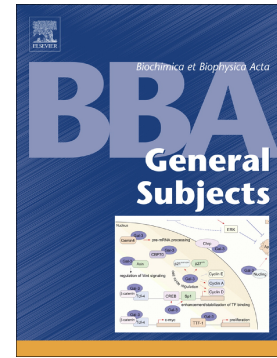# Accepted Manuscript

Structural basis of pyrrole polymerization in human porphobilinogen deaminase

Paula Pluta, Pietro Roversi, Ganeko Bernardo-Seisdedos, Adriana L. Rojas, Jonathan B. Cooper, Shuang Gu, Richard W. Pickersgill, Oscar Millet

Structural basis of pyrrole polymerization in human porphobilinogen deaminase

Paula Pluta,[1] Pietro Roversi,[2,3] Ganeko Bernardo-Seisdedos,[1] Adriana L. Rojas,[4] Jonathan B. Cooper,[5,6] Shuang Gu,[7] Richard W. Pickersgill,[7] Oscar Millet[1]

[1]Protein Stability and Inherited Disease Laboratory, CIC bioGUNE, Derio, Bizkaia, 48160, Spain.

[2]Oxford Glycobiology Institute, Dept. of Biochemistry, University of Oxford, Oxford OX1 3QU, UK.

[3]Leicester Institute of Structural and Chemical Biology, Department of Molecular and Cell Biology, University of Leicester, Henry Wellcome Building, Lancaster Road, Leicester, LE1 7RH, England, UK.

[4]Structural Biology Unit, CIC bioGUNE, Derio, Bizkaia, 48160, Spain.

[5]Department of Biological Sciences, Birkbeck, London WC1E 7HX, UK.

[6]Division of Medicine, University College London, London WC1E 6BT, UK.

[7]School of Biological and Chemical Sciences, Chemistry & Biochemistry Department, Queen Mary University of London, Mile End Road, London, E1 4NS, UK.

Corresponding author:

Oscar Millet

Protein Stability and Inherited Disease Laboratory,

CIC bioGUNE, Derio, Bizkaia, 48160, Spain

omillet@cicbiogune.es

ABSTRACT

Human porphobilinogen deaminase (PBGD), the third enzyme in the heme pathway, catalyzes four times a single reaction to convert porphobilinogen into hydroxymethylbilane. Remarkably, PBGD employs a single active site during the process, with a distinct yet chemically equivalent bond formed each time. The four intermediate complexes of the enzyme have been biochemically validated and they can be isolated but they have never been structurally characterized other than the apo- and holo-enzyme bound to the cofactor. We present crystal structures for two human PBGD intermediates: PBGD loaded with the cofactor and with the reaction intermediate containing two additional substrate pyrrole rings. These results, combined with SAXS and NMR experiments, allow us to propose a mechanism for the reaction progression that requires less structural rearrangements than previously suggested: the enzyme slides a flexible loop over the growing-product active site cavity. The structures and the mechanism proposed for this essential reaction explain how a set of missense mutations result in acute intermittent porphyria.

Introduction

The linear tetrapyrrole hydroxymethylbilane (HMB, also named pre-uroporphyrinogen) is the precursor of a variety of essential metabolites including heme, chlorophylls and cobalamin. In particular, the heme group acts as a prosthetic addition to many enzymes, transporters, and receptors including hemoglobin, myoglobin, catalases, peroxidases, cytochromes P450, and in sensor proteins for diatomic gases such as O_2 or NO [1]. There are three pathways to make heme [2] yet, in eukaryotes, heme synthesis appears to follow the protoporphyrin pathway [3]. Heme synthesis is accomplished by the sequential action of several enzymes (eight in humans, mainly expressed in liver and in erythroid cells). Porphobilinogen deaminase (PBGD, EC 2.5.1.61) is the third enzyme of this biosynthetic route, catalysing the condensation of four units of porphobilinogen (PBG) to yield HMB, which is the precursor to all the tetrapyrroles. Inborn errors in any of these enzymes ultimately lead to severe metabolic disorders family termed as porphyrias [4]. Specifically, mutations in *HMBS* are responsible for Acute Intermittent Porphyria (AIP) with more than 400 pathogenic mutations described so far (including nonsense, missense and other mutations) [5,6].

PBGD is a monomeric enzyme, highly conserved across different species, with a molecular mass ranging between 34 to 45 kDa. *Apo*PBGD is thermodynamically unstable and it spontaneously and covalently attaches a dipyrromethane cofactor (DPM) from the reaction of the apo-enzyme with hydroxymethylbilane to generate the active holoenzyme in its ES_2 form [7,8]. Human PBGD is comprised of 3 distinct domains [9] and shares over 45% primary sequence identity with its counterparts from *E. coli* and others [10]. Two isoforms have been reported: the 44-kDa ubiquitous (housekeeping) enzyme (PBGD1) and the 42-kDa erythrocyte-specific enzyme (PBGD2) [11]. The ubiquitous PBGD1 has 17 extra residues in the N-terminal region when compared to erythrocytic PBGD2. The role for this tail in the N-terminal region is unclear but experimental evidence demonstrates that such residues remain unstructured, do not alter the overall fold, nor the catalytic activity, nor the enzyme's stability [12].

PBGD catalyses the formation of HMB by incorporating four units of PBG in a multistep reaction (Fig. S1) [13]. Reaction initially involves deamination of the PBG, followed by the attachment of the resulting azafulvene to the free alpha position. The DPM cofactor serves as a primer for the loading of further PBG molecules and, remarkably, *holo*PBGD uses the same acidic residue (D99 in the human sequence) for the pyrrole condensation (four times), repeating reactions where each intermediate of the polypyrrole elongation remains covalently attached to the DPM cofactor [14]. The resulting linear hexapyrrole is unstable and cleaves off through an unknown mechanism to yield the linear tetrapyrrole HMB product, also restoring *holo*PBGD.

Two different mechanisms have been proposed to explain how the enzyme deals with the growing oligopyrrole chain during catalysis and how the cleavage of the product occurs [15,16]. The first one assumes that the enzyme structure is flexible, such that the active site is opened for the entry of a new PBG molecule and closed during the ring fusion. Such opening and closing of the active site in human PBGD would be accomplished by movement of protein domains around proposed hinge regions (S96, H120, L238) [9]. The second mechanism speculates that at least four pyrroles could be accommodated without significant movement of the protein domains. Rotation of the cofactor around the covalent bond, linking the methylene C atom of the C1 pyrrole with the Cys261 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 cofactor movement would be aided by the presence of two almost completely invariant glycine residues [17].

To date, 14 crystal structures of the protein are available in the Protein Data Bank (2 of *H. sapiens*, 5 of *E. coli*, 5 of *B. megaterium*, 1 of *A. thaliana* and 1 of *V. cholerae*) [9,10,17-20]. Despite these efforts, the structures are either the *apo-* (50V4, D82A-PBGD of *B. megaterium*) or the *holo*PBGD and they provide little information on the intermediate catalytic stages of PBGD. Here, we present crystal structures of PBGD loaded with the cofactor (E_{holo}, 2.78 Å, PDB-ID: 5M7F) and PBGD with two additional substrate rings (ES_2, 2.73 Å, PDB-ID: 5M6R). Our structures, supported by additional Small Angle X-ray Scattering (SAXS) and NMR experiments, allow us to propose a mechanism for the reaction coordinate, with fewer enzyme structural changes than previously suggested, mostly effected by a change in volume at the active site. The latter change in turn results from a substrate-induced rearrangement of a flexible loop (R255-V263) sliding over the active site during reaction progression, in basic agreement with the static hypothesis. The large number of AIP-causing mutations occurring in this loop confirms the key relevance of this structural element not only for the reaction but also in disease.

Materials & Methods

Protein expression and purification. Freshly transformed *E. coli* BL21 (DE3) cells were used for protein expression. Cells to produce proteins for crystallization tests were grown in regular LB media, whereas cells to produce protein samples required for NMR experiments included ammonium chloride as the sole sources of nitrogen for protein synthesis. Protein purification was achieved as previously described [21], with the following modifications: the fractions containing pure PBGD were then loaded to an anion exchange chromatography column, either a HiTrap Q HP 5ml or MonoQ 5/50 GL. The resulting enzyme-intermediate

complexes were eluted with a linear gradient of buffer B; 0-30% in 35 CV or 0-20% in 50CV for respectively, Q or MonoQ column. Separated complexes, whose concentration were measured by UV-vis spectrometry at 280 nm, can be used immediately or flash frozen in 1 ml aliquots and stored in -80 °C.

SAXS data collection and analysis. Synchrotron SAXS data were collected on beamline BM29 at ESRF (Grenoble, France) with a 2D detector (Pilatus 1M) over an angular range q_{min} = 0.01 Å^{-1} to $_{qmax}$ = 0.5 Å^{-1}. X-ray scattering patterns were recorded for PBGD reaction intermediates at the concentration of 1.92 and 6.40 mg/mL (E_{holo}), 2.63 and 5.24 mg/mL (ES), 2.10 and 4.26 mg/mL (ES_2), 1.53 and 3.43 mg/mL (ES_3) in standard size exclusion chromatography buffer (20mM Tris-HCl pH 8.5, 40mM NaCl). In order to generate an *ab initio* model, ten runs of DAMMIN[22] were performed, and after averaging and filtering a representative bead model with 732 beads was obtained. Superposition of the bead model on the relaxed crystallographic PBGD (3ECR) structure was carried out using the program SUPCOMB [22].

Crystallization and Structure Determination. Diffraction-quality PBGD crystals for holoenzyme (E_{holo}, PBGD2) and one of the reaction intermediate complexes (ES2, PBGD1) were obtained by hanging-drop vapour diffusion at 21 °C in a solution containing 0.1 M sodium cacodylate pH 6.5, 0.2 M sodium chloride, 2 M ammonium sulphate and 0.1 M Na HEPES pH 7.2, 0.9 M Na dihydrogen phosphate/0.9 M K dihydrogen phosphate, respectively. Individual crystals were cryo-protected by a brief soak in mother liquor supplemented with 25 % glycerol and flash frozen in liquid nitrogen. Diffraction data were remotely collected at a wavelength of 0.972422 Å at the ID23-1 beamline at the ESRF synchrotron centre (Grenoble, France). Diffraction data were processed using the XDS and the Aimless programs [23]. Molecular replacement was carried out by the MOLREP program with *E. coli* PBGD (PDB ID: 3ECR) as the search model. Refinement was performed using the autoBUSTER computer program [24] and iterative cycles of model building achieved using COOT from the CCP4 suite [25].

NMR spectroscopy. $^{1}H,^{15}N$-HSQC experiments were recorded at 37 °C on an 800 MHz Bruker Advance III spectrometer from ^{15}N isotopically enriched protein samples in PBGD purification buffer A. 1024x200 points where collected and water suppression was achieved using WATERGATE. The data were processed using NMRPipe [26] software and in-house built scripts. Processed data were analysed with NMRDraw.

Docking simulations. A standard docking procedure was used for a rigid protein and a flexible ligand whose torsion angles were identified; ten independent runs per ligand. A grid

centred on macromolecule of 126, 126, and 126 points in x, y, and z directions was built with a grid spacing of 0.458 Å. The default settings were used for all other parameters. From the resulting ligand clusters those with best (lowest) binding energies and localized within the active centre area were analysed for hydrogen bonding using AutoDockTools and Discovery Studio Visualizer 4.5.

Molecular Dynamics simulations. MD simulations (for structure relaxation before fitting model to the SAXS data) were performed with the NAMD2 program along with the CHARMM27 force field. Simulations were carried out in the constant volume-constant temperature (NVT) ensemble with the Leapfrog Verlet integrator employing an integration time step of 2 fs. A protein was solvated in a water sphere and the system was neutralized with NaCl. Covalent bonds between hydrogen atoms and heavy atoms were constrained by means of the RATTLE algorithm. A switching function was used to truncate the non-bonded interactions (Van der Waals and electrostatics) in the cut-off distance of 12 Å. The conjugated gradient method of minimization was used to relax the structure before the production for 5000 steps of force evaluations. MD run was performed for 17 ns (excluding the minimization). Coordinates were written to trajectory file every 50 or 100 steps and energy value were output every 10 steps.

Results

Isolation of human PBGD reaction intermediates. Investigation of the reaction mechanism requires the isolation of the PBGD reaction intermediates, which is a challenging task because each such species shows inherent reversibility, thus affecting the integrity of each isolated form over time. To address this issue, reaction intermediates of human PBGD (PBGD1 or PBGD2) were heterologously expressed in *E. coli* and purified as previously described [27] but adapting a protocol to ensure maximum purity and stability of the isolated intermediates (see Materials & Methods). Specifically, no ammonium ions were present at any time during the purification, not to shift the reaction balance to E_{holo}; and size exclusion chromatography was used as an early purification step, before optimizing the ion exchange chromatographic separation of the intermediates. As a result, five enzyme reaction intermediates could be isolated from each isoenzyme form, as determined by native gel analysis (Fig. 1). Mass spectrometry confirmed the chemical composition of each enzyme reaction intermediate (ES_{holo}, ES, ES_2 x2 and ES_3, Fig. 1). Two ES_2 intermediates can be separated by gel electrophoresis (Fig. 1) and ion exchange chromatography (Fig. S2), indicating the existence of two ES_2 conformations with different solvent exposed charge. The purified reaction intermediates show long term stability (more than 1 month) and do not scramble upon reinjection onto the column.

Reaction intermediates do not show large conformational changes. We have used a set of biophysical techniques to characterize the conformational changes across the PBGD reaction intermediates. First, small-angle X-ray scattering (SAXS) was collected at a synchrotron source for each of the protein intermediates (obtained from the PBGD1 isoform). For all collected datasets, SAXS data is only compatible with a monomeric form of PBGD, in agreement with ion exchange chromatography elution profiles (Fig. S2). The low-resolution SAXS-derived structural models were compared to the human high-resolution structure (PDB-ID: 3ECR, E_{holo}). To that end, the 3ECR structure was completed with homology modelling to fill the portions missing from the crystal structure (1-20, 60-78, 361-364) and subjected to the MD simulations for structure relaxation (17 ns, see Star Methods section), while the cofactor and reaction intermediates were not included. The SAXS molecular shapes (colorful mesh) are in good agreement with the 3ECR structure (black cartoon) for E_{holo}, ES *and* ES_2 reaction intermediates (Fig. 2A). The radius of gyration (R_g), estimated from the low-angle scattering region by a Guinier plot [28], also yields no significant differences among these enzyme variants (Fig. 2B and Table 1). In case of the last stage of reaction, before product release (PBGD ES_3), SAXS reports a larger R_g value and increases in the maximum linear dimension D_{max} and in the Chi parameter, but the 3ECR structure still fits the E_{holo} SAXS-derived low-resolution envelope rather well.

Isotope-enriched enzyme-intermediate complexes were also analyzed by NMR spectroscopy. Due to the exquisite sensitivity of chemical shift to changes in the nuclear environment the $^1H,^{15}N$-HSQC is a fingerprint spectrum for a protein, suitable to monitor its conformational rearrangements. The $^1H,^{15}N$-HSQC spectra for the different reaction intermediates are very similar (Fig. 2C), with a small number of peaks showing chemical shift perturbation. This is the expected outcome for localized changes instead of segmental domain motion, agrees with the SAXS data and suggests a reaction mechanism with localized segmental dynamics, ruling out the domain-reorienting mechanism. To confirm this hypothesis and to identify the involved structural elements, high-resolution structures are desirable.

High resolution structures for E_{holo} and the ES_2 reaction intermediate. To maximize chances of obtaining diffraction-quality PBGD crystals, trials using both PBGD variants (PBGD1 and PBGD2) were attempted, using crystallization screens based on the conditions reported in the literature or commercially available screens. The temperature was also varied. Crystals were always grown in the dark, to prevent cofactor and substrate photo-oxidation and the protein concentration was finely tuned, since PBGD tends to form multiple

crystal plates that grow in tight stacks (Fig. S3). The yellow color of the crystals indicates partial oxidation of the polypyrrol moiety. Out of the more than 160 crystals tested and more than 40 diffraction datasets collected, two crystals produced good resolution diffraction datasets for two reaction intermediates: E_{holo} (PBGD2, Fig. S4) and (ES_2, PBGD1, Fig. S5).

The diffraction data obtained from these crystals were processed to a resolution of 2.78 Å and 2.73 Å for E_{holo} and ES_2-complex, respectively. The crystallographic data collection and refinement statistics are presented in Table 2. Both crystals belong to the monoclinic space group $P2_1$, with unit-cell parameters a = 70.42 Å, b = 80.95 Å, c = 76.22 Å, β = 95.70° for E_{holo}; and a = 68.91 Å, b = 81.20 Å, c = 79.69 Å, β = 93.01° for the ES_2-complex. The Matthews coefficient[29,30], suggests that there is a single PBGD dimer per crystallographic asymmetric unit, with a solvent content of 57%, in both crystals. Structure solution by molecular-replacement used *H. sapiens* PBGD (PDB ID: 3ECR) as the search model. The final structures were validated using the MolProbity Server [31] and the final statistics were R_{work} = 0.220 (0.260) and R_{free} = 0.197 (0.246) for E_{holo} (ES_2–complex), respectively. Coordinates were deposited in the PDB with the accession codes 5M7F (PBGD2:E_{holo}) and 5M6R (PBGD1:ES_2).

Overall description of E_{holo} and ES_2–complex and comparison with other PBGD structures. The human PBGD structures for E_{holo} and ES_2–complex share the same fold, comprised of three α/β domains each of about 110 residues (domain 1: 1-116 and 216-239; domain 2: 117-215; domain 3: 240-361), linked together by flexible hinge regions (S96, H120, L238). Domains 1 and 2 are structurally close to periplasmic carbohydrate binding proteins [32]. The substrate-binding site is located in a cleft region between domains 1 and 2, while the dipyrromethane cofactor is covalently attached to Cys261, in a loop connecting $\alpha1_3$ and $\beta1_3$ in domain 3 (residues 248-264). Residues S59–S76, previously uncharacterized in human (or *E. coli*) structures, show well defined electron density in the ES_2–complex PBGD structure. These residues lie in a loop in the vicinity of the active site, forming a cap-like structure that covers the entire cavity.

E_{holo} is very similar to the previously elucidated human PBGD holo structures with r.m.s.d. values of 0.5 Å and 0.7 Å when compared 3ECR and 3EQ1 [9,10]. Such structural homology is largely maintained also for ES_2, which compares well with 3ECR (0.8 Å) and 3EQ1 (0.9 Å). A superposition of E_{holo} and ES_2-complex shows their striking structural similarity, which is reflected in a C^α r.m.s.d. of 0.6 Å for the 316 common residues (Fig. 3A). The comparison is made with structures from different isoforms, underlining the structural/functional equivalence between PBGD1 and PBGD2 and we suggest that the 17 extra residues found in PBGD1 might be related to protein trafficking and/or the cellular localization of the

enzyme. When compared domain by domain, the r.m.s.d. between E_{holo} and ES_2 become 0.4, 0.3 and 0.8 Å for domains 1, 2 and 3 respectively. The most noticeable changes are concentrated in the flexible loop that contains Cys261, located in domain 3 (Fig. 3B). In the ES_2 structure the loop is pulled back from the active site, displacing the Cys261 covalently-linked cofactor to a novel binding site. Such rearrangement is key for understanding the mechanism of substrate elongation (*vide infra*) but it also underlines the fact that localized and not domain motions govern the conformational changes from E_{holo} to ES_2, consistent with the SAXS and NMR data.

Structure of the active site and insights into the catalytic mechanism. In the E_{holo} PBGD structure, the dipyrromethane cofactor (pyrrole units C1 and C2, Fig. 4A) is positioned within a deep positively charged cleft formed between domains 1 and 2 (purple and orange volumes in Fig. 4E corresponding to the *frontal* and *rear* sub-cavities of the cleft). This ligand placement is also found in equivalent E_{holo} PBGD high-resolution structures and the cofactor is stabilized by well described interactions with surrounding residues (Fig. 4C) [10,18,33]. The C1 pyrrole unit is covalently attached to C261, which belongs to a loop (R255-V263) located in domain 3. In the ES_2 PBGD structure, the R255-V263 loop pulls away from the active site, displacing the polypyrrole chain by about 5 Å, dragging the C1 and C2 pyrroles into a new pocket (expansion of the rear sub-cavity, orange in Figures 4B-D); while the S1 and S2 units now occupy the positions the cofactor had in the E_{holo} structure, with the S1-S2 pyrrole units in ES_2 PBGD adopting *exactly* the same conformation and being stabilized by the same residues as the C1-C2 cofactor in E_{holo} PBGD. This suggests that the cofactor in E_{holo} is strongly bound and already adopts a catalytically relevant conformation. These observations are consistent with the large number of conserved surrounding residues and with the reaction's stereospecificity [34]. Yet, the ES_2-complex is further stabilized by the extra interactions that the newly displaced C1-C2 cofactor creates with the protein at the rear cavity, mostly with T102, V215 and S262. Such added interactions provide a structural rationale for the empirical observation that ES_2 is the most stable of the purified reaction intermediates [27,35].

In the ES_2-complex, D99, a key residue in the polypyrrole chain elongation, retains the hydrogen bonding with the N-H of S2 (C2 in E_{holo}), required for the elongation reaction, but it also falls at a *reactive* distance from the C2-S1 interface (3.8 Å to the S1 H-N), which is the product cleavage point (Fig. 4E-F). This proximity strongly suggests that D99 may also play a relevant role in the product release by catalyzing the hydrolysis reaction, in agreement with the spontaneous formation of E_{holo} when *apo*PBGD is incubated with HMB [36]. In the

context of our ES_2 PBGD structure, D99 is located at about 8.5 Å from the cofactor C1 ring (Fig. 4F), which now would become *protected* from hydrolysis during the elongation process.

In E_{holo} the active site is large (674 Å^3, Fig. 4A) and rich in positively charged residues (mainly Lys and Arg residues) to compensate-for and attract the negative charges introduced with the ligands' carboxylates during the elongation process. The placement of the cofactor and the location of D99 suggest that the PBG substrate enters the site from the frontal sub-cavity, as indicated by the arrow in Figure 3A. Remarkably, the displacement of the R255-V263 loop in the ES_2 PBGD intermediate enlarges the overall cavity size to a final volume of 787 Å^3 (Fig. 4D), creating a large empty space on the distal end of the catalytic site and dividing the cavity in two similar subspaces around D99: the frontal sub-cavity that remains largely unaltered as compared to E_{holo} (purple in Figures 4C,D, ES_2 frontal sub-cavity volume is 487 Å^3); and the rear sub-cavity (orange in Figures 4C,D), which is very large for ES_2 (300 Å^3) and directly produced by the loop displacement.

To investigate which parts of the cavity are relevant for the elongation reaction, we have used a set of competitive inhibitors that are able to effectively block the reaction ($K_i \leq 100$ µM). Specifically, computational docking of the inhibitors into the E_{holo} and ES_2 structures allows pinpointing the areas that may accommodate the substrate. As is shown in Fig. S6, both sub-cavities are occupied by the inhibitors in ES_2, while only the frontal sub-cavity is hit by the inhibitors in E_{holo}. This is consistent with substrates approaching the active site via the frontal sub-cavity. Interestingly, the extended size of the rear sub-cavity in ES_2 would allow the tetrapyrrole to adopt a different placement, in which the S2 moiety would occupy the S1 pocket and the S1 pyrrole ring would be completely inserted into the rear sub-cavity. Such a conformation would ease the entrance to the S2 site of a subsequent PBG molecule which would then proceed to fuse with ES_2 and elongate the product to ES_3. Based on this hypothesis, a computationally-derived model of the resulting ES_3 complex is shown in Fig. S7. Presumably, further displacement of the R255-V263 loop would enlarge the rear sub-cavity even more so the reaction can progress towards the formation of ES_4. In support of this model, the R26H mutant located in the frontal volume has been associated only with a reduced capacity to incorporate new substrate units [34].

The active-site loop is a lid for the catalytic site. The active site is partially protected by a flap region constituted by residues 57-74 in human PBGD, which is likely relevant in binding the substrate and cofactor as well as in aiding catalysis. This active-site loop is missing in all the deposited structures so far except in the E_{holo} PBGD from *A. thaliana* (4HTG) [17]. In this structure one of two chains in the asymmetric unit of the ES_2-complex crystal (5M6R_A)

displays ordered electron density for the flap region. The latter was revealed to form a cap-like structure that covers the active site. The loop in the remaining human holo-structures described so far is only partially ordered, with residues G60-I71 adopting α-helical secondary structure, while the equivalent region in PBGD ID 4HTG is much more disordered (Fig. 3C). Interestingly, the loop in our ES_2 human structure folds towards the active site, so that residue E73 is able to form a hydrogen bond with the propionate side chain of the C2 ring of the growing product, ultimately stabilizing the complex. In the *A. thaliana* structure, the equivalent residue to E73 is D66, which is solvent exposed and displays no interactions, in agreement with the E_{holo} contents of that active site.

Discussion

In this contribution, we report the first medium resolution structure of a reaction intermediate for PBGD (ES_2 of human PBGD1, PBD ID 5M6R). The combined analysis of the ES_2-complex with the structure of human E_{holo} PBGD2 (PBD ID 5M7F, also presented here) reveals that the structural changes upon polypyrrole elongation are largely constrained to the R255-V263 loop, refuting the hypothesis of significant domain rearrangements during the reaction progression up to the ES_2 intermediate (Fig. 3). It is entirely plausible to suggest that the only significant changes in the structure during the second part of the reaction (formation of ES_4 and product release) would also only affect to the R255-V263 loop, in keeping with our SAXS, NMR and gel filtration chromatography data.

In spite of the limited extent of the reported conformational changes among the reaction intermediates, the progressive increase in the electrophoretic mobility observed in the native polyacrylamide electrophoresis (Fig. 1) and their successful ionic exchange separation (Fig. S2) suggest a gradual increase in net negative charge compared to that of E_{apo} or E_{holo}, as previously suggested [37]. As shown in Table S1, this is actually the case, and it can be mainly attributed to the PBGD Arg residues, which experience a large reduction in their solvent accessible area in ES_2 as compared to E_{holo}, due to salt-bridge interactions with the negatively charged pyrrole units.

The structures reported here provide molecular detail on how the reaction process takes place within the confines of a single active site. At least three previously undescribed observations obtained from the structures appear to be relevant for the reaction mechanism: i) the displacement of the R255-V263 loop, which allows reaction progression up to the ES_2 stage allocating the cofactor (C1-C2 rings) into a new pocket and extracting it from the catalytic pocket; ii) the displacement of the same loop generates a new internal cavity, which is potentially suitable for hosting the further reaction intermediate; and iii) the relative position

of the ligands from the catalytic residue D99 [38], which is always at reactive distance from the S2(C2) elongation pocket, but also close enough to the C2 ring. Therefore, we propose that *D99 is also responsible for product release*. A scheme that embraces all these features into a mechanistic model of reaction progression is shown in Fig. 5. In the E_{holo} structure, the catalytic site is nearby the key residue D99 and occupied by the cofactor. The entrance of new PBG molecules occurs through the frontal sub-cavity and is facilitated by electrostatic complementarity. The entrance of the PBG into the rear sub-cavity likely produces the cofactor shift, mediated by the displacement of the loop containing Cys261. As a result, C1 now occupies a new pocket whilst C2 occupies the former C1 position and the new pyrrole unit is allocated at the condensation catalytic site. The progressive enlargement of the rear sub-cavity produced by the concerted R255-V263 loop displacement would allow for this mechanism to be repeated twice, up to the formation of the ES_4 intermediate.

A plausible mechanism for the HMB release is also shown in Fig. 5. According to our model, cleavage of the bilane attached to the cofactor and polypyrrole elongation are catalyzed by the same aspartic acid, which is always at *reactive distance* from the catalytic pocket (the location of the next PBG unit for condensation) and from the N-H of C2 (likely, the cleavage point in the polymer). We hypothesize that steric tension and the proximity of the C2 ring to D99 may trigger the cleavage hydrolytic reaction and product release. Essentially, the mechanism we propose corresponds to the reverse condensation reaction mechanism, also involving an azafulvene intermediate. In this context, both reactions are reversible and they could spontaneously occur at the appropriate stage of the elongation reaction. The interplay between the two opposite reactions (condensation and hydrolysis) is governed by a delicate balance between the S2 and C2 distances to D99. According to this hypothesis, the condensation reaction would dominate from E_{holo} to ES_4 due to a closer distance of D99 to the reaction site than to C2 (confirmed for ES_2: D99 is at 2.2 Å and 3.8 Å of the N-H of S2 and C2 respectively), with some aid from the chemical driving force of the substrate excess. During the elongation reaction, it is plausible that the D99(O)-C2(NH) distance shortens to favor the hydrolysis reaction, as reflected in the structural model for the ES_3 reaction intermediate (Fig. S7). This feature, combined with the increasing steric tension of the reaction intermediate, may favor the hydrolysis reaction. The active site loop (residues 57-74 in the human isoform) is forming a contact with the C2 ring that could presumably fine tune the distance between D99 and the pyrrole unit. The experimental reconstitution of E_{holo} starting from E_{apo} and HMB [36] supports this model.

The reported structures also allow rationalization of the relative stability of the reaction intermediates. Clearly, ES_2 is more stable than E_{holo}, because it integrally maintains the same cofactor-protein interactions as observed in E_{holo}, whilst being further stabilized by the

interactions generated between the C1/C2 pyrroles and the residues lining the new pocket. Consistent with these observations, the ES reaction intermediate would be more stable than E_{holo} and less stable than ES_2. We predict that progression from ES_2 to ES_4 might generate further interactions between the polypyrrole and the cavity residues but likely at the cost of generating a tensioned and, therefore, ultimately unstable conformation.

Last but not least, the reported structures and the proposed reaction mechanism allow rationalization of previously unexplained missense mutations in human *HMBS* that result in AIP. As it is common in porphyria [21], most of the 168 missense mutations reported in the Human Mutation Database [5] can be associated with reduced protein stability. Yet, a subset of the point mutations has already been associated to a loss of interactions between protein residues and the C1-C2 pyrrole units in the E_{holo} structure, affecting the stability of the reaction intermediates [10,39]. Here we highlight the relevance of the sliding loop, key for reaction elongation, which also contains seven residues (E250, A252, L254, H256, E258, G260 and C261) that produce AIP upon mutation. It is likely that these mutations constrain the loop mobility, ultimately compromising reaction progression. Furthermore, the V215M and V215E mutations, which would abrogate the V215-mediated stabilization of the C2 ring at the shifted position in the ES_2 structure, also results in AIP. Other residues in the rear cavity are also involved in AIP upon mutation: L254, R149, A214 and Y213. These mutants may be amenable to stabilization by pharmacological chaperones, opening new therapeutic intervention avenues against the disease.

Atomic coordinates for the structures of human PBGD have been deposited in the Protein Data Bank (PDB codes 5M7F, 5M6R).

We have no conflict of interest to declare.

References

1. A.R. Battersby, Tetrapyrroles: the pigments of life, Nat. Prod. Rep. 17 (2000) 507-526.

2. H. A. Dailey, T. A. Dailey, S. Gerdes, D. Jahn, M. Jahn, M.R. O'Brian, M. J. Warren, Prokaryotic Heme Biosynthesis: Multiple Pathways to a Common Essential Product, Microbiol. Mol. Biol. Rev. 81 (2017) e00048-16.

3. G. Layer, J. Reichelt, D. Jahn, D.W. Heinz, Structure and function of enzymes in heme biosynthesis, Protein Sci. 19 (2010) 1137-1161.

4. S. Besur, W. Hou, P. Schmeltzer, H.L. Bonkovsky, Clinically important features of porphyrin and heme metabolism and the porphyrias, Metabolites. 4 (2014) 977-1006.

5. P.D. Stenson, M. Mort, E.V. Ball, K. Evans, M. Hayden, S. Heywood, M. Hussain, A.D. Phillips, D.N. Cooper, The Human Gene Mutation Database: towards a comprehensive repository of inherited mutation data for medical research, genetic diagnosis and next-generation sequencing studies, Hum. Genet. 136 (2017) 665-677.

6. B. Chen, C. Solis-Villa, A.L. Erwin, M. Balwani, I. Nazrenko, J.D. Phillips, R.J. Desnick, M. Yasuda, Identification and characterization of 40 novel hydroxymethylbilane synthase mutations that cause acute intermittent porphyria, J. Inherit. Metab. Dis. (2018) on-line publication.

7. P.M. Jordan PM, M.J. Warren, Evidence for a dipyrromethane cofactor at the catalytic site of E. coli porphobilinogen deaminase, FEBS Lett. 225 (1987) 87-92.

8. J. Pluscec, L. Bogorad, A dipyrrylmethane intermediate in the enzymatic synthesis of uroporphyrinogen, Biochemistry. 9 (1970) 4736-4743.

9. G. Song, Y. Li, C. Cheng, Y. Zhao, A. Gao, R. Zhang, A. Joachimiak, N. Shaw, Z.J. Liu, Structural insight into acute intermittent porphyria, FASEB J. 23 (2009) 396-404.

10. R. Gill, S.E. Kolstoe, F. Mohammed, A. Al D-Bass, J.E. Mosely, M. Sarwar, J.B. Cooper, S.P. Wood, P.M. Shoolingin-Jordan, Structure of human porphobilinogen deaminase at 2.8 A: the molecular basis of acute intermittent porphyria, Biochem J. 420 (2009) 17-25.

11. B. Grandchamp, H. De Verneuil, C. Beaumont, S. Chretien, O. Walter, Y. Nordmann, Tissue-specific expression of porphobilinogen deaminase. Two isoenzymes from a single gene, Eur. J. Biochem. 162 (1987) 105-110.

12. F. ben Bdira, E. Gonzalez, P. Pluta, A. Lain, A. Sanz-Parra, J.M. Falcon-Perez, O. Millet, Tuning intracellular homeostasis of human uroporphyrinogen III synthase by enzyme engineering at a single hotspot of congenital erythropoietic porphyria, Hum. Molec. Genet. 23 (2014) 5805-5813.

13. B. Frydman, R.B. Frydman, A. Valasinas, E.S. Levy, G. Feinstein, Biosynthesis of uroporphyrinogens from porphobilinogen: mechanism and the nature of the process, Philos. Trans. R. Soc. Lond. B. Biol. Sci. 273 (1976) 137-160.

14. P.M. Jordan, S.D. Thomas, M.J. Warren, Purification, crystallization and properties of porphobilinogen deaminase from a recombinant strain of Escherichia coli K12, Biochem J. 254 (1988) 427-435.

15. P.M. Jordan, A. Berry, Mechanism of action of porphobilinogen deaminase. The participation of stable enzyme substrate covalent intermediates between porphobilinogen and the porphobilinogen deaminase from Rhodopseudomonas spheroides, Biochem J. 195 (1981) 177-181.

16. M.J. Warren, A.I. Scott, Tetrapyrrole assembly and modification into the ligands of biologically functional cofactors, Trends Biochem. Sci. 15 (1990), 486-491.

17. A. Roberts, R. Gill, R.J. Hussey, H. Mikolajek, P.T. Erskine, J.B. Cooper, S.P. Wood, E.J. Chrystal, P.M. Shoolingin-Jordan, Insights into the mechanism of pyrrole polymerization catalysed by porphobilinogen deaminase: high-resolution X-ray studies of the Arabidopsis thaliana enzyme, Acta Crystallogr. D Biol. Crystallogr. 69 (2013) 471-485.

18. G.V. Louie, P.D. Brownlie, R. Lambert, J.B. Cooper, T.L. Blundell, S.P. Wood, M.J. Warren, S.C. Woodcock, P.M. Jordan, Structure of porphobilinogen deaminase reveals a flexible multidomain polymerase with a single catalytic site, Nature. 359 (1992) 33-39.

19. T. Uchida, T. Funamizu, M. Chen, Y. Tanaka, K. Ishimori, Heme Binding to Porphobilinogen Deaminase from Vibrio cholerae Decelerates the Formation of 1-Hydroxymethylbilane, ACS Chem. Biol. 13 (2018) 750-760.

20. N. Azim, E. Deery, M.J. Warren, B.A. Wolfenden, P. Erskine, J.B. Cooper, A. Coker, S.P. Wood, M. Akhtar, Structural evidence for the partially oxidized dipyrromethene and dipyrromethanone forms of the cofactor of porphobilinogen deaminase: structures of the Bacillus megaterium enzyme at near-atomic resolution, Acta Crystallogr. D Biol. Crystallogr. 70 (2014) 744-751.

21. A. Fortian, D. Castano, G. Ortega, A. Lain, M. Pons, O. Millet, Uroporphyrinogen III Synthase Mutations Related to Congenital Erythropoietic Porphyria Identify a Key Helix for Protein Stability, Biochemistry 48 (2009) 454-461.

22. D.I. Svergun, Restoring low resolution structure of biological macromolecules from solution scattering using simulated annealing, Biophys. J. 76 (1999) 2879-2886.

23. W. Kabsch, XDS, Acta Crystallogr. D Biol. Crystallogr. 66 (2010) 125-132.

24. E. Blanc, P. Roversi, C. Vonrhein, C. Flensburg, S.M. Lea, G. Bricogne, Refinement of severely incomplete structures with maximum likelihood in BUSTER-TNT, Acta Crystallogr. D Biol. Crystallogr. 60 (2004) 2210-2221.

25. W.D. Winn, C.C. Ballard, K.D. Cowtan, E.J. Dodson, P. Emsley, P.R. Evans, R.M. Keegan, E.B. Krissinel, A.G. Leslie, A. McCoy, et al., Overview of the CCP4 suite and current developments, Acta Crystallogr. D Biol. Crystallogr. 67 (2011) 235-242.

26. F. Delaglio, S. Grzesiek, G.W. Vuister, G. Zhu, J. Pfeifer, A. Bax, NMRPipe: a multidimensional spectral processing system based on UNIX pipes, J Biomol. NMR. 6 (1995) 277-293.

27. M.J. Warren, P.M. Jordan, Investigation into the nature of substrate binding to the dipyrromethane cofactor of Escherichia coli porphobilinogen deaminase, Biochemistry 27 (1988) 9020-9030.

28. M.H. Koch, P. Vachette, D.I. Svergun DI, Small-angle scattering: a view on the properties, structures and structural changes of biological macromolecules in solution, Q. Rev. Biophys. 36 (2003) 147-227.

29. K.A. Kantardjieff, B. Rupp, Matthews coefficient probabilities: Improved estimates for unit cell contents of proteins, DNA, and protein-nucleic acid complex crystals, Protein Sci. 12 (2003) 1865-1871.

30. B. W. Matthews, Solvent content of protein crystals, J. Mol. Biol. 33 (1968) 491-497.

31. V.B. Chen, W.B. Arendall, J.J. Headd, D.A. Keedy, R.M. Immormino, G.J. Kapral, L.W. Murray, J.S. Richardson, D.C. Richardson, MolProbity: all-atom structure validation for macromolecular crystallography, Acta Crystallogr. D. Biol. Crystallogr. 66 (2010) 12-21.

32. G. Ortega, D. Castano, T. Diercks, O. Millet, Carbohydrate Affinity for the Glucose-Galactose Binding Protein Is Regulated by Allosteric Domain Motions, J. Am. Chem.Soc. 134 (2012) 19869-19876.

33. M. Lander, A.R. Pitt, P.R. Alefounder, D. Bardy, C. Abell, A.R. Battersby, Studies on the mechanism of hydroxymethylbilane synthase concerning the role of arginine residues in substrate binding, Biochem J. 275 (1991) 447-452.

34. P.M. Jordan, S.C. Woodcock, Mutagenesis of arginine residues in the catalytic cleft of Escherichia coli porphobilinogen deaminase that affects dipyrromethane cofactor assembly and tetrapyrrole chain initiation and elongation, Biochem J. 280 (1991) 445-449.

35. A.I. Scott, C.A. Roessner, N.J. Stolowich, P. Karuso, H.J. Williams, S.K. Grant, M.D. Gonzalez, T. Hoshino, Site-directed mutagenesis and high-resolution NMR spectroscopy of the active site of porphobilinogen deaminase, Biochemistry. 27 (1988) 7984-7990.

36. S.J. Awan, G. Siligardi, P.M. Shoolingin-Jordan, M.J. Warren, Reconstitution of the holoenzyme form of Escherichia coli porphobilinogen deaminase from apoenzyme with porphobilinogen and preuroporphyrinogen: a study using circular dichroism spectroscopy, Biochemistry. 36 (1997) 9273-9282.

37. A. Berry, P.M. Jordan, J.S. Seehra, The isolation and characterization of catalytically competent porphobilinogen deaminase-intermediate complexes, FEBS Lett. 129 (1981) 220-224.

38. S.C. Woodcock, P.M. Jordan, Evidence for participation of aspartate-84 as a catalytic group at the active site of porphobilinogen deaminase obtained by site-directed mutagenesis of the hemC gene from Escherichia coli, Biochemistry. 33 (1994) 2688-2695.

39. P.D. Brownlie, R. Lambert, G.V. Louie, P.M. Jordan, T.L. Blundell, M.J. Warren, J.B. Cooper, S.P. Wood, The three-dimensional structures of mutants of porphobilinogen deaminase: toward an understanding of the structural basis of acute intermittent porphyria, Protein Sci. 3 (1994) 1644-1650.

Figure Captions

Figure 1. Native polyacrylamide electrophoresis for the different reaction intermediates isolated for human PBGD. The band downshift suggests a gradual increase in net negative charge as compared to E_{apo}. ES_2 and ES_2' both correspond to the reaction intermediate with two PBG condensed units. The presence of multiple bands probably indicates incomplete separation between the intermediates. The asterisk indicates the intermediate that was successfully crystallized.

Figure 2. Biophysical properties of the PBGD reaction intermediates. (**A**) Superimposition of the relaxed model structure (3ECR, after 17 ns MD simulation) with the SAXS-derived low-resolution protein volumes. (**B**) Fitting of the experimental scattering spectra with a simulated fit obtained from the relaxed structure. (**C**) $^1H,^{15}N$-HSQC spectra for the isolated PBGD reaction intermediates show chemical shift perturbations consistent with localised structural changes.

Figure 3. (**A**) Overall structure of human PBGD. Cartoon representation of human PBGD bound with dipyromethane cofactor (E_{holo}, erythrocytic PBGD isoform, green) overlaid with the reaction intermediate (ES_2, housekeeping isoform, blue). The arrow indicates the likely pathway for a new PBG unit as well as the direction of the loop displacement. (**B**) Root mean square deviation values between the residues of E_{holo} (reference) and ES_2-complex structures. Values below the average (r.m.s.d_{av} = 0.434) covered with grey color. (**C**) Cartoon representation of ES_2 PBGD showing the loop region covering the active site, red highlighted. The two orientations correspond to a 90º rotation along the ordinate reference axis.

Figure 4. Active site of human PBGD. (**A-B**) Electron density for the cofactor/ligand in E_{holo} and ES_2. The interactions between the cofactor or the reaction intermediate and the residues from the active site are shown in (**C**) and (**D**) for E_{holo} and ES_2, respectively. The cavity volume enclosing the active site is shown in (**E**) and (**F**) for E_{holo} and ES_2. For clarity, the cavity has been divided into the frontal (purple) and rear (orange) sub-cavities, as indicated. Distances between the catalytic residue D99 and some atoms relevant for the reaction are shown in (**G**) and (**H**) for E_{holo} and ES_2-complex.

Figure 5. Scheme showing the active site cavity as a function of the reaction progression. The R255-V263 loop hosts C261 which covalently binds the cofactor. The loop displaces during polypyrrole elongation. At all times, D99 is at reactive distance of the S2 and C2 positions in the ES_2 complex. According to our model, once the ES_4 complex is reached, the hydrolysis occurs spontaneously and mediated by D99 (reverse reaction), as indicated. The

cavities for E_{holo} and ES_2 (dash gray line highlight) are depicted from the high-resolution structures while the other cavities are theoretical models.

Table 1. SAXS data analysis statistics

Protein sample	R_g [Å]	D_{max} [Å]	Chi	NSD[1]
PBGD1 E$_{holo}$	2.39 +/- 0.01	8.38	1.094	0.8212
PBGD1 ES	2.45 +/- 0.07	8.59	2.04	0.8169
PBGD1 ES$_2$	2.43 +/- 0.04	8.51	1.788	0.8117
PBGD1 ES$_3$	2.92 +/- 0.12	10.23	5.037	0.8114

[1] Normalized Spatial Discrepances.

Table 2. PBGD crystallographic data collection and refinement statistics

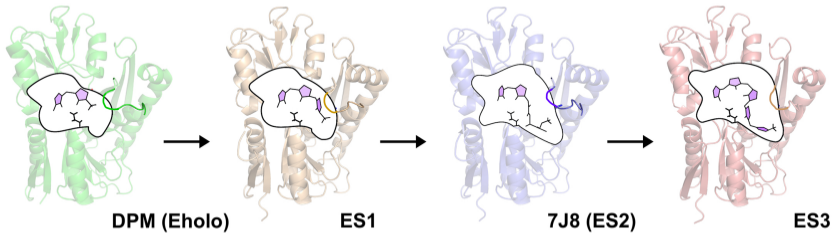| PDB ID	5M7F (E_{holo})	5M6R (ES_2)
X-ray data collection statistics		
X-ray source	ESRF ID23-1	ESRF ID23-1
Detector	Dectris Pilatus 6M-F	Dectris Pilatus 6M-F
Wavelength (Å)	0.972422	0.972422
Space group	P 1 2_1 1	P 1 2_1 1
Cell constants **a, b, c, α, β, γ**	70.42Å, 80.95Å, 76.22Å, 90.00º, 95.70º, 90.00º	68.91Å, 81.20Å, 79.69Å, 90.00º, 93.01º, 90.00º
Resolution limits (Å)	80.95–2.78 (2.83–2.78)	81.20–2.73 (2.78–2.73)
Completeness (%)	98.9 (83.1)	97.7 (65.7)
Unique reflections measured	21353	22971
Multiplicity	3.3 (2.6)	9.2 (3.8)
$R_{merge}(I)$	0.247 (1.265)	0.194 (1.377)
CC_{half}	0.946 (0.477)	0.996 (0.655)
Average I/σ(I)	3.1 (0.9)	11.7 (1.2)
Refinement statistics		
Resolution range (Å)	29.77–2.78 (2.92–2.78)	56.84–2.73 (2.86–2.73)
Completeness (%)	98.5 (91.77)	95.8 (77.95)
Unique reflections refined against	21279 (2509)	22510 (2306)
Free set	1112 (112)	1104 (127)
R	0.222	0.200
R_{work}	0.220 (0.250)	0.197 (0.239)
R_{free}	0.260 (0.301)	0.246 (0.289)
RMSD bond lengths (Å)	0.01	0.01
RMSD bond angles (º)	1.16	1.14
Ramachandran outliers (protein backbone)	0	3
Ramachandran favoured (protein backbone)	97%	95%
Residues modelled (range)	A: 20–57, 75–353, B: 20–56, 75–353	A: 18–357, B: 18–61, 76–353
Waters modelled	49	23
Non-protein molecules	2 DPM, 6 SO4	2 7J8, 3 PO4, 1 EPE
Total number of atoms	10034	10597
Average B, all atoms ($Å^2$)	45.0	91.0

Values in parentheses refer to the outer resolution range

Highlights

- First crystal structure of a PBGD reaction intermediate.
- Structural data reveal molecular basis for tetrapyrrole elongation by PBGD.
- Identification of a mobile loop and a cavity of variable size, key for reaction.
- Rationalization of a set of mutations causing acute intermittent porphyria.

PBGD Reaction Progression

DPM (Eholo)　　　ES1　　　7J8 (ES2)　　　ES3

Graphics Abstract

Figure 1

A

E_{holo} ES ES$_2$ ES$_3$

B

Exp. (PBGD1 Eholo)
The. (3ECR MD17ns)
Chi: 1.094

Exp. (PBGD ES1)
The. (3ECR MD17ns)
Chi: 2.040

Exp. (PBGD ES2)
The. (3ECR MD17ns)
Chi: 1.788

Exp. (PBGD1 ES3)
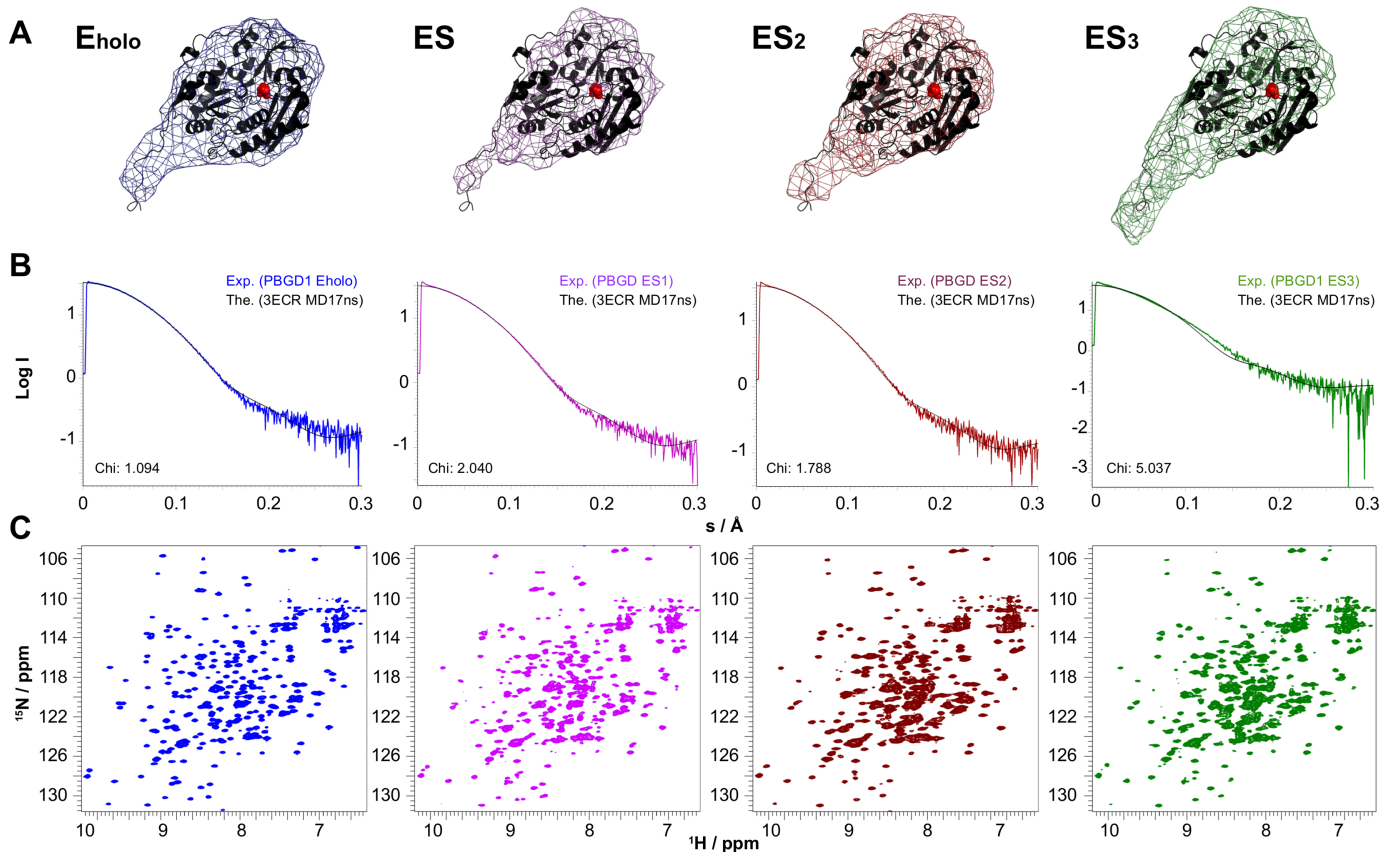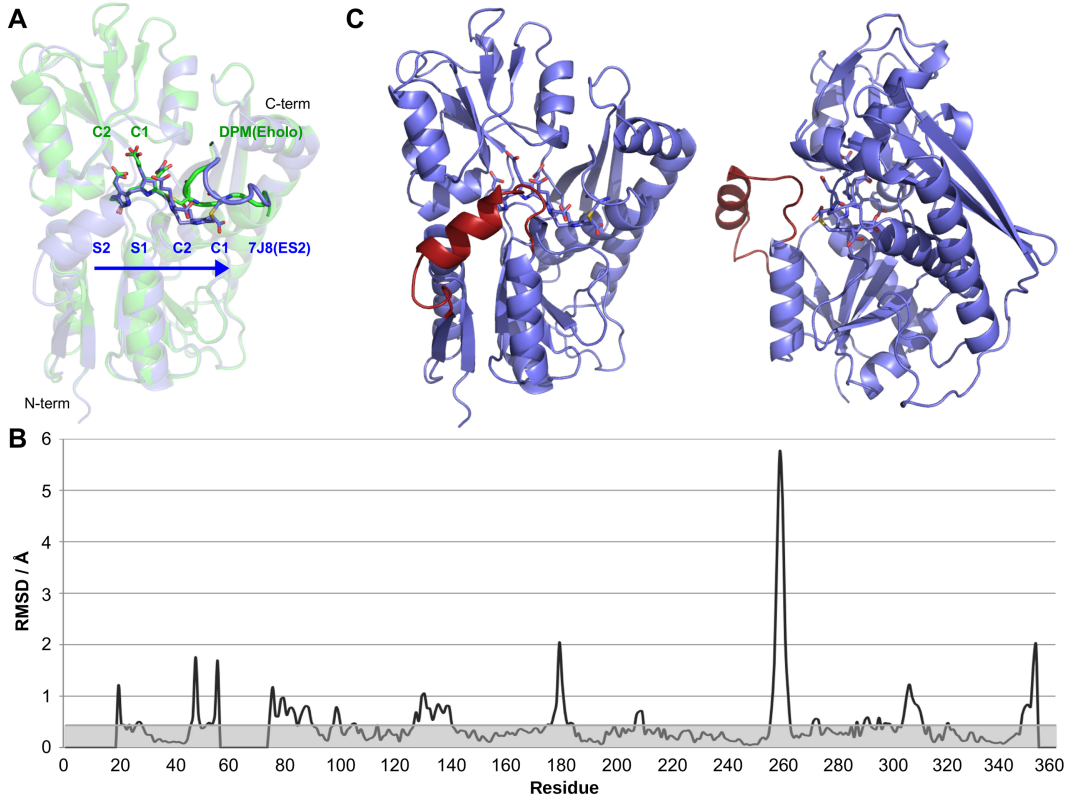The. (3ECR MD17ns)
Chi: 5.037

Log I

s / Å

C

^{15}N / ppm

^1H / ppm
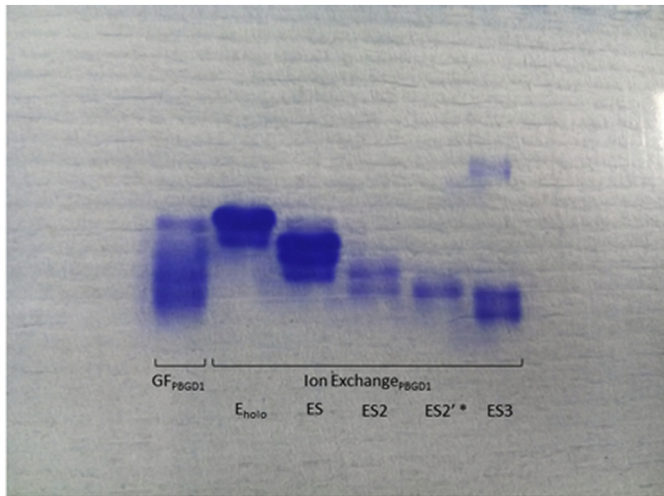
Figure 2

Figure 3

Figure 4

Figure 5