Structure of human porphobilinogen deaminase at 2.8 Å: the molecular basis of acute intermittent porphyria

Raj GILL*, Simon E. KOLSTOE*, Fiyyaz MOHAMMED†, Abeer AL-D-BASS†‡, Julie E. MOSELY†‡, Mohammed SARWAR†, Jonathan B. COOPER*, Stephen P. WOOD* and Peter M. SHOOLINGIN-JORDAN†‡

*Centre for Amyloidosis and Acute Phase Proteins, Department of Medicine, UCL Medical School, Rowland Hill Street, London NW3 2PF, U.K., and †Biomolecular Sciences Group, School of Biological Sciences, University of Southampton, Bassett Crescent East, Southampton SO16 7PX, U.K.

INTRODUCTION

PBGD (porphobilinogen deaminase; EC 4.3.4.8, also referred to as hydroxymethylbilane synthase, uroporphyrinogen synthase or uroporphyrinogen I synthase) is the third enzyme in the haem biosynthesis pathway in mammals [1]. The reaction catalysed by PBGD involves the formation of porphobilinogen, a linear tetrapyrole (bilane), by the extension of an enzyme-bound DPM (dipyrromethane) cofactor that acts as a reaction primer [2,3]. This is achieved by the sequential binding, deamination and condensation of four molecules of the substrate(S) PBG (porphobilinogen) through covalently bound enzyme (E) intermediates ES, ES2, ES3 and ES4 linked to the DPM cofactor. The resulting enzyme-bound hexapyrrole (ES6) is then hydrolysed, releasing the unstable tetrapyrole product, porphobilinogen (also called 1-hydroxymethylbilane) with regeneration of the enzyme with the DPM cofactor still bound at the active site.

PBGD possesses several novel features. First, the initial protein translation product is an apoenzyme that has the ability to form its own DPM cofactor using two of the pyrrole units of porphobilinogen [4]. Secondly, each of the four substrate condensation steps occurs at a single catalytic site [5] and the enzyme is therefore able to translocate the growing polypyrrole chain to vacate the substrate-binding site for the next incoming substrate. Thirdly, the enzyme can ‘count’ precisely to four and terminate the polymerization reaction by hydrolysis when the hexapyrrole chain (ES6) has been assembled [3].

Mutations in the human PBGD (porphobilinogen deaminase) gene cause the inherited defect AIP (acute intermittent porphyria). In the present study we report the structure of the human uPBGD (ubiquitous PBGD) mutant, R167Q, that has been determined by X-ray crystallography and refined to 2.8 Å (1 Å = 0.1 nm) resolution (Rfree = 0.26, Rwork = 0.29). The protein crystallized in space group F222 with two molecules in the asymmetric unit (a = 81.0 Å, b = 104.4 Å and c = 109.7 Å). Phases were obtained by molecular replacement using the Escherichia coli PBGD structure as a search model. The human enzyme is composed of three domains each of approx. 110 amino acids and possesses a dipyrromethane cofactor at the active site, which is located between domains 1 and 2. An ordered sulfate ion is hydrogen-bonded to Arg26 and Ser25 at the proposed substrate-binding site in domain 1. An insert of 29 amino acid residues, present only in mammalian PBGD enzymes, has been modelled into domain 3 where it extends helix αS and forms a β-hairpin structure that contributes to a continuous hydrogen-bonding network spanning domains 1 and 3. The structural and functional implications of the R167Q mutation and other mutations that result in AIP are discussed.

Key words: acute intermittent porphyria, dipyrromethane, haem, human porphobilinogen deaminase, X-ray structure.

Comparison of the primary amino acid sequences of PBGDs from a wide range of organisms demonstrates a high degree of conservation (see Supplementary Figure S1 at http://www.BiochemJ.org/bj/420/bj420017add.htm). For example, there is 60% similarity and 43% identity between the deaminases from Escherichia coli and humans [6] (see Figure 1). PBGD from E. coli was the first haem biosynthesis enzyme structure to be solved, initially at 1.90 Å (1 Å = 0.1 nm) resolution [5] and subsequently refined to 1.76 Å [8]. The E. coli enzyme is composed of three α/β domains of approximately equal size, linked together by flexible hinge regions. Domains 1 (N-terminal) and 2 are linked by two hinge segments and each possess a similar fold based on a five-stranded, mixed β-sheet. There are relatively few direct interactions between the two domains, which form an extensive active site cleft at their interface. Both domains 1 and 2 have the same overall topology as found in the transferrins and a number of bacterial periplasmic binding proteins, which are known to adopt ‘open’ and ‘closed’ states in response to ligand binding [6]. Domain 3 (C-terminal) possesses an open-faced anti-parallel β-sheet of three strands, one face of which is covered by three α-helices. A loop from domain 3 carries an invariant cysteine that forms the covalent attachment site for the DPM cofactor. Domain 3 interacts equally with both domains 1 and 2, the packing being mediated primarily through polar interactions of loop regions.

Mutations in human PBGD cause AIP (acute intermittent porphyria), an inherited autosomal dominant disorder characterized by colicky abdominal pain and both peripheral and central
Figure 1 Structural alignment of the amino acid sequences of human uPBGD and E. coli PBGD

Identical residues are boxed and the secondary structural elements are coloured according to domain. N- and C-terminal residues, active site loop residues and the human domain 3 insert residues not visible in the electron density map are highlighted in grey. AIP-associated single-residue mutations are displayed beneath the human uPBGD sequence. Residues conserved across PBGDs from all species are indicated by $/$, conserved substitutions are indicated by $\times$ and semi-conservative substitutions are indicated by $\circ$. The sequence alignment was depicted with ALSCRIPT [7]. Amino acids are colour-coded such that red indicates acidic residues, blue indicates basic residues, green indicates neutral residues, pink indicates hydrophobic residues, yellow indicates cysteine residues and white indicates the conformationally important glycine, alanine and proline residues.

neurological symptoms accompanied by elevated levels of the haem precursors ALA (5-aminolevulinic acid) and PBG in the urine. AIP is speculated to have been the cause of the madness of King George III [9]. In humans, alternative splicing of the PBGD mRNA gives rise to two monomeric isoenzymes, uPBGD (ubiquitous PBGD) [10] that is expressed in all cells (361 residues; $M_r$ 39620) and ePBGD (erythroid PBGD) [11] that lacks the first 17 amino acids (344 residues; $M_r$ 38120). To date, over 250 mutations resulting in human PBGD deficiency are listed in The Human Gene Mutation Database.
(http://www.hgmd.cf.ac.uk/ac/gene.php?gene=HMBS) and references therein. Over one-third of these mutations are nucleotide substitutions resulting either in single amino acid substitutions or premature chain termination. The likely impact of mutations on the structure and activity of human PBGD was originally assessed from a homology model based on the E. coli enzyme structure [12,13]; however, in regions of low homology and particularly in the vicinity of the 29-residue insert in the human enzyme sequence, a confident prediction of the functional consequences of mutations was not possible.

In this paper, we present the X-ray structure of the R167Q mutant of human uPBGD. This was one of the first PBGD mutations to be identified [14] and is particularly interesting because the enzyme is weakly active (8.5%), has a reduced pH optimum and accumulates long-lived ES, ES₂, ES₃ and ES₄ enzyme–intermediate complexes that are potentially able to provide insights into the mechanism of pyrrole polymerization.

MATERIALS AND METHODS

Protein expression and purification

Human uPBGD was expressed in E. coli strain BL21(DE3) pLysS after transformation with the pUHD2 plasmid, a derivative of pT7-7 [15]. The uPBGD insert in pUHD2 was constructed from a cDNA kindly provided by Professor Bernard Grandchamp (Hôpital Bichat, Paris, France). The R167Q mutant of human PBGD was constructed by site-directed mutagenesis and expressed and purified using a purification procedure similar to that used for the native enzyme as described below [16].

Six 800 ml flasks of 2TY containing 100 μg/ml ampicillin were inoculated with overnight cultures of transformed E. coli and grown at 37 °C with shaking to a Dₗ₀₀ (attenuance) of 0.6. IPTG (isopropyl β-D-thiogalactoside) was then added to 1 mM and bacterial growth was continued for a further 3 h. All subsequent steps were performed at 4 °C except where noted. The cells were pelleted by centrifugation at 4500 g, washed with 0.9 M NaCl and re-centrifuged. The bacterial pellet was then resuspended in 80 ml of 20 mM Tris/HCl buffer (pH 8.2) containing 5 mM DTT (dithiothreitol) and 100 μM PMSF and sonicated for 20 cycles of 30 s bursts using an MSE Soniprep 150 sonicator, interspersed with periods of cooling for 90 s. The sonicate was placed in a flask filled with N₂ gas, heated rapidly to 60 °C with periods of cooling for 90 s. The sonicate was then applied to a Pharmacia PD10 column. The specific activities of the purified recombinant wild-type and R167Q uPBGD proteins were determined as described previously [17]. Protein concentrations were determined from the calculated molar absorption coefficient (ε) [18] (ε₂₈₀ 0.40 for a 0.1 % solution).

Protein crystallization

Human R167Q uPBGD was crystallized using the vapour diffusion method [19]. Hanging drops were prepared by mixing equal volumes [4 μl of protein solution (20 mg/ml) in 20 mM Tris/HCl buffer (pH 8.2) containing 5 mM DTT] and reservoir solution [0.6 M ammonium sulfate, 1.2 M lithium sulfate, 5% (v/v) ethylene glycol, 50 mM sodium citrate (pH 5.6) and 50 mM DTT]. The hanging drops were equilibrated against 1 ml of reservoir solution and placed at room temperature in the dark.

X-ray data collection and processing

Crystals of human R167Q uPBGD were transferred from mother liquor to cryoprotectant [30% (v/v) glycerol] and flash-cooled in liquid ethane. Diffraction data were collected at the ESRF (European Synchrotron Radiation Facility, Grenoble, France) at station BM14 linked to a MAR CCD (charge-coupled device) detector at 100 K using an Oxford Cryosystems Cryostream cooler. A total of 180 0.5° oscillations were collected with an exposure time of 120 s per frame at a crystal-to-detector distance of 145 mm. Intensity data were processed with MOSFLM [20] and sorted, scaled and merged using the CCP4 (Collaborative Computational Project 4) suite of programs [21].

Structure determination and refinement

The structure of human R167Q uPBGD was solved by molecular replacement using the program MOLREP [22]. The search model consisted of the E. coli PBGD structure (PDB accession code 1PDA) with the DPM cofactor and water molecules omitted.

The model was refined using CNS [23] and PHENIX [24] with 21934 unique reflections in the 46.8–2.8 Å resolution range. The model was initially subjected to a round of rigid body refinement and torsion angle simulated annealing (slow cool from 5000 to 300 K in 25 K decrements) using CNS. This was followed by a number of rounds of model building in COOT [25], grouped B-factor and restrained co-ordinate refinement in the program phenix.refine [26]. Two-fold NCS (non-crystallographic symmetry) restraints, bulk solvent and anisotropy correction were used throughout the restrained refinement process, which was monitored using the R_factor and R_free statistics [27].

RESULTS

Specific activity

The specific activity at pH 8.0 of the purified R167Q mutant deaminase was 0.12 μmol of porphyrin formed/h per mg compared with the wild-type deaminase specific activity of 1.4 μmol of porphyrin formed/h per mg.

X-ray structure analysis

Thin colourless plate-like crystals grew after several weeks. The best crystal diffracted X-rays to a resolution of 2.8 Å and the
space group was determined to be orthorhombic (space group \(P2_12_12\)) with cell dimensions of \(a = 81.0 \, \text{Å}, \quad b = 104.4 \, \text{Å} \) and \(c = 109.7 \, \text{Å}\). The calculated crystal packing parameter \(V_M\), assuming two uPBGD molecules (84 kDa) per asymmetric unit, is 2.63 \(\text{Å}^3\cdot\text{Da}^{-1}\), which corresponds to a solvent content of 53% [28]. Although the data extended beyond 2.8 \(\text{Å}\) resolution, the intensity of the diffraction pattern for this crystal form was anisotropic (being particularly weak in the direction of the \(k\) reciprocal axis); nevertheless the completeness of the data with \(I/\sigma > 3\) was good.

The rotation function yielded two significant peaks at 7.5\(\sigma\) and 6.7\(\sigma\), corresponding to the two molecules in the crystal asymmetric unit. The top rotation peak yielded a translation function peak of 13.2\(\sigma\) with the next peak being at 6.9\(\sigma\). This solution was fixed and a second translation search was performed to determine the relative position of the second molecule, yielding a top peak of 6.0\(\sigma\), with a correlation coefficient of 0.43 and an \(R_{\text{factor}}\) of 0.52. After the positioning of both molecules in the target asymmetric unit, the crystal packing of this solution was viewed using MOLPACK [29] and displayed sensible crystal contacts. After rigid body refinement and simulated annealing, the \(R_{\text{factor}}\) decreased to 0.36 (\(R_{\text{free}} = 0.40\)). Initial examination of the averaged electron density maps revealed prominent \(F_o-F_c\) density for the DPM cofactor and for most of the 29-residue insert in domain 3, confirming the validity of the molecular replacement solution. The model was improved by manual rebuilding, during which the uPBGD residues were fitted to the \(2F_o-F_c\) density. Twenty-four residues of the 29-residue insert in domain 3 and the DPM cofactor were built in during the later stages of refinement.

Electron density was only visible from Val20 onwards at the N-terminus and residues 56–76 of the active-site loop were also disordered. This loop becomes disordered in some deaminase inactivated. Indeed, the rigid oxidized cofactor generally leads to a considerable improvement in crystal order. Nevertheless, the electron density here was of sufficient quality to build most of the uPBGD residues which the uPBGD residues were fitted to the \(2F_o-F_c\) density.

### Overall structure

The overall dimensions of the human uPBGD molecule are 57 \(\times\) 43 \(\times\) 32 \(\text{Å}\). The topologies of domain 1 (residues 20–115 and 214–238) and domain 2 (residues 116–213) are broadly similar with both consisting of a doubly wound, five-stranded \(\beta\)-sheet with three \(\beta\)-sheets. In contrast, domain 3 (residues 239–356) comprises an open faced three-dimensional structure. Table 1 contains the data-collection and refinement statistics for human uPBGD.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unit cell and symmetry</td>
<td></td>
</tr>
<tr>
<td>Radiation source</td>
<td>ESRF BM14</td>
</tr>
<tr>
<td>Radiation wavelength (Å)</td>
<td>1.000</td>
</tr>
<tr>
<td>Space group</td>
<td>(P2_12_12)</td>
</tr>
<tr>
<td>Unit-cell parameters</td>
<td></td>
</tr>
<tr>
<td>(a) (Å)</td>
<td>81.03</td>
</tr>
<tr>
<td>(b) (Å)</td>
<td>104.44</td>
</tr>
<tr>
<td>(c) (Å)</td>
<td>109.73</td>
</tr>
<tr>
<td>(\alpha) (°)</td>
<td>90.00</td>
</tr>
<tr>
<td>(\beta) (°)</td>
<td>90.00</td>
</tr>
<tr>
<td>(\gamma) (°)</td>
<td>90.00</td>
</tr>
<tr>
<td>Estimated solvent content (%)</td>
<td>53</td>
</tr>
<tr>
<td>Subunits in asymmetric unit</td>
<td>2</td>
</tr>
<tr>
<td>Dataset statistics</td>
<td></td>
</tr>
<tr>
<td>Resolution range (Å)</td>
<td>46.8–2.8 (\sigma)</td>
</tr>
<tr>
<td>Number of reflections in free set</td>
<td>1711</td>
</tr>
<tr>
<td>Number of reflections in free set</td>
<td>1711</td>
</tr>
<tr>
<td>Number of reflections in free set</td>
<td>1711</td>
</tr>
<tr>
<td>Number of protein atoms</td>
<td>4830</td>
</tr>
<tr>
<td>Average (B)-factor (Å(^2))</td>
<td>64.01</td>
</tr>
<tr>
<td>Residues in most favoured regions (%)</td>
<td>94.75</td>
</tr>
<tr>
<td>Residues in most favoured regions (%)</td>
<td>94.75</td>
</tr>
<tr>
<td>Residues in generously allowed regions (%)</td>
<td>5.25</td>
</tr>
</tbody>
</table>

**Table 1** Data-collection and refinement statistics for human uPBGD

Values in parentheses are for the outer shell.

### Notes

The apo-enzyme and which is positioned after strand \(\beta_3\), leading to a \(\beta\)-hairpin with residues 312–315 (\(\beta_4\)) and residues 318–321 (\(\beta_5\)) making up the antiparallel strands. Main-chain hydrogen bonds (Leu\(^{114}\)-NH-Ile\(^{116}\)-CO and Leu\(^{154}\)Ile\(^{156}\)-NH) give way to a main chain–side chain interaction (Asp\(^{122}\)-CO-Arg\(^{211}\)-Ne) as the strands diverge. The apical residue of the hairpin, Val\(^{316}\), hydrogen-bonds through its carbonyl oxygen with the side chain of Arg\(^{211}\). The hairpin packs on top of \(\alpha_1\) (residues 241–256) and \(\alpha_2\) (residues 325–343) with the hydrophobic side chains Leu\(^{115}\), Val\(^{316}\), Ile\(^{318}\) and Ala\(^{320}\) that protrude beneath the hairpin interacting with Leu\(^{292}\), Ile\(^{294}\) and Leu\(^{264}\) of the domain 3 helices. There are also a range of interactions between the hairpin and domain 1, in particular hydrogen bonds between the main chain of the hairpin and the terminal residue of \(\beta_4\), the last strand of domain 1 (Gly\(^{137}\)CO-Ile\(^{138}\)-NH and Thr\(^{119}\)NH-Ile\(^{110}\)-CO). This hydrogen-bonding pattern comes close to defining an extended seven-stranded sheet linking domains 1 and 3. Beyond the hydrogen-bonded region the relative strand twist is too large to define such a sheet formally. However, there are additional interactions, notably the Gln\(^{114}\) side chain hydrogen-bonds with the backbone carbonyl of Phe\(^{108}\) and the OG atoms of Thr\(^{109}\) and Thr\(^{118}\) are hydrogen-bonded with one another, further stabilizing the interaction between domains 1 and 3. Following this excursion into the \(\beta\)-hairpin, helix \(\alpha_2\) (residues 325–343) extends for 19 residues (5 residues longer than the equivalent helix in the E. coli enzyme).

The \(\beta\)-hairpin is also involved in a range of packing interactions with neighbouring molecules in the crystal. The elongated loops that lead into the \(\beta\)-hairpin between strands \(\beta_4\) and \(\beta_5\) nestle in a solvent channel between two symmetry-related neighbours.
Structure of human porphobilinogen deaminase

Figure 2  Tertiary and secondary structures of human uPBGD

(A) The structure of human uPBGD with the three domains coloured differently: domain 1 (green), domain 2 (blue) and domain 3 (yellow). The DPM cofactor (red) and sulfate ion are located at the catalytic site between domains 1 and 2. (B) The topology of the secondary structure of human uPBGD. Domain 1 (green) extends from residues 20 to 115 and residues 214 to 238. Domain 2 (blue) comprises residues 116–213. Domain 3 (yellow) extends from residue 239 to residue 356 and contains the cofactor-binding Cys^{251} residue. The disordered residues (56–76) and the location of the DPM cofactor are indicated.

Intermolecular contacts in the crystal

Human uPBGD crystallizes with two molecules in the asymmetric unit, related by a local 2-fold axis. Both molecules interact through a substantial interface (1472 Å²) stabilized by a series of hydrophobic contacts and a number of salt bridges and hydrogen bonds (Figure 3). The contacts arise from helix α11 (residues 29–46), the loop (residues 126–141) connecting strands β52 (residues 121–125) and β12 (residues 142–144), the C-terminal of helix α32 and the proceeding loop region (residues 189–196). Analysis by ProFace [32] demonstrates that the dimer interface area is only 5% of the total solvent accessible surface area of the individual protomers, whereas the average value for a stable dimer is 16% [33]. Furthermore, the non-polar fraction of the interface area (881 Å²) accounts for only 14% of the fully buried atoms in the structure, whereas the average value for a stable dimer is 65% [34]. Indeed there is no evidence to suggest the formation of stable dimers of human uPBGD on gel filtration at neutral pH [35].

Interactions between domains

There are mainly polar contacts between the three domains in the human deaminase molecule, suggesting a flexible structure similar to that proposed for E. coli PBGD [5]. With the exception of the main-chain connection between domains 1 and 2, there is also a hydrophobic between Gln^{29} and Met^{196}. The 29-residue insert following the third strand in domain 3 (β3), which is absent in the E. coli deaminase structure, is involved in packing against domain 1, the interaction being mediated primarily by hydrogen-bonding.

In addition to the mainly hydrophilic interactions at the interdomain regions of the human structure, there are also several hydrophobic interactions. Residues at the interface between domains 1 and 3 include Leu^{59}, Leu^{244} and Cys^{247}, and between domains 2 and 3, residues Ala^{192}, Met^{112} and Leu^{30} are involved. The interface between domains 2 and 3 also contains the bulky aromatic Trp^{283} that packs against Gln^{153}.

Structural comparison between human and E. coli PBGDs

The three-dimensional structure of human PBGD has many features in common with the E. coli PBGD structure, most notably the presence of three protein domains. The main differences between the human and E. coli PBGD structures occur as insertions in loop regions. For example, there is a three-residue insertion occurring between strands β52 and β12 (residues 129–131), a single-residue insertion between helix α23 and strand β32 (residue 180) and a 29-residue insertion between strand β3 and α22 (residues 296–324). Of the 360 residues in human uPBGD there are only 148 sequence identities with the E. coli enzyme. These identities partly cluster around the active site but otherwise are distributed evenly over the protein structure. The differences are not especially concentrated in loop regions connecting secondary structure elements but often involve hydrophobic residues of the core. The close conservation of the tertiary structure is achieved by a subtle blend of compensatory substitutions arising from different parts of the sequence that come together in the fold.

Superposition of the conserved secondary structural regions of human uPBGD on to those of E. coli PBGD yields an RMSD (root mean square deviation) of 0.97 Å over 160 Cα atoms. Superposition of individual domains demonstrates that domain 2 of the human uPBGD structure superposes well on to the corresponding domain of E. coli PBGD, yielding an RMSD of 0.61 Å (over 41 Cα atoms). Domains 1 and 3 yield RMSD of 0.75 Å (over 62 Cα atoms) and 0.81 Å (over 86 Cα atoms) respectively. When corresponding domains are individually overlaid, the other two domains are also well superposed, demonstrating that there is no major perturbation of domain relationships between the bacterial and human enzymes. However,
Although both molecules (labelled A and B) interact through a substantial interface stabilized by hydrophobic contacts, salt bridges and hydrogen bonds, they probably do not constitute a physiological dimer.

The close equivalence of the E. coli and human enzyme structures demonstrated by their superposition implies that the volumes of their respective hydrophobic cores are very similar in spite of the many substitutions that occur in these regions. In domain 1, Leu, Ile, Ala, Val, Leu, Ile, Leu, Ile, Cys and Leu of E. coli PBGD are replaced by Ile, Val, Thr, Phe, Ile, Leu, Phe, Val and Val respectively in the human uPBGD structure. In domain 2, Ile, Val, Tyr and Leu in the E. coli deaminase structure are replaced by Phe, Leu, Phe and Trp respectively in the human deaminase structure. In all cases where there is a substantial change in side chain volume, this is accompanied by compensatory changes and adjustments in adjacent residues. Most of the residues that contribute towards the hydrophobic core in domain 3 of the human enzyme possess small aliphatic side chains that are also conserved in the E. coli deaminase structure, occurring between two roughly parallel α-helices (α13 and α23). There are two notable differences in this region, whereby Thr and Met in the E. coli PBGD structure are replaced by Ile and Phe respectively in the human uPBGD structure. Side chain substitution patterns at domain interfaces are similarly compensatory and contribute to the conservation of domain relationships.

The active site of human uPBGD and the impact of the Gln167 mutation

The active site cleft (15 × 13 × 12 Å) containing the DPM cofactor is located at the interface between domains 1 and 2 in a similar arrangement to that in the E. coli enzyme (Figure 5). The DPM cofactor is covalently bound to the protein via a thioether linkage with Cys in a type-I turn in the sequence Gly-Gly-Cys-Ser-Val-Pro (residues 259–264) located at the end of a loop connecting helix α1 and strand β1 in domain 3. The electron density for the DPM cofactor indicates that it occupies the native reduced conformation, consistent with the colourless nature of the crystals. Most of the cofactor interactions with the protein are conserved between the human and E. coli enzymes. The acetate and propionate moieties of the DPM cofactor contribute most of the ionic interactions and hydrogen bonds to neighbouring
active site side chains (Figure 5). The interactions between the carboxylate groups of ring C1 and the enzyme involve several invariant arginine residues (Arg149, Arg150 and Arg173), the backbone amide nitrogen of Ser146 and the hydroxy and backbone carbonyl group of Ser147. The DPM cofactor ring C2 is positioned towards the back of the active site cleft and interacts with Ser65 and the backbone amide nitrogen atoms of both Ala189 and Gly218. The invariant Lys98 and Asp99 residues interact with both C1 and C2 rings; Lys98 forms salt bridges with the acetate cofactor side chains, whereas a carboxylate oxygen of the catalytic Asp99 residue hydrogen-bonds to both pyrrole NH atoms of the DPM cofactor.

A prominent peak of electron density was observed within hydrogen-bonding distance of Arg26 and Ser28 within the active site region and this has been refined as a sulfate ion derived from the crystal mother liquor (Figure 5). This site is occupied by the DPM ring C2 propionate in E. coli PBGD when the cofactor is in the oxidized linear conformation, and ring C2 occupies the proposed substrate-binding site [5]. When this site is vacant in the reduced E. coli enzyme, an acetate ion, again derived from the crystal mother liquor, is found at the equivalent position as the sulfate ion observed here [8]. The electron density map allowed modelling of 24 residues of the 29-residue insert in domain 3 of human uPBGD (residues 296–324) that is not present in the E. coli enzyme and is positioned after strand β3. This insert elongates helix α2, and extends into a β-hairpin (strands β4, and β5,) and comes close to forming an extended seven-stranded sheet linking domains 1 and 3. Because of low sequence homology in this region the insert had previously been modelled before strand β3, at residue positions 289–317. Although the insert is distant from the active site, its interaction with domain 1 may modulate conformational fluctuations associated with enzyme action and perhaps promote the weak dimerization exhibited in the crystal. Although human uPBGD crystallized with two monomers in the asymmetric unit, the dimer interface is not as extensive as seen in stable (physiological) dimers of other proteins, and there is no evidence to suggest the formation of dimeric human uPBGD at neutral pH in solution. However, the possibility of a weak interaction between subunits of the human enzyme cannot be excluded, perhaps playing a role in the channelling of the unstable preuroporphyrinogen product to uroporphyrinogen III synthase or some other regulatory process.

In humans the rate-limiting reaction in haem biosynthesis is governed by ALAS (5-aminolaevulinate synthase), the first enzyme of the eight-step Shemin pathway [37]. Subsequently, PBGD has the lowest specific activity, but under normal conditions 50% of the enzyme is adequate to sustain normal demands for haem. However, under conditions where the demand for haem is high, the induction of ALAS is triggered. Any compromise in the activity of PBGD therefore results in the accumulation of the immediate precursors, ALA and PBG, and may precipitate an attack of AIP.

The human uPBGD structure facilitates an understanding of the large number of inherited mutations that result in AIP. The original homology model of human PBGD based on the E. coli PBGD X-ray structure [12,13] yielded structural insights into how the many AIP-associated mutations could affect enzyme activity. Interestingly, many of the porphyria-associated mutation sites are sites that also differ between the human and E. coli deaminases. However, in the human PBGD structure, these isolated AIP-associated substitutions cannot sustain the active form of the enzyme.

The R167Q mutation is of particular interest since it provides insights into important aspects of the pyrrole polymerization with the DPM cofactor in the native reduced conformation with an ordered sulfate ion hydrogen-bonded to Arg26 and Ser28 at the proposed substrate-binding site.

The electron density map allowed modelling of 24 residues of the 29-residue insert in domain 3 of human uPBGD (residues 296–324) that is not present in the E. coli enzyme and is positioned after strand β3. This insert elongates helix α2, and extends into a β-hairpin (strands β4, and β5,) and comes close to forming an extended seven-stranded sheet linking domains 1 and 3. Because of low sequence homology in this region the insert had previously been modelled before strand β3, at residue positions 289–317. Although the insert is distant from the active site, its interaction with domain 1 may modulate conformational fluctuations associated with enzyme action and perhaps promote the weak dimerization exhibited in the crystal. Although human uPBGD crystallized with two monomers in the asymmetric unit, the dimer interface is not as extensive as seen in stable (physiological) dimers of other proteins, and there is no evidence to suggest the formation of dimeric human uPBGD at neutral pH in solution. However, the possibility of a weak interaction between subunits of the human enzyme cannot be excluded, perhaps playing a role in the channelling of the unstable preuroporphyrinogen product to uroporphyrinogen III synthase or some other regulatory process.

In humans the rate-limiting reaction in haem biosynthesis is governed by ALAS (5-aminolaevulinate synthase), the first enzyme of the eight-step Shemin pathway [37]. Subsequently, PBGD has the lowest specific activity, but under normal conditions 50% of the enzyme is adequate to sustain normal demands for haem. However, under conditions where the demand for haem is high, the induction of ALAS is triggered. Any compromise in the activity of PBGD therefore results in the accumulation of the immediate precursors, ALA and PBG, and may precipitate an attack of AIP.

The human uPBGD structure facilitates an understanding of the large number of inherited mutations that result in AIP. The original homology model of human PBGD based on the E. coli PBGD X-ray structure [12,13] yielded structural insights into how the many AIP-associated mutations could affect enzyme activity. Interestingly, many of the porphyria-associated mutation sites are sites that also differ between the human and E. coli deaminases. However, in the human PBGD structure, these isolated AIP-associated substitutions cannot sustain the active form of the enzyme.

The R167Q mutation is of particular interest since it provides insights into important aspects of the pyrrole polymerization...
mechanism. The mutant enzyme is weakly active in samples taken from human blood and the recombinant human enzyme expressed in *E. coli* also exhibits an enzymatic activity of only 5–10% of that of the native enzyme [34]. The R167Q uPBGD structure presented here demonstrates that the considerable loss of activity associated with this mutation arises from a discrete perturbation of the enzyme mechanism rather than from any major structural perturbation. Most interestingly, when incubated with PBG the R167Q mutant generates stable catalytic intermediates (ES, ES₂, ES₃ and ES₄), whereas the native enzyme turns these intermediates over rapidly. Accordingly, the R167Q mutant is CRIM (cross-reacting immunological material) positive, reflecting the accumulation of these substantially longer lived enzyme–intermediate complexes. Arg¹⁶⁷ is thought to play a key role in deaminase binding to PBG by breaking the internal salt bridge between the amino and acetic acid side chains of the unbound substrate. The combination of Asp⁹⁹ and Arg¹⁶⁷ may facilitate the cleavage of the intra-molecular salt bridge to form new interactions, with the substrate amino group interacting with Asp⁹⁹ and the substrate acetyl group interacting with Arg¹⁶⁷. It has previously been observed that the equivalent Arg¹⁴⁹ residue of the *E. coli* enzyme is well ordered only when interacting with the acetyl side chain of a pyrrole bound at the proposed substrate-binding site [5]. The proposed role of Arg¹⁶⁷ in breaking the intra-molecular salt bridge in the PBG substrate would account for the lowering of the pH optimum of human uPBGD from pH 8 to pH 6 when it is replaced by a glutamine residue, since the salt bridge is much weaker at lower pH and can break without the assistance of the enzyme. The structure presented here suggests that the mechanistic lesions associated with this mutation are only felt when the Arg¹⁶⁷ side chain is required to visit the substrate site. Additionally, activity may also be reduced by the neutral Gln¹⁶⁷ side chain disrupting the positively charged active site surface over which the negatively charged polypyrrole chain may ‘slide’ during polymerization.

PBGD mutations associated with porphyria are widely dispersed over all three domains of the enzyme structure. Over 100 of these mutations involve single base changes that result in the substitution of one amino acid residue for another (missense; summarized in Figure 1) or the formation of a truncated protein (nonsense) and it is the former group that is especially interesting for providing structure/function information for the enzyme. While the R167Q mutation discussed above exemplifies those affecting pyrrole polymerization, mutations that lead to attacks of AIP may be divided into three more broad groups according to their molecular basis, including mutations that affect protein stability and folding, cofactor assembly and the catalytic process.

Mutations of residues that are essential for protein folding and stability include those occurring in the hydrophobic core, in tight turns and other conformationally restricted areas. Their mutation may result in an abnormally folded or unstable protein that is rapidly degraded in *vivo*. Such proteins typically exhibit CRIM phenotype (CRIM negative) with antibody raised against the native protein. Because of their intrinsic instability it is not possible to study the equivalent recombinant proteins in *vivo*; however, the impact of the mutation may be assessed by inspecting its location in the human uPBGD structure. For example, Leu²⁴⁵ is a conserved buried hydrophobic side chain that interacts with a cluster of proline residues (Pro²⁴¹, Pro³⁰⁶ and Pro³³⁷) in domain 1. Replacement of Leu²⁴⁵ by the AIP-associated Arg²⁴⁵ would bury a large positively charged side chain with resulting adverse consequences for the stability of the domain fold. Clearly, mutations that lead to the premature introduction of a stop codon will result in a truncated protein, affecting both folding and stability. For example, the most common mutation within the PBGD gene in Swedish patients (Trp¹⁹⁸Term) results in a severely truncated, probably inactive, and easily degraded protein, with patients exhibiting a CRIM negative phenotype [38]. The incidence of attacks of AIP in patients carrying this lesion is one of the highest, considering the low penetrance (8%) of the disease [39]. Two truncating AIP-associated mutations also target the domain 3 insert (Gln²⁹⁶Term and Gln³¹⁴Term) and both of these would be expected to destabilize domain 3 and interfere with the correct folding of the enzyme during biosynthesis.

Other mutations do not adversely affect apoenzyme formation but do prevent the binding, reaction and assembly of the cofactor precursor uroporphyrinogen (1-hydroxymethylbilane). In the absence of the stabilizing influence of the DPM cofactor, the intrinsically unstable apoenzyme is denatured and broken down intracellularly. The importance of the DPM cofactor not only for enzyme activity but also for enhancing the stability of the protein is exemplified by the dramatic finding that the apoenzyme is rapidly denatured at 60°C, whereas the holoenzyme loses little activity over 15 min at this temperature [40]. A close inspection of the structure in the vicinity of the cofactor demonstrates that it makes some 25 direct contacts with the protein. Phenotypes in this class are therefore normally CRIM negative and include, for example, the well-characterized ‘base 500 mutations’ that target Arg³⁴⁶ (R149Q and R149L), a residue intimately involved in cofactor assembly. Expression of recombinant R149Q human uPBGD in *E. coli* yields apoprotein with less than 5% of wild-type activity in crude extracts and which is heat labile and completely inactive after purification [35]. A negative reaction with Ehrlich’s reagent [17] confirmed that the purified mutant is devoid of the DPM cofactor, and, furthermore, attempts to incorporate uroporphyrinogen in *vivo* were unsuccessful [34].

Finally, mutations that affect catalytic residues or substrate-binding residues result in a correctly folded but inactive enzyme. Mutations at the substrate-binding site include R26H and R173Q. Mutagenesis studies at the equivalent positions in the *E. coli* enzyme (R11H and R155H [41] or R11L and R155L [42]) demonstrate that these mutations prevent the enzyme from binding the substrate. Both of these human mutant PBGDs appear to be correctly folded since they exhibit a CRIM positive phenotype and contain the DPM cofactor. However, the disruption of the PBG-binding site clearly prevents ES complex formation and both mutants are essentially inactive. Mutations involving the catalytic Asp⁹⁹ residue are of particular interest since they trap the deaminase in the ES₂ form. Although reaction of D99G human uPBGD apoenzyme with the tetrapyrrole cofactor precursor, uroporphyrinogen, appears to be unaffected, the resulting ES₂ complex is unable to catalyse the addition of further substrate to complete the catalytic cycle and generate the product and the DPM cofactor [43]. The related D84A mutation in *E. coli* PBGD also appears to exist as an inactive ES₂ complex [44]. Both findings support the hypothesis that the cofactor arises not from two single molecules of substrate but from a preformed tetrapyrrole [4].

All of these mutants of human PBGD have been identified on the basis that their *in vivo* activity is low enough to cause symptoms of AIP. When the CRIM status confirms that normal amounts of enzyme are produced, these mutations provide an informative repertoire of lesions in protein function that contribute towards a deeper understanding of the enzyme’s mechanism of action.

**ACKNOWLEDGEMENTS**

We are grateful to Professor Joël Janin (Laboratoire d’Enzymologie et Biochimie Structurales, Centre National de la Recherche Scientifique, Paris, France) and Dr Ranjit Bahadur (Department of Biochemistry, Bose Institute, Kolkata, India) for performing the dimer interface calculations. We thank Mr Barry Lockyer and Miss Vashi Maharaj for assisting with the preparation of the Figures.
REFERENCES


7 Barton, G. J. (1993) ASCRIBER: a tool to format multiple sequence alignments. Protein Eng. 6, 37–40


FUNDING

The Wellcome Trust provided grants for the X-ray and MS equipment and the Biotechnology and Biological Sciences Research Council provided support for M.S. and J.M. A.A.D-B thanks the Royal Embassy of Saudi Arabia for the award of a scholarship.

Received 13 October 2008/10 February 2009; accepted 11 February 2009

Published as BJ Immediate Publication 11 February 2009, doi:10.1042/BJ20082077

© The Authors Journal compilation © 2009 Biochemical Society
SUPPLEMENTARY ONLINE DATA
Structure of human porphobilinogen deaminase at 2.8 Å: the molecular basis of acute intermittent porphyria

Raj GILL*, Simon E. KOLSTOE*, Fiyaz MOHAMMED†1, Abeer AL D-BASS†2, Julie E. MOSELY†3, Mohammed SARWAR†, Jonathan B. COOPER*, Stephen P. WOOD* and Peter M. SHOOLINGIN-JORDAN†4

*Centre for Amyloidosis and Acute Phase Proteins, Department of Medicine, UCL Medical School, Rowland Hill Street, London NW3 2PF, U.K., and †Biomolecular Sciences Group, School of Biological Sciences, University of Southampton, Bassett Crescent East, Southampton SO16 7PX, U.K.

Please see the following pages for Figure S1

1 Present address: Cancer Research UK Institute for Cancer Studies, School of Cancer Sciences, University of Birmingham, Vincent Drive, Edgbaston, Birmingham B15 2TT, U.K.
2 Present address: Department of Biochemistry, College of Sciences, King Saud University, P.O. Box 2455, Riyadh 11451, Saudi Arabia.
3 Present address: Neurology and Gastrointestinal Centre of Excellence for Drug Discovery, GlaxoSmithKline, New Frontiers Science Park, Third Avenue, Harlow, Essex CM19 5AW, U.K.
4 To whom correspondence should be addressed (email pmsj@soton.ac.uk).

The structure of human ubiquitous porphobilinogen deaminase has been deposited in the PDB under accession code 3EQ1.
Figure S1  PBGD sequence alignments for representative bacteria, archaea and eukarya species

Representative PBGD sequences from the bacterial, archaeal and eukaryotic taxa deposited in the UniProt Knowledgebase (UniProtKB; http://www.uniprot.org) were aligned with CLUSTALW [1] and depicted with ALSCRIPT [2]. Residues conserved across all species are indicated by ●, conserved substitutions are indicated by □ and semi-conservative substitutions are indicated by ×. Amino acids are colour-coded such that red indicates acidic residues, blue basic, green neutral, pink hydrophobic, yellow cysteine and white indicates the conformationally important glycine, alanine and proline residues. The representative species and sequence codes are as follows. Bacterial: Q1INI6 Acidobacteria/ Acidothermus strain, B2V8Q3 Aquificales/Sulfurihydrogenibium, P06983 Proteobacteria/ Escherichia coli K12, A1ZNC8 Bacteroidetes/Microscilla marina, B3QWI1 Chlorobiacea/Chloroherpeton thalassium, ABAYV9 Cyanobacteria/Microcystis aeruginosa, A5UU37 Chloroflexi/Roseiflexus sp. RS-1, Q0RRI0 Deinococci/Deinococcus radiodurans, A6C1S2 Plantomycetes/Plantomyces maris DSM 8797, B3DQ16 Verrucomicrobia/Methylacidiphilum infernorum V4, Q7P7E7 Fusobacteria/ Fusobacterium nucleatum, A6LKX0 Thermus/ Thermosiphon marisvita (strain C7/ATCC BAA-1331), A9A1A1 Thermotoga/ Nitrosopumilus maritimus (strain SCM1). Archaeal: A6VFG4 Euryarchaeota/ Methanococcus maripaludis (strain C7/ATCC BAA-1331), A0A1A1 Thermalectros/  Thermoplasma Sulfolobus, P26789 Ascomyta/ Saccharomyces cerevisiae, Q5P3G3 Mycetozoa/Dictyostelium discoideum, F13446 Euglenozoa/ Euglena gracilis, A8CBH8 Apicomplexa/ Plasmodium falciparum, Q43316 Viridiplantae/ Arabidopsis thaliana, P08397 Coelomata/ Homo sapiens.

REFERENCES

2 Barton, G. J. (1993) ALSCRIPT: a tool to format multiple sequence alignments. Protein Eng. 6, 37–40

Received 13 October 2008/10 February 2009; accepted 11 February 2009
Published as BJ Immediate Publication 11 February 2009, doi:10.1042/BJ20082077

© The Authors Journal compilation © 2009 Biochemical Society