STRUCTURAL BIOLOGY COMMUNICATIONS

Structure of the 4-hydroxy-tetrahydrodipicolinate synthase from the thermoacidophilic methanotroph *Methylacidiphilum fumariolicum* SolV and the phylogeny of the aminotransferase pathway

Rob A. Schmitz,[a] Andreas Dietl,[b] Melanie Müller,[b] Tom Berben,[a] Huub J. M. Op den Camp[a]* and Thomas R. M. Barends[b]*

[a]Department of Microbiology, Institute for Water and Wetland Research, Radboud University Nijmegen, 6525 AJ Nijmegen, The Netherlands, and [b]Department of Biomolecular Mechanisms, Max Planck Institute for Medical Research, 69120 Heidelberg, Germany. *Correspondence e-mail: h.opdencamp@science.ru.nl, thomas.barends@mpimf-heidelberg.mpg.de

The enzyme 4-hydroxy-tetrahydrodipicolinate synthase (DapA) is involved in the production of lysine and precursor molecules for peptidoglycan synthesis. In a multistep reaction, DapA converts pyruvate and L-aspartate-4-semialdehyde to 4-hydroxy-2,3,4,5-tetrahydrodipicolinic acid. In many organisms, lysine binds allosterically to DapA, causing negative feedback, thus making the enzyme an important regulatory component of the pathway. Here, the 2.1 Å resolution crystal structure of DapA from the thermoacidophilic methanotroph *Methylacidiphilum fumariolicum* SolV is reported. The enzyme crystallized as a contaminant of a protein preparation from native biomass. Genome analysis reveals that *M. fumariolicum* SolV utilizes the recently discovered aminotransferase pathway for lysine biosynthesis. Phylogenetic analyses of the genes involved in this pathway shed new light on the distribution of this pathway across the three domains of life.

1. Introduction

Animals rely on prokaryotes and plants for the production of the essential amino acid lysine. Two dissimilar lysine biosynthesis pathways have evolved independently over time: (i) the α-aminoadipate pathway and (ii) the diaminopimelate (DAP) pathway (Pearce *et al.*, 2017). In the DAP pathway, lysine biosynthesis is initiated by the enzyme 4-hydroxy-tetrahydrodipicolinate synthase (DapA; EC 4.3.3.7), which catalyses a multistep reaction. In the first step of this reaction, a conserved catalytic lysine residue in the active site of DapA reacts with pyruvate, forming a Schiff base linkage and resulting in a covalent enamine intermediate (Blickling & Knäblein, 1997). This enamine subsequently reacts with L-aspartate-4-semialdehyde (ASA), creating a second covalent intermediate. Finally, this intermediate is cyclized to form (2*S*,4*S*)-4-hydroxy-2,3,4,5-tetrahydrodipicolinic acid (HTPA) (Fig. 1). After this, a water molecule is removed from HTPA and the resulting double bond is reduced by the enzyme 4-hydroxy-tetrahydrodipicolinate reductase (DapB; EC 1.17.1.8) to form 2,3,4,5-tetrahydrodipicolinate (THDP) (Devenish *et al.*, 2010). THDP is then converted into *meso*-DAP, the direct precursor for lysine and for the peptidoglycan synthesis pathway (Pillai *et al.*, 2009), *via* a variety of routes. One of these, the aminotransferase pathway, was discovered only relatively recently in plants and subsequently in a few microorganisms such as methanococci (Hudson *et al.*, 2006;
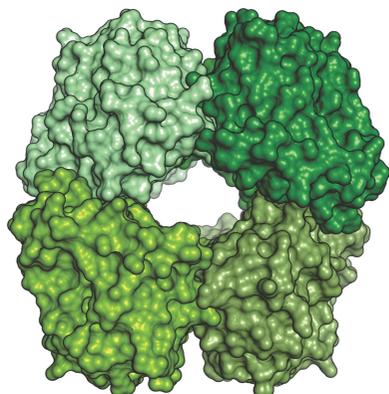
Liu *et al.*, 2010). In this pathway, the enzyme L,L-diamino-pimelate (LL-DAP) aminotransferase (DapL; EC 2.6.1.83) converts THDP into LL-DAP in a single step (Hudson *et al.*, 2006, 2008). LL-DAP is then converted into the final lysine precursor *meso*-DAP. DapA is of pivotal importance for the regulation of the lysine biosynthesis pathway by means of negative feedback, since this enzyme is inhibited allosterically by lysine, which binds in a dedicated pocket close to the interface between two adjacent monomers (Mazelis *et al.*, 1977; Geng *et al.*, 2013). Given its central role in bacterial amino-acid biosynthesis, structures of DapA from a broad range of prokaryotic species have been determined, such as *Escherichia coli* (Mirwaldt *et al.*, 1995), *Clostridium botulinum* (Atkinson *et al.*, 2009), *Aquifex aeolicus* (Sridharan *et al.*, 2014) and several others (Christensen *et al.*, 2016; Conly *et al.*, 2014; Devenish *et al.*, 2009; Girish *et al.*, 2008; Kaur *et al.*, 2011; Kefala *et al.*, 2008; Mank *et al.*, 2015; Naqvi *et al.*, 2016; Padmanabhan *et al.*, 2009; Pearce *et al.*, 2011; Phenix *et al.*, 2008; Rice *et al.*, 2008; Voss *et al.*, 2010). Moreover, DapA is a potential target for antibiotic development, since humans lack this enzyme (Skovpen *et al.*, 2016). Structural studies have shown that DapA displays a TIM-barrel fold and occurs in microorganisms either as tetramers of approximate 222 point-group symmetry (Griffin *et al.*, 2008, 2010, 2012; Voss *et al.*, 2010) or, in some cases, as dimers (Burgess *et al.*, 2008; da Costa *et al.*, 2016). Here, we report the first crystal structure of DapA from a methanotroph, *Methylacidiphilum fumariolicum* SolV. This thermoacidophile was isolated from a hot and extremely acidic volcanic ecosystem and belongs to the phylum Verrucomicrobia, which mainly represents (volcanic) soil bacteria. It can grow below pH 1 and at up to 65°C, and is dependent on rare-earth elements for growth (Pol *et al.*, 2007, 2014). We also show that *M. fumariolicum* SolV encodes DapL and thus is likely to use the aminotransferase pathway for lysine biosynthesis. Phylogenetic analyses were conducted to assess the phylogenies of DapA and DapL among prokaryotic and eukaryotic phyla.

2. Materials and methods

2.1. Macromolecule production

During the purification procedure of an [NiFe] hydrogenase from *M. fumariolicum* SolV, DapA copurified as a contaminant. *M. fumariolicum* SolV was grown as a pure culture in a chemostat as described previously (Pol *et al.*, 2007). Cell lysis and protein purification were performed as described previously (Schmitz *et al.*, 2020). Briefly, the cell membranes were homogenized and gently stirred with 1%(*w/v*) *n*-dodecyl-β-D-maltoside for 1 h at room temperature. The resulting mixture was clarified by ultracentrifugation (1 h, 137 000*g*, 4°C) and the supernatant was used for further purification by three chromatographic steps, each in the presence of 0.02%(*w/v*) *n*-dodecyl-β-D-maltoside. Firstly, the solution was loaded onto a Q Sepharose column (GE Healthcare, Chicago, Illinois, USA) equilibrated with 20 m*M* bis-Tris pH 7.0. After washing with 20 m*M* bis-Tris, 100 m*M* NaCl pH 7.0 and with 20 m*M* bis-Tris, 200 m*M* NaCl pH 7.0, the most active fractions in terms of hydrogenase activity as determined by the method described by Schmitz *et al.* (2020) were pooled. After exchanging the buffer for 20 m*M* potassium phosphate pH 7.0 and concentrating the sample, it was loaded onto a ceramic hydroxyapatite column (Bio-Rad, Hercules, California, USA)
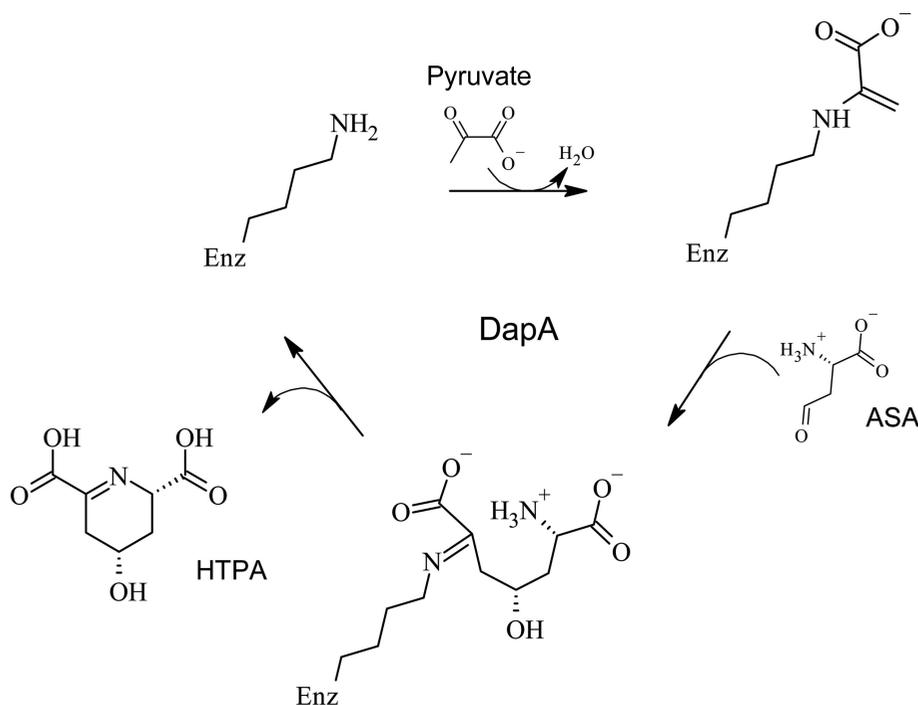

Figure 1
The multistep reaction catalysed by DapA. The catalytic Lys162 is condensed with pyruvate, forming an enamine that reacts with L-aspartic semialdehyde (ASA). The resulting covalent intermediate is cyclized, resulting in the production of (2*S*,4*S*)-4-hydroxy-2,3,4,5-tetrahydrodipicolinic acid (HTPA).

Table 1
Crystallization.

Method	Sitting-drop vapour diffusion
Plate type	Greiner XTL low profile
Temperature (K)	293
Protein concentration	$\sim$9 mg ml^{-1} (total for all proteins combined)
Buffer composition of protein solution	20 mM bis-Tris pH 7.0, 200 mM NaCl, 0.02%(w/v) n-dodecyl-β-D-maltoside (DDM)
Composition of reservoir solution	0.05 M citric acid, 0.1 M lithium sulfate, 0.05 M sodium phosphate dibasic dihydrate, 19%(w/v) PEG 1000
Volume and ratio of drop	100 + 100 nl
Volume of reservoir (µl)	70

which had been pre-equilibrated with 20 mM potassium phosphate pH 7.0. Elution was carried out with a gradient to 500 mM potassium phosphate pH 7.0 over 30 column volumes. After combining the fractions with the strongest hydrogenase activity, the buffer of the sample was exchanged to 20 mM bis-Tris pH 7.0. The final chromatographic step was then carried out on a TSKgel DEAE-5PW column (Merck, Darmstadt, Germany) that had been pre-equilibrated with 20 mM bis-Tris pH 7.0 using a gradient to 20 mM bis-Tris, 300 mM NaCl pH 7.0 over 20 column volumes. The resulting protein mixture was flash-frozen in liquid nitrogen and stored at −80°C. To assess the protein purity, SDS–PAGE (Precast Mini-Protean TGX Tris–Glycine 4–20% Gradient gel; Bio-Rad) was performed. The protein bands were manually excised and digested with trypsin. Stable proteolytic fragments were analyzed by matrix-assisted laser desorption/ionization time-of-flight (MALDI-TOF) mass spectrometry on an Axima Performance mass spectrometer (Shimadzu Biotech, Duisburg, Germany) using α-cyano-4-hydroxycinnamic acid as the matrix compound. Characteristic tryptic peptides were identified by *MASCOT* (Matrix Science, Massachusetts, USA).

2.2. Crystallization

Crystallization trials with this preparation were performed using a Mosquito pipetting robot (SPT Labtech, Melbourne, UK). Colourless, needle-shaped crystals grew within one day in condition G4 of the commercially available MemGold I screen (Molecular Dimensions, Newmarket, UK). These were flash-cooled in liquid nitrogen after soaking for a few seconds in mother liquor supplemented with 25%(v/v) ethylene glycol. Details of the crystallization conditions are given in Table 1.

2.3. Data collection and processing

X-ray diffraction data were collected at the Swiss Light Source (SLS), Paul Scherrer Institute, Villigen, Switzerland. The data were processed with *XDS* (Kabsch, 2010), which indicated $P4_x$ space-group symmetry. Systematic absences allowed the identification of the space group as $P4_2$. Details and statistics of data collection and processing are given in Table 2.

2.4. Structure solution and refinement

As the original target of the investigation was an iron-containing protein, the lack of colour in the crystals was puzzling. X-ray fluorescence measurements at the beamline (not shown) also failed to detect any iron, leaving the possibility that the crystals contained a proteolysis product of the [NiFe] hydrogenase without the cofactors. Electrophoretic separation of the protein preparation by SDS–PAGE showed several bands, including one at an apparent molecular weight of $\sim$35 kDa. Excision of this band followed by MALDI peptide-mass fingerprinting identified several proteins from *M. fumariolicum* SolV: the chaperone protein ClpB (MfumV2_0146), the large and small subunits of the [NiFe] hydrogenase (MfumV2_0979 and MfumV2_0978, respectively), DapA (MfumV2_0415) and the elongation factor Ts (MfumV2_2094) (Fig. 2).

For some of these, the structures of closely related proteins were available in the Protein Data Bank. These were therefore used as search models for molecular replacement with *Phaser* (McCoy *et al.*, 2007). A very clear solution was found (translation-function Z score of 44.9) using the structure of dihydrodipicolinate synthase from *Methanocaldococcus jannaschii* (PDB entry 2yxg; Padmanabhan *et al.*, 2009; 45.3% sequence identity), revealing the identity of the protein in the crystals to be DapA. A starting model was quickly and conveniently constructed by *MOLREP* using the known sequence (Vagin & Teplyakov, 2010). The final structure was obtained through iterative cycles of rebuilding in *Coot* (Emsley & Cowtan, 2004; Emsley *et al.*, 2010) and refinement with *Phenix* (Adams *et al.*, 2010; Liebschner *et al.*, 2019).
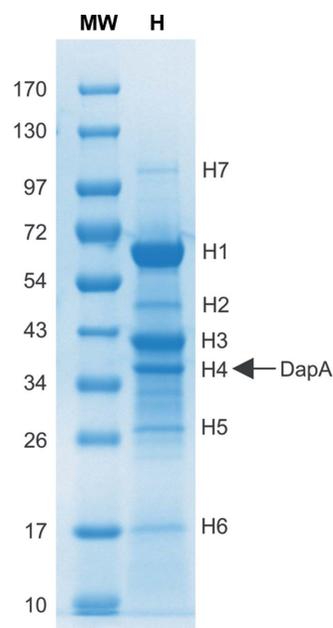
Figure 2
A Tris–glycine 4–20% gradient gel (Coomassie staining) of the protein preparation (lane H). Bands H1–H7 were excised for peptide mass fingerprinting. Band H4 was identified as *M. fumariolicum* SolV DapA (MfumV2_0415), bands H1 and H3 as the large and small subunits of the [NiFe] hydrogenase, respectively (MfumV2_0979 and MfumV2_0978), and band H2 as a degradation product of the large subunit. Band H5 was identified as elongation factor Ts (MfumV2_2094) and band H7 as the ClpB chaperone (MfumV2_0146). No match was found for band H6. Lane MW contains a molecular-weight ladder, with the molecular masses indicated in kDa on the left.

Table 2
Data collection and processing.

Values in parentheses are for the outer shell.

Diffraction source	X10SA, SLS
Wavelength (Å)	1.0000
Temperature (K)	100
Detector	PILATUS 6M
Crystal-to-detector distance (mm)	350
Rotation range per image (°)	0.15
Total rotation range (°)	120
Exposure time per image (s)	0.1
Space group	$P4_2$
a, b, c (Å)	109.1, 109.1, 107.8
α, β, γ (°)	90, 90, 90
Mosaicity (°)	0.1
Resolution range (Å)	50–2.10 (2.15–2.10)
Total No. of reflections	338313 (24379)
No. of unique reflections	74317 (5359)
Completeness (%)	99.8 (99.5)
Multiplicity	4.6 (4.5)
$\langle I/\sigma(I)\rangle$	9.2 (2.4)
$R_{\text{r.i.m.}}$	0.156 (0.739)
R_{merge}	0.138 (0.654)
$CC_{1/2}$	0.994 (0.721)
Overall B factor from Wilson plot (Å^2)	22.92

Table 3
Structure solution and refinement.

Values in parentheses are for the outer shell.

Resolution range (Å)	48.77–2.10 (2.13–2.10)
Completeness (%)	99.8
σ Cutoff	$F > 1.36\sigma(F)$
No. of reflections, working set	69658 (2628)
No. of reflections, test set	3740 (2)
Final R_{cryst}	0.181 (0.214)
Final R_{free}	0.232 (0.317)
Maximum-likelihood-based DPI (Å)	0.24
No. of non-H atoms	
Protein	9288
Ligand	20 [4 SO$_4^{2-}$]
Water	751
R.m.s. deviations	
Bonds (Å)	0.011
Angles (°)	1.411
Average B factors (Å^2)	
Protein	23.8
Ligands	35.1
Water	31.1
Ramachandran plot	
Most favoured (%)	97.6
Allowed (%)	2.1
Outliers (%)	0.3†

† These are the Tyr107 residues, which are part of the catalytic triads (see text).

Several tetrahedrally shaped density peaks were identified that could be either sulfate or phosphate ions, given that both were present during crystallization. The data do not allow distinction between the two ions, but since sulfate was present at a somewhat higher concentration than phosphate, these entities were modelled as sulfate ions. The final model contains four molecules in the asymmetric unit and has excellent statistics, as shown in Table 3. Structural figures were prepared with *PyMOL* (Schrödinger, New York, USA). The model and structure-factor amplitudes have been deposited in the Protein Data Bank as entry 6t3t.

2.5. Distribution and phylogeny of DapL and DapA

DapL sequence counts per domain and per phylum were obtained by using the following query in UniProt (The UniProt Consortium, 2019): 'LL-diaminopimelate amino-transferase NOT acetyl NOT succinyl fragment:no' and sorting the results based on taxonomy. DapL amino-acid sequences were retrieved from UniProt in *FASTA* format, aligned using *MUSCLE* 3.8.31 (Edgar, 2004) and then filtered using *trimAl* 1.2 revision 59 (Capella-Gutiérrez *et al.*, 2009) to only include positions with at most 5% gaps. Neighbour-joining trees were generated using *MEGA*7 (Kumar *et al.*, 2016) with a Poisson substitution model, gamma-distributed rates (gamma parameter 5) and 500 bootstrap replicates. *N*-Acetyl-L,L-DAP-aminotransferase (DapX) sequences from Firmicutes were used as an outgroup (not shown in the figure). The DapA tree was generated using the same method, but with *N*-acetylneuraminate lyase (NanA) sequences from Firmicutes as the outgroup instead.

3. Results and discussion

A protein mixture was obtained from native biomass of *M. fumariolicum* SolV with a view to studying the [NiFe] hydrogenase from this organism. Upon screening for crystallization conditions for the hydrogenase, the contaminating protein DapA from *M. fumariolicum* SolV crystallized and its crystal structure was determined at 2.1 Å resolution. The crystallization of contaminants has been reported by several researchers before, as reviewed in, for example, Niedzial-kowska *et al.* (2016), and a database of often-encountered, easily crystallizing contaminants has been set up to enable the rapid identification of the most common ones (Hungler *et al.*, 2016). In this case, the predicted isoelectric point of DapA (pI = 6.4) is close to that of the large hydrogenase subunit (pI = 6.7), which is likely to have contributed to its copurification, as the purification protocol relied heavily on separation by charge. However, the structure of *M. fumariolicum* SolV DapA represents a structure that has not previously been described and is the first structure available of a DapA from a methanotroph. The enzyme complex consists of a dimer of dimers forming a homotetramer with a large central cavity, as is typical for DapA enzymes. In the tetramer, each monomer contacts only two adjacent monomers, *via* either a strong or a weak dimer interface (Fig. 3). This feature results in a large central cavity, similar to the DapA structures from other microorganisms and certain plants (Pearce *et al.*, 2017). The DapA structures from the wild tobacco plant (*Nicotiana sylvestris*) and the common grapevine (*Vitis vinifera*), however, form a so-called 'back-to-back' quaternary structure (Blickling & Knäblein, 1997; Atkinson *et al.*, 2012). In the strong dimer interfaces of the DapA enzymes from most organisms, both the active site and the binding site for allosteric inhibition by lysine can be found (Fig. 3). Initially, it was believed that only DapA enzymes from Gram-negative bacteria possess this allosteric inhibition site and that those from Gram-positive organisms do not. However, this was recently disproved, and the presence of a histidine or

glutamate residue in position 56 (*E. coli* numbering) was introduced as an accurate predictor of the occurrence of allosteric inhibition (da Costa *et al.*, 2016).

Each monomer displays an eightfold α/β-barrel (TIM barrel) and a C-terminal α-helix, as is typical for DapA enzymes (Christensen *et al.*, 2016; Conly *et al.*, 2014; Devenish *et al.*, 2009; Girish *et al.*, 2008; Kaur *et al.*, 2011; Kefala *et al.*, 2008; Mank *et al.*, 2015; Mirwaldt *et al.*, 1995; Naqvi *et al.*, 2016; Padmanabhan *et al.*, 2009; Pearce *et al.*, 2011; Phenix *et al.*, 2008; Rice *et al.*, 2008; Sridharan *et al.*, 2014; Voss *et al.*, 2010). However, in *M. fumariolicum* SolV DapA the C-terminal helix is longer by three turns in comparison to other DapA structures. This helix is thought to be important in maintaining the tetramer, since the quaternary structure of DapA from *E. coli* is disrupted upon truncation of the C-terminus (Guo *et al.*, 2009). As is the case in the DapA structure from *Acinetobacter baumannii* (PDB entry 3pue; O. Jithesh, S. Yamini, N. Kaur, A. Gautam, R. Tewari, G. S. Kushwaha, P. Kaur, A. Srinivasan, S. Sharma & T. P. Singh, unpublished work), a sulfate ion is bound to the active site of each monomer. This ion assumes the position of the carboxylate group of the enamine intermediate produced by incubating the enzyme with pyruvate, as was found in the DapA structures from *Campylobacter jejuni* (Conly *et al.*, 2014) and *Clostridium botulinum* (Atkinson *et al.*, 2009). The residues around the catalytic Lys162 are identical to those in other DapA structures, including the catalytic triad (Fig. 4*a*) identified in *E. coli* DapA which functions as a proton relay between the active site and the outside of the protein (Dobson *et al.*, 2004). Moreover, superposition of the DapA structure from *M. fumariolicum* SolV with the DapA structure from *C. jejuni* reveals a pocket that is virtually identical to the allosteric binding site in the *C. jejuni* enzyme (Conly *et al.*,

2014). All of the residues that coordinate the allosteric lysine in DapA from *C. jejuni* are present in DapA from *M. fumariolicum* SolV (Conly *et al.*, 2014; Fig. 4*b*). This includes His56, which was identified as diagnostic of allosteric regulation (da Costa *et al.*, 2016), and Glu84, which is partially conserved in allosterically inhibited enzymes (Supplementary Fig. S1). However, no lysine was present during crystallization, and no density for a copurified lysine molecule could be observed. Investigation of the domain dynamics of DapA from *C. jejuni* using *DynDom* (Hayward & Lee, 2002) revealed a subtle hinging motion upon lysine binding that has been implicated in allosteric regulation (Conly *et al.*, 2014). We therefore superimposed the DapA structure from *M. fumariolicum* SolV with one of the two domains identified in the *C. jejuni* enzyme (distal to the allosteric pocket) and observed that the mutual orientation of the domains in DapA from *M. fumariolicum* SolV corresponded most closely to the structure of uninhibited (*i.e.* lysine-free) DapA from *C. jejuni* (Fig. 5).

Since *M. fumariolicum* SolV is a thermophile that can grow at temperatures of up to 65°C, it is conceivable that the enzymes in this bacterium are more thermostable compared with homologous enzymes in mesophilic microorganisms (Pol *et al.*, 2007). This was recently shown for one of the hydrogenases encoded by *M. fumariolicum* SolV (Schmitz *et al.*, 2020). Therefore, we searched for structural motifs that could enhance thermostability. The crystal structure of DapA from *A. aeolicus* shows a unique disulfide bond involving the Cys139 residues of two monomers (Sridharan *et al.*, 2014). This bond is located at the allosteric interface, but distal to the allosteric binding site, and aids in maintenance of the quaternary structure at high temperatures. Interestingly, in the sequence of *M. fumariolicum* DapA a cysteine residue is
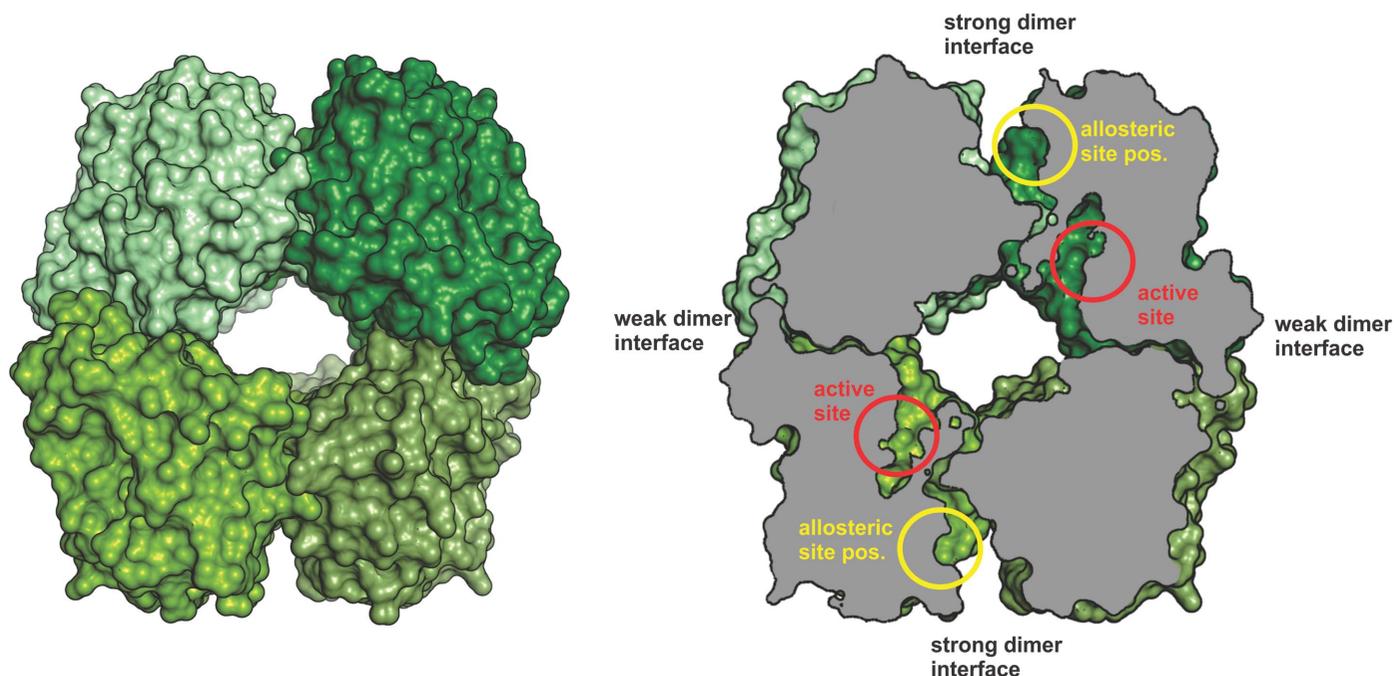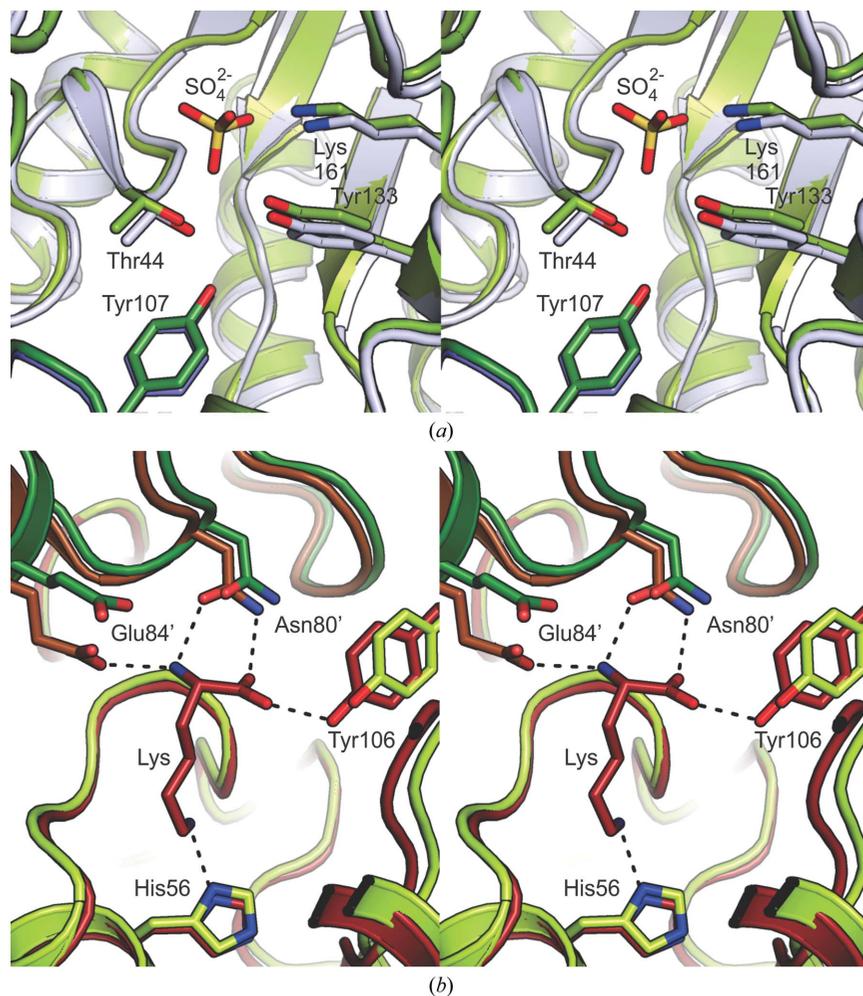

Figure 3
Overview of the DapA tetramer from *M. fumariolicum* SolV. The four monomers are coloured various shades of green. The right panel shows a cut through the tetramer. In two of the four monomers, the positions of the active sites (red), as well as the locations corresponding to the lysine-binding sites for allosteric regulation (yellow), are indicated.

found just prior to this position. However, the crystal structure shows that the side chain of this residue points away from the allosteric interface, and a considerable change in the local structure would be required to bring these residues to within bonding distance (Fig. 6). Since *A. aeolicus* grows at temperatures of up to 95°C, significantly higher than the maximum growth temperature of *M. fumariolicum* SolV, it is conceivable that such a disulfide bond is only required for stabilization at very high temperatures (Deckert *et al.*, 1998).

Following the production of HTPA by DapA, organisms use various routes to convert this molecule into lysine. Investigating the *M. fumariolicum* SolV genome revealed the presence of a DapL-encoding gene (Mfumv2_2169, previously annotated as a hypothetical protein). It is thus highly likely that this organism uses the aminotransferase pathway for this purpose. Moreover, genes encoding enzymes involved in other lysine-biosynthesis pathways were not found. We therefore investigated the phylogenies of *M. fumariolicum* SolV DapA and DapL. Although described only relatively recently, the aminotransferase pathway for lysine biosynthesis may be used by a diverse group of organisms. In this pathway, L,L-diaminopimelate is produced from 2,3,4,5-tetrahydrodipicolinate in a single step catalysed by DapL (Hudson *et al.*, 2006). We investigated the distribution of DapL across the three domains of life (Fig. 7a). In Eukarya the occurrence of DapL is mostly limited to green plants (Viridiplantae) and in Archaea mostly to the Euryarchaeota, a phylum containing the methanogens. Indeed, DapL was first discovered in the model plant organism *Arabidopsis thaliana* and subsequently also in methanococci (Liu *et al.*, 2010). DapL is found in a broader range of phyla among the Bacteria, including the phylum Verrucomicrobia, to which *M. fumariolicum* SolV belongs. Importantly, this analysis does not take into account the amount of sequencing data that is available for each group, which will inevitably be higher for bacteria (and to a lesser extent archaea) than for eukaryotes. Accordingly, this phylogenetic analysis strictly represents the current knowledge of the prevalence of the aminotransferase pathway, and the distribution shown may change dramatically as more sequence data become available. Surprisingly, a clear difference in the
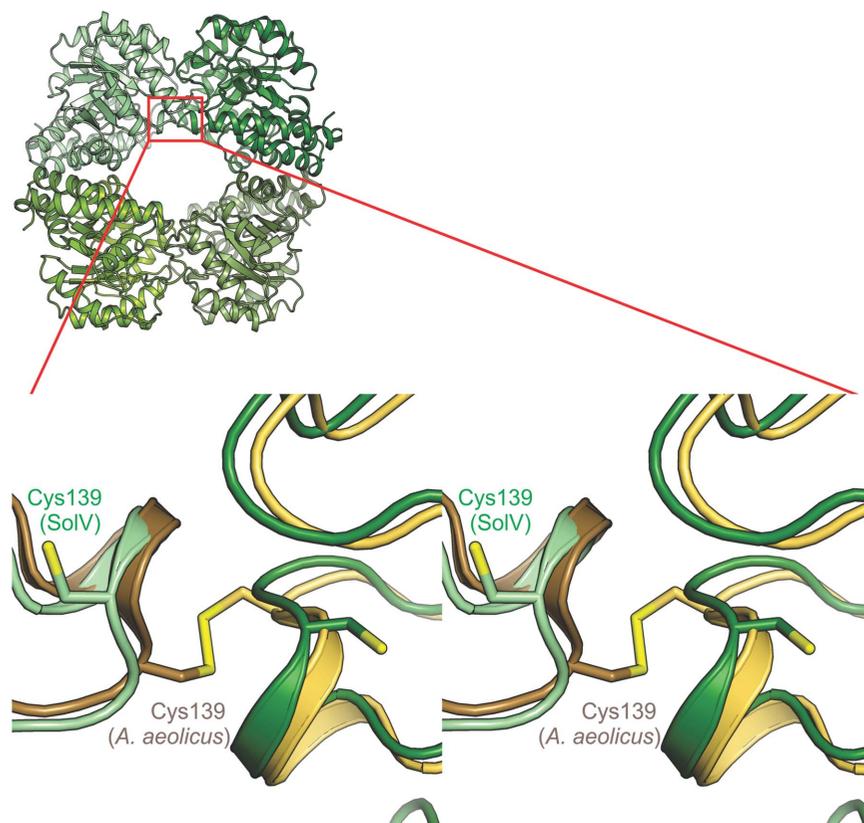

Figure 4
(a) Stereo figure showing the active site of *M. fumariolicum* SolV (shades of green) superimposed with that of *E. coli* DapA (light blue/grey). A sulfate (or phosphate; see text) ion is bound in the active site. The residues of the catalytic triad, as well as the catalytic lysine, superimpose closely. (b) Stereo figure showing the putative allosteric pocket in *M. fumariolicum* SolV (shades of green) superimposed with that in the lysine-inhibited *C. jejuni* DapA structure (dark red/brown). The residues interacting with the allosteric lysine (Lys) in the *C. jejuni* structure are conserved in *M. fumariolicum* SolV DapA, including His56, which is believed to be diagnostic for allosteric regulation. Residues from an adjacent subunit are marked with an apostrophe.
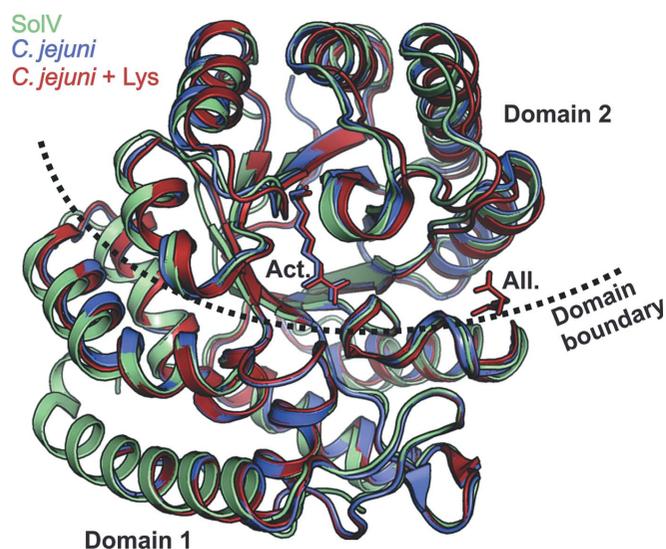
Figure 5
Superposition of DapA from *M. fumariolicum* SolV (SolV, green) with DapA from *C. jejuni* without (*C. jejuni*, blue) and with (*C. jejuni* + Lys, red) lysine in the allosteric pocket. The protein is shown as a cartoon; the covalent adduct in the active site (Act.) and the lysine in the allosteric pocket (All.) are shown as sticks. The approximate boundary between the domains is indicated by the dashed line. Domain 1 of the *C. jejuni* structure was used for superposition. Domain 2 of the SolV protein superimposes closely with the corresponding domain in the *C. jejuni* protein without lysine in the allosteric site and less well with the structure of the *C. jejuni* enzyme with a lysine bound in the allosteric pocket.

phylogeny of DapA and DapL from *M. fumariolicum* SolV was found (Fig. 7*b*). The tree for DapA mostly follows the expected phylogenetic divergence, with the exception of the second Verrucomicrobia clade (and the lone 'Candidatus Moanabacter tarae'). All of the taxa in this clade contain multiple DapA sequences, at least one of which is part of the Verrucomicrobia I group. This is expected as DapA is essential for lysine biosynthesis, whereas DapL catalyses a reaction for which multiple alternatives exist. The DapL phylogenetic tree suggests one or more horizontal gene-transfer events, as several groups of organisms are broken up into multiple clades, all of which are strongly supported by bootstrap analysis. Interestingly, the verrucomicrobial methanotrophs (which belong to the order Methylacidiphilales) form a clade with the euryarchaeotal methanococci. These methanogenic methanococci were shown not to cluster with the typical phylogenetic groups DapL1 and DapL2 (Liu *et al.*, 2010; Hudson *et al.*, 2008). Furthermore, the bacterial order Methylacidiphilales clusters with the archaeal class Methanomicrobia and members of the archaeal phylum Lokiarchaeota. All other Verrucomicrobia instead cluster with the euryarchaeotal Methanobacteria. While this could reflect factors such as the co-occurrence of different phylogenetic groups in specific habitats, a detailed analysis of this falls outside the scope of this study. In conclusion, we present the 2.1 Å resolution crystal structure of the essential 4-hydroxy-tetrahydrodipicolinate synthase (DapA) from the

Figure 6
Stereofigure showing the cysteine residues at the allosteric interface in DapA from *M. fumariolicum* SolV (green) and *A. aeolicus* (brown/yellow). In the *A. aeolicus* enzyme the residues form an intersubunit disulfide bond, whereas in *M. fumariolicum* DapA the side chains point away from each other.

thermoacidophilic methanotroph *M. fumariolicum* SolV. This homotetrameric enzyme is structurally highly conserved in comparison to the DapA structures from other microorganisms and possesses the crucial His56 residue, strongly

suggesting allosteric inhibition of DapA by lysine as a regulatory mechanism for lysine biosynthesis in *M. fumariolicum* SolV. Analysis of the gene encoding DapL indicates that *M. fumariolicum* SolV uses the aminotransferase pathway to
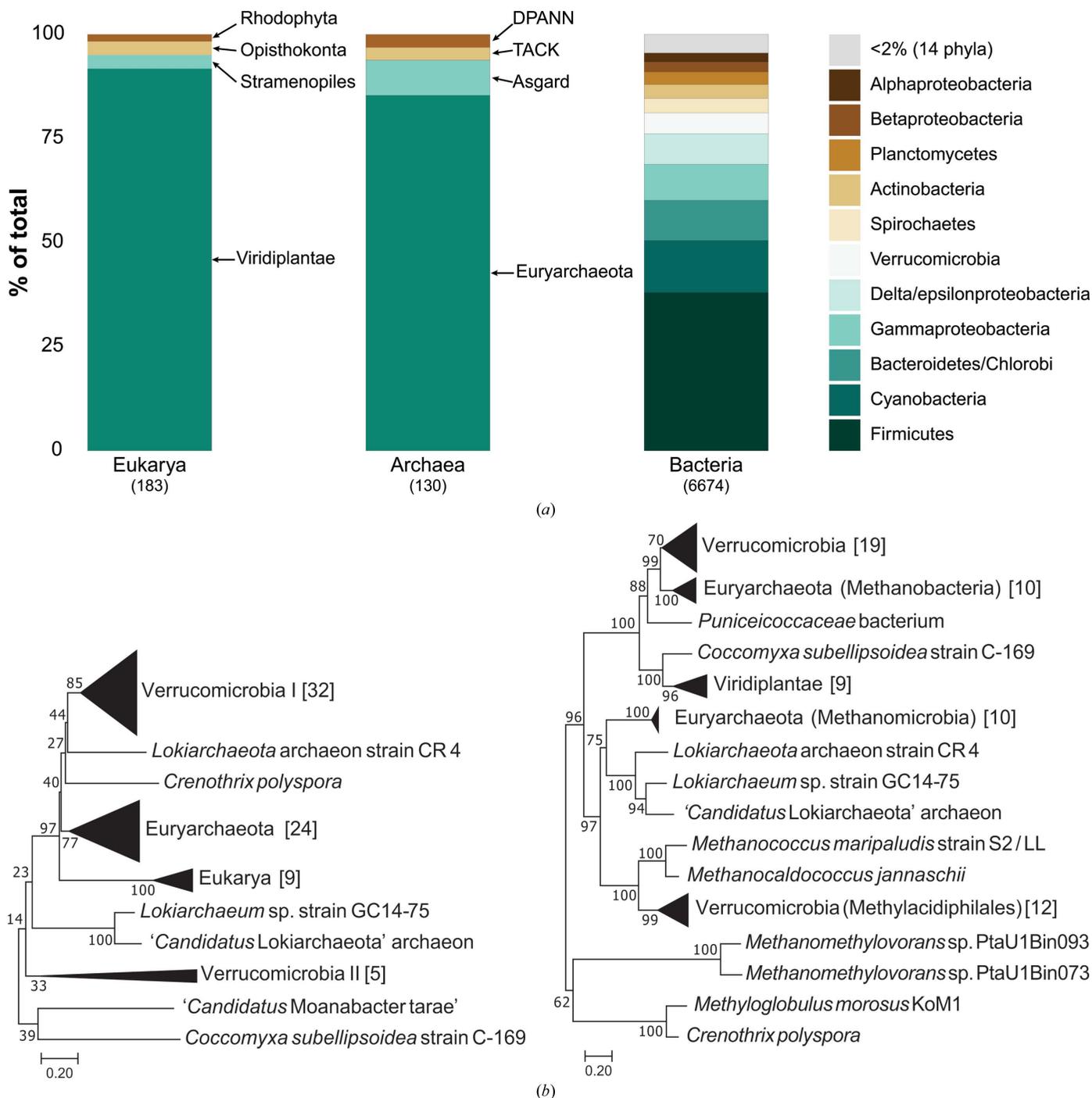
Figure 7
(*a*) Distribution of DapL among Eukarya, Archaea and Bacteria according to sequences present in the UniProt database. The total number of sequences included in each bar is shown in parentheses. The grey bar represents all bacterial phyla that individually account for less than 2% of the total number of bacterial DapL sequences (14 phyla in total). DPANN: archaeal superphylum consisting of Diapherotrites, Parvarchaeota, Aenigmarchaeota, Nanoarchaeota and Nanohaloarchaeota. TACK: archaeal superphylum consisting of Thaumarchaeota, Aigarchaeota, Crenarchaeota and Korarchaeota. (*b*) Neighbour-joining phylogenetic trees of DapA (left) and DapL (right), showing the relationships between methanotrophic and non-methanotrophic Verrucomicrobia, DapL-containing methanotrophic Proteobacteria, methanogenic Euryarchaeota, Asgardarchaeota and Eukarya. Bracketed numbers indicate the number of sequences in a collapsed branch. Both trees contain sequences from the same taxa. For accession numbers, see Supplementary Table S1.

synthesize lysine and that this pathway is found across the three domains of life.

References

Adams, P. D., Afonine, P. V., Bunkóczi, G., Chen, V. B., Davis, I. W., Echols, N., Headd, J. J., Hung, L.-W., Kapral, G. J., Grosse-Kunstleve, R. W., McCoy, A. J., Moriarty, N. W., Oeffner, R., Read, R. J., Richardson, D. C., Richardson, J. S., Terwilliger, T. C. & Zwart, P. H. (2010). *Acta Cryst.* D**66**, 213–221.

Atkinson, S. C., Dobson, R. C. J., Newman, J. M., Gorman, M. A., Dogovski, C., Parker, M. W. & Perugini, M. A. (2009). *Acta Cryst.* F**65**, 253–255.

Atkinson, S. C., Dogovski, C., Downton, M. T., Pearce, F. G., Reboul, C. F., Buckle, A. M., Gerrard, J. A., Dobson, R. C. J., Wagner, J. & Perugini, M. A. (2012). *PLoS One*, **7**, e38318.

Blickling, S. & Knäblein, J. (1997). *Biol. Chem.* **378**, 207–210.

Burgess, B. R., Dobson, R. C. J., Bailey, M. F., Atkinson, S. C., Griffin, M. D. W., Jameson, G. B., Parker, M. W., Gerrard, J. A. & Perugini, M. A. (2008). *J. Biol. Chem.* **283**, 27598–27603.

Capella-Gutiérrez, S., Silla-Martínez, J. M. & Gabaldón, T. (2009). *Bioinformatics*, **25**, 1972–1973.

Christensen, J. B., Soares da Costa, T. P., Faou, P., Pearce, F. G., Panjikar, S. & Perugini, M. A. (2016). *Sci. Rep.* **6**, 37111.

Conly, C. J. T., Skovpen, Y. V., Li, S., Palmer, D. R. J. & Sanders, D. A. R. (2014). *Biochemistry*, **53**, 7396–7406.

Deckert, G., Warren, P. V., Gaasterland, T., Young, W. G., Lenox, A. L., Graham, D. E., Overbeek, R., Snead, M. A., Keller, M., Aujay, M., Huber, R., Feldman, R. A., Short, J. M., Olsen, G. J. & Swanson, R. V. (1998). *Nature*, **392**, 353–358.

Devenish, S. R. A., Huisman, F. H. A., Parker, E. J., Hadfield, A. T. & Gerrard, J. A. (2009). *Biochim. Biophys. Acta*, **1794**, 1168–1174.

Devenish, S. R. A., Blunt, J. W. & Gerrard, J. A. (2010). *J. Med. Chem.* **53**, 4808–4812.

Dobson, R. C. J., Valegård, K. & Gerrard, J. A. (2004). *J. Mol. Biol.* **338**, 329–339.

Edgar, R. C. (2004). *Nucleic Acids Res.* **32**, 1792–1797.

Emsley, P. & Cowtan, K. (2004). *Acta Cryst.* D**60**, 2126–2132.

Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. (2010). *Acta Cryst.* D**66**, 486–501.

Geng, F., Chen, Z., Zheng, P., Sun, J. & Zeng, A. P. (2013). *Appl. Microbiol. Biotechnol.* **97**, 1963–1971.

Girish, T. S., Sharma, E. & Gopal, B. (2008). *FEBS Lett.* **582**, 2923–2930.

Griffin, M. D. W., Billakanti, J. M., Wason, A., Keller, S., Mertens, H. D. T., Atkinson, S. C., Dobson, R. C. J., Perugini, M. A., Gerrard, J. A. & Pearce, F. G. (2012). *PLoS One*, **7**, e40318.

Griffin, M. D. W., Dobson, R. C. J., Gerrard, J. A. & Perugini, M. A. (2010). *Arch. Biochem. Biophys.* **494**, 58–63.

Griffin, M. D. W., Dobson, R. C. J., Pearce, F. G., Antonio, L., Whitten, A. E., Liew, C. K., Mackay, J. P., Trewhella, J., Jameson, G. B., Perugini, M. A. & Gerrard, J. A. (2008). *J. Mol. Biol.* **380**, 691–703.

Guo, B. B. B., Devenish, S. R. A., Dobson, R. C. J., Muscroft-Taylor, A. C. & Gerrard, J. A. (2009). *Biochem. Biophys. Res. Commun.* **380**, 802–806.

Hayward, S. & Lee, R. A. (2002). *J. Mol. Graph. Model.* **21**, 181–183.

Hudson, A. O., Gilvarg, C. & Leustek, T. (2008). *J. Bacteriol.* **190**, 3256–3263.

Hudson, A. O., Singh, B. K., Leustek, T. & Gilvarg, C. (2006). *Plant Physiol.* **140**, 292–301.

Hungler, A., Momin, A., Diederichs, K. & Arold, S. T. (2016). *J. Appl. Cryst.* **49**, 2252–2258.

Kabsch, W. (2010). *Acta Cryst.* D**66**, 125–132.

Kaur, N., Gautam, A., Kumar, S., Singh, A., Singh, N., Sharma, S., Sharma, R., Tewari, R. & Singh, T. P. (2011). *Int. J. Biol. Macromol.* **48**, 779–787.

Kefala, G., Evans, G. L., Griffin, M. D. W., Devenish, S. R. A., Pearce, F. G., Perugini, M. A., Gerrard, J. A., Weiss, M. S. & Dobson, R. C. J. (2008). *Biochem. J.* **411**, 351–360.

Kumar, S., Stecher, G. & Tamura, K. (2016). *Mol. Biol. Evol.* **33**, 1870–1874.

Liebschner, D., Afonine, P. V., Baker, M. L., Bunkóczi, G., Chen, V. B., Croll, T. I., Hintze, B., Hung, L.-W., Jain, S., McCoy, A. J., Moriarty, N. W., Oeffner, R. D., Poon, B. K., Prisant, M. G., Read, R. J., Richardson, J. S., Richardson, D. C., Sammito, M. D., Sobolev, O. V., Stockwell, D. H., Terwilliger, T. C., Urzhumtsev, A. G., Videau, L. L., Williams, C. J. & Adams, P. D. (2019). *Acta Cryst.* D**75**, 861–877.

Liu, Y., White, R. H. & Whitman, W. B. (2010). *J. Bacteriol.* **192**, 3304–3310.

Mank, N., Arnette, A., Klapper, V., Offermann, L. & Chruszcz, M. (2015). *Acta Cryst.* F**71**, 449–454.

Mazelis, M., Whatley, F. R. & Whatley, J. (1977). *FEBS Lett.* **84**, 236–240.

McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C. & Read, R. J. (2007). *J. Appl. Cryst.* **40**, 658–674.

Mirwaldt, C., Korndörfer, I. & Huber, R. (1995). *J. Mol. Biol.* **246**, 227–239.

Naqvi, K. F., Staker, B. L., Dobson, R. C. J., Serbzhinskiy, D., Sankaran, B., Myler, P. J. & Hudson, A. O. (2016). *Acta Cryst.* F**72**, 2–9.

Niedzialkowska, E., Gasiorowska, O., Handing, K. B., Majorek, K. A., Porebski, P. J., Shabalin, I. G., Zasadzinska, E., Cymborowski, M. & Minor, W. (2016). *Protein Sci.* **25**, 720–733.

Padmanabhan, B., Strange, R. W., Antonyuk, S. V., Ellis, M. J., Hasnain, S. S., Iino, H., Agari, Y., Bessho, Y. & Yokoyama, S. (2009). *Acta Cryst.* F**65**, 1222–1226.

Pearce, F. G., Dobson, R. C. J., Jameson, G. B., Perugini, M. A. & Gerrard, J. A. (2011). *Biochim. Biophys. Acta*, **1814**, 1900–1909.

Pearce, F. G., Hudson, A. O., Loomes, K. & Dobson, R. C. J. (2017). *Subcell. Biochem.* **83**, 271–289.

Phenix, C. P., Nienaber, K., Tam, P. H., Delbaere, L. T. & Palmer, D. R. (2008). *ChemBioChem*, **9**, 1591–1602.

Pillai, B., Moorthie, V. A., van Belkum, M. J., Marcus, S. L., Cherney, M. M., Diaper, C. M., Vederas, J. C. & James, M. N. G. (2009). *J. Mol. Biol.* **385**, 580–594.

Pol, A., Barends, T. R. M., Dietl, A., Khadem, A. F., Eygensteyn, J., Jetten, M. S. M. & Op den Camp, H. J. M. (2014). *Environ. Microbiol.* **16**, 255–264.

Pol, A., Heijmans, K., Harhangi, H. R., Tedesco, D., Jetten, M. S. M. & Op den Camp, H. J. M. (2007). *Nature*, **450**, 874–878.

Rice, E. A., Bannon, G. A., Glenn, K. C., Jeong, S. S., Sturman, E. J. & Rydel, T. J. (2008). *Arch. Biochem. Biophys.* **480**, 111–121.

Schmitz, R. A., Pol, A., Mohammadi, S. S., Hogendoorn, C., van Gelder, A. H., Jetten, M. S. M., Daumann, L. J. & Op den Camp, H. J. M. (2020). *ISME J.*, https://doi.org/10.1038/s41396-020-0609-3.

Skovpen, Y. V., Conly, C. J., Sanders, D. A. & Palmer, D. R. (2016). *J. Am. Chem. Soc.* **138**, 2014–2020.

Soares da Costa, T. P., Desbois, S., Dogovski, C., Gorman, M. A., Ketaren, N. E., Paxman, J. J., Siddiqui, T., Zammit, L. M., Abbott, B. M., Robins-Browne, R. M., Parker, M. W., Jameson, G. B., Hall, N. E., Panjikar, S. & Perugini, M. A. (2016). *Structure*, **24**, 1282–1291.

Sridharan, U., Ebihara, A., Kuramitsu, S., Yokoyama, S., Kumarevel, T. & Ponnuraj, K. (2014). *Extremophiles*, **18**, 973–985.

The UniProt Consortium (2019). *Nucleic Acids Res.* **47**, D506–D515.

Vagin, A. & Teplyakov, A. (2010). *Acta Cryst.* D**66**, 22–25.

Voss, J. E., Scally, S. W., Taylor, N. L., Atkinson, S. C., Griffin, M. D. W., Hutton, C. A., Parker, M. W., Alderton, M. R., Gerrard, J. A., Dobson, R. C. J., Dogovski, C. & Perugini, M. A. (2010). *J. Biol. Chem.* **285**, 5188–5195.