

ARTICLE

Received 9 May 2012 | Accepted 19 Jul 2012 | Published 21 Aug 2012

DOI: 10.1038/ncomms2009

# Structures of Pup ligase PafA and depupylase Dop from the prokaryotic ubiquitin-like modification pathway

Dennis Özcelik<sup>1,\*</sup>, Jonas Barandun<sup>1,\*</sup>, Nikolaus Schmitz<sup>1</sup>, Markus Sutter<sup>1</sup>, Ethan Guth<sup>1</sup>, Fred F. Damberger<sup>1</sup>, Frédéric H.-T. Allain<sup>1</sup>, Nenad Ban<sup>1</sup> & Eilika Weber-Ban<sup>1</sup>

Pupylation is a posttranslational protein modification occurring in mycobacteria and other actinobacteria that is functionally analogous to ubiquitination. Here we report the crystal structures of the modification enzymes involved in this pathway, the prokaryotic ubiquitin-like protein (Pup) ligase PafA and the depupylase/deamidase Dop. Both feature a larger amino-terminal domain consisting of a central  $\beta$ -sheet packed against a cluster of helices, a fold characteristic for carboxylate-amine ligases, and a smaller C-terminal domain unique to PafA/Dop members. The active site is located on the concave surface of the  $\beta$ -sheet with the nucleotide bound in a deep pocket. A conserved groove leading into the active site could have a role in Pup-binding. Nuclear magnetic resonance and biochemical experiments determine the region of Pup that interacts with PafA and Dop. Structural data and mutational studies identify crucial residues for the catalysis of both enzymes.

<sup>1</sup>ETH Zurich, Institute of Molecular Biology & Biophysics, CH-8093, Switzerland. \*These authors contributed equally to this work. Correspondence and requests for materials should be addressed to E.W.-B. (email: eilika@mol.biol.ethz.ch).

The pupylation pathway<sup>1–3</sup> has an important role during the intracellular persistence of *Mycobacterium tuberculosis* (*Mtb*). It supports resistance of this pathogen towards oxidative and nitrosative stress encountered inside the host macrophages<sup>4,5</sup>. *Mtb* kills 2 million people every year and, with the emergence of multi-drug-resistant strains, new therapeutic approaches are urgently needed. The molecular components of the pupylation pathway, therefore, are promising targets for drug development.

Pupylation is a posttranslational protein-tagging system that marks target proteins for degradation by proteasomes in *Mtb* and other actinobacteria<sup>1,2</sup>. The pupylation gene locus also occurs sporadically in other bacterial lineages<sup>6,7</sup>. In analogy to ubiquitination, a small (60–70 residues), prokaryotic ubiquitin-like protein (Pup) is covalently attached to lysine residues of target proteins via an isopeptide bond<sup>1–3</sup>. Pup shares no structural homology with ubiquitin and is disordered in its free form<sup>8–10</sup>. Pup is recognized by the amino-terminal coiled-coil domains of the proteasomal ATPase Mpa<sup>10,11</sup>, leading to ATPase-driven unfolding of pupylated substrate proteins followed by degradation inside the proteasome<sup>12,13</sup>.

Although functionally analogous to ubiquitination, pupylation occurs by a chemically distinct pathway<sup>3</sup>. In mycobacteria, pupylation involves the sequential action of two homologous but catalytically different enzymes. First, the enzyme Dop (deamidase of Pup) converts the C-terminal glutamine of Pup (PupQ) to a glutamate (PupE), thereby rendering it ligation-competent<sup>3</sup>. Then, the enzyme PafA (proteasome accessory factor A) catalyses the formation of an isopeptide bond between the side chain of PupE's

C-terminal glutamate and the  $\epsilon$ -amino group of a substrate lysine<sup>3,14</sup>. To carry out the respective reactions, both enzymes, which were bioinformatically classified as belonging to the carboxylate-amine ligase family<sup>7</sup>, use ATP as a cofactor, but only the Pup ligase PafA turns over ATP<sup>3</sup>. In addition to performing the deamidation of PupQ to PupE<sup>3</sup>, Dop has an important role as depupylase by removing Pup from modified lysine residues<sup>15,16</sup>. PafA catalyses a two-step reaction proceeding through a phosphorylated intermediate formed at the C-terminal glutamate of Pup before transfer of Pup to the substrate acceptor lysine<sup>14,17</sup>. The detailed reaction mechanism for Dop remains unknown.

In this work, we present the crystal structures of the two enzymes involved in the pupylation/depupylation pathway. Biochemical analysis of PafA and Dop variants, designed on the basis of the molecular architecture of the active sites, provides mechanistic insight. We demonstrate that the C-terminal half of Pup interacts with Dop and PafA and propose that a conserved groove observed in both enzymes may bind this part of Pup, thereby placing Pup's C-terminal residue near the active site.

## Results

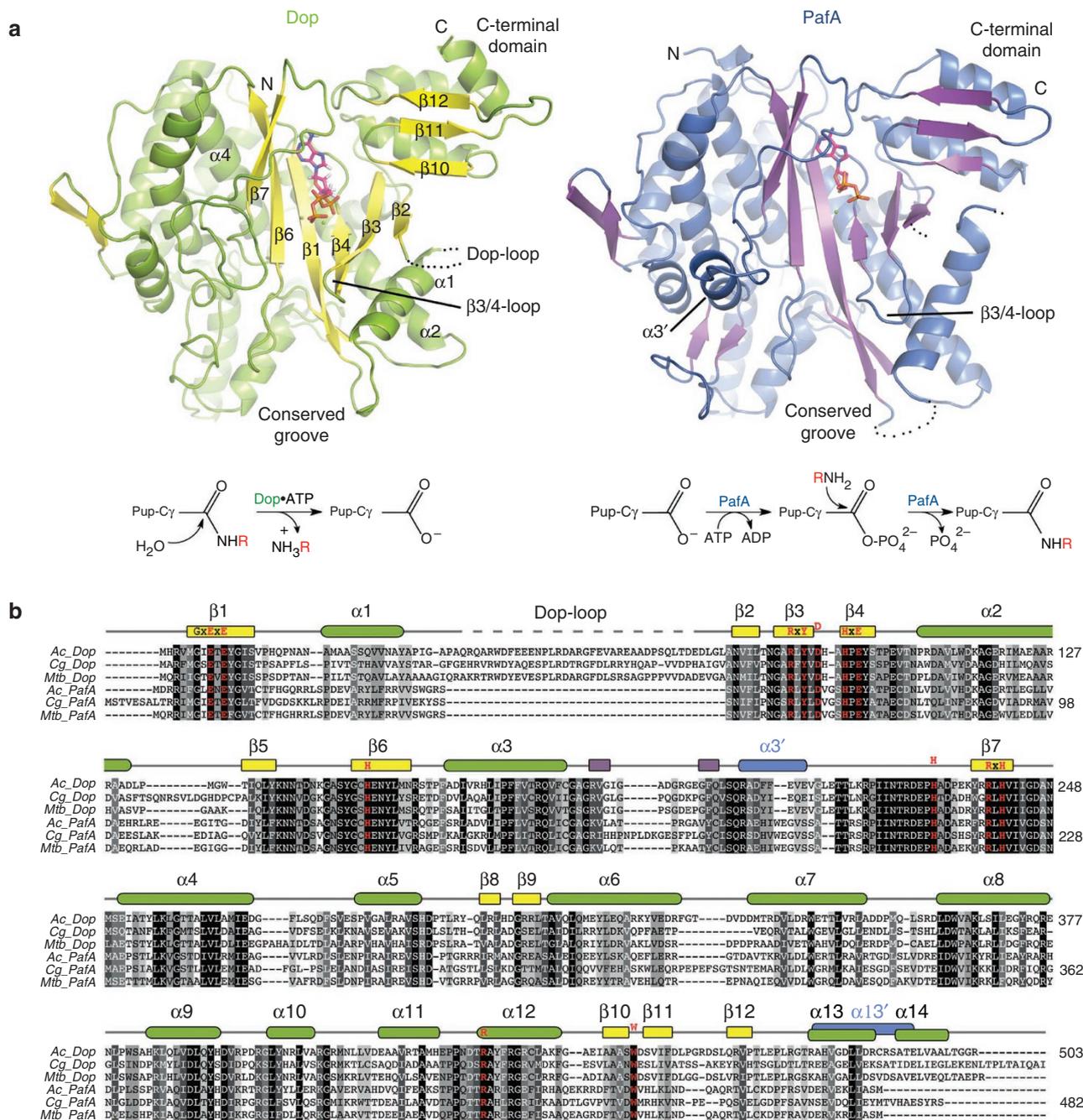
**Crystallization and structure determination.** We solved the crystal structures of full-length Dop (57 kDa) from *Acidothermus cellulolyticus* (Dop<sub>Accl</sub>) and full-length PafA (54 kDa) from *Corynebacterium glutamicum* (PafA<sub>Cglu</sub>). The structure of Dop was solved by multiple isomorphous replacement and refined to a resolution of 2.6 Å without ATP and 2.85 Å with ATP (Table 1). The final electron density map showed continuous density except for a

**Table 1 | Data collection and refinement statistics.**

	Dop PtCl <sub>4</sub>	Dop UOAc	Dop	Dop <sub>ATP</sub>	PafA <sub>ADP</sub>
<i>Data collection</i>					
Radiation source	SLS X06SA				
Radiation wavelength (Å)	1.0722	1.6000	1.00	1.00	1.00
Space group	P2 <sub>1</sub> 2 <sub>1</sub> 2 <sub>1</sub>				
Cell dimensions					
<i>a</i> , <i>b</i> , <i>c</i> (Å)*	66.68	66.43	66.55	65.36	61.85
	71.01	71.73	71.47	70.41	118.09
	97.24	95.93	96.12	93.52	161.41
Resolution (Å)	57–3.5	49–3.5	48–2.6	47–2.85	49–2.15
<i>R</i> <sub>merge</sub> (%)	5.4 (17.2)	4.6 (13.5)	7.1 (61.5)	6.2 (43.6)	4.0 (51.2)
$\langle I \rangle / \langle \sigma(I) \rangle$	21.9 (8.5)	25.3 (10.0)	15.59 (2.01)	15.69 (2.89)	18.47 (2.70)
Completeness (%)	99.9 (100)	100.0 (100)	99.2 (99.7)	98.4 (98.4)	98.4 (99.7)
Redundancy	8.97	8.94	3.6	3.4	3.7
<i>Refinement</i>					
Resolution (Å)			2.6	2.85	2.15
No. of reflections			14513	10376	65412
<i>R</i> <sub>work</sub> / <i>R</i> <sub>free</sub>			0.194/0.238	0.189/0.246	0.177/0.219
No. of atoms					
Protein			3615	3615	7388
Ligand/ion			1 Mg <sup>2+</sup>	1 ATP	2 ADP
				2 Mg <sup>2+</sup>	2 Mg <sup>2+</sup>
Water			29	5	433
<i>B</i> -factors					
Protein			57.27	56.18	53.8
Nucleotide/Mg <sup>2+</sup>			—	94.38/61.92	43.38/51.34
Water			46.92	33.0	54.8
RMSD					
Bond lengths (Å)			0.003	0.003	0.008
Bond angles (°)			0.808	0.779	1.202
PDB code			4BOR	4BOS	4BOT

Values in parentheses are for highest-resolution shell.

\*Unit cell angles:  $\alpha$ ,  $\beta$ ,  $\gamma$  = 90.0°, 90.0°, 90.0°.



**Figure 1 | Dop and PafA are close structural homologues.** (a) The crystal structures of deamidase/depupylase Dop (green) from *A. cellulolyticus* with ATP and Pup ligase PafA (blue) from *C. glutamicum* with ADP. Front views are shown, looking into the  $\beta$ -sheet cradle (yellow for Dop and purple for PafA). The dotted lines indicate structurally undefined regions. Selected  $\beta$ -strands and  $\alpha$ -helices are labelled in the Dop structure. Other structural features are indicated in both Dop and PafA. The catalysed reaction mechanism of Dop and PafA is depicted below the corresponding structures.

(b) Schematic depiction of secondary structure elements above the amino-acid sequence alignments of Dop and PafA from *A. cellulolyticus*, *C. glutamicum* and *M. tuberculosis*, respectively. The backbone is indicated by a grey line, linking  $\alpha$ -helices (green,  $\alpha$ 1-14) and  $\beta$ -strands (yellow,  $\beta$ 1-12) common to both enzymes, whereas differing structural elements in PafA are indicated as purple  $\beta$ -strands or blue  $\alpha$ -helices, respectively. Backbone regions not visible in the structures are shown as dashed grey lines.

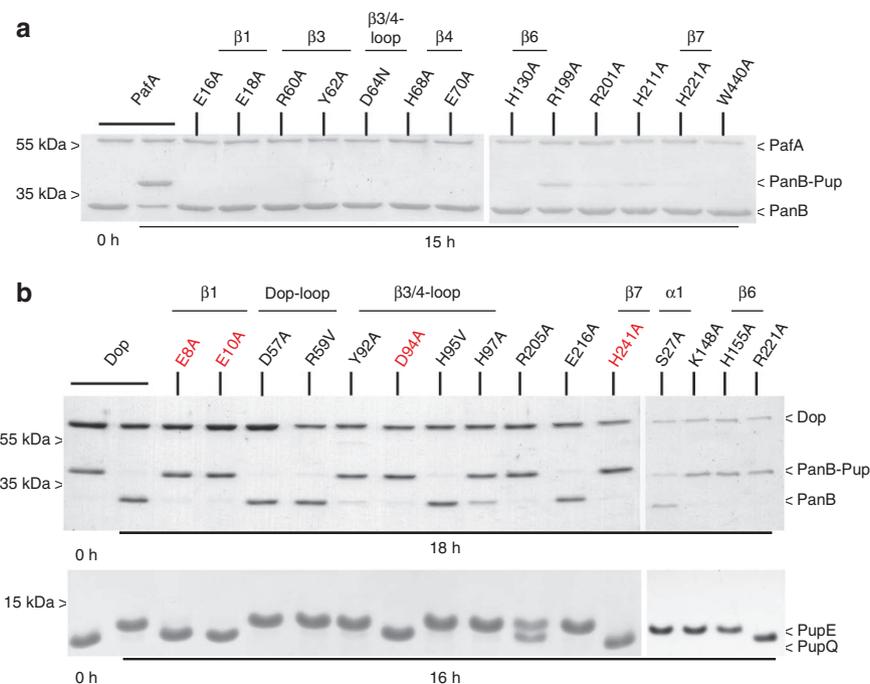
disordered region between residues 36 to 80. The structure of PafA with ADP bound was solved by molecular replacement using the Dop structure and refined to a resolution of 2.15 Å (Table 1).

PafA crystals contain two molecules per asymmetric unit, forming a stable dimer by swapping of the N-terminal strand-helix motif ( $\beta$ 1 and  $\alpha$ 1, Supplementary Fig. S1). However, the position of the exchanged strand and helix is equivalent to the position of the cor-

responding strand and helix in the monomer, based on comparison with the structure of Dop. Therefore, the PafA structure is displayed and discussed throughout this manuscript as the catalytically active monomer (chain A: N52-A477 and chain B: S2-S51).

**Dop and PafA feature two tightly interacting domains.** Dop<sub>Acel</sub> and PafA<sub>Cglu</sub> (38% sequence identity) are globular proteins with





**Figure 3 | Structure-based active site variants of Dop and PafA.** (a) Formation of a covalent PanB-Pup conjugate analysed by SDS-PAGE and Coomassie staining after incubation of PanB<sub>Mtb</sub> (6 μM) and PupE<sub>Cglu</sub> (16 μM) with PafA<sub>Cglu</sub> (1 μM) in the presence of ATP (5 mM). (b) Depupylation of a covalent PanB-Pup conjugate (upper gel) and deamidation of PupQ (lower gel) by Dop after incubation of PanB<sub>Mtb</sub>-PupE<sub>Acel</sub> (2 μM) or PupE<sub>Acel</sub> (25 μM), respectively, with Dop<sub>Acel</sub> (0.5 μM) in the presence of ATP (5 mM). Variants, fully inactive for both depupylation and deamidation, are depicted in red.

in our crystal structures likely due to the absence of a bound substrate (Pup). Alanine substitution of either of these residues (E8A and E10A) renders Dop unable to perform the deamidation and depupylation reaction (Fig. 3b) and PafA (E16A and E18A) unable to catalyse the ligation (Fig. 3a). An additional glutamate (E99 in Dop, E70 in PafA) located in strand β4 also contributes to coordination of the Mg<sup>2+</sup> ions, either directly (Dop) or via a water molecule (PafA).

Two histidine residues on strands β6 and β7 (H155 and H241 in Dop; H130 and H221 in PafA) coordinate the Mg<sup>2+</sup> ion in position n2 in the Dop structure. Changing either one of these residues to an alanine abolishes depupylation in Dop (H155A or H241A) and ligation in PafA (H130A or H221A). Deamidation by Dop, however, can still occur in the case of the Dop variant H155A (Fig. 3b).

**Active site features of PafA and Dop.** The putative binding site for Pup's C-terminal glutamate residue is located at the accessible end of the β-sheet cradle, close to the N-terminal end of β6, based on the location of the glutamate molecule in the structure of glutamine synthetase<sup>19</sup> (Fig. 2b; Supplementary Fig. S3). A conserved arginine serves to coordinate the α-carboxylate group of glutamate in glutamine synthetase<sup>20</sup>. PafA and Dop have an equivalent arginine residue (R205 in Dop, R185 in PafA) in the same location (Fig. 2b; Supplementary Fig. S3 and Supplementary Table S1), which is likely to help position the C-terminal residue of Pup in the active site. In Dop, changing this residue to an alanine (R205A) impairs depupylation and significantly slows deamidation (Fig. 3b).

The mechanism of Dop and PafA differs particularly in one aspect: PafA must activate the γ-carboxylate in the C-terminal glutamate of Pup through formation of a phosphorylated intermediate. This is necessary, because the hydroxide anion is a poor leaving group. Deamidation or depupylation on the other hand can proceed without such activation, because ammonium or substituted amines are much better leaving groups. PafA forms a surprisingly stable phosphorylated Pup intermediate, where the γ-phosphate of

ATP has been transferred to the γ-carboxylate of Pup's C-terminal glutamate<sup>17</sup>. In contrast, the amide group of the γ-carboxamide of Pup's C-terminal Gln is a very poor nucleophile and could only serve this function under very special steric circumstances (for example, when forced by enzyme constraints into a sp<sup>3</sup> hybridization state<sup>21</sup>). It is therefore not expected to attack the γ-phosphate of ATP, which is in agreement with the observation that Dop does not hydrolyse ATP<sup>3</sup>. The inability of Dop to support attack of the carboxylate oxygen of the PupE deamidation product on the γ-phosphate of ATP might be caused by a different relative positioning between the γ-carboxylate of Pup's C-terminal Glu residue and the γ-phosphate of ATP. Alternatively, ATP might not remain bound in the active site, once the product has been formed and could be required to rebind after PupE has been released.

Although Dop and PafA catalyse different reactions, parts of the mechanism are expected to require similar catalytic assistance. In both reactions, a nucleophilic attack must occur on the carbonyl-carbon of the C-terminal glutamine/glutamate side chain of Pup. This attack would be facilitated by active site residues acting as a catalytic base (H-acceptor) to activate the nucleophile, which in the case of Dop is water and in case of PafA the ε-amino group of lysine. The ε-amino group of the target lysine is expected to be located in a position equivalent to the ammonium ion-binding site in glutamine synthetase<sup>20</sup>, where residues in the loop between strands β3 and β4 coordinate the nucleophile. On the basis of their position in the active site and their chemical properties, possible candidates for proton abstraction in PafA are D64 or H68 in the β3/4-loop (Fig. 2b). Although in PafA, these residues are part of a flexible loop and point away from the active site, amino-acid substitutions at these positions (D64N or H68A) result in pupylation-inactive (D64N) and only marginally active (H68A) variants of PafA (Fig. 3a). The equivalent loop in Dop adopts a closed conformation bringing D94 and H97 towards the active site. The Dop variant D94A is completely inactivated, while the H97A variant is still

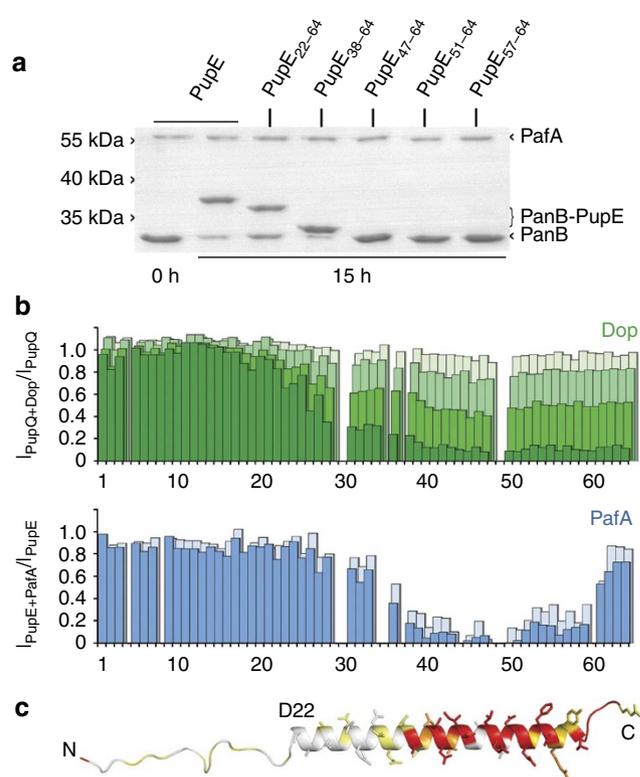
able to deamidate PupQ and exhibits some depupylation activity (Fig. 3b). This suggests that the catalytic bases supporting nucleophilic attack by the  $\epsilon$ -amino group and by water in PafA and Dop are D64 and D94, respectively, whereas the conserved histidine residue may have a role in binding and positioning of Pup in the active site. This would also agree with the fact that in Dop, the aspartate is located close to the putative position of the glutamate  $\gamma$ -carboxylate, although the histidine is located farther away.

The nucleophilic attack could be further supported by coordination of the carbonyl oxygen to a positively charged side chain of the enzyme, resulting in a more electrophilic carbonyl-carbon in the C-terminal Pup glutamine/glutamate side chain. In PafA, this role could be played by one of the arginine side chains located above the putative glutamate-binding site following the  $\alpha 3'$ -helix, R199 or R201. Alanine variants at this position in PafA exhibit marginal (R201A) or reduced (R199A) ligase activity (Fig. 3a). Because the second arginine (R201) is also conserved in Dop (R221) and an alanine mutation of that residue (R221A) leads to inactivity as well (Fig. 3b), it is the more likely candidate for this role.

**Unique features of Dop compared with PafA.** To identify residues that specifically act in deamidation/depupylation, we looked for residues or sequence stretches present or conserved only in Dop but not in PafA (Supplementary Fig. S4). One such region is the loop preceding strand  $\beta 2$  in Dop, termed the Dop-loop. Unexpectedly, deletion of this loop from Dop results in a Dop variant (Dop $\Delta$ Dop-loop) that still shows full deamidase and PanB-Pup depupylase activity (Supplementary Fig. S4c). Removal of this Dop-specific loop also did not convert Dop into a ligase, because Dop $\Delta$ Dop-loop was unable to modify PanB with Pup (Supplementary Fig. S4d). Additionally, three residues that are strictly conserved in Dop, but not in PafA, were identified: S27, H95 and K148. We produced variants replacing these residues individually with the respective residues found in PafA from the same organism (*A. cellulolyticus*). Two of these variants, S27A and H95V, exhibited wild-type behaviour. The variant K148A could no longer carry out depupylation of PanB-Pup, but was still deamidation-active (Fig. 3b). Significant differences between Dop and PafA are also observed in the region between helix  $\alpha 3$  and strand  $\beta 7$  (Dop $_{Acel}$ : 208–216, PafA $_{Acel}$ : 174–182). In PafA, this region is more structured, forming an additional  $\alpha$ -helix ( $\alpha 3'$ ). Swapping this region in Dop with the equivalent stretch of sequence from PafA $_{Acel}$  created the variant Dop $_{\alpha 3'PafA}$  (Supplementary Fig. S4). Dop $_{\alpha 3'PafA}$  was still able to deamidate PupQ and depupylate PanB-Pup *in vitro*. Even when this sequence swap was combined with deletion of the Dop-loop to form Dop $\Delta$ Dop-loop- $\alpha 3'PafA$ , no effect on these two activities was observed in comparison to wild-type Dop.

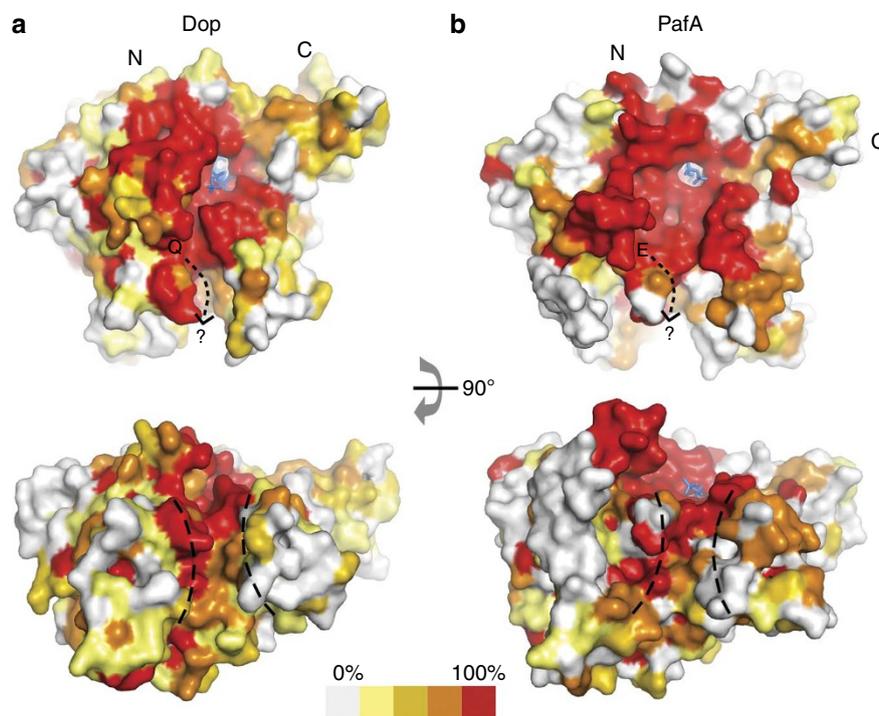
The short loop segment between strands  $\beta 3$  and  $\beta 4$  ( $\beta 3/4$ -loop) adopts a different conformation in Dop and PafA. A Dop variant, wherein this region was replaced with the equivalent region of PafA, Dop $\beta 3/4$ -loop-PafA, could still deamidate, but could not depupylate PanB-Pup (Supplementary Fig. S4). Two alanine-substituted Dop variants in this region (Y92A and H97A) showed similar behaviour (Fig. 3b), a defect in depupylation but not deamidation. To exclude a concentration-dependent effect, we repeated the depupylation assay with the same concentration of PanB-Pup as used for PupQ in deamidation and still observed that depupylation was impaired for the variants in comparison with wild type (Supplementary Fig. S5). This suggests that this region of Dop has a role in controlling accessibility to the isopeptide bond or in binding the target protein portion (in this case PanB) of pupylated substrates.

**Binding of Pup to Dop and PafA.** Dop and PafA catalyse distinct reactions in the pupylation pathway<sup>3,15,16</sup>. However, both must recognize and bind Pup. We performed a gel-based pupylation assay in combination with electrospray ionisation mass spectrometry



**Figure 4 | Interaction of Dop and PafA with Pup.** (a) Gel-based PanB $_{Mtb}$  pupylation assay using C-terminal PupE fragments of varying size. Formation of covalent conjugates between PanB $_{Mtb}$  and the indicated Pup-fragments analysed by Coomassie-stained SDS-PAGE after incubation of PanB $_{Mtb}$  (6  $\mu$ M) and corynebacterial Pup-fragments (16  $\mu$ M) with PafA $_{Cglu}$  (1  $\mu$ M) in the presence of ATP (5 mM). (b) Binding of Pup $_{Q_{Mtb}}$  to Dop $_{Cglu}$  at ratios of 1:0.1, 1:0.2, 1:0.5 and 1:1 (in increasingly dark shades of green) or Pup $_{E_{Mtb}}$  to PafA $_{Bang}$  at ratios of 1:0.1 and 1:0.25 (light and dark blue, respectively) analysed by NMR. Relative peak heights of backbone resonances from assigned residues in [ $^{15}$ N,  $^1$ H]-HSQC or [ $^{15}$ N,  $^1$ H]-HMQC spectra recorded at each ratio of Pup to enzyme were determined by normalizing to their intensity in the spectrum of Pup alone, resulting in a 'footprint' on the Pup sequence indicating the area of interaction. (c) Structural model of Pup generated with Modeller using PDB entry 3m91 (ref. 11) and helical restraints for residues 52 to 60 coloured according to residue conservation. High conservation is indicated by red, low conservation by white. The model is approximately aligned to the residue numbers in panel b.

using C-terminal PupE fragments of varying lengths. PupE $_{38-64}$  was identified as the shortest peptide fragment tested that can still be attached to PanB by PafA (Fig. 4a; Supplementary Fig. S6). Having established residues 38–64 of Pup as sufficient for conjugation by PafA, we determined the residues of Pup involved in binding to both enzymes using nuclear magnetic resonance (NMR). Considering the high degree of sequence conservation in residues 38–64 of Pup, NMR titrations with  $^{15}$ N-labelled Pup $_{E_{Mtb}}$  or Pup $_{Q_{Mtb}}$  were performed with PafA from *Bifidobacterium angulatum* (PafA $_{Bang}$ ) and Dop from *C. glutamicum* (Dop $_{Cglu}$ ), which both exhibited sufficient solubility unlike their *Mtb* homologues. Titrations were monitored in  $^{15}$ N- $^1$ H correlation spectra after addition of the unlabelled enzymes (Fig. 4b). Resonance positions of Pup signals are insensitive to the addition of Dop, whereas the intensity of specific signals gradually decreased with increasing amounts of Dop added (Fig. 4b; Supplementary Fig. S7), indicating slow exchange on the NMR timescale. This is consistent with a submicromolar affinity of Dop for Pup



**Figure 5 | Conserved surface representation of Dop and PafA.** (a) Deamidase/depupylase Dop. (b) Pup ligase PafA. Highly conserved residues are outlined in red, nonconserved residues in white. Nucleotides are shown in stick representation (blue). Termini of the polypeptide chains are depicted with either N or C, respectively. The potential Pup-binding groove (indicated by black dotted lines in the lower panels) shows intermediate to high conservation.

measured by isothermal titration calorimetry (ITC) (Supplementary Fig. S8). Pup signal positions showed more significant chemical shift changes when titrated with PafA and rapidly decreased in intensity at very low ratios of enzyme to Pup indicating that the complex is in intermediate exchange on the NMR timescale (Fig. 4b; Supplementary Fig. S7). This is consistent with the micromolar affinity measured for the Pup–PafA complex (Supplementary Fig. S8).

The interaction of PafA with Pup results in a footprint spanning residues 36 to 60 of Pup. The interaction profile of Dop on Pup is somewhat wider, extending from residues ~28 to 64. These data also suggest that in PafA, the four C-terminal residues of Pup are less constrained than in Dop, which agrees with the more narrow shape of the active site cradle in the putative glutamine-/glutamate-binding region in Dop.

The NMR experiments indicate that both Dop and PafA bind to the conserved C-terminal half of Pup, whereas the N-terminal half does not participate. The interacting stretch of residues overlaps with a region of Pup that has previously been predicted to have a propensity for forming coiled-coils<sup>7,10</sup> and adopts a helical conformation when bound to the proteasomal ATPase Mpa<sup>11</sup> (Fig. 4c). It is possible that Pup may also adopt a helical conformation when it is bound by Dop or PafA.

Surface representations of Dop and PafA, coloured according to residue conservation, reveal a conserved groove leading into the active site where the glutamate is bound in glutamine synthetases<sup>19,20</sup> (Fig. 5a,b). Pup could bind this groove either in an extended or a helical conformation. The mutation of L376E in PafA near this groove abolishes Pup ligation activity (Supplementary Fig. S9a). In Dop, mutations that introduce a negatively charged residue (Q139E or R400E) in this conserved groove strongly impair binding of PupQ as tested by analytical gel filtration and abrogate depupylation of PanB–Pup (Supplementary Fig. S9b,c). These results are consistent with a putative role of the groove in Pup binding.

## Discussion

Pupylation is an ubiquitin-like modification that has evolved in a subset of bacteria, amongst them the highly pathogenic *M. tuberculosis* (*Mtb*)<sup>1,2</sup>. Persistence of *Mtb* in the host is supported by this tagging pathway<sup>5</sup>, making it interesting from a medical perspective. The molecular details of the Pup modification system are also interesting from a mechanistic point of view, because pupylation occurs by a pathway chemically distinct from ubiquitination<sup>3</sup>. To provide the molecular framework for the mechanism of pupylation, we have determined the X-ray structures of both the Pup ligase PafA and the depupylase/deamidase Dop.

Structural alignment of Dop with PafA shows that the two enzymes are similar in overall structure and fold with differences occurring in loop regions between the central  $\beta$ -sheet and the helix cluster packed against the convex side of the sheet (Fig. 1a; Supplementary Figs S2 and S4). The active site is located in a broad  $\beta$ -sheet cradle accessible at one end (Fig. 1a). The ATP binding site, situated in both enzymes at the closed end of the  $\beta$ -sheet cradle, is well shielded from solution (Figs 1a and 2). This is achieved at least in part by the small, C-terminal domain, in addition to the adjacent loop preceding strand  $\beta$ 7. Because of this arrangement, the nucleobase and the ribose of the nucleotide are completely buried and only the phosphates are exposed.

On the basis of their bioinformatic classification as carboxylate-amine ligases, PafA and Dop were both predicted to act as Pup ligases<sup>7</sup>. However, despite the high homology between Dop and PafA, biochemical analysis clearly established distinct enzymatic activities for PafA and Dop, with PafA acting canonically in isopeptide-bond formation whereas Dop surprisingly provides the opposing peptidase/deamidase activity<sup>3,15,16</sup>. There are several differences in the molecular architecture of both enzymes potentially contributing to the distinct enzyme activities (Supplementary Fig. S4). One obvious difference is a flexible region with multiple conserved

residues downstream of helix  $\alpha 1$ , which is present only in Dop members (Dop-loop, Fig. 1b). Although it is disordered in our structure, this region could become more ordered in the presence of substrates or could have a role in the interaction with other potential binding partners. Another major difference between the two enzymes occurs in the region between helix  $\alpha 3$  and strand  $\beta 7$ . In PafA, this region is more conserved, and our structures show that it contains two short  $\beta$ -strands and one short  $\alpha$ -helix in PafA ( $\alpha 3'$ ), whereas in Dop it is less structured. This region flanks and partially covers the putative glutamyl residue-binding region and could, therefore, contribute to substrate binding and positioning of the Pup C-terminal end. Furthermore, the  $\beta 3/4$ -loop in Dop seems more rigid than in PafA and could constrain the accessibility of the active site and also influence the positioning of Pup's C-terminus. Interestingly, the exchange of any of these regions in Dop with the analogous parts of PafA does not abolish the ability of Dop to depupylate/deamidate substrates *in vitro* and, more importantly, it does not lead to a gain of function allowing pupylation to be carried out by Dop (Supplementary Fig. S4). This excludes each of these structural elements as a single determining factor for Dop activity. The deamidated form of Pup (PupE) is generated as a product of both the deamidation and the depupylation reaction. However, the carboxylate oxygen of the PupE product in Dop is apparently unable to attack the  $\gamma$ -phosphate of ATP. This might be caused by a different positioning of this carboxylate with respect to the  $\gamma$ -phosphate of ATP, preventing close enough approach or, alternatively, by asynchronous binding of PupE and ATP.

A common feature of the pupylation/depupylation enzymes is a conserved groove running along the surface of the large domain and into the  $\beta$ -sheet cradle (Fig. 5). It leads directly to the site where in homologous ligase structures the glutamate substrate is bound<sup>19</sup>. Essential features of this glutamate-binding site are conserved in Dop and PafA. For example, in the case of glutamine synthetase, the conserved arginine (R321) is involved in binding of the  $\alpha$ -carboxylate of glutamate<sup>20</sup>. The same residue is also present in the corresponding position in Dop (R205) and PafA (R185) (Fig. 2b; Supplementary Table S1 and Supplementary Fig. S3cd), indicating that both Dop and PafA probably bind Pup's C-terminal glutamine/glutamate at the same position. The dimensions of the conserved groove and the fact that it leads towards the putative glutamyl binding site, makes it an intriguing possibility that Pup might bind in the groove with its C-terminal residue positioned close to the  $\beta$ -sheet cradle. Although free Pup was shown to be mostly disordered<sup>8–10</sup>, the C-terminal part of Pup adopts an  $\alpha$ -helical structure on binding to the proteasomal ATPase Mpa<sup>10,11</sup>. Hence, the C-terminal half of Pup, identified by NMR and biochemical experiments as interacting with Dop and PafA, could also bind in the groove in a helical conformation. As this region of Pup overlaps in part with the interaction site for Mpa<sup>10,11</sup>, the depupylase Dop and Mpa are likely to be sterically prevented from binding Pup simultaneously, and thus they compete for pupylated substrates, ultimately exerting influence on the fate of the substrate as degradation or depupylation target. It is interesting that Dop variants with changes in the  $\beta 3/4$  strand region are capable of catalysing deamidation but are impaired in depupylation. This region of Dop might be involved in binding of the target protein portion of the pupylated substrate or act to control access to the isopeptide bond.

Proteomic studies have shown that the Pup ligase PafA has a large array of target proteins of varying size, oligomeric state and fold<sup>22,23</sup>. Similarly, the depupylase Dop was shown to remove Pup from numerous pupylated targets<sup>15,16</sup>. This promiscuity and low selectivity is reflected in a rather open approach to the active site allowing easy access of protein substrates from the concave face of the  $\beta$ -sheet cradle.

Dop and PafA are both strictly required for pupylation in mycobacteria<sup>2,3,24</sup>, and thus represent attractive targets for drug development. On the basis of the presented structural data, potential target

sites are the Pup-binding groove or the nucleotide-binding pocket. The nucleotide-binding pocket has unique features for PafA/Dop members of the carboxylate-amine ligase family because of the contribution of the C-terminal domain. Both targets are thus expected to be specific to the pupylation pathway, so that compounds binding to them should not inhibit other members of this family. The structural and biochemical data presented here provide a framework for future mechanistic experiments on the pupylation pathway.

## Methods

**Cloning.** To produce *A. cellulolyticus* Dop for crystallization experiments, the *dop* gene was amplified from genomic DNA of *Acidothermus cellulolyticus* ATCC 43068 and cloned with a C-terminal TEV-EGFP-His<sub>6</sub> tag via NdeI/EcoRI into a modified pET-vector (Novagen), resulting in the expression of a Dop<sub>Accl</sub>-TEV-EGFP-His<sub>6</sub> fusion protein. This construct was used to generate the Dop variants Dop<sub>ADop-loop</sub>, Dop <sub>$\alpha 3'$ PafA</sub>, Dop<sub>ADop-loop- $\alpha 3'$ PafA</sub> and Dop <sub>$\beta 3/4$ -loop-PafA</sub> via fusion PCR by exchanging the selected DNA sequence from Dop<sub>Accl</sub> with the targeted PafA<sub>Accl</sub> sequence (primer sequences are provided in Supplementary Table S2).

For all other studies in this work, a variant of Dop<sub>Accl</sub> was generated in a modified pET-vector (Novagen) lacking the last two amino acids of the wild-type sequence and carrying a C-terminal His<sub>5</sub>-tag resulting in the dop<sub>AGR</sub>-His<sub>5</sub> expression vector. Mutations of this gene were introduced by site-directed mutagenesis according to manufacturer's instructions (Stratagene).

To prepare PafA protein used in the NMR study, the *pafA* gene from *Bifidobacterium angulatum* DSM 20098 was amplified from genomic DNA via PCR. Cloning was performed with a C-terminal TEV-EGFP-His<sub>6</sub> tag via NdeI/EcoRI into a modified pET-vector (Novagen), resulting in the expression of a PafA<sub>Bang</sub>-TEV-EGFP-His<sub>6</sub> fusion protein. All corynebacterial genes were generated by PCR from *Corynebacterium glutamicum* ATCC 13032 genomic DNA in a previous study<sup>16</sup>.

The gene for PupE<sub>Accl</sub> was obtained from genomic DNA of *Acidothermus cellulolyticus* ATCC 43068 as described for the mycobacterium constructs<sup>3</sup>. Similarly, the PupQ<sub>Accl</sub> variant was obtained by using a modified reverse primer encoding a terminal glutamine.

**Expression and purification of proteins.** Dop<sub>Accl</sub> was expressed in *Escherichia coli* Rosetta (DE3) cells (Invitrogen) and PafA<sub>Bang</sub> was expressed in *E. coli* BL21 (DE3) (Invitrogen) from IPTG-inducible plasmids at 25 °C. Both were expressed as Dop<sub>Accl</sub>-TEV-EGFP-His<sub>6</sub> or PafA<sub>Bang</sub>-TEV-EGFP-His<sub>6</sub> fusion proteins and purified by Ni-affinity chromatography (HiTrap IMAC HP, GE Healthcare). After cleavage of the fusion protein with TEV-protease (Invitrogen), EGFP-His<sub>6</sub> and TEV-protease were removed by Ni-affinity chromatography. The same protocol was used to produce the variants Dop<sub>ADop-loop</sub>, Dop <sub>$\alpha 3'$ PafA</sub>, Dop<sub>ADop-loop- $\alpha 3'$ PafA</sub> and Dop <sub>$\beta 3/4$ -loop-PafA</sub>. PafA<sub>Bang</sub> was further purified by size-exclusion-chromatography on a Superdex200 column (GE Healthcare) in 25 mM Tris-HCl, pH 7.5, 300 mM NaCl and 5 mM 2-mercaptoethanol. The final size exclusion chromatography for Dop<sub>Accl</sub> was performed in 20 mM Tris-HCl, pH 7.5, 300 mM NaCl and 5 mM 2-mercaptoethanol.

The Dop<sub>AGR</sub>-His<sub>5</sub> from *A. cellulolyticus* and PafA<sub>Cglu</sub> were expressed and purified as described<sup>16</sup>. Briefly, the proteins were expressed in *E. coli* Rosetta cells from an IPTG-inducible plasmid at 23 °C and were purified by Ni-affinity chromatography and subsequent size-exclusion-chromatography. The final size exclusion chromatography for all wild-type Dop<sub>Accl</sub> and variants of Dop<sub>Accl</sub> was performed on a Superdex200 column in 50 mM Tris-HCl, pH 7.5, 300 mM NaCl and 5 mM 2-mercaptoethanol. The generated Dop variants all elute at the same retention volume as the wild-type protein.

For crystallization of PafA<sub>Cglu</sub>, the final gel filtration step was performed on a Superose6 column in 20 mM Tris-HCl, pH 7.5, 50 mM NaCl and 5 mM DTT.

For biochemical experiments with wild-type PafA<sub>Cglu</sub> and variants of PafA<sub>Cglu</sub>, the final gel-filtration step was performed on a Superdex75 column in 50 mM Tris-HCl, pH 7.5, 300 mM NaCl, 20 mM MgCl<sub>2</sub> and 1 mM DTT. The generated PafA variants all elute at the same retention volume as the wild-type protein.

PanB<sub>Mtb</sub>-His<sub>6</sub> was purified using Ni<sup>2+</sup> NTA affinity chromatography, as described<sup>3</sup>.

Pup<sub>Accl</sub>, full-length Pup and Pup<sub>22–64</sub> from *C. glutamicum* were expressed and purified as described previously<sup>3</sup>. Other corynebacterial Pup truncations were synthesized (GeneScript). All protein concentrations were determined spectrophotometrically.

<sup>15</sup>N-labelled Pup from *M. tuberculosis* H37Rv was produced by growing the cells in M9 minimal medium supplemented with <sup>15</sup>N (98%) ammonium chloride obtained from Cambridge Isotope Laboratories or Sigma-Aldrich and purified as described<sup>3,14</sup>. For NMR studies, cell lysis and Ni-affinity purification were performed in 50 mM Na<sub>2</sub>HPO<sub>4</sub>, pH 7.8 and 300 mM NaCl followed by gel filtration on a Superose75 column in 20 mM Na<sub>2</sub>HPO<sub>4</sub>, pH 6.0 and 50 mM NaCl and then exchanged into the final NMR buffer using centricon concentrators.

**Deamidation assay.** PupQ<sub>Accl</sub> (25  $\mu$ M) was incubated with 0.5  $\mu$ M Dop in reaction buffer (50 mM Tris-HCl, pH 7.5 (23 °C), 150 mM NaCl, 10% Glycerol, 1 mM DTT, 20 mM MgCl<sub>2</sub>) supplemented with 5 mM ATP at 30 °C for 16 h as

described in ref. 3. The formation of PupE was analysed by SDS–PAGE and Coomassie staining.

**Pup-conjugation assay.** *C. glutamicum* Pup variants (Pup<sub>57–64</sub>, Pup<sub>51–64</sub>, Pup<sub>47–64</sub>, Pup<sub>38–64</sub>, Pup<sub>22–64</sub> or full-length PupE; 16 μM) were incubated with 6 μM PanB<sub>Mtb</sub> and 1 μM PafA<sub>Cglu</sub> in reaction buffer (50 mM Tris–HCl, pH 7.5 (23 °C), 300 mM NaCl, 10% Glycerol, 1 mM DTT, 5 mM MgCl<sub>2</sub>) supplemented with 5 mM ATP at 23 °C for 15 h as described in ref. 3. The formation of covalent PanB–Pup conjugates was analysed by SDS–PAGE and electrospray ionisation mass spectrometry.

**Depupylation assay.** PanB<sub>Mtb</sub>–Pup<sub>Acel</sub> conjugate (2 μM or 25 μM) (produced as described in ref. 16) was incubated with 0.5 μM Dop at 30 °C for 18 h in reaction buffer (50 mM Tris–HCl, pH 7.5 (23 °C), 150 mM NaCl, 10% Glycerol, 1 mM DTT, 20 mM MgCl<sub>2</sub>) supplemented with 5 mM ATP. The formation of PanB<sub>Mtb</sub> and PupE<sub>Acel</sub> was analysed by SDS–PAGE and Coomassie staining.

**Crystallization and derivatization.** Purified Dop<sub>Acel</sub> was supplemented with 20 mM MgCl<sub>2</sub>, 5 mM 2-mercaptoethanol and 3 mM ATP. Crystallization was carried out in sitting-drop vapour diffusion plates at a protein concentration of 15 mg ml<sup>−1</sup> at 26 °C. Samples were mixed with an equal volume of reservoir solution (0.1 M HEPES–HCl, pH 7.2, and 14% (v/v) PEG-3350). Before flash freezing, the crystals were stabilized by increasing the PEG concentration by 1–3% (v/v) and adding ethylene glycol to a final concentration of 20% (v/v). Soaking of Dop crystals with ATP was performed in cryo-stabilization solution with 5 mM ATP for 30 min. Crystals were derivatized in cryo-stabilization solution by soaking with 5 mM K<sub>2</sub>PtCl<sub>4</sub> or 5 mM UO<sub>2</sub>(CH<sub>3</sub>COO)<sub>2</sub> for 3–4 h before flash-freezing.

Crystallization of PafA<sub>Cglu</sub> was carried out at a protein concentration of 15–20 mg ml<sup>−1</sup> at 4 °C. 2 μl of protein was mixed with 1 μl of reservoir solution (0.1 M CHES–NaOH, pH 9.0, 200 mM Li<sub>2</sub>SO<sub>4</sub> and 22% (v/v) PEG-4000). Before flash freezing, crystals were stabilized by adding 20% PEG-400. Soaking of PafA crystals with ADP was performed in cryo-stabilization solution with 5 mM ADP for 30 min.

**Data collection and structure determination.** The structure of DOP was solved using Pt and uranyl acetate derivatives, whereas PafA was solved via molecular replacement with the DOP structure. Data sets of both DOP<sub>Acel</sub> and PafA<sub>Cglu</sub> were collected at beamline X06SA of the Swiss Light Source (Paul Scherrer Institute, Villigen, Switzerland). For data indexing and integration, X-ray detector software (XDS)<sup>25</sup> was used. Scaling and merging of diffraction data as well as calculation of structure factor amplitudes was performed with the CCP4 program SCALA. Locations of Pt and uranyl acetate sites in DOP<sub>Acel</sub> were determined with programs from the SHELX software package<sup>26</sup> using single-wavelength anomalous dispersion (SAD). Initial phases were obtained using AUTOSOL of the Phenix package<sup>27,28</sup>.

A preliminary model was generated using BUCCANEER<sup>29</sup> and manually completed and corrected in COOT<sup>30</sup>. Manual rebuilding was alternated with refinement with phenix.refine<sup>28</sup>. To obtain the preliminary model of PafA<sub>Cglu</sub> molecular replacement was performed using the program PHASER with DOP<sub>Acel</sub> as a search model. The 2Fo–Fc maps were of good quality (Supplementary Fig. S10) except for residues 36–80 of DOP and residues 26–31 of PafA; these regions were omitted from the final model. Structures of ATP and ADP ligands for Dop and PafA, respectively, were built into difference Fourier maps obtained by refining the structures of apo enzymes against data obtained from nucleotide-soaked crystals. For PafA, only the ADP-bound structure is presented here because of better diffraction properties. Following refinement of the nucleotides, additional density was interpreted as coordinated Mg<sup>2+</sup> ions. In one of the two copies of PafA in the asymmetric unit of the crystal, the di-phosphate was modelled with two alternative conformations, but only the major one is discussed in the text. Structure determination of enzyme–nucleotide complexes was carried out via COOT and phenix.refine. Molecular graphics representations were created using the software PyMOL (<http://www.pymol.org/>).

## References

- Burns, K. E., Liu, W. T., Boshoff, H. I., Dorrestein, P. C. & Barry, C. E. III. Proteasomal protein degradation in Mycobacteria is dependent upon a prokaryotic ubiquitin-like protein. *J. Biol. Chem.* **284**, 3069–3075 (2009).
- Pearce, M. J., Mintseris, J., Ferreyra, J., Gygi, S. P. & Darwin, K. H. Ubiquitin-like protein involved in the proteasome pathway of *Mycobacterium tuberculosis*. *Science* **322**, 1104–1107 (2008).
- Striebel, F. *et al.* Bacterial ubiquitin-like modifier Pup is deamidated and conjugated to substrates by distinct but homologous enzymes. *Nat. Struct. Mol. Biol.* **16**, 647–651 (2009).
- Darwin, K. H., Ehrst, S., Gutierrez-Ramos, J. C., Weich, N. & Nathan, C. F. The proteasome of *Mycobacterium tuberculosis* is required for resistance to nitric oxide. *Science* **302**, 1963–1966 (2003).
- Gandotra, S., Schnappinger, D., Monteleone, M., Hillen, W. & Ehrst, S. In vivo gene silencing identifies the *Mycobacterium tuberculosis* proteasome as essential for the bacteria to persist in mice. *Nat. Med.* **13**, 1515–1520 (2007).
- De Mot, R. Actinomycete-like proteasomes in a Gram-negative bacterium. *Trends Microbiol.* **15**, 335–338 (2007).
- Iyer, L. M., Burroughs, A. M. & Aravind, L. Unraveling the biochemistry and provenance of pupylation: a prokaryotic analog of ubiquitination. *Biol. Direct* **3**, 45 (2008).
- Chen, X. *et al.* Prokaryotic ubiquitin-like protein pup is intrinsically disordered. *J. Mol. Biol.* **392**, 208–217 (2009).
- Liao, S. *et al.* Pup, a prokaryotic ubiquitin-like protein, is an intrinsically disordered protein. *Biochem. J.* **422**, 207–215 (2009).
- Sutter, M., Striebel, F., Damberger, F. F., Allain, F. H. & Weber-Ban, E. A distinct structural region of the prokaryotic ubiquitin-like protein (Pup) is recognized by the N-terminal domain of the proteasomal ATPase Mpa. *FEBS Lett.* **583**, 3151–3157 (2009).
- Wang, T., Darwin, K. H. & Li, H. Binding-induced folding of prokaryotic ubiquitin-like protein on the Mycobacterium proteasomal ATPase targets substrates for degradation. *Nat. Struct. Mol. Biol.* **17**, 1352–1357 (2010).
- Wang, T. *et al.* Structural insights on the Mycobacterium tuberculosis proteasomal ATPase Mpa. *Structure* **17**, 1377–1385 (2009).
- Striebel, F., Hunkeler, M., Summer, H. & Weber-Ban, E. The mycobacterial Mpa-proteasome unfolds and degrades pupylated substrates by engaging Pup's N-terminus. *EMBO J.* **29**, 1262–1271 (2010).
- Sutter, M., Damberger, F. F., Imkamp, F., Allain, F. H. & Weber-Ban, E. Prokaryotic ubiquitin-like protein (Pup) is coupled to substrates via the side chain of its C-terminal glutamate. *J. Am. Chem. Soc.* **132**, 5610–5612 (2010).
- Burns, K. E. *et al.* 'Depupylation' of prokaryotic ubiquitin-like protein from mycobacterial proteasome substrates. *Mol. Cell* **39**, 821–827 (2010).
- Imkamp, F. *et al.* Dop functions as a depupylase in the prokaryotic ubiquitin-like modification pathway. *EMBO Rep.* **11**, 791–797 (2010).
- Guth, E., Thommen, M. & Weber-Ban, E. Mycobacterial ubiquitin-like protein ligase PafA follows a two-step reaction pathway with a phosphorylated pup intermediate. *J. Biol. Chem.* **286**, 4412–4419 (2011).
- Pearce, M. J. *et al.* Identification of substrates of the Mycobacterium tuberculosis proteasome. *EMBO J.* **25**, 5423–5432 (2006).
- Eisenberg, D., Gill, H. S., Pfluegl, G. M. & Rotstein, S. H. Structure-function relationships of glutamine synthetases. *Biochim. Biophys. Acta* **1477**, 122–145 (2000).
- Gill, H. S. & Eisenberg, D. The crystal structure of phosphinothricin in the active site of glutamine synthetase illuminates the mechanism of enzymatic inhibition. *Biochemistry* **40**, 1903–1912 (2001).
- Lizak, C., Gerber, S., Numao, S., Aebi, M. & Locher, K. P. X-ray structure of a bacterial oligosaccharyltransferase. *Nature* **474**, 350–355 (2011).
- Festa, R. A. *et al.* Prokaryotic ubiquitin-like protein (Pup) proteome of *Mycobacterium tuberculosis* [corrected]. *Plos One* **5**, e8589 (2010).
- Watrous, J. *et al.* Expansion of the mycobacterial 'PUPylome'. *Mol. Biosyst.* **6**, 376–385 (2010).
- Imkamp, F. *et al.* Deletion of dop in *Mycobacterium smegmatis* abolishes pupylation of protein substrates in vivo. *Mol. Microbiol.* **75**, 744–754 (2010).
- Kabsch, W. XDS. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 125–132 (2010).
- Sheldrick, G. M. A short history of SHELX. *Acta Crystallogr. A* **64**, 112–122 (2008).
- Adams, P. D. *et al.* PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 213–221 (2010).
- Adams, P. D. *et al.* The Phenix software for automated determination of macromolecular structures. *Methods* **55**, 94–106 (2011).
- Cowtan, K. The Buccaneer software for automated model building. 1. Tracing protein chains. *Acta Crystallogr. D Biol. Crystallogr.* **62**, 1002–1011 (2006).
- Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 486–501 (2010).

## Acknowledgements

We acknowledge the staff of X06SA at the Swiss Light Source for support. We thank B. Blattmann and C. Stutz-Ducommun for help with initial screening, M. Leibundgut and T. Maier for help with data collection and structure refinement, F. Striebel, F. Imkamp and members of the Weber-Ban group for reading the manuscript. This work was supported by the Swiss National Science Foundation (SNSF), the National Center of Excellence in Research (NCCR) Structural Biology program of the SNSF, an ETH research grant and an ETH postdoctoral fellowship to EG.

## Author contributions

D.Ö. crystallized Dop and performed the biochemical analysis of Dop variants. M.S., D.Ö., and N.S. contributed to structure solution of Dop. J.B. crystallized PafA, determined its structure and carried out the biochemical analysis of PafA variants. F.F.D., M.S. and F.A. are responsible for NMR experiments and their analysis. N.B. oversaw the crystallographic aspects of the study. E.G. purified the PafA used in the NMR titration. D.Ö., J.B. and E.W.B. conceived the study and wrote the paper. All authors provided editorial input.

**Additional information**

**Accession codes:** Coordinates and structure factors for Dop, Dop<sub>ATP</sub> and PafA<sub>ADP</sub> have been deposited in the Protein Data Bank under the accession codes 4B0R, 4B0S and 4B0T, respectively.

**Supplementary Information** accompanies this paper at <http://www.nature.com/naturecommunications>

**Competing financial interests:** The authors declare no competing financial interests.

**Reprints and permission** information is available online at <http://npg.nature.com/reprintsandpermissions/>

**How to cite this article:** Özcelik, D. *et al.* Structures of Pup ligase PafA and depupylase Dop from the prokaryotic ubiquitin-like modification pathway. *Nat. Commun.* 3:1014 doi: 10.1038/ncomms2009 (2012).