Structure of LP2179, the first representative of Pfam family PF08866, suggests a new fold with a role in amino-acid metabolism

Constantina Bakolitsa,[a,b] Abhinav Kumar,[a,c] Dennis Carlton,[a,d] Mitchell D. Miller,[a,c] S. Sri Krishna,[a,b,e] Polat Abdubek,[a,f] Tamara Astakhova,[a,e] Herbert L. Axelrod,[a,c] Hsiu-Ju Chiu,[a,c] Thomas Clayton,[a,d] Marc C. Deller,[a,d] Lian Duan,[a,e] Marc-André Elsliger,[a,d] Julie Feuerhelm,[a,f] Slawomir K. Grzechnik,[a,e] Joanna C. Grant,[a,f] Gye Won Han,[a,d] Lukasz Jaroszewski,[a,b,e] Kevin K. Jin,[a,c] Heath E. Klock,[a,f] Mark W. Knuth,[a,f] Piotr Kozbial,[a,b] David Marciano,[a,d] Daniel McMullan,[a,f] Andrew T. Morse,[a,e] Edward Nigoghossian,[a,f] Linda Okach,[a,f] Silvya Oommachen,[a,c] Jessica Paulsen,[a,f] Ron Reyes,[a,c] Christopher L. Rife,[a,c] Henry J. Tien,[a,d] Christina V. Trout,[a,d] Henry van den Bedem,[a,c] Dana Weekes,[a,b] Qingping Xu,[a,c] Keith O. Hodgson,[a,g] John Wooley,[a,e] Ashley M. Deacon,[a,c] Adam Godzik,[a,b,e] Scott A. Lesley[a,d,f] and Ian A. Wilson[a,d]*

[a]Joint Center for Structural Genomics, http://www.jcsg.org, USA, [b]Program on Bioinformatics and Systems Biology, Burnham Institute for Medical Research, La Jolla, CA, USA, [c]Stanford Synchrotron Radiation Lightsource, SLAC National Accelerator Laboratory, Menlo Park, CA, USA, [d]Department of Molecular Biology, The Scripps Research Institute, La Jolla, CA, USA, [e]Center for Research in Biological Systems, University of California, San Diego, La Jolla, CA, USA, [f]Protein Sciences Department, Genomics Institute of the Novartis Research Foundation, San Diego, CA, USA, and [g]Photon Science, SLAC National Accelerator Laboratory, Menlo Park, CA, USA

Correspondence e-mail: wilson@scripps.edu

The structure of LP2179, a member of the PF08866 (DUF1831) family, suggests a novel $\alpha+\beta$ fold comprising two β-sheets packed against a single helix. A remote structural similarity to two other uncharacterized protein families specific to the *Bacillus* genus (PF08868 and PF08968), as well as to prokaryotic *S*-adenosylmethionine decarboxylases, is consistent with a role in amino-acid metabolism. Genomic neighborhood analysis of LP2179 supports this functional assignment, which might also then be extended to PF08868 and PF08968.

1. Introduction

The Pfam database (Finn *et al.*, 2008) contains over 2000 domains of unknown function (DUFs), which are protein families for which the biological function is unknown and cannot be deduced by homology. Currently, DUFs are the best source for the discovery of new folds (Jaroszewski *et al.*, submitted), followed by large families with no structural representatives. DUF structures provide the first step towards establishing functional hypotheses and extending our understanding of the protein universe. In an effort to sample and understand the diversity of protein-fold and structure space, targets were selected from Pfam protein family PF08866 (DUF1831). Here, we report the crystal structure of LP2179, the first structural representative of this family, which was determined using the semiauto-mated high-throughput pipeline of the Joint Center for Structural Genomics (JCSG; Lesley *et al.*, 2002) as part of the NIGMS Protein Structure Initiative (PSI; http://www.nigms.nih.gov/Initiatives/PSI/). The *LP2179* gene of *Lactobacillus plantarum*, a lactic acid-producing bacterium found in human saliva and intestinal flora, encodes a protein with a molecular weight of 12.6 kDa (residues 1–113) and a calculated isoelectric point of 8.9. LP2179 appears to adopt a novel fold with remote similarities to proteins with a TBP-like fold (TATA-binding protein), including *S*-adenosyl-L-methionine decarboxylase (EC 4.1.1.50), an enzyme implicated in the urea cycle and the cata-bolism of methionine and amino groups. Analysis of the genomic neighborhood of DUF1831 homologs reveals the systematic presence of other enzymes implicated in amino-acid and amino-group meta-bolism, suggesting a similar role for other members of the DUF1831 family and for two other functionally uncharacterized families that show partial structural similarity to LP2179.

2. Materials and methods

2.1. Protein production and crystallization

Clones were generated using the Polymerase Incomplete Primer Extension (PIPE) cloning method (Klock *et al.*, 2008). The gene encoding LP2179 (GenBank NP_785678, gi:28378786, Swiss-Prot Q88V95) was amplified by polymerase chain reaction (PCR) from *L. plantarum* WCFS1 NCIMB8826 genomic DNA using *PfuTurbo* DNA polymerase (Stratagene) and I-PIPE (Insert) primers (forward primer, 5′-ctgtacttccagggcATGGCATACACAACAACGGTTAAAC-

Table 1
Summary of crystal parameters, data-collection and refinement statistics for LP2179 (PDB code 2iay).

Values in parentheses are for the highest resolution shell.

	λ_1 Se	λ_2 MADSe	λ_3 MADSe	λ_4 MADSe
Data collection				
Space group	$P2_12_12_1$	$P2_12_12_1$		
Unit-cell parameters (Å)	$a = 36.29$, $b = 47.90$, $c = 58.01$	$a = 36.41$, $b = 47.99$, $c = 57.83$		
Wavelength (Å)	0.9798	0.9493	0.9794	0.9793
Resolution range (Å)	28.9–1.2 (1.23–1.20)	29.0–1.3 (1.36–1.33)	29.0–1.4 (1.41–1.37)	29.0–1.4 (1.41–1.37)
No. of observations	181436	69961	63756	63463
No. of unique reflections	29115	21577	19702	19718
Completeness (%)	90.4 (50.7)	90.4 (49.2)	90.2 (46.3)	90.3 (47.2)
Mean $I/\sigma(I)$	16.4 (2.1)	13.3 (2.0)	12.9 (2.0)	12.9 (1.7)
R_{merge} on I† (%)	6.9 (20.4)	6.2 (38.8)	6.2 (37.9)	7.2 (46.7)
R_{meas} on I‡ (%)	7.3 (26.7)	7.3 (52.4)	7.3 (51.1)	8.5 (63.1)
Model and refinement statistics				
Resolution range (Å)	28.9–1.2			
No. of reflections (total)	29080§			
No. of reflections (test)	1488			
Completeness (%)	90.1			
Data set used in refinement	λ_1 Se			
Cutoff criterion	$|F| > 0$			
R_{cryst}¶	0.120			
R_{free}¶	0.147			
Stereochemical parameters				
Restraints (r.m.s.d. observed)				
Bond angles (°)	1.65			
Bond lengths (Å)	0.016			
Average isotropic B value (Å²)	8.86			
ESU†† based on R_{free} (Å)	0.038			
Protein residues/atoms	114/944			
Water molecules/other solvent	195/2			

† $R_{merge} = \sum_{hkl} \sum_i |I_i(hkl) - \langle I(hkl)\rangle| / \sum_{hkl} \sum_i I_i(hkl)$. ‡ $R_{meas} = \sum_{hkl} [N/(N-1)]^{1/2} \times \sum_i |I_i(hkl) - \langle I(hkl)\rangle| / \sum_{hkl} \sum_i I_i(hkl)$ (Diederichs & Karplus, 1997). § Typically, the number of unique reflections used in refinement was slightly less that the total number that were integrated and scaled. Reflections were excluded owing to systematic absences, negative intensities and rounding errors in the resolution limits and unit-cell parameters. ¶ $R_{cryst} = \sum_{hkl} ||F_{obs}| - |F_{calc}|| / \sum_{hkl} |F_{obs}|$, where F_{calc} and F_{obs} are the calculated and observed structure-factor amplitudes, respectively. R_{free} is the same as R_{cryst} but for 5.1% of the total reflections chosen at random and omitted from refinement †† Estimated overall coordinate error (Collaborative Computational Project, Number 4, 1994; Tickle *et al.*, 1998).

3′; reverse primer, 5′-aattaagtcgcgttaGTCCGTCGTGAGGATATC-CCGTTC-3′; target sequence in upper case) that included sequences for the predicted 5′ and 3′ ends. The expression vector pSpeedET, which encodes an amino-terminal tobacco etch virus (TEV) protease-cleavable expression and purification tag (MGSDKIHHHHHH-ENLYFQ/G), was PCR-amplified with V-PIPE (Vector) primers (forward primer, 5′-taacgcgacttaattaactcgtttaaacggtctccagc-3′; reverse primer, 5′-gccctggaagtacaggttttcgtgatgatgatgatgatg-3′). V-PIPE and I-PIPE PCR products were mixed to anneal the amplified DNA fragments together. *Escherichia coli* GeneHogs (Invitrogen) competent cells were transformed with the V-PIPE/I-PIPE mixture and dispensed onto selective LB–agar plates. The cloning junctions were confirmed by DNA sequencing. Expression was performed in a selenomethionine-containing medium. At the end of fermentation, lysozyme was added to the culture to a final concentration of 250 µg ml⁻¹ and the cells were harvested and frozen. After one freeze–thaw cycle, the cells were sonicated in lysis buffer [50 m*M* HEPES pH 8.0, 50 m*M* NaCl, 10 m*M* imidazole, 1 m*M* tris(2-carboxyethyl)phosphine–HCl (TCEP)] and the lysate was clarified by centrifugation at 32 500g for 30 min. The soluble fraction was passed over nickel-chelating resin (GE Healthcare) pre-equilibrated with lysis buffer, the resin was washed with wash buffer [50 m*M* HEPES pH 8.0, 300 m*M* NaCl, 40 m*M* imidazole, 10%(*v*/*v*) glycerol, 1 m*M*

TCEP] and the protein was eluted with elution buffer [20 m*M* HEPES pH 8.0, 300 m*M* imidazole, 10%(*v*/*v*) glycerol, 1 m*M* TCEP]. The eluate was buffer-exchanged with TEV buffer (20 m*M* HEPES pH 8.0, 200 m*M* NaCl, 40 m*M* imidazole, 1 m*M* TCEP) using a PD-10 column (GE Healthcare) and incubated with 1 mg of TEV protease per 15 mg of eluted protein. The protease-treated eluate was run over nickel-chelating resin (GE Healthcare) pre-equilibrated with HEPES crystallization buffer (20 m*M* HEPES pH 8.0, 200 m*M* NaCl, 40 m*M* imidazole, 1 m*M* TCEP) and the resin was washed with the same buffer. The flowthrough and wash fractions were combined and concentrated to 10 mg ml⁻¹ by centrifugal ultrafiltration (Millipore) for crystallization trials. LP2179 was crystallized by mixing 200 nl protein solution with 200 nl crystallization solution in a sitting-drop format over a 50 µl reservoir volume using the nanodroplet vapor-diffusion method (Santarsiero *et al.*, 2002) with standard JCSG crystallization protocols (Lesley *et al.*, 2002). Crystals from two different crystallization conditions were used for data collection and structure determination. The crystallization reagent yielding a cube-like crystal (0.1 × 0.1 × 0.1 mm) used for MAD phasing consisted of 20.0%(*w*/*v*) PEG 6000 and 0.1 *M* Bicine pH 9.0 as the precipitant. A long rod-like crystal (0.3 × 0.1 × 0.1 mm) used for refinement was obtained using 0.2 *M* NaCl, 20.0%(*w*/*v*) PEG 8000 and 0.1 *M* CAPS pH 10.5. Crystallization was carried out at 277 K for both conditions. Glycerol was added to both crystals as a cryoprotectant to a final concentration of 15%(*v*/*v*). Initial screening for diffraction was carried out using the Stanford Automated Mounting system (SAM; Cohen *et al.*, 2002) at the Stanford Synchrotron Radiation Light-source (SSRL, Menlo Park, California, USA). Both sets of diffraction data were indexed in the orthorhombic space group $P2_12_12_1$ (Table 1). The oligomeric state of LP2179 was determined using a 0.8 × 30 cm² Shodex Protein KW-803 column (Thomson Instruments) pre-calibrated with gel-filtration standards (Bio-Rad).

2.2. Data collection, structure solution and refinement

Multiple-wavelength anomalous diffraction (MAD) data were collected at the Advanced Photon Source (APS, Argonne, Illinois, USA) on beamline 23-ID-D at wavelengths corresponding to the high-energy remote (λ_2), inflection (λ_3) and peak (λ_4) of a selenium MAD experiment. Higher resolution data from a different crystal were collected at the Advanced Light Source (ALS, Berkeley, California, USA) on beamline 8.2.2. The data sets were collected at 100 K using a MAR Mosaic 300 detector (APS) and an ADSC Quantum-315 CCD detector (ALS). The MAD data were integrated and reduced using *XDS* and then scaled with the program *XSCALE* (Kabsch, 1993). The higher resolution (λ_1) data were integrated and reduced using *MOSFLM* (Leslie, 1992) and then scaled with the program *SCALA* (Collaborative Computational Project, Number 4, 1994). Phasing of the MAD data was performed with *SOLVE* (Terwilliger & Berendzen, 1999; four selenium sites per asymmetric unit, mean FOM = 0.52) and automated model building was performed with *ARP/wARP* (Cohen *et al.*, 2004). The resulting model was used for model completion and refinement against the higher resolution (λ_1) data with *Coot* (Emsley & Cowtan, 2004) and *REFMAC* 5.2 (Murshudov *et al.*, 1999). Data reduction and refinement statistics are summarized in Table 1.

2.3. Validation and deposition

Analysis of the stereochemical quality of the model was accomplished using *AutoDepInputTool* (Yang *et al.*, 2004), *MolProbity* (Davis *et al.*, 2004), *SFCHECK* 4.0 (Collaborative Computational Project, Number 4, 1994) and *WHATIF* 5.0 (Vriend, 1990). Protein

quaternary-structure analysis used the *PISA* server (Krissinel & Henrick, 2007). Fig. 1(*c*) was adapted from an analysis using *PDBsum* (Laskowski *et al.*, 2005) and all other figures were prepared with *PyMOL* (DeLano Scientific). Atomic coordinates and experimental structure factors for LP2179 at 1.20 Å resolution have been deposited in the PDB under accession code 2iay.

3. Results and discussion

3.1. Overall structure

The crystal structure of LP2179 (Fig. 1*a*) was initially determined to 1.33 Å resolution using the multiple-wavelength anomalous dispersion (MAD) method and was further refined to 1.20 Å resolution using data collected from a different crystal. Data-collection, model and refinement statistics are summarized in Table 1. The final model includes 114 residues (*i.e.* the residual Gly0 from the expression tag followed by residues 1–113 of LP2179), one glycerol molecule, one chloride ion and 195 water molecules in the asymmetric unit. The side chains of Lys8, Lys59 and Lys86 were not modeled owing to poor electron density. The Matthews coefficient (V_M; Matthews, 1968) is 2.0 $Å^3$ Da^{-1} and the estimated solvent content is 37.2%. The Ramachandran plot produced by *MolProbity* (Davis *et al.*, 2004) shows that 98.2% and 100% of the residues are in favored and in favored and additionally allowed regions, respectively.

LP2179 forms a single domain composed of two antiparallel β-sheets packed against a long C-terminal helix H3 (Fig. 1). A second helix, H1, links strand β2 from the first β-sheet (order 127), which is assembled from the two N-terminal and the C-terminal β-strands, to the second β-sheet (order 3456) and packs parallel to H3. *Pre-SCOP* classifies LP2179 as a novel fold termed LP2179-like (http://www.mrc-lmb.cam.ac.uk/agm/pre-scop/999384.html). Analysis of the crystallographic packing of LP2179 using the *PISA* server (Krissinel & Henrick, 2007) and analytical size-exclusion chromatography in combination with static light scattering indicate that a monomer is the likely quaternary form.

3.2. Comparison with other structures

A search with *FATCAT* (Ye & Godzik, 2004) revealed a remote structural similarity of LP2179 to members of the YugN-like family (PF08868), which are characterized by a TBP-like fold (http://www.mrc-lmb.cam.ac.uk/agm/pre-scop/55944.html). Superposition of LP2179 onto ABC2387 (PDB code 2pww; U. A. Ramagopal, J. Freeman, C. Lau, R. Toro, K. Bain, L. Rodgers, J. M. Sauder, S. K. Burley & S. C. Almo, unpublished work), a YugN-like homolog from *Bacillus clausii*, clearly reveals that both proteins share the same fold and topology over all of the helices and strands β3–β5 from the second β-sheet (strands β3–β6; Fig. 2*a*). The structural similarity involves a main-chain r.m.s.d. of 2.5 Å over 81 residues, although the
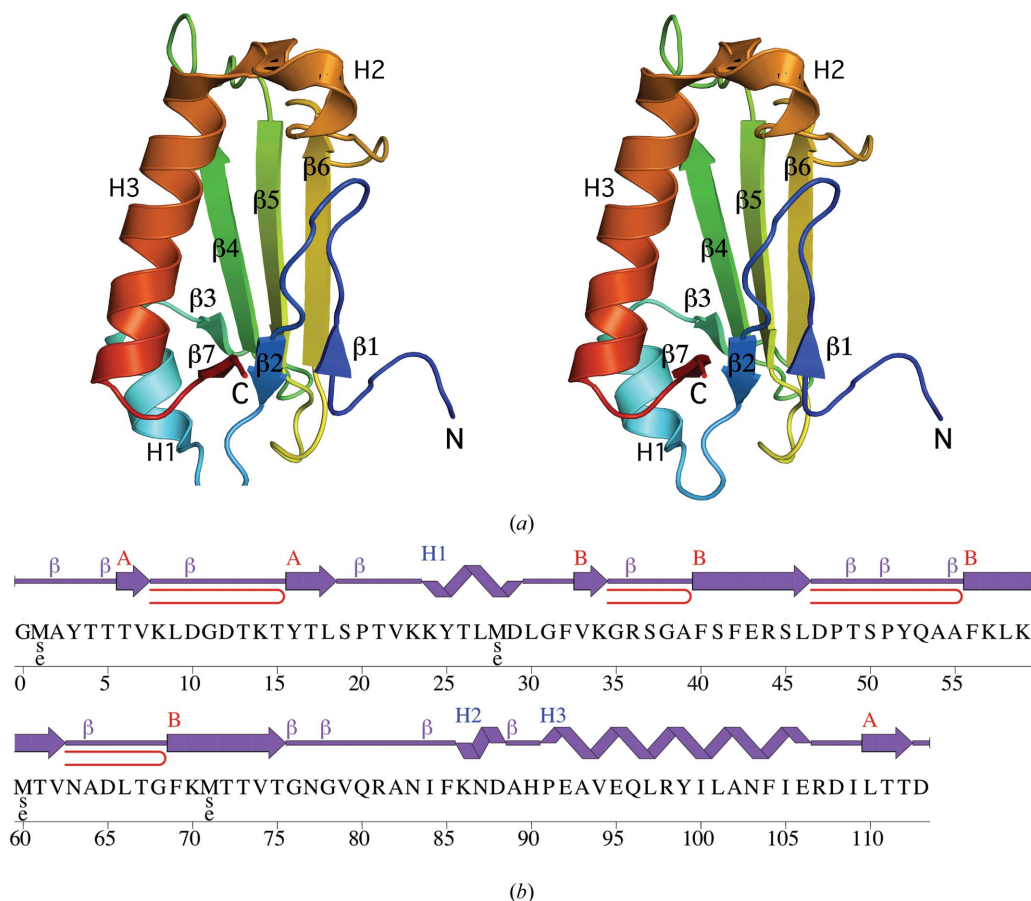

Figure 1
Crystal structure of LP2179 from *L. plantarum*. (*a*) Stereo ribbon diagram of the LP2179 monomer color-coded from the N-terminus (blue) to the C-terminus (red). Helices H1–H3 and β-strands (β1–β7) are indicated. (*b*) Diagram showing the secondary-structure elements of LP2179 superimposed on its primary sequence. The labeling of secondary-structure elements is in accord with *PDBsum* (http://www.ebi.ac.uk/pdbsum), where α-helices are sequentially labeled (H1, H2, H3 *etc*), β-strands are labeled (A, B, C *etc*.) according to the β-sheets to which they are assigned, β-turns and γ-turns are designated by Greek letters (β, γ) and β-hairpins by red loops. For LP2179, the α-helices (H1–H3), β-sheets (A, B) and β-turns (β) are indicated. Selenomethionine residues used for phasing are labeled MSE.

sequence identity is only 7%. Similar values are obtained for GK1089, another YugN homolog from *Geobacillus kaustophilus* (PDB code 2r5x), with an r.m.s.d. of 2.9 Å and 10% sequence identity over 87 aligned residues. Both YugN-like homologs show an interruption in the regular hydrogen-bonding pattern of strand β6 in the β-sheet, resulting in two shorter, collinear strands that hydrogen bond separately to β5. However, as the TBP-like fold is characterized by a β-α-β₄-α topology, the main topological difference between the two families involves the first β-sheet in LP2179, which is replaced in YugN-like homologs by a β-strand that forms part of the single β-sheet (Fig. 2a). The H2 helix, which is absent in both YugN-like and DUF1885 homologs, might constitute an additional difference, but owing to its short size (one helical turn) and its involvement in crystal

contacts (Asp88–Arg107′ and Phe85–Arg107′) it might not represent a biologically relevant conformation of this region in solution.

A search with *FFAS* (Jaroszewski *et al.*, 2005) showed no significant sequence similarity of LP2179 to any protein family other than PF08866. However, significant sequence similarity (*FFAS* score −11 with 20% sequence identity) was observed between ABC2387 and RBSTP2229, a member of the protein family PF08968 (DUF1885) from *B. stearothermophilus*. Like the YugN-like homologs, RBSTP2229 also exhibits a TBP-like fold. A structural superposition of ABC2387 (PDB code 2pww) with RBSTP2229 (PDB code 1t6a) shows a backbone r.m.s.d. of 2.8 Å over 57 residues. Over the same residue range, LP2179 has a backbone r.m.s.d. of 3.3 Å with RBSTP2229 (Fig. 2b). However, the length and orientation of helix
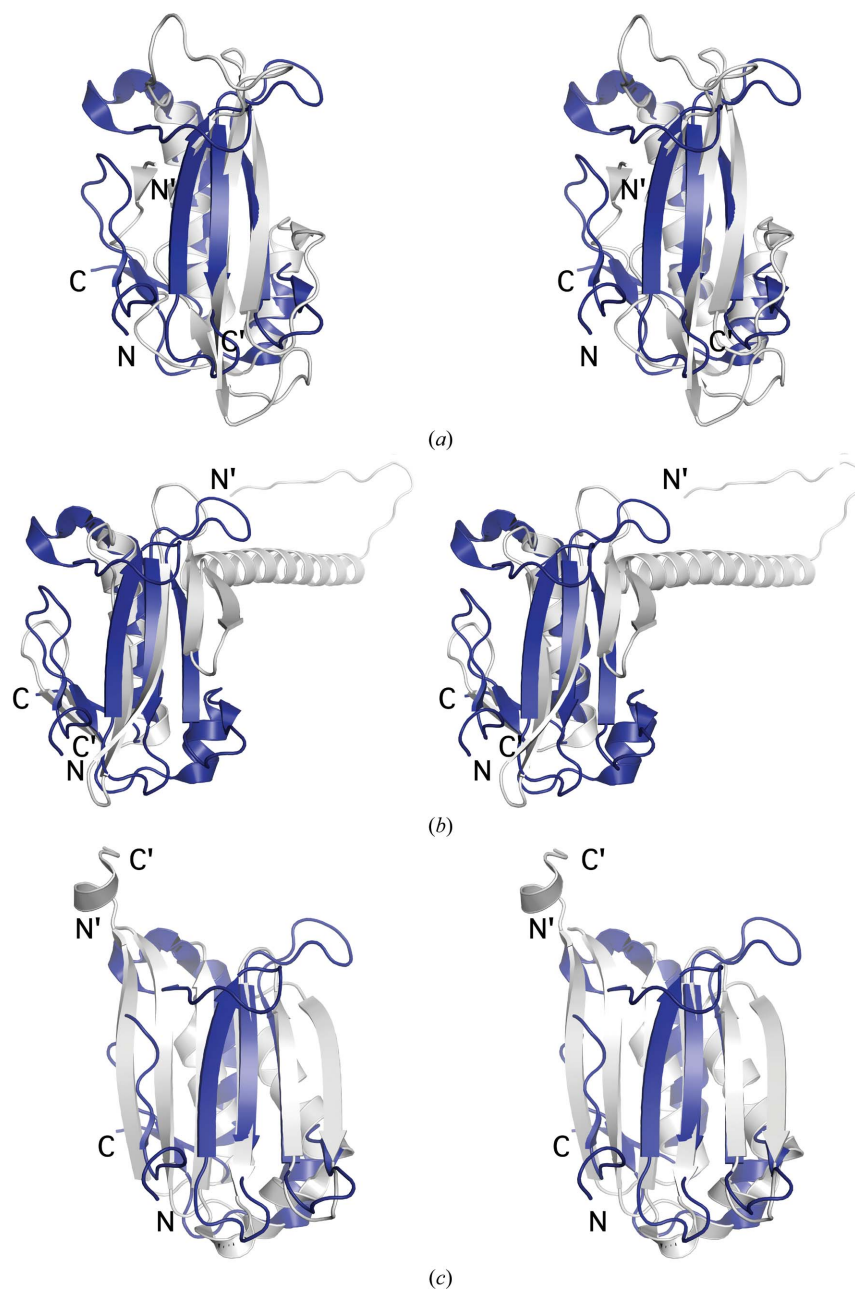

(a)

(b)

(c)

Figure 2
LP2179 exhibits structural similarity to members of the YugN-like family, DUF1185 and *S*-adenosylmethionine decarboxylases. Stereoviews of the structural superposition of LP2179 (PDB code 2iay, in blue) with (in gray) (a) a YugN-like homolog from *B. clausii* (PDB code 2pww), (b) a DUF1885 homolog from *B. stearothermophilus* (PDB code 1t6a) and (c) *S*-adenosylmethionine decarboxylase proenzyme (TM0655) from *Thermotoga maritima* (PDB code 1vr7). N- and C-termini are indicated for LP2179 and are indicated with primes (N′, C′) for other structures.

H1 in RBSTP2229 (pointing outwards from the structure instead of packing against the central β-sheet) differs substantially from that observed in ABC2387 and LP2179, while the subsequent β-strand is positioned differently with respect to helix H3 in all three structures (Figs. 2a and 2b). Among these TBP-like variants, LP2179 is unique in that the N- and C-terminal β-strands are combined to form an additional β-sheet that is situated between the central β-sheet and helix H3. However, both YugN-like and DUF1885 homologs display shorter variants of this secondary-structure element in the same region (YugN-like homologs contain a single β-strand; DUF1185 forms a C-terminal hairpin), raising the possibility that this region might represent a locus in this family for structural and possibly functional drift (Krishna & Grishin, 2005).

Structural similarities of LP2179 to prokaryotic S-adenosyl-methionine decarboxylases (AdoMetDCs; EC 4.1.1.50) were also observed. Superposition of LP2179 onto the AdoMetDC from *Thermotoga maritima* results in a backbone r.m.s.d. of 3.3 Å over 82 residues with 3% sequence identity (Fig. 2c). Similar values (an r.m.s.d of 3.3 Å over 67 residues with 3% sequence identity) were obtained for the AdoMetDC from *Aquifex aeolicus* (PDB code 2iii). As with the YugN-like homologs, prokaryotic AdoMetDCs share a similar fold and topology as LP2179 that includes the main β-sheet ($\beta3$–$\beta6$) and helices (H1–H3) in addition to the C-terminal β-strand ($\beta7$) of LP2179. The main differences involve the arrangement of the N- and C-terminal β-strands in prokaryotic AdoMetDCs that hydrogen bond to form a single six-stranded antiparallel β-sheet, as opposed to the two separate sheets in LP2179, and a C-terminal helix that is absent in LP2179 (Fig. 2c).

Structural comparison between these four Pfam families reveals the conservation of a core β-α-β_4-α (TBP-like) fold with β-strand additions at the N- or C-terminus or both. In LP2179, a strand is added at both the N- and C-termini, while YugN-like homologs contain an extra β-strand at the N-terminus (topology β_2-α-β_4-α) and PF08968 homologs contain an additional β-strand at the C-terminus that follows a circular permutation of the core fold (topology α-β_4-α-β_2). AdoMetDCs contain an additional β-strand at the C-terminus that hydrogen bonds to the N-terminal strand to form an antiparallel six-stranded β-sheet (topology β-α-β_4-α-β).

It is widely accepted that protein structure is more conserved than amino-acid sequence, suggesting that structural relationships between proteins might provide information that is not available from sequence alone (see review by Kolodny *et al.*, 2006). Both the PF08866 (DUF1831) and PF08868 (YugN-like) protein families are currently functionally uncharacterized. AdoMetDC is a pyruvoyl-dependent amino-acid decarboxylase that is involved in methionine metabolism and is essential for polyamine biosynthesis (Pegg *et al.*, 1998). The structure of prokaryotic AdoMetDC proenzyme (Toms *et al.*, 2004) reveals that despite the lack of any detectable sequence similarity between the eukaryotic and prokaryotic forms of the enzyme (13% sequence identity), the two structures can be super-imposed with an r.m.s.d. of 2.0 Å for 156 backbone residues. The catalytic site residues are also conserved (Toms *et al.*, 2004).

The AdoMetDC proenzyme undergoes an autocatalytic intra-molecular self-cleavage reaction that generates a pyruvoyl group in a loop between two β-strands ($\beta3$ and $\beta4$ in Fig. 2c). Although the catalytic residues (Ser and Glu) of the AdoMetDC proenzyme are not conserved in LP2179 and YugN-like or Pfam08968 homologs, sequence alignment reveals the conservation of charged and aromatic residue clusters between LP2179 and YugN-like homologs (Fig. 3). In the respective structures, these clusters occur along the first two strands and intervening loop of the central β-sheet ($\beta3$ and $\beta4$ in Figs. 1a and 2c) surrounding the AdoMetDC catalytic site and may serve a similar functional role.

3.3. Genomic neighborhood analysis

The genomic neighborhood (http://string.embl.de) of LP2179 shows a high degree of confidence in a predicted functional association with cysteine desulfurase (LP2180, score 0.81) and methylthioadenosine nucleosidase (LP2181, score 0.64). Cysteine desulfurase (EC 2.8.1.7) catalyzes the production of alanine from cysteine, while methylthioadenosine nucleosidase (EC 3.2.2.16) also participates in the metabolism of amino groups. These two enzymes are found in the genomic context or neighborhood of every member of the DUF1831 family, supporting a role for DUF1831 in amino-acid metabolism.

In Gram-positive bacteria, such as the *Bacillus* genus, amino-acid metabolism is directly coupled to several other metabolic pathways, including trans-sulfuration, polyamine synthesis and recycling, the activated methyl cycle and quorum sensing (Lebeer *et al.*, 2007). As previously indicated, AdoMetDC is a central regulator of these pathways. Modified amino acids, such as homocysteine, or their catabolic products, such as polyamines, can serve both pathogenic and probiotic roles. In pathogenic bacteria, polyamines and homo-cysteine are involved in biofilm formation (Shah & Swiatlo, 2008; Abraham, 2006), with polyamines also being implicated in bacterio-cin production and protection from acid and oxidative stress (Shah & Swiatlo, 2008). The probiotic role of lactobacilli has been well documented (Ryan *et al.*, 2008); their antimicrobial activity results from the production of bacteriocins and antifungal peptides (De Vuyst & Leroy, 2007). Further work will be required to determine whether the fold similarities observed between the *Bacillus* protein families described in this paper translate into similarities in function and whether this function might involve a probiotic role.

The availability of more DUF1831 sequences and structures might shed light on the evolutionary history of this intriguing protein family.

```
Q88V95_LACPL  MAYTTTVKLDG----DTKTYTLSPTVKKYTLMDLGFVKGRSGAFSFERSLDPTSPYQAAF  56
Q5WFD8_BACSK  ----MSLKFPD---TGLEEKEVAFSIVNHAAKSLGFIHVDQWDYERVMFDYKIVHHEGTF  53
Q5L106_GEOKA  ----MSLKFEN---TGLENQTVELSRLDDIMERLGFVRAAQWDYERVTYDRKYVVKEGTY  53

Q88V95_LACPL  KLKMTVNADLTGFKMTTVTGNGVQRANIFKNDAHPEAVEQ-----LRYILANFIERDILT  111
Q5WFD8_BACSK  YLRVPA-YAVKC-EIPRPSTIVQIMTPILGKYYYPHGVEY-EGETFPQAVIDKCNNKLAL  110
Q5L106_GEOKA  YLRVQG-YAIEC-NVDSRYALIKLLTPIMGKHYYPHGVEYGDDEHFPSSLVSQCQNVLAQ  111

Q88V95_LACPL  TD--------- 113
Q5WFD8_BACSK  LAKTIKAEWE- 120
Q5L106_GEOKA  VKSELEKIKEE 122
```

Figure 3
Sequence alignment of LP2179 and members of the YugN-like family. UniProt abbreviations are as follows: Q88V95_LACPL, gene locus lp_2179 from *L. plantarum*; Q5WFD8_BACSK, gene locus ABC2387 from *B. clausii*; Q5L106_GEOKA, gene locus GK1089 from *G. kaustophilus*. Residues are shaded by identity (black) and similarity (gray).

structural communications

The information presented here, in combination with further biochemical and biophysical studies, should yield valuable insights into the functional role of LP2179. Models for LP2179 homologs can be accessed at http://www1.jcsg.org/cgi-bin/models/get_mor.pl?key=2iayA.

Additional information about the protein described in this study is available from TOPSAN (Krishna *et al.*, 2010) http://www.topsan.org/explore?PDBid=2iay.

4. Conclusions

The first structural representative of the DUF1831 family reveals a potential new fold with remote similarities to TBP-like structures. This similarity, in combination with genomic context analysis, leads us to propose an involvement in amino-acid metabolism that might also be extended to two other families of unknown function.

References

Abraham, W. R. (2006). *Curr. Med. Chem.* **13**, 1509–1524.

Cohen, A. E., Ellis, P. J., Miller, M. D., Deacon, A. M. & Phizackerley, R. P. (2002). *J. Appl. Cryst.* **35**, 720–726.

Cohen, S. X., Morris, R. J., Fernandez, F. J., Ben Jelloul, M., Kakaris, M., Parthasarathy, V., Lamzin, V. S., Kleywegt, G. J. & Perrakis, A. (2004). *Acta Cryst.* D**60**, 2222–2229.

Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* D**50**, 760–763.

Davis, I. W., Murray, L. W., Richardson, J. S. & Richardson, D. C. (2004). *Nucleic Acids Res.* **32**, W615–W619.

De Vuyst, L. & Leroy, F. (2007). *J. Mol. Microbiol. Biotechnol.* **13**, 194–199.

Diederichs, K. & Karplus, P. A. (1997). *Nature Struct. Biol.* **4**, 269–275.

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

Finn, R. D., Tate, J., Mistry, J., Coggill, P. C., Sammut, S. J., Hotz, H. R., Ceric, G., Forslund, K., Eddy, S. R., Sonnhammer, E. L. & Bateman, A. (2008). *Nucleic Acids Res.* **36**, D281–D288.

Jaroszewski, L., Rychlewski, L., Li, Z., Li, W. & Godzik, A. (2005). *Nucleic Acids Res.* **33**, W284–W288.

Kabsch, W. (1993). *J. Appl. Cryst.* **26**, 795–800.

Klock, H. E., Koesema, E. J., Knuth, M. W. & Lesley, S. A. (2008). *Proteins*, **71**, 982–994.

Kolodny, R., Petrey, D. & Honig, B. (2006). *Curr. Opin. Struct. Biol.* **16**, 393–398.

Krishna, S. S. & Grishin, N. V. (2005). *Bioinformatics*, **21**, 1308–1310.

Krishna, S. S., Weekes, D., Bakolitsa, C., Elsliger, M.-A., Wilson, I. A., Godzik, A. & Wooley, J. (2010). *Acta Cryst.* F**66**, 1143–1147.

Krissinel, E. & Henrick, K. (2007). *J. Mol. Biol.* **372**, 774–797.

Laskowski, R. A., Chistyakov, V. V. & Thornton, J. M. (2005). *Nucleic Acids Res.* **33**, D266–D268.

Lebeer, S., De Keersmaecker, S. C., Verhoeven, T. L., Fadda, A. A., Marchal, K. & Vanderleyden, J. (2007). *J. Bacteriol.* **189**, 860–871.

Lesley, S. A. *et al.* (2002). *Proc. Natl Acad. Sci. USA*, **99**, 11664–11669.

Leslie, A. G. W. (1992). *Jnt CCP4/ESF–EACBM Newsl. Protein Crystallogr.* **26**.

Matthews, B. W. (1968). *J. Mol. Biol.* **33**, 491–497.

Murshudov, G. N., Vagin, A. A., Lebedev, A., Wilson, K. S. & Dodson, E. J. (1999). *Acta Cryst.* D**55**, 247–255.

Pegg, A. E., Xiong, H., Feith, D. J. & Shantz, L. M. (1998). *Biochem. Soc. Trans.* **26**, 580–586.

Ryan, K. A., Jayaraman, T., Daly, P., Canchaya, C., Curran, S., Fang, F., Quigley, E. M. & O'Toole, P. W. (2008). *Lett. Appl. Microbiol.* **47**, 269–274.

Santarsiero, B. D., Yegian, D. T., Lee, C. C., Spraggon, G., Gu, J., Scheibe, D., Uber, D. C., Cornell, E. W., Nordmeyer, R. A., Kolbe, W. F., Jin, J., Jones, A. L., Jaklevic, J. M., Schultz, P. G. & Stevens, R. C. (2002). *J. Appl. Cryst.* **35**, 278–281.

Shah, P. & Swiatlo, E. (2008). *Mol. Microbiol.* **68**, 4–16.

Terwilliger, T. C. & Berendzen, J. (1999). *Acta Cryst.* D**55**, 849–861.

Tickle, I. J., Laskowski, R. A. & Moss, D. S. (1998). *Acta Cryst.* D**54**, 243–252.

Toms, A. V., Kinsland, C., McCloskey, D. E., Pegg, A. E. & Ealick, S. E. (2004). *J. Biol. Chem.* **279**, 33837–33846.

Vriend, G. (1990). *J. Mol. Graph.* **8**, 52–56.

Yang, H., Guranovic, V., Dutta, S., Feng, Z., Berman, H. M. & Westbrook, J. D. (2004). *Acta Cryst.* D**60**, 1833–1839.

Ye, Y. & Godzik, A. (2004). *Nucleic Acids Res.* **32**, W582–W585.