Jordan, P. M. & Warren, M. J. (1997). *Biochemistry*, **36**, 9273–9282.
- Evans, P. (2006). *Acta Cryst.* **D62**, 72–82.
- Evans, P. (2012). *Science*, **336**, 986–987.
- 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.
- 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.* **D55**, 631–643.
- Jones, R. M. & Jordan, P. M. (1994). *Biochem. J.* **299**, 895–902.
- Jordan, P. M. (1991). *Biosynthesis of Tetrapyrroles*, edited by P. M. Jordan, pp. 1–66. Amsterdam: Elsevier.
- Jordan, P. M. & Warren, M. J. (1987). *FEBS Lett.* **225**, 87–92.
- Kantardjieff, K. A. & Rupp, B. (2003). *Protein Sci.* **12**, 1865–1871.
- Karplus, P. A. & Diederichs, K. (2012). *Science*, **336**, 1030–1033.
- Leslie, A. G. W. (2006). *Acta Cryst.* **D62**, 48–57.
- 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.
- Matthews, B. W. (1968). *J. Mol. Biol.* **33**, 491–497.
- Roberts, A. (2008). PhD Thesis. University of Southampton, England.
- 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.
- Vagin, A. & Teplyakov, A. (2010). *Acta Cryst.* **D66**, 22–25.
- Winn, M. D. *et al.* (2011). *Acta Cryst.* **D67**, 235–242.
- Wood, S., Lambert, R. & Jordan, P. M. (1995). *Mol. Med. Today*, **1**, 232–239.