The FEBS Journal

FEBS PRESS science publishing by scientists

One ring closer to a closure: the crystal structure of the ES_3 hydroxymethylbilane synthase intermediate

Helene J. Bustad[1], Marthe S. Christie[1], Mikko Laitaoja[2], Aasne K. Aarsand[1,3], Aurora Martinez[4], Janne Jänis[2] and Juha P. Kallio[4]

1 Norwegian Porphyria Centre (NAPOS), Department for Medical Biochemistry and Pharmacology, Haukeland University Hospital, Bergen, Norway
2 Department of Chemistry, University of Eastern Finland, Joensuu, Finland
3 Norwegian Organization for Quality Improvement of Laboratory Examinations, Haraldsplass Deaconess Hospital, Bergen, Norway
4 Department of Biomedicine, University of Bergen, Norway

Keywords
acute intermittent porphyria; haem biosynthesis; hydroxymethylbilane synthase; porphobilinogen; pyrrole elongation

Correspondence
J. P. Kallio, Department of Biomedicine, University of Bergen, Jonas Lies vei 91, 5009 Bergen, Norway
Tel: +47 45848923
E-mail: juha.kallio@uib.no

Helene J. Bustad and Marthe S. Christie contributed equally to this article

(Received 7 July 2023, revised 8 September 2023, accepted 17 October 2023)

doi:10.1111/febs.16982

Hydroxymethylbilane synthase (HMBS), involved in haem biosynthesis, catalyses the head-to-tail coupling of four porphobilinogens (PBGs) via a dipyrromethane (DPM) cofactor. DPM is composed of two PBGs, and a hexapyrrole is built before the tetrapyrrolic 1-hydroxymethylbilane product is released. During this elongation, stable enzyme (E) intermediates are formed from the holoenzyme, with additional PBG substrates (S): ES, ES_2, ES_3 and ES_4. Native PAGE and mass spectrometry of the acute intermittent porphyria (AIP)-associated HMBS variant p.Arg167Gln demonstrated an increased amount of ES_3. Kinetic parameters indicated catalytic dysfunction, however, the product release was not entirely prevented. Isolation and crystal structure analysis of the ES_3 intermediate (PDB: 8PND) showed that a pentapyrrole was fully retained within the active site, revealing that polypyrrole elongation proceeds within the active site via a third interaction site, intermediate pyrrole site 3 (IPS3). The AIP-associated HMBS variant p.Arg195Cys, located on the opposite side to p.Arg167Gln in the active site, accumulated the ES_4 intermediate in the presence of excess PBG, implying that product hydrolysis was obstructed. Arg167 is thus involved in all elongation steps and is a determinant for the rate of enzyme catalysis, whereas Arg195 is important for releasing the product. Moreover, by substituting residues in the vicinity of IPS3, our results indicate that a fully retained hexapyrrole could be hydrolysed in a novel site in proximity of the IPS3.

Introduction

Haem is a tetrapyrrole essential to all life, and yet how the haem precursor and linear tetrapyrrole 1-hydroxymethylbilane (HMB) is formed is still not fully understood. HMB is the product of the third enzyme of the haem biosynthesis, hydroxymethylbilane synthase (HMBS; also known as porphobilinogen deaminase, EC2.5.1.61) and arises from the stepwise polymerisation of four pyrrole substrate units: porphobilinogen (PBG) molecules containing a pyrrole moiety [1,2]. The head-to-tail coupling of PBG molecules involves protonation

Abbreviations
Ac, acetate; AIP, acute intermittent porphyria; ALA, δ-aminolevulinic acid; ASU, asymmetric unit.*At*HMBS, *Arabidopsis thaliana* HMBS; CPS, constant pyrrole site; DPM, dipyrromethane; E, enzyme; E_{apo}, apoenzyme; ESI FT-ICR MS, electrospray ionisation Fourier transform ion cyclotron resonance mass spectrometry; hHMBS, human HMBS; HMB, 1-hydroxymethylbilane; HMBS, hydroxymethylbilane synthase; IEX, anion-exchange chromatography; IPS, intermediate pyrrole site; PBG, porphobilinogen; Pr, propionate; S, substrate; SES, substrate entry site.

The FEBS Journal **291** (2024) 510–526 © 2023 The Authors. *The FEBS Journal* published by John Wiley & Sons Ltd on behalf of Federation of European Biochemical Societies.

of the substrate before a nucleophilic attack with ammonia being cleaved off, followed by deprotonation of the newly formed elongation product. This occurs through several enzyme (E) intermediates, which are formed with additional pyrrole substrates (S): ES, ES_2, ES_3 and ES_4 [3,4]. The overall HMBS fold is a three-domain structure, with the interface between domains 1 and 2 constituting a cleft that serves as the active site [5,6]. The cofactor, dipyrromethane (DPM), is composed of two PBG molecules and is covalently bound to the enzyme via Cys261 on domain 3 and is located within the active site cleft [5–8]. In the holoenzyme of HMBS (E_{holo}), DPM serves as the template for the polymerisation of a hexapyrrole, and after the release of HMB by hydrolysis, the enzyme is regenerated with the cofactor intact [9]. The cofactor is either created by the apoenzyme from two PBG molecules, or via the ES_2 intermediate by direct incorporation of HMB in newly synthesised apoenzyme of HMBS [10].

So far, the X-ray crystal structures of human HMBS (hHMBS) E_{holo} (PDBs: 3ECR, 5M7F, 7AAJ, 7CCX, 3EQ1) and the enzyme intermediate ES_2 (PDBs: 5M6R, 7CCZ, 7AAK) are solved [6,11–14]. The hHMBS structure confirmed the high degree of fold similarity with the *Escherichia coli* structure already published in 1994 (PDB: 1PDA [5]).

Variants in the *HMBS* gene yielding HMBS with reduced enzymatic activity or incapability of product turnover are associated with the inherited disorder acute intermittent porphyria (AIP), an autosomal dominant inborn error of metabolism. AIP is characterised by acute neurovisceral attacks of severe abdominal pain and various neurological and psychiatric symptoms, typically triggered by haem-demanding environmental, hormonal, and metabolic factors [15–17]. These factors may induce an upregulation of the haem biosynthesis in the liver [15], and in *HMBS* gene variant carriers, HMBS becomes rate-limiting causing PBG accumulation. The haem precursor δ-aminolevulinic acid (ALA) accumulates concurrent with PBG as ALA dehydratase, the second enzyme involved in the haem biosynthesis, is inhibited by PBG [18]. ALA is considered likely to be causing symptoms by a neurotoxic effect, but much is still unknown both regarding disease mechanisms and factors that influence natural history and patient outcome.

Hydroxymethylbilane synthase is highly conserved across species, with most of the absolutely conserved residues located in and around the active site [6]. Arg167 is one of several highly conserved arginines in hHMBS and is located in the active site cleft where it has been identified as an important residue, suggested to have a key role in substrate binding and catalytic activity [19]. Results from molecular dynamics simulations lead to the proposal that Arg167 is an essential residue for proton donation when PBG molecules are bound [20]. Functional studies have demonstrated that the catalytic activity is severely affected, and the elongation mechanism is disturbed [14,19,21] when Arg167 is substituted with glutamine (c.500G>A; p.Arg167Gln). Furthermore, p.Arg167Gln in the E_{holo} state was the first disease-associated variant to be crystallised [14]. Unfortunately, the electron density for the substituted residue was missing, which was also the case in the other crystal structures of hHMBS, and the role of Arg167 could not be deduced. Another missense substitution reported at the same location, p.Arg167Trp, caused severely impaired enzyme activity [22–24], and, compared to p.Arg167Gln, it also demonstrated lower stability [25].

Arg195 is another conserved residue in the active site cleft in domain 2. Structural studies have shown that Arg195 interacts with the cofactor and variants at this residue are expected to reduce the enzymatic activity. The missense variant c.583C>T (p.Arg195Cys) occurs at a CpG dinucleotide hotspot and was described in 1995 as the first *de novo* variant in hHMBS [26]. This variant causes a markedly reduced enzymatic activity [21,27]. Another disease-related variant at the same location, p.Arg195His, has also been reported [28,29]. Furthermore, as demonstrated by site-directed mutagenesis, p.Arg195Ala is inactive [6].

Both Arg167 and Arg195 are implicated in the polypyrrole elongation by hHMBS. Arg167 possibly by interacting with the incoming substrate without direct contact with the polypyrrole [12]. Additionally, the corresponding arginine ($Arg161^{At}$) in *Arabidopsis thaliana* HMBS (*At*HMBS) interacts with the active-site loop in the crystal structure (PDB: 4HTG) [30]. Arg195, on the other hand, interacts directly with the pyrroles in the active site [6,11]. As introduced above, disease-associated variants of both residues have diminished catalytic activity, and yet their effect on the elongation mechanism is not fully understood. The decreased activity together with an indirect role of Arg167 makes the p.Arg167Gln variant an interesting target for studying the HMBS enzyme intermediates. p.Arg195Cys is also relevant for further investigations as the possible loss of the direct interaction between Arg195 and the formed polypyrrole in the active site may also affect the elongation and intermediate distribution of HMBS.

In this work, we present the crystal structure of the ES_3 intermediate for the HMBS p.Arg167Gln variant (PBD: 8PND), which reveals how the polypyrrole elongation proceeds within the enzyme active site via a

third, and previously undesignated pyrrole interaction site. By using a combination of electrospray ionisation Fourier transform ion cyclotron resonance mass spectrometry (ESI FT-ICR MS), native PAGE and enzyme activity assays, we have characterised the distribution of the enzyme intermediates of p.Arg167Gln and p.Arg195Cys variants. Our results indicate that Arg167 is a determinant for the rate of enzyme catalysis in all elongation steps and delays the product release, whereas Arg195 is mainly involved in the release of the HMB product. Further investigations of the residues Arg255, Glu258 and Arg355 in the vicinity of the new interaction site demonstrate that the HMB product might be hydrolysed off when the hexapyrrole is fully retained within the active site.

Results

The p.Arg167Gln enzyme intermediates

The recombinant p.Arg167Gln variant was purified in similar amounts to wild-type (wt) hHMBS. Preparative for further purification and crystallography attempts, the spectrum of p.Arg167Gln intermediate mixture was analysed using far-UV circular dichroism, indicating that it contained folded protein (Fig. S1). The thermostability of each individual enzyme intermediate of p.Arg167Gln analysed by differential scanning fluorimetry resembled the intermediates of wt-hHMBS (Table 1).

The distribution of the enzyme intermediates was confirmed using native ESI FT-ICR MS, with the peaks corresponding to the apoenzyme (E_{apo}), E_{holo}, ES, ES_2 and ES_3 (Fig. 1A). Furthermore, the amount of the ES_3 intermediate was increased as compared to the amount seen in wt-HMBS, which is in accordance with our previous findings for another variant at the same position, p.Arg167Trp [13]. Isolation of the enzyme intermediates by anion-exchange chromatography (IEX) demonstrated an elution profile similar to wt-hHMBS, and an increased amount of the peak corresponding to ES_3 was also noticed here for p.Arg167Gln (Fig. 1C). The isolated intermediates were analysed by native-PAGE, which confirmed well-separated intermediates (Fig. 1C; inset). The variant eluted later than wt-hHMBS in IEX, while it migrated faster in native-PAGE, both findings which can be attributed to the change from the positively charged arginine to the polar and uncharged glutamine, as has also been reported previously [19].

The catalytic cycle of p.Arg167Gln

The enzyme activity was assayed at standard conditions (reaction time 4 min at 37 °C), which confirmed the expected dysfunction of p.Arg167Gln [19], with very little remaining activity (0.6% relative to wt-hHMBS; Table 2). The steady-state kinetics at standard conditions also demonstrated severe catalytic defects for this variant. Although the kinetic constants could not be properly estimated from the fitting of the data to the Michaelis–Menten equation, the results indicated a very high apparent K_m and decreased V_{max} (Table 2, Fig. 2; blue circles), concurrent with lower affinity for the substrate and low k_{cat} for the p.Arg167Gln variant. Increasing the reaction time from 4 to 10 min, led to measurable amounts of product, however, mainly at higher substrate concentrations (Fig. 2; blue triangles). This suggests that the Arg-to-Gln substitution may result in a decreased catalytic constant by affecting all elongation steps and, notably, the last $ES_3 \rightarrow ES_4$ step, which becomes rate-limiting for the overall elongation reaction.

To investigate whether the product release of p.Arg167Gln was compromised, the variant was analysed with saturating concentrations of PBG using native-PAGE. As already shown in Fig. 1C; inset, p.Arg167Gln intermediates migrated in a pattern similar to wt-hHMBS. Upon incubation with five-fold molar excess of PBG the migration pattern remained the same (Fig. 3A, inset; Ctrl), however, the distribution of the intermediates was altered over time (Fig. 3A, inset): Quantification of the distribution within rows (Table 3) showed that after 1 h, the highest band (a) had a 1.5-fold increase and 1.8-fold after 18 h. The second band (b) rapidly decreased with four-fold after 1 h

Table 1. Melting temperatures (T_m) for enzyme intermediates of wt-HMBS and p.Arg167Gln determined by differential scanning fluorimetry.

	T_m (°C)[a]				
	E_{apo}	E_{holo}	ES	ES_2	ES_3
wt-HMBS	43.9 ± 0.3	79.2 ± 0.7	73.3 ± 0.7	72.3 ± 0.7	71.0 ± 0.2
p.Arg167Gln	–	78.8 ± 0.1	73.2 ± 0.0	73.0 ± 0.1	70.4 ± 0.0

[a] T_m-values are provided as mean ± SD.

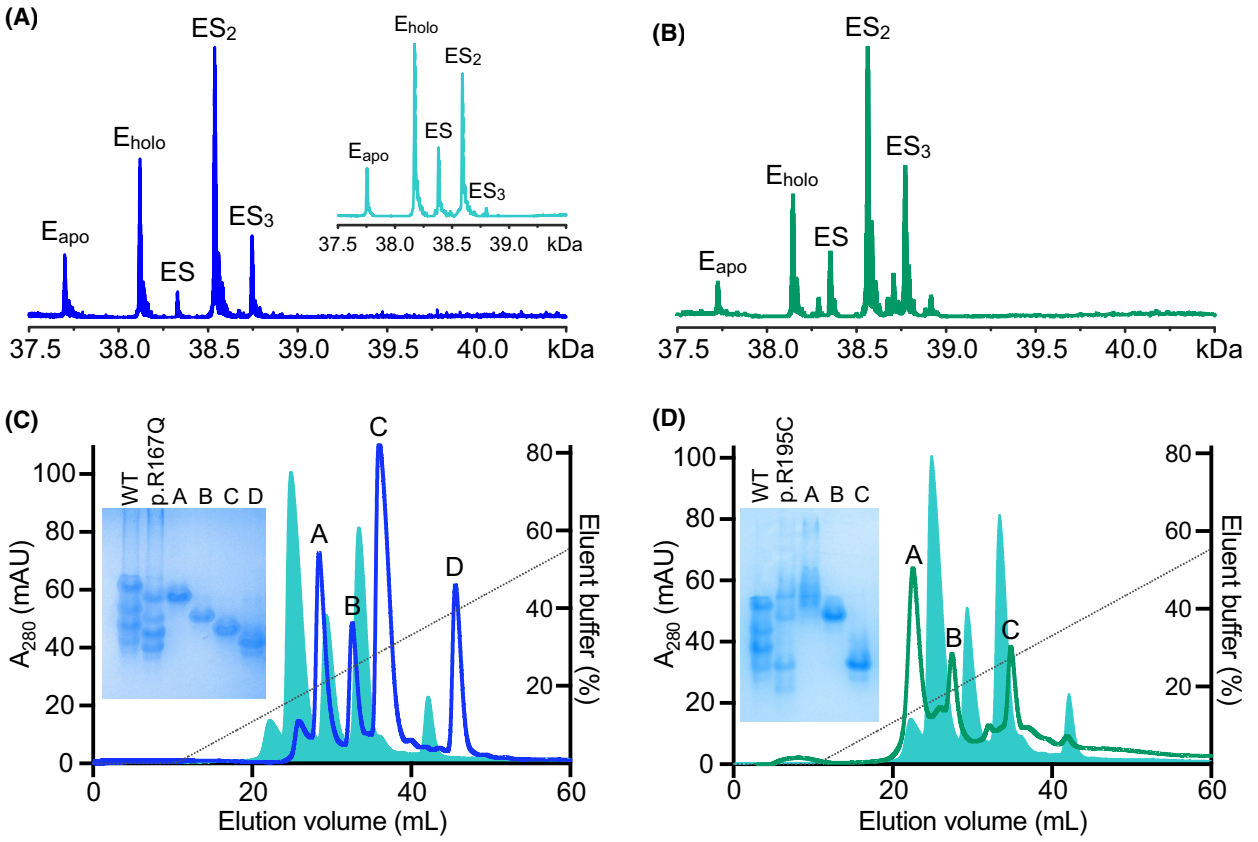

Fig. 1. Intermediates of HMBS variants p.Arg167Gln and p.Arg195Cys. Charge-deconvoluted ESI FT-ICR mass spectrum of (A) p.Arg167Gln and (B) p.Arg195Cys in denaturing conditions at 1 μM concentration. Wt-HMBS (aqua) as inset for reference. All peaks are denoted with respective intermediate. Enzyme intermediates are separated using anion exchange chromatography. Wt-HMBS (aqua) and variants (C) p.Arg167Gln (blue) and (D) p.Arg195Cys (green). The insets in C and D show native PAGE analysis of the well-separated fractions containing IEX peaks A, B, C or D of the variants (p.R167Q or p.R195C) after isolation and concentration, with wt-HMBS (WT) as control. The wells were loaded with ~ 3 μg protein per well. The annotated peaks do not correspond with the intermediate distribution seen in ESI FT-ICR mass spectrum, only to the isolated species from the solution. Representative mass spectra, chromatograms and native PAGE images are presented from at least two independent experiments (n = 2).

Table 2. Steady-state kinetic parameters of expressed HMBS variants. HMBS, hydroxymethylbilane synthase.

	Relative activity (%)	Specific activity[a] (nmol × h^{-1} × mg^{-1})	V_{max}[b] (nmol × h^{-1} × mg^{-1})	K_m[b] (μM)
wt-HMBS	100	2262 ± 40	2963 ± 17	27 ± 1
p.Arg167Gln	0.6	14 ± 0	—	—
p.Arg195Cys	—	n.d.	—	—
p.Arg255Ser	4	92 ± 11	664 ± 12	20 ± 2
p.Glu258Thr	27	602 ± 39	188 ± 17	130 ± 46
p.Arg355Ala	5	109 ± 30	565 ± 64	939 ± 302

[a]The specific activity of HMBS was defined as nmol of uroporphyrinogen I /h per mg of enzyme at 37 °C under the given assay conditions.; [b]Enzyme kinetic analyses were conducted using a range of 3–4000 μM PBG with 5 μg enzyme. Steady-state kinetic parameters *V*max and *K*m were analysed from three independent replicates (n = 3) ± SEM using the Michaelis–Menten enzyme kinetics model in GRAPHPAD PRISM.

to a minor contribution (1% of within row intensity) after 18 h. The two lowest bands indicated less change; however, the third band (c) decreased to 75% after 1 h and to 50% after 18 h. The last band (d) demonstrated a stable change with a 1.1 to 1.2-fold increase in intensity from 1 to 18 h incubation time, respectively.

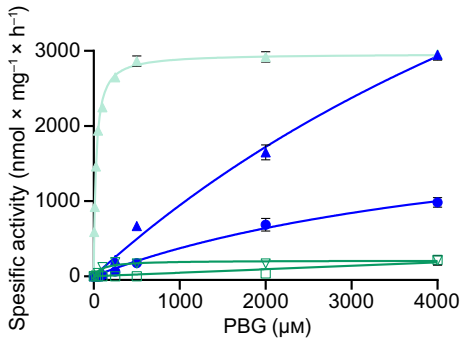

Fig. 2. Enzyme kinetics with different reaction times for HMBS variants. The Michael–Menten substrate-velocity plots for p.Arg167Gln at standard conditions (4 min, 37 °C; blue circles) and with prolonged reaction time (10 min, 37 °C; blue triangles), and p.Arg195Cys at standard conditions (4 min, 37 °C; green open squares) and with prolonged reaction time (10 min, 37 °C; green open diamond triangles). Results for wt-HMBS at standard conditions are shown as reference (aqua triangles). Activity is expressed as nmol of Uroporphyrinogen I produced per mg of HMBS per h. The mean values from three independent replicates ($n = 3$) are plotted with error bars representing SEM.

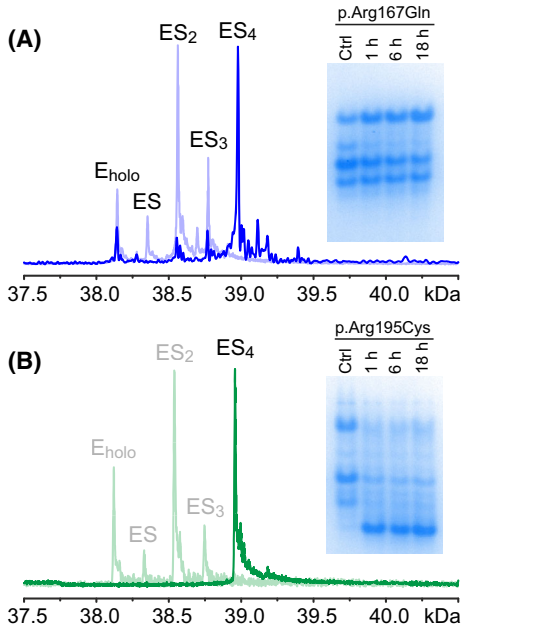

Fig. 3. ESI FT-ICR mass spectra of HMBS variants p.Arg167Gln and p.Arg195Cys as purified and with substrate excess. (A) Charge-deconvoluted ESI FT-ICR mass spectrum of p.Arg167Gln as purified in native conditions (light blue) and after 1 h incubation with 10-fold molar excess of PBG (blue). All peaks are denoted with respective intermediate. Inset: Native PAGE of p.Arg167Gln incubated with five-fold molar excess of PBG (130 μM) at three different time points. Lane 1; as purified (Ctrl), lanes 2–4; 1, 6 and 18 h incubation, respectively. (B) Charge-deconvoluted ESI FT-ICR mass spectrum of p.Arg195Cys as purified in native conditions (light green) and after 1 h incubation with 10-fold molar excess of PBG (green). All peaks are denoted with respective intermediate. Inset: Native PAGE of p.Arg195Cys incubated with five-fold molar excess of PBG at three different time points. Lane 1; as purified (Ctrl), lanes 2–4; 1, 6 and 18 h incubation, respectively. Representative mass spectra and native PAGE images are presented from at least two independent experiments ($n = 2$).

Analysing the reaction after 1 h incubation with 10-fold molar excess of PBG using native ESI FT-ICR MS confirmed that all the intermediates were present as seen in native PAGE (Fig. 3A). The distribution was, however, different from that seen in the native PAGE, as the ES_4 intermediate was also detected. With this method, the fast time-course does not allow for a complete product release; however, product was also detected. On the other hand, the time (2.5 h) and heat generated during electrophoresis probably facilitate activity and continuous product release despite reduced residual activity, which prevent the detection of ES_4 with native PAGE.

Crystal structure of the HMBS ES_3 intermediate

The small changes in the surface properties caused by stepwise addition of PBG in the HMBS mechanism allow for the isolation of the reaction intermediates using IEX, and the chromatography of p.Arg167Gln yielded well-separated peaks (Fig. 1C). Peak D, mainly containing one single intermediate as confirmed by native PAGE (Fig. 1C, inset), was isolated and collected for crystallography. Crystals were successfully grown in 15–17.5% polyethylene glycol 3350, 100 mM ammonium citrate at pH 5.5.

The crystal structure was determined to 1.9 Å resolution (Table 4, PDB: 8PND). The protein crystallised in space group $P22_12_1$ having two molecules in the asymmetric unit (ASU), with the biological assembly clearly being monomeric, as expected. This was confirmed with PDBePISA, finding no specific interactions that could result in the formation of stable quaternary structures. Furthermore, no oligomeric

Table 3. Quantification of native PAGE of HMBS variants incubated with five-fold molar excess of PBG.

	Intensity (%)							
	p.Arg167Gln				p.Arg195Cys			
Band	Ctrl	1 h	6 h	18 h	Ctrl	1 h	6 h	18 h
(a)	34	51	54	61	45	25	24	21
(b)	12	3	2	1	3	2	2	2
(c)	37	28	25	18	28	11	8	8
(d)	17	18	19	20	18	7	11	10
(e)	—	—	—	—	6	55	56	59

Table 4. Data collection and refinement statistics. Statistics for the highest-resolution shell are shown in parentheses.

Resolution range	41.12–1.9 (1.968–1.9)
Space group	$P22_12_1$
Unit cell	53.6 81.136 192.347 90 90 90
Total reflections/unique reflections	500 825/66 997
Multiplicity	7.5 (7.7)
Completeness (%)	99.80 (99.95)
Mean I/sigma (I)	22.50 (1.36)
R-merge	0.04154 (1.492)
R-pim	0.01623 (0.5688)
CC1/2	1 (0.627)
Refinement statistics	
R_{work}/R_{free}	18.6/21.7
RMSD of bond length (Å)	0.007
RMSD of bond angles (°)	0.77
Ramachandran statistics	
Ramachandran favoured (%)	96.14
Ramachandran allowed (%)	3.55
Ramachandran outliers (%)	0.31
Average B-factors	
Protein	70.46
Ligands	66.09
Water	59.59
Number of non-hydrogen atoms	
Protein	5088
Ligands	188
Water	272

proteins were detected with native ESI FT-ICR MS measurements conducted earlier. Both monomers explicitly revealed the presence of the ES_3 intermediate, enabling its crystallographic observation for the first time. A polder map was calculated to confirm an unbiased presence of the pentapyrrole (Fig. 4A).

The two monomers in the structure are practically identical, albeit the quality of the electron density map for chain B is lower than for chain A. More concretely, the region around the human-specific loop between residues Ile299 and Ile323 is poorly defined in chain B. Crystal packing may explain this, as chain A of ASU is surrounded by symmetry-related molecules, whereas chain B is more exposed to solvent, possibly allowing slight movement within the structure.

The architecture of the HMBS ES_3 intermediate

The HMBS protein structure is organised by three distinct domains, and the previous crystal structures agree well with each other, showing little differences. When superimposing C_{alpha}-atoms for wt-E_{holo} (PDB: 7AAJ) to ES_2 (PDB: 7AAK) and ES_3 (PDB: 8PND), the RMSD values are 0.34 and 0.29 Å, respectively. The RMSD analysis does not consider the flexible active-site loop, which for hHMBS is only defined in the crystal structures of the ES_2 intermediate (PDBs: 7AAK and 5M6R); however, it indicates enlargement of the active site cavity without major changes in the overall domain structure during the catalysis (Fig. 4B).

The structures of the ES_2 intermediate demonstrate that the cofactor with pyrrole rings C1 and C2 (Fig. 4C) relocate from the initial sites in E_{holo}, herein referred to as constant pyrrole sites 1 and 2 (CPS1 and CPS2; Fig. 4C), into two new interaction sites, denoted as intermediate pyrrole sites 1 and 2 (IPS1 and IPS2; Fig. 4C). The growing polypyrrole chain is thus pulled through the active site cleft, and the PBGs S1 and S2 (Fig. 4C) are positioned in CPS1 and CPS2 in ES_2. Moreover, for the ES_3 intermediate presented here, the incoming PBG (S3) occupies CPS2 creating a novel intermediate pyrrole site (IPS3; Fig. 4C). This occurs by elegant localised conformational adjustments where the cofactor-binding loop ^{258}EGGCSV^{263} rearranges from the original conformation: during pyrrole elongation from E_{holo} to ES_2 (via the uncharacterised ES), the cofactor-binding loop retracts ca. 4.5 Å making space for the growing polypyrrole (Fig. 5A). In ES_3, the conformational change consequently allows the Cys261 side chain to achieve a new orientation, with translocation of Cys261 Cα being approximately 2.5 Å (Fig. 5B).

The twisting of the cofactor-binding loop also induces a rearrangement of the C-terminal helix $\alpha3_3$ upon which the helix is slightly shifted in ES_3 compared to E_{holo} and ES_2 (Fig. 5C). The adopted helix conformation is stabilised by Arg355 hydrogen-bonding to Glu258 and Gly259 main chain carbonyls. Gly259 and Glu258 provide van der Waals interactions in the vicinity of the pyrrole acetate (Ac) and propionate (Pr) side chains, $C2_{Ac}$ and $C1_{Pr}$ respectively, whereas Ser262 creates a hydrogen bond to $C1_{Ac}$. In this position, $C1_{Pr}$ side chain is solvent accessible, with less stabilising interactions than in the previous positions for cofactor and substrate interactions.

The entire active-site loop (residues 57–75) is not defined by the electron density but both ends of the loop, residues 57–59 and 72–75, clearly adopt a different conformation compared to known hHMBS-ES_2 structures showing the ordered active-site loop. The C-terminal end of the loop is in ES_3 now covering pyrrole C1 at the novel IPS3. This conformation also requires partial unfolding of the N-terminal end of the $\alpha3_1$ helix involving residues 75–79 and mostly resembles the conformation only observed in the *At*HMBS structure (PDB: 4HTG) where the active-site loop is in closed conformation (Fig. 6A). In *At*HMBS, the active-site loop is stabilised by hydrogen bonding between Arg161^{At} side chain and main chain carbonyls of Gln62^{At}, Ile59^{At} and Gly56^{At} (Fig. 6B). This part of

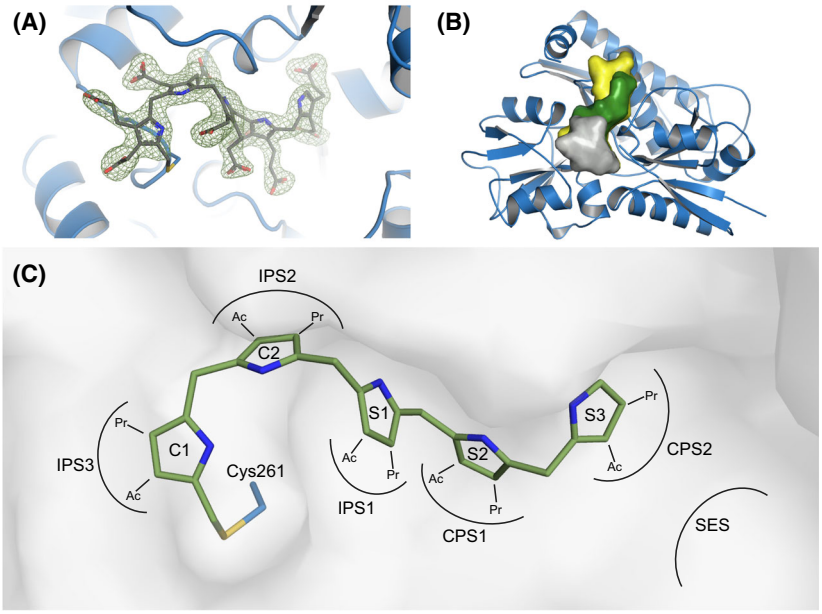

Fig. 4. The pentapyrrole in the active site of hydroxymethylbilane synthase (HMBS) ES_3 intermediate. (A) Polder omit-map (green mesh) was calculated to confirm the presence of the pentapyrrole without model bias that would be present in the conventional 2Fo-Fc density. The map is presented on level 5 s. (B) The active site cavity in the HMBS structures enlarges through the elongation. The cavity surface of E_{holo} (PDB: 7AAJ, wt-HMBS; grey), ES_2 (PDB: 7AAK, p.Arg173Trp; green), and ES_3 (PDB: 8PND), p.Arg167Gln; yellow). (C) Semi-schematic presentation of the pyrrole elongation presenting the intermediate binding site terminology in the ES_3 enzyme intermediate of p.Arg167Gln. C1 and C2 denote the two PBG pyrrole molecules constituting the DPM cofactor, where C1 is always bound to Cys261 in the cofactor-binding loop. S1, S2 and S3 denote the substrate PBG pyrroles elongated from the cofactor. The Constant Pyrrole Sites 1 and 2 (CPS1 and CPS2) are the sites where C1 and C2 sit in E_{holo}, and where the incoming PBG molecules are elongated through. The Intermediate Pyrrole Sites (IPS) 1, 2 and 3 are the sites created for the pyrroles through elongation. The Substrate Entry Site (SES) is the positioning site for the incoming pyrrole substrate [12]. The pyrroles are presented without PBG side chains. Ac, acetate and Pr, propionate represent side chains. The illustrations for the specified structures were created in PyMOL.

the active-site loop is not described in the electron density map of the p.Arg167Gln ES_3 structure presented here, making it difficult to predict whether Gln167 is able to interact with the loop.

Pentapyrrole interactions with the enzyme

Most interactions between the pentapyrrole and the enzyme in CPS1 and CPS2 were similar to both the E_{holo} and ES_2-states, as well as the interactions in IPS1 and IPS2 in the ES_2-state. However, in our p.Arg167Gln structure in the novel ES_3-state, the S1 pyrrole in IPS1 is tilted approximately 35° compared with the conformation observed for C2 in the ES_2-state (PDB: 7AAK) (Fig. 7A). Despite new pyrrole conformation in IPS1, the interactions are similar to the ones observed before for ES_2 [11–13]. Only the reorientation of the active-site loop described above leads to the loss of interactions between $S1_{Pr}$ and Lys74, and to the formation of hydrogen bond between the $S1_{Ac}$ and main chain of Leu76, which has not been previously seen to interact with the polypyrrole (Fig. 7B). The C2-pyrrole in IPS2 clearly forms unseen interactions in the ES_3-structure. C2 is stabilised through hydrogen bonding between $C2_{Ac}$ to Gly260 (main chain) in addition to previously observed hydrogen bonding to Thr102. Another new interaction is between main chain carbonyl of Lys74 and pyrrole nitrogen (Fig. 7B).

The novel IPS3 site now accommodates the C1 pyrrole, which is forming hydrogen bonds to Ser262 (both side chain and main chain) and to Cys261 main chain via $C1_{Ac}$. The C1 ring interacts with Lys74 through π-alkyl interactions but no clear interactions are made via $C1_{Pr}$ side chain. $C1_{Pr}$ is located between Glu258 and Lys74 with side chains being partially surface exposed. Arg355 appears as a new attractive charge interaction for $C1_{Ac}$ and in different conformations could be involved in the hydrogen bonding network (Fig. 7B).

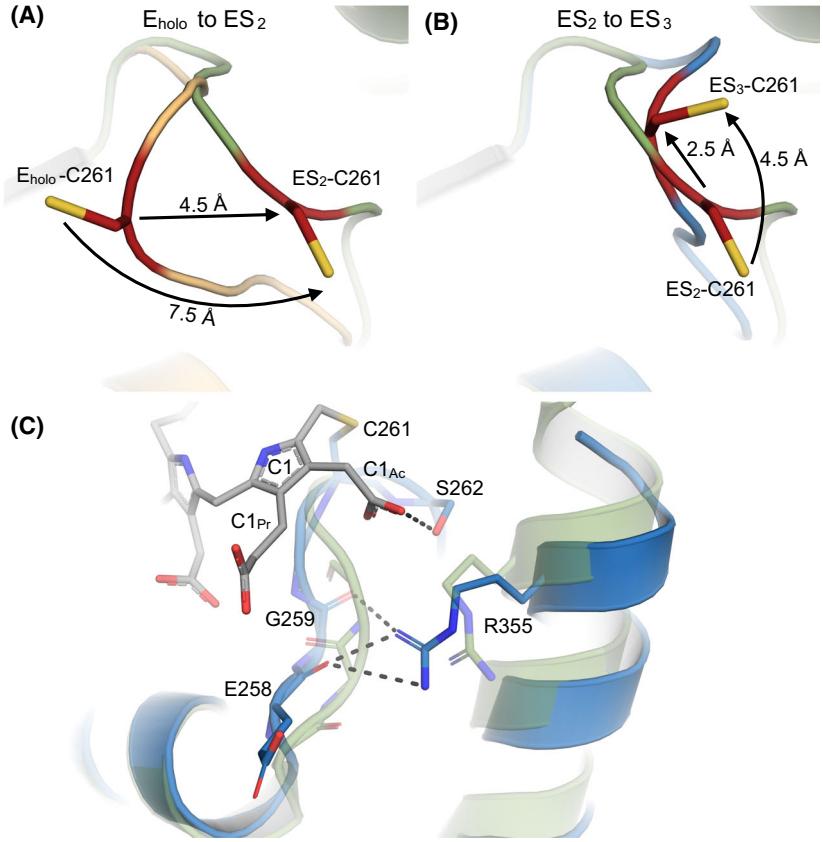

Fig. 5. The rearrangement of the cofactor-binding loop in ES_3. Comparison of the cofactor-binding loop in E_{holo} (yellow, PDB: 7AAJ), ES_2 (green, PDB: 7AAK) and ES_3 (blue, PDB: 8PND) intermediates. (A) Rearrangement from E_{holo} to ES_2 happens by a retraction of the entire loop. (B) From ES_2 to ES_3, the loop is twisted, and Cys261 side chain conformation changes and points in the opposite direction compared with E_{holo}. (C) Reorientation of the cofactor-binding loop induces new conformation to the C-terminal helix. Arg355 forms stabilising interactions with residues Asp258 and Gly259 in the cofactor-binding loop and is in distance to also interact with $C1_{Ac}$. The illustrations for the specified structures were created in PyMOL.

The role of p.Arg195Cys in the elongation

In light of the inefficient catalytic activity of p.Arg167Gln, we sought to elucidate the mechanism of HMBS further by looking at Arg195 located on the opposite side of the active-site cleft. Arg195 is a well-conserved residue implicated in the elongation mechanism [6], and p.Arg195Cys is a disease-associated variant known to be catalytically defect [21,27]. ESI FT-ICR MS analysis in native conditions confirmed the presence of E_{holo}, ES, ES_2 and ES_3 (Fig. 1B). Native-PAGE showed that p.Arg195Cys migrated as mainly four bands (Fig. 1D, inset), corresponding to the ones in wt-HMBS and p.Arg167Gln.

Interestingly, when p.Arg195Cys was incubated with excess PBG, only the ES_4 intermediate accumulated (Fig. 3B). This was also seen by native-PAGE, with mainly one distinct band migrating further than without PBG, concurrent with the additional charge from PBG (Fig. 3B, inset). The p.Arg195Cys variant did not show activity at standard conditions (4 min, 37 °C; Fig. 2, green open circles), which is in line with the previously reported relative activity for the variant (0–3% of wt activity *in vitro*) [21,27]. A small amount of product formation was detected when the reaction time was increased to 10 min (Fig. 2, green open diamonds); however, the activity was not at the level obtained for p.Arg167Gln with the same conditions. Traces of product were also detected by MS for the p.Arg195Cys (data not shown). This could be a result of co-purification of HMB with the enzyme.

Prediction of a novel site for hydrolysis

The crystal structure solved in this study provides the conformation of the ES_3 intermediate where we can observe that the C1 and C2 pyrroles start to rotate around the covalent bond to Cys261. The analysis of the structure opens the possibility of how the product is released other than via Asp99 where it might be envisioned that the hydrolysis happens with the hexapyrrole fully retained within the active site, with the two last incoming PBGs positioned in CPS1 and CPS2. A new site for hydrolysis would be needed; however, such a site is already indicated in the ES_3 intermediate structure between IPS2 and IPS3. By examining the crystal structure for amino acids with properties appropriate for hydrolysis, we found Arg255, Glu258 and Arg355 (Fig. 7B) in the proximity

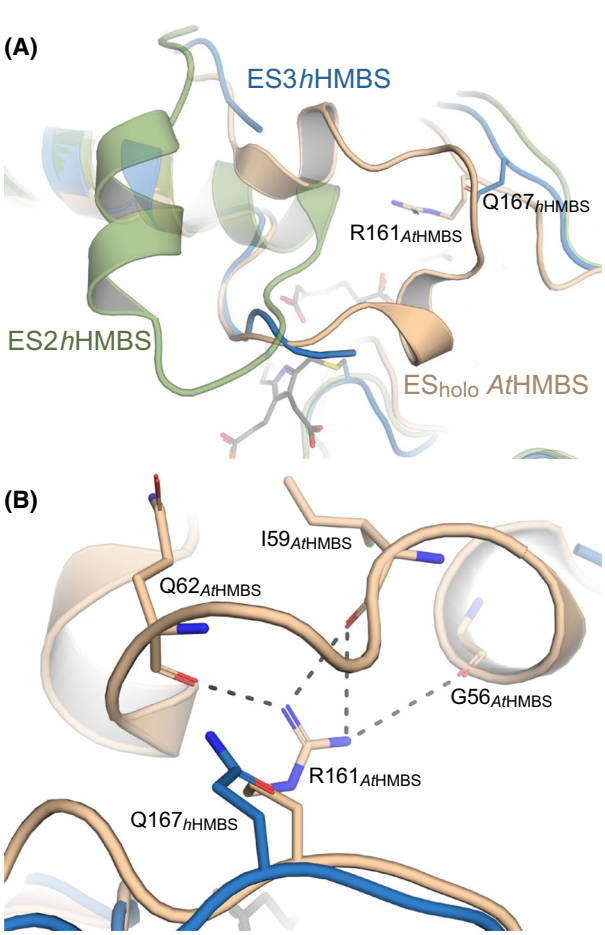

Fig. 6. The active-site loop in human hydroxymethylbilane synthase (hHMBS) p.Arg167Gln ES_3 intermediate resembles *Arabidopsis thaliana* (*At*HMBS). (A) Comparison of the active-site loop in *At*HMBS in ES_{holo}-state (beige, PDB: 4HTG), hHMBS in ES_2-state (green, PDB: 7AAK) and in ES_3-state (blue, PBB: 8PND) presented in this study. (B) Previously undescribed interactions between active site loop and Arg161 in *At*HMBS (corresponding to Arg167 in hHMBS). The illustrations for the specified structures were created in PyMOL.

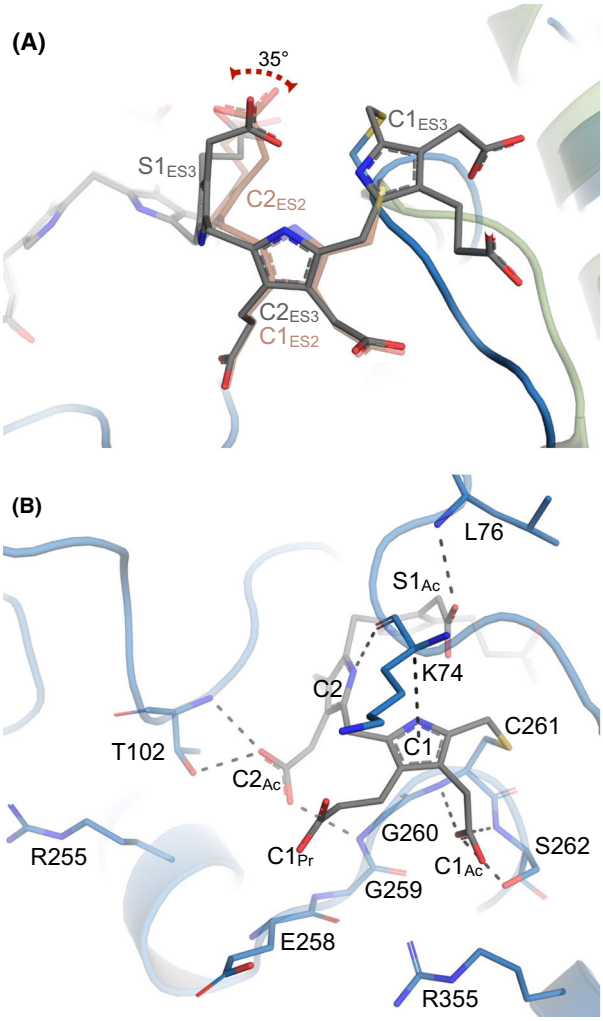

Fig. 7. Orientation of the ES_3-state polypyrrole and the description of the novel C1 interactions in IPS3. (A) The S1 pyrrole in ES_3 ($S1_{ES3}$, grey) tilts 35° compared to C2 in ES_2 ($C2_{ES2}$, brown) to enable the newly formed torsion angles in the growing polypyrrole. The hHMBS in the ES_2-state is in green (PDB: 7AAK) and the ES_3-state is in blue (PDB: 8PND). (B) In IPS3, the C1 pyrrole in in the ES_3 intermediate (C1, grey) forms a stabilising interaction via the $C1_{Ac}$ side chain to Cys261 (main chain) and Ser262 (both main chain and side chain), whereas the $C1_{Pr}$ is surface exposed. The $S1_{Ac}$ side chain forms a new interaction to Leu76 while the previously reported $S1_{Pr}$ interaction to Lys74 is lost. Interactions between the ES_3 polypyrrole and protein are marked with dashed lines. Ac, acetate and Pr, propionate in the PBG side chains. The illustrations for the specified structures were created in PyMOL.

of the putative hydrolysis site. Arg255, Glu258 and Arg355 are all conserved in vertebrates; moreover, Glu258 is semi-conserved (Glu/Asp) among eukaryotes. In humans, there are reported AIP-associated variants for all; however, in both population databases and in disease-associated reports, the significance of these is unclear [21,31,32].

We selected substitution residues avoiding similar properties yet non-deleterious, expressed and purified variants p.Arg255Ser, p.Glu258Thr and p.Arg355Ala, at similar yields as wt-HMBS. The enzymatic activity relative to wt was 4%, 27% and 5%, respectively, at standard conditions (Table 2). The apparent K_m for p.Arg255Ser was in the range of that obtained for wt-HMBS; however, a four-fold reduction in V_{max} was measured. The p.Glu258Thr variant presented a five-fold increased apparent K_m, and a 16-fold decrease in V_{max}, and for p.Arg355Ala, a 35-fold increase in apparent K_m and five-fold decrease in V_{max} were

measured. Thus, all variants presented catalytic dysfunction, including a remarkable defect in apparent affinity for the substrate for p.Glu258Thr and p.Arg355Ala. Analysis with native ESI FT-ICR MS detected varying amounts of enzyme intermediates (from E_{holo} to ES_4) for all the variants (Fig. 8, lighter colours). Upon incubation of these variants with a 10-fold molar excess of PBG, unexpected enzyme-intermediate patterns were observed, where all the variants yielded covalent intermediates apparently exceeding the number of pyrrole units in ES_4 (Fig. 8, darker colours). This is a surprising observation as enzyme intermediates beyond ES_4 have not been detected for any other HMBS variant. For all the variants, intermediates up to ES_{10} were observed. This result suggests that residues Arg255, Glu258 and Arg355 are involved in the product release (hydrolysis) and their substitutions allow over-elongation of the oligopyrrole residing in the active site, especially when the PBG substrate is present in excess amounts.

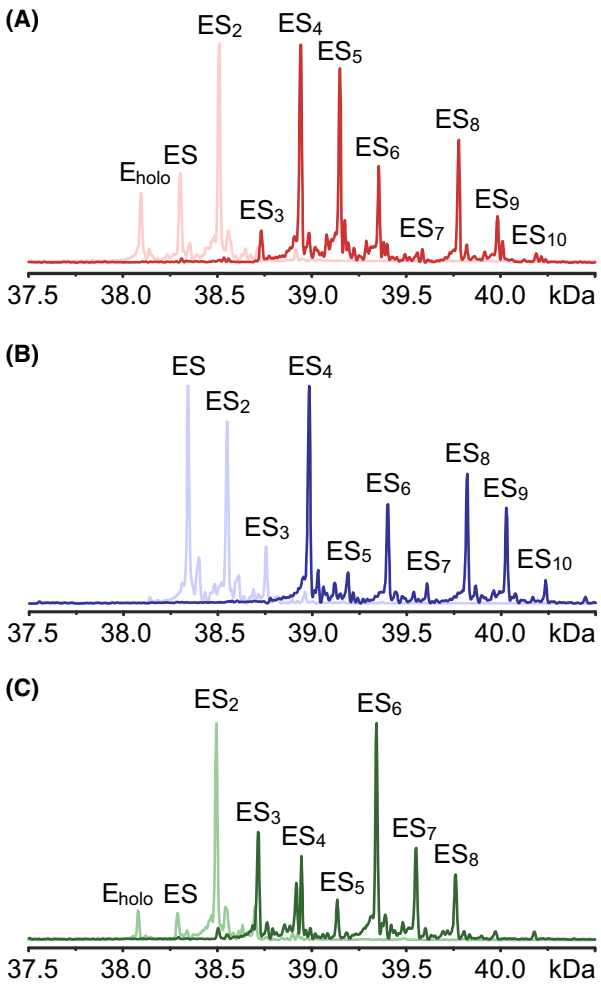

Fig. 8. Hydroxymethylbilane synthase variants in the vicinity of intermediate pyrrole site 3 over-elongate with excess PBG. The HMBS variants (A) p.Arg255Ser, (B) p.Glu258Thr, and (C) p.Arg355Ala were analysed with ESI FT-ICR MS directly as purified in native conditions (lighter colour) and after 1 h incubation with 10-fold molar excess of PBG (darker colour). All peaks are annotated with the respective enzyme intermediate. Representative mass spectra are presented from at least two independent experiments ($n = 2$).

Discussion

To understand the elongation process of the polypyrrole substrate by HMBS, crystal structures and careful characterisation of variants have been essential. Despite a structure with three well-defined domains connected through hinges and the obvious hypothesis involving rearrangement of the domains to accommodate pyrrole elongation [5,6,20], the structures of HMBS reaction intermediate ES_2 revealed very little overall movement compared to E_{holo} [11–13]. Instead, the cofactor-binding loop moved back and inwards within the enzyme core pulling the C1-C2 pyrroles stepwise into two new interaction sites, IPS2 and IPS1, respectively, with CPS1 and CPS2 available for the consecutively incoming S1 and S2 pyrroles (Fig. 4C and Fig. 9). This was possible with a repositioning of the cofactor-loop without significant domain movements or rearrangements of the active site architecture; the rear sub-cavity behind was large enough to accommodate both C1 and C2. Previous hypotheses, where the pyrrole elongation was assumed to occur from a statically positioned cofactor either by curling up inside the active site cleft or as a more stretched out polypyrrole before exiting [20], were thus disproved.

Our previous work on the crystal structure of the p.Arg173Trp variant demonstrated the potential of using variants to obtain snapshots of reaction intermediates, in that case the ES_2 [13], however, risking that the variant conformation could be forced in an unnatural conformation. This is also applicable for the p.Arg167Gln ES_3 structure presented here; nonetheless, the justification for using substitution variants has recently been discussed in a review by Helliwell [33]. The structural homology showed remarkably consistent least-square values of superimposed models of the known wt-HMBS and variant ES_2 structures and supports that the overall structure of the variant is correct. It remains to elucidate whether the tetrapyrrole of p.Arg173Trp is a product of the direct reaction of HMB with apo-HMBS, or if the variant has activity up to the point of ES_2 but not further. In contrast to p.Arg173Trp, amino acid substitution of Arg167 does

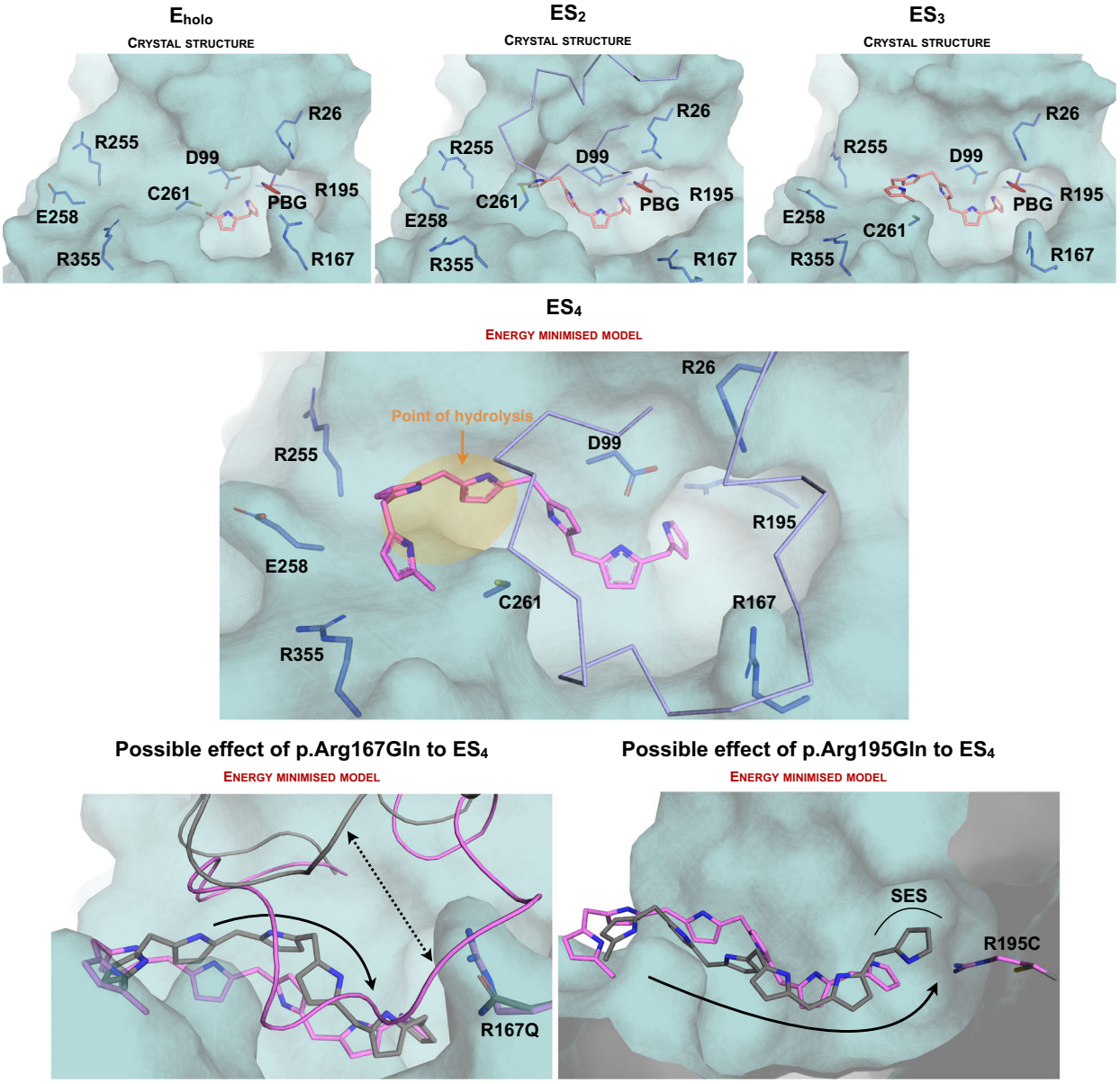

Fig. 9. Schematic presentation of known intermediate structures and suggested model for the final elongation step before product hydrolysis. The growing polypyrrole chain in light brown (side chains removed) in intermediate states based on the known crystal structures; E_{holo} wt-HMBS; PDB: 7AAJ, ES_2 p.Arg173Trp; PDB: 7AAK, and ES_3 p.Arg167Gln; PDB: 8PND (top panels). Incoming substrate is presented in brown. The energy-minimised modelled polypyrrole in the ES_4-state (pink) is shown in the middle panel. The active-site loop is presented as blue ribbon to show the observed conformation in the ES_2 structure and hypothetical conformation in ES_4. The ES_4 model indicates a novel site for product hydrolysis (highlighted in orange) outside the traditional active site. The active-site loop is needed to maintain the correct conformation of the hexapyrrole. The possible effects of the two characterised variants have been presented as hypothetical models (bottom panels). Here, the polypyrrole in the ES_4 model is shown in pink with the possible outcome for the polypyrrole caused by the substitution variant in grey. The p.Arg167Gln variant causes the active-site loop to lose the stabilising hydrogen bonding to the residue, which may allow the loop to have a higher degree of flexibility and thus altered orientations (dashed arrow). This could allow the hexapyrrole to bulge up from the optimal position (black arrow). The active-site loop could be needed to stabilise the last transfer of pyrroles to create the IPS4. p.Arg195Cys introduces the loss of hydrogen bonding between the S4 propionate side chain and the arginine. This could cause S4, the last added PBG, to remain in SES, holding the rest of the pyrroles back (black arrows), and thus not reaching the point of hydrolysis and creating a state where the ability for product release is lost. The illustrations for the specified structures were created in PyMOL.

not cause a block in the elongation, keeping a distribution of the intermediates albeit with extremely slow product turnover [14,24]. This is also demonstrated herein where our results indicate that time and temperature, as detected during electrophoresis, can compensate for the slow kinetics (Table 2, Figs 1 and 2). In this work, we have taken advantage of the higher relative amounts of expressed p.Arg167Gln ES_3 intermediate, compared to both wt-HMBS and p.Arg167Trp, also implying a distortion in the overall elongation kinetics. Furthermore, although the kinetic parameters are significantly altered in p.Arg167Gln (Table 2), the PBG saturation (Fig. 3A, inset) indicates that Arg167 is not crucial for product release, albeit tetrapyrrole release is slowed down. On the contrary, Arg195 proves to play an important role in the release of HMB as ES_4 accumulates when Arg195 is substituted for a cysteine.

The role of the active-site loop in HMBS catalysis is still unclear. Recently, Sato *et al.* [12] proposed that this loop in not only involved in intermediate stabilisation via basic residues (Lys70 and Lys74) but also in recruitment of the incoming PBG substrates. The Arg167 interactions with the active-site loop could be important if the loop is needed in a closed conformation for stabilisation of the present intermediate state. As mentioned, $Arg161^{At}$ in *At*HMBS equals Arg167 in hHMBS. The crystal structure (PDB: 4HTG) indicates interactions between $Arg161^{At}$ and backbone carbonyl oxygens of $Gly56^{At}$ (Gly60 in hHMBS), $Ile59^{At}$ (Ile63 in hHMBS) and $Gln62^{At}$ (Tyr66 in hHMBS) (Fig. 6B). The latter is part of a non-conserved region in the active-site loop (^{61}SQPLAD^{66} in *At*HMBS and ^{65}DTALSK^{70} in hHMBS); however, the interactions might only be dependent on the backbone oxygens and not the side chains. Neither Ile63 nor Tyr66 are known to have natural variants. Nevertheless, the Arg167 to Gln substitution in the structure seems to prevent these loop-stabilising interactions. Although there is no electron density to describe the entire loop, the C- and N-terminal sections of the active site loop can be modelled and those match the ones in *At*HMBS. $Arg161^{At}$ also interacts with the two oxygens in $C2_{Pr}$ through NE and NH_2; however, this is only possible because of the state of the cofactor in *At*HMBS where DPM is oxidised to dipyrromethenone. The interactions in CPS2 are then lost [30], and a closed active-site loop may possibly compensate for the lost CPS2 interactions. In the case of a growing polypeptide, the formed torsional angles might cause tension in the intermediates that could require the active-site loop to close, thus maintaining a correct conformation for the final step in the elongation and subsequently product release.

Arg167 and Arg195 are both highly conserved residues and have undoubtedly important roles in the elongation and release mechanism of HMBS. In this work, we have aimed to shed light on the two roles, and how they differ from each other, with proposed mode of actions. The active-site loop has been observed in a closed (PDB: 4HTG) and open conformation (PDBs: 5M6R and 7AAK). Arg167 is linked to this active-site loop dynamics, where the side chain interactions probably aid in stabilising hydrogen bonding between Arg167 and the loop, and the changes in the catalytic rate of HMB production in p.Arg167Gln are possibly because of the lacking loop-interactions. These would normally function as a guardian by restricting the degree of freedom for the growing polypoyrrole, which undergoes increasing torsion angle constraints beyond ES_2. The hypothetical scenario is presented as an energy-minimised model in the ES_4-state (Fig. 9). Our experimental data demonstrates a variant which affects all elongation steps, including a decreased rate of product release. This supports the hypothesis of a polypyrrole in a suboptimal location for hydrolysis; however, the capability to product turnover is not completely lost, possibly because of the dynamic nature of the active-site loop.

On the other hand, our results indicate that the role of Arg195 is not mainly in the substrate addition, but rather in the final step of elongation and product release, where it seems to be a crucial residue. When Arg195 is substituted for cysteine, the enzyme is incapable of executing this task. This could be caused by (a) obstructing a hydrolysis between pyrroles in IPS1 and CPS2 or (b) the loss of important interactions in CPS2. In the case of (a), this mode of action includes positioning and reaction of the S4 pyrrole in CPS2, upon which steric restraints facilitate relaxation of the structural conformation and hydrolysis of the hexapyrrole by Asp99, which is proposed to be in reactive distance [11]. For (b) on the other hand, if the hydrolysis takes place with a hexapyrrole fully retained in the active site (Fig. 9), we speculate that the final S4 ring may be hindered from moving from SES to CPS2, and thus obstructed from the correct position for hydrolysis, exemplified in Fig. 9. Based on our results, the capability of product release is the main difference between p.Arg167Gln and p.Arg195Cys, demonstrated by native PAGE with excess PBG, where p.Arg167Gln, despite reduced residual activity, is capable of continuous product release during the electrophoresis, whereas p.Arg195Cys accumulates ES_4

intermediate. The overall role of Arg195 is thus likely to accommodate the upcoming pyrrole into the CPS2.

The crystal structure of hHMBS in the ES_3 intermediate state provides novel information on the presence of an interaction site for the C1-pyrrole in HMBS. Furthermore, our results lead to the hypothesis concerning the site of hydrolysis and release mechanism of HMB other than previously predicted [19,34,35]. Traditionally, Asp99 is thought to play an important role in the pyrrole hydrolysis; however, for this to be possible, the hexapyrrole needs to slide through the active site and place C1 and C2 back into the original E_{holo}-positions CPS1 and CPS2. The hydrolysis site would be between IPS1 and CPS1 with Asp99 as the main player. However, in our MS experiments on p.Asp99Asn, HMBS is only detected in ES_2 state (data not shown). The result corresponds to the previous observation where HMBS is inactive and trapped in the ES_2 state after Asp99 is substituted with glycine [19], indicating that HMB reacts directly with the apoenzyme. On the other hand, if the cofactor-loop reorientates one last time to accommodate the final step in polypyrrole elongation, keeping the last two pyrroles S3 and S4 in the hexapyrrole chain stationary in CPS1 and CPS2, a fourth IPS (IPS4) may appear close to the existing IPS3. This is still possible within the cavity around IPS3 and is illustrated in the energy-minimised model of ES_4 (Fig. 9). The site of hydrolysis is then indicated in the present crystal structure between IPS2 and IPS3, and adjacent amino acids, e.g., Arg255, Glu258 and Arg355, could therefore play important roles in the hydrolysis. Reorienting the hexapyrrole through the active-site cleft would need substantial energy input, and in addition, rotating the polypyrrole around the covalent bond of Cys261 likely causes distorted torsion angles between the pyrroles, aiding the hydrolysis reaction. This would make it more plausible for hydrolysis to happen on the suggested site (Fig. 9). Investigating amino acid substitutes revealed over-elongated enzyme-intermediate species, which could indicate a loss of hydrolytic capacity in connection to these residues. However, further studies are necessary to investigate these species.

In conclusion, the crystal structure of p.Arg167Gln provides details of a third pyrrole interaction site, IPS3, and a retraction of the growing polypyrrole even further into the active site of HMBS. Furthermore, this work has revealed the crucial role of Arg195 in product release. The ES_3 crystal structure brings us one step closer to the closure: the structure and the predicted model for ES_4 (Fig. 9) suggests the hydrolysis site in close proximity to IPS3 for a maximally retracted hexapyrrole, including adjacent residues (Arg255, Glu258 and Arg355) rather than in CPS2 with Asp99 as the main proton donor in the hydrolysis action.

Materials and methods

In this work, we used the erythroid-specific HMBS2 isoform for HMBS wt (NM_001024382.2), and disease-associated p.Arg150Gln (NM_001024382.2:c.449G>A) and p.Arg178Cys (NM_001024382.2:c.532C>T). The numbering used throughout is for the corresponding amino acids in the human HMBS1 isoform to keep consistency with the literature. There are minimal structural and no known functional differences between HMBS1 and HMBS2 [11]. Thus, variants p.Arg150Gln and p.Arg178Cys are referred to as p.Arg167Gln and p.Arg195Cys, respectively, throughout the paper.

Also, the variants p.Arg255Ser, p.Glu258Thr and p.Arg355Ala follow the numbering of HMBS1 isoform, although erythroid-specific HMBS2 isoform was used for the constructs.

Protein expression and purification

All HMBS expression constructs were purchased from GenScript Biotech (Piscataway, NJ, US).

Wt-hHMBS and variants were expressed and purified according to previous report [36]. The His-tagged constructs were expressed in *E. coli* BL21 (DE3) cells (Merck KGaA, Darmstadt, Germany). Lennox LBmix™ (Merck) starter cultures supplemented with 50 $\mu g \cdot mL^{-1}$ kanamycin were grown overnight at 200 r.p.m. and 27 °C. The starter cultures were diluted in 1 L Lennox LB cultures and induced with 0.5 mM IPTG at OD_{600} 0.5–0.6 before further incubation overnight. Cells were harvested at 6000 ***g*** for 30 min, and pellets were frozen.

Thawed pellets were resuspended in 400 mM NaCl, 25 mM Tris at pH 8, 20 mM imidazole at pH 7.5, 0.5 mM 125 tris (2-carboxyethyl) phosphine (TCEP) added protease inhibitors [1 tablet cOmplete™ (Roche) w/EDTA, 10 mM benzamidine and 0.2 mM PMSF] before sonication and centrifugation. The His-tagged protein was purified from the supernatant using gravity flow column with nickel-charged affinity resin (Ni-NTA Agarose; Qiagen, Hilden, Germany) before the His-tag was cleaved off using tobacco etched virus (TEV) protease during dialysis overnight in 150 mM NaCl, 25 mM Tris pH 8 and 0.5 mM TCEP. The cleaved protein was further purified using size exclusion chromatography (SEC) on a GE HiLoad Superdex 75 16/60 PG column connected to an Äkta Pure Protein Purification System (Cytiva Europe GmbH, Freiburg, Germany) at 4 °C.

Isolation of the enzyme intermediate species was performed using an elution gradient from 0 to 65% 400 mM NaCl, 25 mM Tris pH 8 at 4 °C using an anion-exchange chromatography (IEX) Mono Q 4.6/100 PE column.

Circular dichroism spectroscopy

Far-UV spectra were obtained using a J-810 Jasco spectropolarimeter (Jasco Europe S.R.L., Cremella, Italy) and enzyme at 5 μM in 10 mM K_2HPO_4, 100 mM NaF. Three scans were obtained for each spectrum ($n = 3$), and the buffer scans were subtracted using the integrated JASCO Spectra M software (Jasco Europe). The graph is presented as mean and prepared with GRAPHPAD PRISM version 8.0.1, GraphPad Software, San Diego, CA, USA, www.graphpad.com.

Native PAGE

Approximately 5 μg enzyme was mixed 1:1 with native PAGE sample buffer containing 62.5 mM Tris–HCl, pH 6.8, 40% glycerol, 0.01% bromophenol blue (Bio-Rad Laboratories, Hercules, CA, USA) and loaded on a 10% Mini-PROTEAN® TGX™ Precast Protein Gel (Bio-Rad). The gel ran at 140 V for 2 ½ h at 4 °C with native PAGE running buffer (25 mM Tris, 192 mM glycine at pH 8.3) before staining with Coomassie blue staining, and it was visualised by a Molecular Imager ChemiDoc XRS+ with ImageLab software (Bio-Rad).

To further investigate the variants on an intermediate level, p.Arg167Gln and p.Arg195Cys were loaded with excess amount of PBG at 5× the protein concentration and incubated at 4 °C for 1, 6 and 18 h before analysis by native PAGE as described above. The reaction was done using 26 μM enzyme supplemented with 130 μM PBG substrate.

Mass spectrometry

All proteins were buffer-exchanged to 20 mM NH_4OAc (pH 6.8) using a PD Miditrap G-25 column (Cytiva), and the protein concentration was determined at absorbance at 280 nm using the sequence derived extinction coefficient (15 470 M^{-1} cm^{-1}) prior to the MS experiments.

The ESI FT-ICR mass spectra measurements were performed using a Bruker 12-T solariX XR instrument (Bruker Daltonics GmbH, Bremen, Germany) with a standard Apollo-II electrospray ionisation source. The instrument was calibrated externally using sodium perfluoroheptanoic acid (NaPFHA) clusters before measurements, and all samples were measure using direct infusion at 2 $\mu L \cdot min^{-1}$. The mass spectrometer was operated using FTMSCONTROL 2.2 software (Bruker), and the data were processed using the BRUKER DATAANALYSIS 5.1 software (Bruker). Deconvolution (i.e., zero-charge) spectra were calculated using a built in Maximum Entropy (MaxEnt) deconvolution algorithm. All masses are reported as average masses.

Differential scanning fluorimetry

Thermal denaturing scans for wt-HMBS and variant p.Arg167Gln were obtained using a Light Cycler 480 Real-Time PCR (Roche) as previously described [24,36]. In brief, 5 μM protein with 5× SYPRO Orange (Agilent Technologies, Santa Clara, CA, USA) was prepared in phosphate-buffered saline and analysed in 384-well plates. The T_m-values are provided as mean ± SD of four between-plates measurements for wt-HMBS and four in-plate measurements for the mutants. All data were analysed using HTSDSF Explorer [37].

Enzyme activity and kinetic characterisation

The enzyme activity was measured by adding 2 μg enzyme in 100 mM Tris–HCl at pH 8. The assay was started by adding pre-warmed PBG to the enzyme and stopped by the addition of 5 M HCl and 0.1% benzoquinone. For specific activity, the production of uroporphyrinogen III was measured with 100 μM PBG after 4 min reaction time. For steady-state kinetics, the reaction was stopped and measured at 4 and 10 min with a range of 3–4000 μM PBG. The kinetic parameters, K_m and V_{max}, were obtained using the Michaelis–Menten enzyme kinetics model in GraphPad Prism.

X-ray crystallography

Prior to screening for crystallisation conditions, the isolated p.Arg167Gln ES_3 intermediate was concentrated using an Amicon Ultra centrifugal filter (Merck) with a 30 kDa cut-off and filtered with a 0.22 μM Costar® Spin-X centrifuge tube filter (Merck). Crystallisation was performed using the vapour diffusion method with MRC SD2 sitting drop plates (Molecular Dimensions Limited, Rotherham, UK), and the pipetting was done using a Mosquito LCP crystallisation robot (SPT LabTech, Melbourn, UK).

The initial screenings were carried out using two commercial screens, PACT premier and JSCG plus, at +20 °C under additional conditions where HMBS was crystallised before addition [13]. Hits from the initial screenings were optimised, and plates were incubated at two temperatures, +8 and +20 °C, with two different protein-to-condition ratios; 1 : 1 and 1 : 2. Crystals were observed after 3 days from conditions containing 15–17.5% polyethylene glycol 3350, 100 mM ammonium citrate at pH 5.5.

Crystals were cryoprotected by soaking in crystallisation solution supplemented with 25% glycerol prior to flash freezing in liquid N_2. Diffraction data for the p.Arg167Gln hHMBS-ES_3 crystals were collected at 100 K using synchrotron radiation ($\lambda = 1.03$ Å) at the P11 beamline at PETRA III/DESY in Hamburg, Germany. Diffraction data were processed using XDS [38] and scaled using AIMLESS [39]. Crystals belonged to the space group $P22_12_1$ (Table 4). The hHMBS p.Arg167Gln structure was determined using the molecular replacement phaser and the hHMBS ES_2-structure (PDB: 7AAK) [13] as a search model. The

structure of the protein was refined with phenix.refine in the PHENIX package [40] in alternating cycles with manual adjustment performed in Coot [41]. Polder map for ES_3-polypyrrole was calculated using phenix.polder [42], model for the ES4 was created in coot and structure was energy minimised in Yasara [43], and the figures were created in PyMOL (The PyMOL Molecular Graphics System, Version 2.0, Schrödinger, LLC.). The diffraction and refinement statistics are summarised in Table 4.

Acknowledgement

We are grateful to Prof John R. Helliwell for the discussion as well as for pre-reviewing the manuscript and crystallography data. This work was supported by grants from the Western Norway Regional Health Authority (project F-12142 and F-11928 to AKA), the Norwegian National Advisory Unit on Rare Disorders (to AKA, MSC and HJB), and Helse Bergen Haukeland University Hospital Health Trust (project F11737-D11022 to AKA). The FT-ICR facility is supported by the European Regional Development Fund (Grant A70135), Biocenter Kuopio, Instruct-FI (Biocenter Finland) and the EU Horizon 2020 Research and Innovation Programme (European Network of Fourier Transform Ion Cyclotron Resonance Mass Spectrometry Centers, EU FT-ICR MS; Grant Agreement 731077). We acknowledge the use of the Core Facility for Biophysics, Structural Biology, and Screening (BiSS) at the University of Bergen, which has received infrastructure funding from the Research Council of Norway (RCN) through NORCRYST (grant number 245828) and NOR–OPENSCREEN (grant number 245922). Furthermore, we acknowledge DESY (Hamburg, Germany), a member of the Helmholtz Association HGF, for the provision of experimental facilities. Parts of this research were carried out at beamline P11 at PETRAIII, and we would like to thank the beamline staff for assistance in using the P11. Beamtime was allocated for proposal BAG-20211049 EC. One of the authors of this publication is a member of the European Reference Network for Rare Hereditary Metabolic Disorders (MetabERN) – Project ID No. 739543.

Conflict of interest

The authors declare no conflict of interest.

Author contributions

AKA, AM, HJB, JJ and JPK conceived and supervised the study. HJB, JPK, ML and MSC planned experiments. JPK, ML and MSC performed experiments. HJB, JPK, ML and MSC analysed the data. HJB and JPK wrote the manuscript. AKA, JPK, HJB, ML, AM and JJ made critical revisions and finalised the manuscript for publication.

Peer review

The peer review history for this article is available at https://www.webofscience.com/api/gateway/wos/peer-review/10.1111/febs.16982.

Data availability statement

The atomic coordinates and crystallographic structure factors of the p.Arg167Gln hHMBS-ES_3 were deposited in RCSB PDB with the accession code PDB: 8PND. The raw images for the crystallographic data can be downloaded from DOI: https://doi.org/10.5281/zenodo.10007385. Other data will be made available upon reasonable request.

References

1 Battersby AR, Fookes CJR, Gustafson-Potter KE, Matcham GWJ & McDonald E (1979) Proof by synthesis that unrearranged hydroxymethylbilane is the product from deaminase and the substrate for cosynthetase in the biosynthesis of uro'gen-III. *J Chem Soc Chem Commun* 1155–1158. doi: 10.1039/C39790001155

2 Burton G, Fagerness PE, Hosozawa S, Jordan PM & Scott AI (1979) 13C n.m.r. evidence for a new intermediate, pre-uroporphyrinogen, in the enzymic transformation of porphobilinogen into uroporphyrinogens I and III. *J Chem Soc Chem Commun* 202–204. doi: 10.1039/C39790000202

3 Anderson PM & Desnick RJ (1980) Purification and properties of uroporphyrinogen I synthase from human erythrocytes. Identification of stable enzyme-substrate intermediates. *J Biol Chem* **255**, 1993–1999.

4 Berry A, Jordan PM & Seehra JS (1981) The isolation and characterization of catalytically competent porphobilinogen deaminase-intermediate complexes. *FEBS Lett* **129**, 220–224.

5 Louie GV, Brownlie PD, Lambert R, Cooper JB, Blundell TL, Wood SP, Warren MJ, Woodcock SC & Jordan PM (1992) Structure of porphobilinogen deaminase reveals a flexible multidomain polymerase with a single catalytic site. *Nature* **359**, 33–39.

6 Song G, Li Y, Cheng C, Zhao Y, Gao A, Zhang R, Joachimiak A, Shaw N & Liu ZJ (2009) Structural insight into acute intermittent porphyria. *FASEB J* **23**, 396–404.

7 Jordan PM, Warren MJ, Williams HJ, Stolowich NJ, Roessner CA, Grant SK & Scott AI (1988) Identification of a cysteine residue as the binding site for the dipyrromethane cofactor at the active site of *Escherichia coli* porphobilinogen deaminase. *FEBS Lett* **235**, 189–193.

8 Miller AD, Hart GJ, Packman LC & Battersby AR (1988) Evidence that the pyrromethane cofactor of hydroxymethylbilane synthase (porphobilinogen deaminase) is bound to the protein through the sulphur atom of cysteine-242. *Biochem J* **254**, 915–918.

9 Warren MJ & Jordan PM (1988) Investigation into the nature of substrate binding to the dipyrromethane cofactor of *Escherichia coli* porphobilinogen deaminase. *Biochemistry* **27**, 9020–9030.

10 Shoolingin-Jordan PM, Warren MJ & Awan SJ (1996) Discovery that the assembly of the dipyrromethane cofactor of porphobilinogen deaminase holoenzyme proceeds initially by the reaction of preuroporphyrinogen with the apoenzyme. *Biochem J* **316** (Pt 2), 373–376.

11 Pluta P, Roversi P, Bernardo-Seisdedos G, Rojas AL, Cooper JB, Gu S, Pickersgill RW & Millet O (2018) Structural basis of pyrrole polymerization in human porphobilinogen deaminase. *Biochim Biophys Acta Gen Subj* **1862**, 1948–1955.

12 Sato H, Sugishima M, Tsukaguchi M, Masuko T, Iijima M, Takano M, Omata Y, Hirabayashi K, Wada K, Hisaeda Y *et al.* (2021) Crystal structures of hydroxymethylbilane synthase complexed with a substrate analog: a single substrate-binding site for four consecutive condensation steps. *Biochem J* **478**, 1023–1042.

13 Bustad HJ, Kallio JP, Laitaoja M, Toska K, Kursula I, Martinez A & Janis J (2021) Characterization of porphobilinogen deaminase mutants reveals that arginine-173 is crucial for polypyrrole elongation mechanism. *iScience* **24**, 102152.

14 Gill R, Kolstoe SE, Mohammed F, Al DBA, Mosely JE, Sarwar M, Cooper JB, Wood SP & Shoolingin-Jordan PM (2009) Structure of human porphobilinogen deaminase at 2.8 a: the molecular basis of acute intermittent porphyria. *Biochem J* **420**, 17–25.

15 Bissell DM, Anderson KE & Bonkovsky HL (2017) Porphyria. *N Engl J Med* **377**, 862–872.

16 Bonkovsky HL, Dixon N & Rudnick S (2019) Pathogenesis and clinical features of the acute hepatic porphyrias (AHPs). *Mol Genet Metab* **128**, 213–218.

17 Stein PE, Badminton MN & Rees DC (2017) Update review of the acute porphyrias. *Br J Haematol* **176**, 527–538.

18 San Juan I, Pereira-Ortuzar T, Cendoya X, Lain A, To-Figueras J, Mateos B, Planes FJ, Bernardo-Seisdedos G, Mato JM & Millet O (2022) ALAD inhibition by porphobilinogen rationalizes the accumulation of delta-aminolevulinate in acute porphyrias. *Biochemistry* **61**, 2409–2416.

19 Shoolingin-Jordan PM, Al-Dbass A, McNeill LA, Sarwar M & Butler D (2003) Human porphobilinogen deaminase mutations in the investigation of the mechanism of dipyrromethane cofactor assembly and tetrapyrrole formation. *Biochem Soc Trans* **31**, 731–735.

20 Bung N, Roy A, Chen B, Das D, Pradhan M, Yasuda M, New MI, Desnick RJ & Bulusu G (2018) Human hydroxymethylbilane synthase: molecular dynamics of the pyrrole chain elongation identifies step-specific residues that cause AIP. *Proc Natl Acad Sci USA* **115**, E4071–E4080.

21 Chen B, Solis-Villa C, Hakenberg J, Qiao W, Srinivasan RR, Yasuda M, Balwani M, Doheny D, Peter I, Chen R *et al.* (2016) Acute intermittent porphyria: predicted pathogenicity of HMBS variants indicates extremely low penetrance of the autosomal dominant disease. *Hum Mutat* **37**, 1215–1222.

22 Chen CH, Astrin KH, Lee G, Anderson KE & Desnick RJ (1994) Acute intermittent porphyria: identification and expression of exonic mutations in the hydroxymethylbilane synthase gene. An initiation codon missense mutation in the housekeeping transcript causes "variant acute intermittent porphyria" with normal expression of the erythroid-specific enzyme. *J Clin Invest* **94**, 1927–1937.

23 Andersson C, Floderus Y, Wikberg A & Lithner F (2000) The W198X and R173W mutations in the porphobilinogen deaminase gene in acute intermittent porphyria have higher clinical penetrance than R167W. A population-based study. *Scand J Clin Lab Invest* **60**, 643–648.

24 Bustad HJ, Vorland M, Ronneseth E, Sandberg S, Martinez A & Toska K (2013) Conformational stability and activity analysis of two hydroxymethylbilane synthase mutants, K132N and V215E, with different phenotypic association with acute intermittent porphyria. *Biosci Rep* **33**, e00056.

25 Solis C, Martinez-Bermejo A, Naidich TP, Kaufmann WE, Astrin KH, Bishop DF & Desnick RJ (2004) Acute intermittent porphyria: studies of the severe homozygous dominant disease provides insights into the neurologic attacks in acute porphyrias. *Arch Neurol* **61**, 1764–1770.

26 Kauppinen R, Mustajoki S, Pihlaja H, Peltonen L & Mustajoki P (1995) Acute intermittent porphyria in Finland: 19 mutations in the porphobilinogen deaminase gene. *Hum Mol Genet* **4**, 215–222.

27 Lenglet H, Schmitt C, Grange T, Manceau H, Karboul N, Bouchet-Crivat F, Robreau AM, Nicolas G, Lamoril J, Simonin S *et al.* (2018) From a dominant to an oligogenic model of inheritance with environmental modifiers in acute intermittent porphyria. *Hum Mol Genet* **27**, 1164–1173.

28 Whatley SD, Mason NG, Woolf JR, Newcombe RG, Elder GH & Badminton MN (2009) Diagnostic strategies for autosomal dominant acute porphyrias: retrospective analysis of 467 unrelated patients referred for mutational analysis of the HMBS, CPOX, or PPOX gene. *Clin Chem* **55**, 1406–1414.
29 Fukuda Y, Cheong PL, Lynch J, Brighton C, Frase S, Kargas V, Rampersaud E, Wang Y, Sankaran VG, Yu B *et al.* (2016) The severity of hereditary porphyria is modulated by the porphyrin exporter and Lan antigen ABCB6. *Nat Commun* **7**, 12353.
30 Roberts A, Gill R, Hussey RJ, Mikolajek H, Erskine PT, Cooper JB, Wood SP, Chrystal EJ & Shoolingin-Jordan PM (2013) 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**, 471–485.
31 Karczewski KJ, Francioli LC, Tiao G, Cummings BB, Alfoldi J, Wang Q, Collins RL, Laricchia KM, Ganna A, Birnbaum DP *et al.* (2020) The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature* **581**, 434–443.
32 Landrum MJ, Lee JM, Benson M, Brown GR, Chao C, Chitipiralla S, Gu B, Hart J, Hoffman D, Jang W *et al.* (2018) ClinVar: improving access to variant interpretations and supporting evidence. *Nucleic Acids Res* **46**, D1062–D1067.
33 Helliwell JR (2021) The crystal structures of the enzyme hydroxymethylbilane synthase, also known as porphobilinogen deaminase. *Acta Crystallogr F Struct Biol Commun* **77**, 388–398.
34 Bung N, Roy A, Priyakumar UD & Bulusu G (2019) Computational modeling of the catalytic mechanism of hydroxymethylbilane synthase. *Phys Chem Chem Phys* **21**, 7932–7940.
35 Guo J, Erskine P, Coker AR, Wood SP & Cooper JB (2017) Structural studies of domain movement in active-site mutants of porphobilinogen deaminase from bacillus megaterium. *Acta Crystallogr F Struct Biol Commun* **73**, 612–620.
36 Christie MS, Laitaoja M, Aarsand AK, Kallio JP & Bustad HJ (2022) Characterisation of a common hotspot variant in acute intermittent porphyria sheds light on the mechanism of hydroxymethylbilane synthase function. *FEBS Open Bio* **12**, 2136–2146.
37 Martin-Malpartida P, Hausvik E, Underhaug J, Torner C, Martinez A & Macias MJ (2021) HTSDSF Explorer, a novel tool to analyze high-throughput DSF screenings. *J Mol Biol* **434**, 167372.
38 Kabsch W (2010) Xds. *Acta Crystallogr D Biol Crystallogr* **66**, 125–132.
39 Evans PR & Murshudov GN (2013) How good are my data and what is the resolution? *Acta Crystallogr D Biol Crystallogr* **69**, 1204–1214.
40 Liebschner D, Afonine PV, Baker ML, Bunkoczi G, Chen VB, Croll TI, Hintze B, Hung LW, Jain S, McCoy AJ *et al.* (2019) Macromolecular structure determination using X-rays, neutrons and electrons: recent developments in phenix. *Acta Crystallogr D Struct Biol* **75**, 861–877.
41 Emsley P & Cowtan K (2004) Coot: model-building tools for molecular graphics. *Acta Crystallogr D Struct Biol* **60**, 2126–2132.
42 Liebschner D, Afonine PV, Moriarty NW, Poon BK, Sobolev OV, Terwilliger TC & Adams PD (2017) Polder maps: improving OMIT maps by excluding bulk solvent. *Acta Crystallogr D Struct Biol* **73**, 148–157.
43 Land H & Humble MS (2018) YASARA: a tool to obtain structural guidance in biocatalytic investigations. *Methods Mol Biol* **1685**, 43–67.

Supporting information

Additional supporting information may be found online in the Supporting Information section at the end of the article.

Fig. S1. Far-UV circular dichroism spectroscopy of HMBS variants.