Post-translational modification in the archaea: structural characterization of multi-enzyme complex lipoylation

Mareike G. Posner*[1], Abhishek Upadhyay*[1], Susan J. Crennell*,[†], Andrew J. A. Watson[‡2], Steve Dorus*[3], Michael J. Danson*,[†], Stefan Bagby*[4]

*Department of Biology & Biochemistry, University of Bath, UK

[†]Centre for Extremophile Research, University of Bath, UK

[‡]Department of Chemistry, University of Bath, UK

[1]These authors contributed equally to this work

[2]Present address: Department of Chemistry, University of Canterbury, New Zealand

[3]Present address: Department of Biology, Syracuse University, USA

Page heading title: Structure of an archaeal lipoylation system

[4]To whom correspondence should be addressed: s.bagby@bath.ac.uk; +44 (0)1225 386436

Synopsis

Lipoylation, the covalent attachment of lipoic acid to 2-oxoacid dehydrogenase multi-enzyme complexes, is essential for metabolism in aerobic bacteria and eukarya. In *Escherichia coli*, lipoylation is catalysed by lipoate protein ligase (LplA) or by lipoic acid synthetase (LipA) and lipoyl(octanoyl) transferase (LipB) combined. Whereas bacterial and eukaryotic LplAs comprise a single, two-domain protein, archaeal LplA function typically involves two proteins, LplA-N and LplA-C. In the thermophilic archaeon *Thermoplasma acidophilum*, LplA-N and LplA-C are encoded by overlapping genes in inverted orientation (*lpla-c* is upstream of *lpla-n*). The structure of *Thermoplasma acidophilum* LplA-N is known, but the structure of LplA-C and its role in lipoylation are unknown. We have determined the structures of the substrate-free LplA-N+LplA-C complex and the dihydrolipoyl acyltransferase lipoyl domain (E2lipD) that is lipoylated by LplA-N+LplA-C, and carried out biochemical analyses of this archaeal lipoylation system. Our data reveal the following: LplA-C is disordered but folds upon association with LplA-N; LplA-C induces a conformational change in LplA-N involving substantial shortening of a loop that could repress catalytic activity of isolated LplA-N; the adenylate binding region of LplA-N+LplA-C includes two helices rather than the purely loop structure of varying order observed in other LplA structures; LplA-N+LplA-C and E2lipD do not interact in the absence of substrate; LplA-N+LplA-C undergoes a conformational change (the details of which are currently undetermined) during lipoylation; LplA-N+LplA-C can utilize octanoic acid as well as lipoic acid as substrate. The elucidated functional inter-dependence of LplA-N and LplA-C is consistent with their evolutionary co-retention in archaeal genomes.

Key words: binding-induced folding, lipoate protein ligase, lipoyl domain, NMR spectroscopy, protein-protein interaction, X-ray crystallography

Abbreviations used: BCOADHC, branched-chain 2-oxoacid dehydrogenase complex; CCD, charge coupled device; DSS, 2,2-dimethyl-2-silapentane-5-sulfonate, sodium salt; E1, 2-oxoacid decarboxylase; E2, dihydrolipoyl acyltransferase; E2lipD, dihydrolipoyl acyltransferase lipoyl domain; E3, dihydrolipoamide dehydrogenase; GCS, glycine cleavage system; HSQC, heteronuclear single-quantum coherence; IPTG, isopropyl β-D-1-thiogalactopyranoside; LA, lipoic acid; LipA, lipoic acid synthetase; LipB, lipoyl(octanoyl) transferase; LplA, lipoate protein ligase; LPT, LplA-like lipoyltransferase; MPD, 2-methyl-2,4-pentanediol; OADHC, 2-oxoacid dehydrogenase multi-enzyme complex; OA, octanoic acid; OGDHC, 2-oxoglutarate dehydrogenase complex; PDB, protein data bank; PDHC, pyruvate dehydrogenase complex; RMSD, root mean square deviation.

Introduction

Aerobic metabolism of 2-oxoacids and C1 metabolism are dependent on lipoic acid (LA) in a highly conserved manner [1]. LA is an essential co-factor of the 2-oxoacid dehydrogenase complexes (OADHCs), which include the pyruvate (PDHC), 2-oxoglutarate (OGDHC) and branched-chain 2-oxoacid (BCOADHC) dehydrogenase complexes, and of the glycine cleavage system (GCS). OADHCs comprise multiple copies of three proteins: 2-oxoacid decarboxylase (E1), dihydrolipoyl acyltransferase (E2), and dihydrolipoamide dehydrogenase (E3). E2 comprises E2 lipoyl domain(s) (E2lipD), a peripheral subunit-binding domain (PSBD), and a catalytic domain. E2lipD is the post-translational modification target: LA is covalently attached to E2lipD via an amide linkage to the ε-amino group of a specific lysine located at the tip of a β-turn. Once attached, the lipoyl moiety acts as a swinging arm that shuttles substrates/intermediates between the active sites of E1, E2 and E3. In the GCS, LA is attached to a lysine of the H protein that is structurally homologous to E2lipD.

In *Escherichia coli*, E2lipD lipoylation is catalysed by lipoic acid synthetase (LipA) and lipoyl(octanoyl) transferase (LipB) or, if LA is present in the medium/environment, by lipoate protein ligase (LplA) [2-4]. LipB and LipA work in tandem: LipB catalyses the covalent attachment of octanoic acid (OA, derived from the fatty acid biosynthetic pathway) to E2lipD, and LipA introduces sulphur atoms at the C6 and C8 positions. LplA is typically a single polypeptide comprising an N-terminal domain (approximately 250 residues) that has an LA binding site, and a smaller C-terminal domain (approximately 90 residues) [5]. In *E. coli* and *Oryza sativa*, LplA can catalyse both steps of the lipoylation process: conversion of LA to lipoyl-AMP (lipoate adenylation) and subsequent covalent attachment of the lipoyl moiety to E2lipD (lipoate transfer) [6, 7]. Mammals achieve the equivalent process using two enzymes, lipoate activating enzyme and a LplA-like lipoyltransferase (LPT) [8]. In yeast, four enzymes are involved in lipoylation, including homologues of LipA, LipB and LplA, but there are major differences compared to lipoylation in *E. coli* [9]. Whilst LA metabolism in bacteria other than *E. coli* is not fully understood, details elucidated to date indicate that numerous variations exist [10-12].

Available structures include *E. coli* LplA [13] (PDB entries 1X2G, 1X2H, 3A7A, 3A7R), Streptococcal LplAs (PDB: 2P0L, 1VQZ) and a mammalian LPT [8] (PDB: 2E5A, 3A7U). LplA N-terminal domain belongs to the α/β class of proteins [3, 5] and is structurally homologous and evolutionarily related to the central catalytic domain of biotin protein ligase and class II aminoacyl-tRNA synthetase [14, 15]. LplA C-terminal domain comprises three α-helices and two 3_{10}-helices packed against a three-stranded β-sheet [5]. Bovine LPT resembles LplA in that it comprises a larger N-terminal domain and a smaller C-terminal domain, both with similar folds to their respective *E. coli* counterparts. The overall conformation of lipoyl-AMP-bound LPT is, however, stretched relative to unliganded *E. coli* LplA due to rotation of the C-terminal domain by about 180° with respect to the N-terminal domain [8]. A similar rotation of the C-terminal domain relative to its apo orientation was observed in the crystal structure of lipoyl-AMP-bound *E. coli* LplA [13]. In the same structure, it was noted that two important loops also undergo conformational change upon lipoate adenylation: the adenylate binding loop (residues 165-184), which is either partially disordered or not close to the active site in apo-LplA, covers the adenylate of lipoyl-AMP and interacts intimately with it, and the lipoate binding loop (residues 69-76) is pulled towards lipoyl-AMP [13]. In the reaction scheme proposed by Fujiwara et al. [13], these conformational changes allow *E. coli* LplA to accommodate the lipoate acceptor domain/protein (E2lipD or H protein) and hence to catalyse the lipoate transfer step. These authors note, however, that in their LplA-apoH complex crystal structure with octanoyl-AMP rather than lipoyl-AMP, the distance between octanoyl-AMP and the acceptor lysine ($Lys_{ApoH}64$) is too great for initiation of lipoyl transfer. In the same study, Fujiwara et al. also observed that, unlike *E. coli* LplA, bovine apo-LPT adopts the same relative N- and C-terminal domain orientations as lipoyl-AMP-bound LPT [13].

Archaeal LplA studies have been conducted largely in *Thermoplasma acidophilum*, a species that possesses genes encoding individual proteins that resemble the N- and C-terminal domains of non-archaeal LplA. We term these gene products LplA-N and LplA-C. Structures of *T. acidophilum* LplA-N in unliganded (PDB: 2ARS and 2C7I), lipoyl-AMP-bound (PDB: 2ART), lipoic acid-bound (PDB: 2C8M), and ATP-bound (2ARU) forms exhibit the same overall fold as the non-archaeal LplA N-terminal domain [3, 16]. It was shown in crystal soaking experiments (one day soaks) that LplA-N can catalyse lipoate adenylation to form lipoyl-AMP [16], but LplA-N is unable to catalyse lipoate transfer *in vitro* and an accessory protein was suggested [3]. We subsequently showed that LplA-N requires LplA-C to carry out lipoylation (corroborated using complementation assays in *E. coli* [17]) and that lipoylation occurs *in vivo* [18].

Comparative genomic analyses across 115 archaeal genomes (Dorus, Bagby et al, Syracuse University and University of Bath, unpublished data) show that archaeal species capable of lipoylation retain either the LplA or LipA-LipB system with 81% (61 of 75 species) retaining LplA. Despite the evolutionary predominance of LplA in the archaea, and the fact that LplA-C is essential for lipoate transfer, no mechanistic information exists concerning coordination of LplA-N and LplA-C function. Here we have used structural and biochemical methods to investigate the role of LplA-C. We present structures of the *T. acidophilum* LplA-N+LplA-C complex and of E2lipD, show that LplA-C folding is driven by association with LplA-N and that LplA-C induces localised conformational change in LplA-N, and use NMR to monitor LplA-N+LplA-C interactions with lipoic acid/ATP and E2lipD, and to monitor E2lipD lipoylation.

Experimental

Expression and purification *of T. acidophilum* LplA-N+LplA-C and *T. acidophilum* E2lipD. pET19b-*lpla* and pET24a-*ctd* [18] were co-transformed into *E. coli* BL21(DE3) and expression was induced with 0.25 mm IPTG at 16°C overnight. Harvested cells were sonicated, the lysate was centrifuged at 21000 g for 40 min and LplA-N+LplA-C complex was purified using His MultiTrap™ FF and His MultiTrap™ HP columns (GE Healthcare Life Sciences, Amersham, UK). The final LplA-N+LplA-C complex purity was >95% as judged by SDS-PAGE. *T. acidophilum* E2lipD was expressed and purified as described previously [18].

Expression and purification of *E. coli* LplA and E2lipD. *E. coli* LplA and E2lipD were expressed using TM202 and pET11c plasmids. Expression in BL21(DE3) was induced with 0.5 mM IPTG for 3 h at 37°C. Cells were sonicated and proteins were purified using HiTrap QFF with a 0-0.5 M NaCl gradient in 20 mM Tris-HCl, pH 7.5.

Crystallization of *T. acidophilum* LplA-N+LplA-C complex, data collection and structural analysis. *T. acidophilum* LplA-N+LplA-C was exchanged into 10 mM Tris-HCl, pH 7.5, concentrated to 20 mg/ml and centrifuged at 13000 g for 20 min at 4°C. Sitting drop, vapour diffusion crystallization screens were set up at 18°C using Molecular Dimensions screens with a Phoenix robot (Art Robbins Instruments, Sunnyvale, California). Crystals in 40 % (v/v) MPD, 0.1 M sodium acetate (pH 4.6), and 0.02 M $CaCl_2$ were suitable for X-ray diffraction without further cryo-protectant. Diffraction data were collected at Diamond Light Source (Oxon, UK) on an ADSC Q315 CCD detector on station IO2 (λ = 0.9795 Å). 360 images were collected at an oscillation angle of 1°. Raw data images were processed using HKL2000 [19].

Model building. Molecular replacement using BALBES [20] was followed by model building with Coot [21] and rounds of refinement using Refmac5, part of CCP4 [22]. Other software included Molprobity [23] and Procheck [24].

Structural analysis. Hydrogen bonds and ionic interactions were evaluated with Contact CCP4 [22], ProtorP [25] and PISA [26]. Molecular graphics figures were prepared in PyMOL (The PyMOL Molecular Graphics System, Version 1.2r3pre, Schrödinger, LLC).

NMR spectroscopy. ^{15}N-labelled LplA-C and ^{15}N- and ^{15}N^{13}C-labelled E2lipD were produced by expression in M9 minimal medium supplemented with 1 g/L $^{15}NH_4Cl$ as the sole nitrogen source or 1 g/L $^{15}NH_4Cl$ and 2 g/L ^{13}C glucose, respectively. His-tagged proteins were purified as described previously [18]. Most E2lipD and all LplA-N+LplA-C NMR data were acquired at 37°C or 50°C on a 600 MHz Varian Unity INOVA spectrometer with an ambient temperature probe, processed using NMRPipe/NMRDraw [27] and analyzed using CCPN Analysis [28]. ^{15}N-edited NOESY and ^{13}C-edited NOESY spectra of E2lipD were acquired on an 800 MHz Varian Inova spectrometer at the MRC Biomedical NMR Centre, Mill Hill. ^{1}H, ^{15}N and ^{13}C chemical shifts were referenced to DSS [29]. Structures were calculated as described previously [30]. ^{1}H-^{15}N HSQC spectra of uniformly ^{15}N-labelled LplA-C, both with and without unlabelled LplA-N, were recorded in 20 mM Tris pH 7.5, 150 mM NaCl. E2lipD spectra were recorded in 50 mM HEPES pH 7.5, 50 mM NaCl. All ^{1}H-^{15}N HSQC spectra in this study were recorded with 128 increments in the nitrogen dimension, unless otherwise stated.

NMR titration of LplA-N+LplA-C with lipoic acid, ATP and Mg^{2+}, then with E2lipD. Lipoic acid (racemic mixture unless stated otherwise), ATP and Mg^{2+} were titrated in combination against an NMR sample containing LplA-N+LplA-C (unlabelled LplA-N and uniformly ^{15}N-labelled LplA-C) in 20 mM Tris pH 7.5, 150 mM NaCl. The molar ratio of LplA-N+LplA-C to lipoic acid at each titration point was 1:0, 1:0.25, 1:0.50, and 1:1.25; a ^{1}H-^{15}N HSQC spectrum (32 scans, 128 min recording time) was recorded at each titration point. Unlabelled E2lipD was then added to the same NMR sample with ratios of LplA-N+LplA-C to E2lipD of 1:0.25, 1:0.50, and 1:1.25; a ^{1}H-^{15}N HSQC spectrum (32 scans, 128 minutes recording time) was recorded after each E2lipD addition.

NMR titration of LplA-N+LplA-C with E2lipD, and then with lipoic acid, ATP and Mg^{2+}. Unlabelled E2lipD was added to an NMR sample containing LplA-N+LplA-C (unlabelled LplA-N and uniformly ^{15}N-labelled LplA-C with no lipoic acid/ATP/Mg^{2+}) in 20 mM Tris pH 7.5, 150 mM NaCl. The molar ratio of LplA-N+LplA-C to E2lipD at each titration point was 1:0, 1:0.25, 1:0.50, 1:0.75, and 1:1; a ^{1}H-^{15}N HSQC spectrum (24 scans, 160 increments, 120 min recording time) was recorded at each titration point. Lipoic acid (2 mM), ATP (2.5 mM) and Mg^{2+} (1 mM) were then added together; a ^{1}H-^{15}N HSQC spectrum (24 scans, 160 increments) was recorded both directly after this addition and on the following day after overnight storage at 4 °C.

NMR titration of E2lipD with LplA-N+LplA-C, lipoic acid, and ATP. Catalytic quantities of LplA-N+LplA-C were added to ^{15}N-labelled 1.1 mM E2lipD in four steps (molar ratio of E2lipD to LplA-N+LplA-C of 1:0.0025, 1:0.005, 1:0.0075 and 1:0.01), followed by two additions of lipoic acid to a final concentration of 2.25 mM, and then by two additions of ATP to a final concentration of 2.25 mM (1.5 mM Mg^{2+} was present in initial NMR sample). A ^{1}H-^{15}N HSQC spectrum was recorded at each titration point (8 scans, 34 min recording time). Chemical shift perturbations were calculated as a weighted average of ^{1}H and ^{15}N chemical shift changes, $\Delta\delta_{av}$ ($\Delta\delta_{av}$ (ppm) = $[(\Delta\delta^2HN + \Delta\delta^2N/25)/2]^{1/2}$) [31].

Lipoylation/octanoylation activity assay. The electrophoretic mobility of E2lipD before and after lipoylation/octanoylation was analysed by non-denaturing PAGE as described previously [18]. In the lipoylation/octanoylation assays and *T. acidophilum*/*E. coli* enzyme cross-reactivity assays, the ratio of lipoylated to non-lipoylated E2lipD was quantified by mass spectrometry.

Synthesis of octanoyl-AMP. This was carried out as described previously [32, 33].

Model of LplA-N+LplA-C:E2lipD complex. In order to model a possible end-point of conformational change in LplA-N+LplA-C that permits lipoylation of E2lipD, the relative orientation of LplA-N and LplA-C was first changed to that observed between the N- and C-terminal domains of *E. coli* LplA in its complex with apo H protein and octanoyl-AMP (PDB:

3A7A). E2lipD was then docked with the reoriented LplA-N+LplA-C complex using ClusPro 2.0 [34, 35]; in the resulting models, E2lipD orientation and acceptor lysine ($Lys_{E2lipD}42$) position were compared to those in 3A7A of apo H protein and $Lys_{ApoH}64$ respectively. Models comparable to 3A7A (i.e. with $Lys_{E2lipD}42$ in proximity to and oriented towards the LplA-N active site) were selected and their quality assessed using QMean [36]. The model from this subset with the highest QMean score was selected as a representative structure.

PDB accession codes. Coordinates and structure factor file for the *T. acidophilum* LplA-N+LplA-C structure and coordinates for the E2lipD structures are in the Protein Data Bank (PDB) under accession codes 3R07 and 2L5T.

Results

Structure of *T. acidophilum* LplA-C. Despite the fact that LplA-C is essential for lipoylation by archaeal LplA [17, 18], the functional and structural relationship between LplA-N and LplA-C is poorly understood. We examined whether LplA-N and LplA-C exist independently or form a stable complex by first studying LplA-C structure without LplA-N. In crystallization screens, LplA-C showed a high propensity to precipitate. Poor chemical shift dispersion, variable peak intensity, and low peak count (approximately 50 peaks observed versus 84 expected based on the LplA-C amino acid sequence) in ^{1}H-^{15}N HSQC NMR spectra showed that LplA-C is disordered and heterogeneous over a range of pH values (pH 6-8) and NaCl concentrations (50-150 mM NaCl) (Figure 1A). Upon stepwise addition of unlabelled LplA-N to ^{15}N labelled LplA-C (final LplA-N+LplA-C molar ratio of 1:1), the observed dramatic increase in dispersion, homogeneity and number of LplA-C ^{1}H-^{15}N HSQC peaks indicates that LplA-C undergoes LplA-N binding-induced folding (Figure 1B). Seventy-six distinct backbone amide NH peaks were observed in the LplA-N-bound LplA-C ^{1}H-^{15}N HSQC spectrum; this close correspondence with the expected total of 84 peaks indicates that LplA-N-bound LplA-C adopts a single dominant conformation on average, and indirectly supports the presence of a single predominant LplA-N+LplA-C complex conformation in solution. The LplA-N-induced LplA-C fold, and by inference the LplA-N+LplA-C complex, is stable to at least 50°C (Figure 1B).

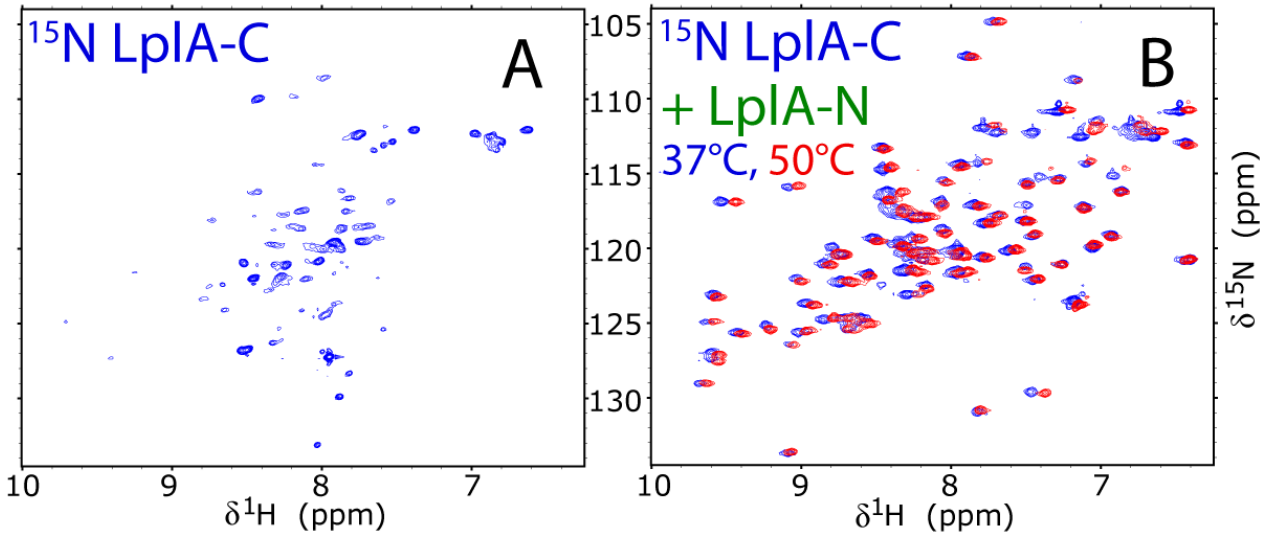

Figure 1 ^{1}H-^{15}N HSQC NMR spectra of *T. acidophilum* LplA-C in the absence and presence of LplA-N. (**A**) ^{1}H-^{15}N HSQC spectrum at 37°C recorded on uniformly ^{15}N-labelled LplA-C by itself and (**B**) in a 1:1 molar ratio with unlabelled LplA-N; an overlay of ^{1}H-^{15}N HSQC spectra recorded at 37°C (blue) and 50°C (red) is shown.

***T. acidophilum* LplA-N+LplA-C X-ray crystal structure: comparison with other LplAs.** Subsequent to the LplA-C NMR studies described above, crystallization screens of co-expressed

LplA-N and LplA-C produced LplA-N+LplA-C complex crystals in 40% (v/v) MPD, 0.1 M sodium acetate (pH 4.6), 0.02 M $CaCl_2$. The LplA-N+LplA-C structure (Figure 2A) was determined to 2.7 Å resolution by molecular replacement with LplA-N (PDB: 2ARS) (Table 1). The overall fold of LplA-N [3, 16] is maintained in the presence of LplA-C, as confirmed using Dali [37] (Table 2). The β-strands in LplA-N are β1 (residues 1-7), β2 (35-39), β3 (44-47), β4 (68-71), β5 (79-81), β6 (85-93), β7 (121-123), β8 (128-131), β9 (144-154), and β10 (157-165); the α-helices are α1 (13-27), α2 (59-65), α3 (98-117), α4 (181-183), α5 (184-194), α6 (206-223), α7 (232-246), and α8 (249-254); and the 3_{10}- or η-helices are η1 (52-56), η2 (171-176) and η3 (198-202). The β-strands in LplA-C are β11 (2-10), β12 (15-23), and β13 (26-35), and the α-helices are α9 (42-52), α10 (58-68), and α11 (79-86), where the LplA-C strand and helix numbering continues from the LplA-N numbering but the residue numbers start anew at the LplA-C N-terminus.

LplA-N+LplA-C is structurally similar to single polypeptide apo-LplAs from *Streptococcus pneumoniae* and *E. coli*, including similar domain orientations (Figure 2C; Table 2). Bovine (*Bos taurus*) LPT, however, has a different arrangement of domains in both apo and lipoyl-AMP-bound forms, as does lipoyl-AMP-bound *E. coli* LplA; in these cases, the C-terminal domain has undergone a 180° rotation (Figure 2C) [13]. The structures of *E. coli* LplA C-terminal domain and *T. acidophilum* LplA-C agree well, with both forming a canopy above the tunnel-like entry to the active site. With respect to the inverted gene orientation in *T. acidophilum* (*lpla-c* is upstream of *lpla-n* with a TATA box upstream of *lpla-c* but no *cis*-regulatory sequence in the proximity of *lpla-n*) [18], it is important to note that the LplA-C C-terminus and LplA-N N-terminus are located at opposite ends of the LplA-N+LplA-C complex, approximately 56 Å apart (Figure 2A), confirming that LplA-N and LplA-C are made as separate polypeptides.

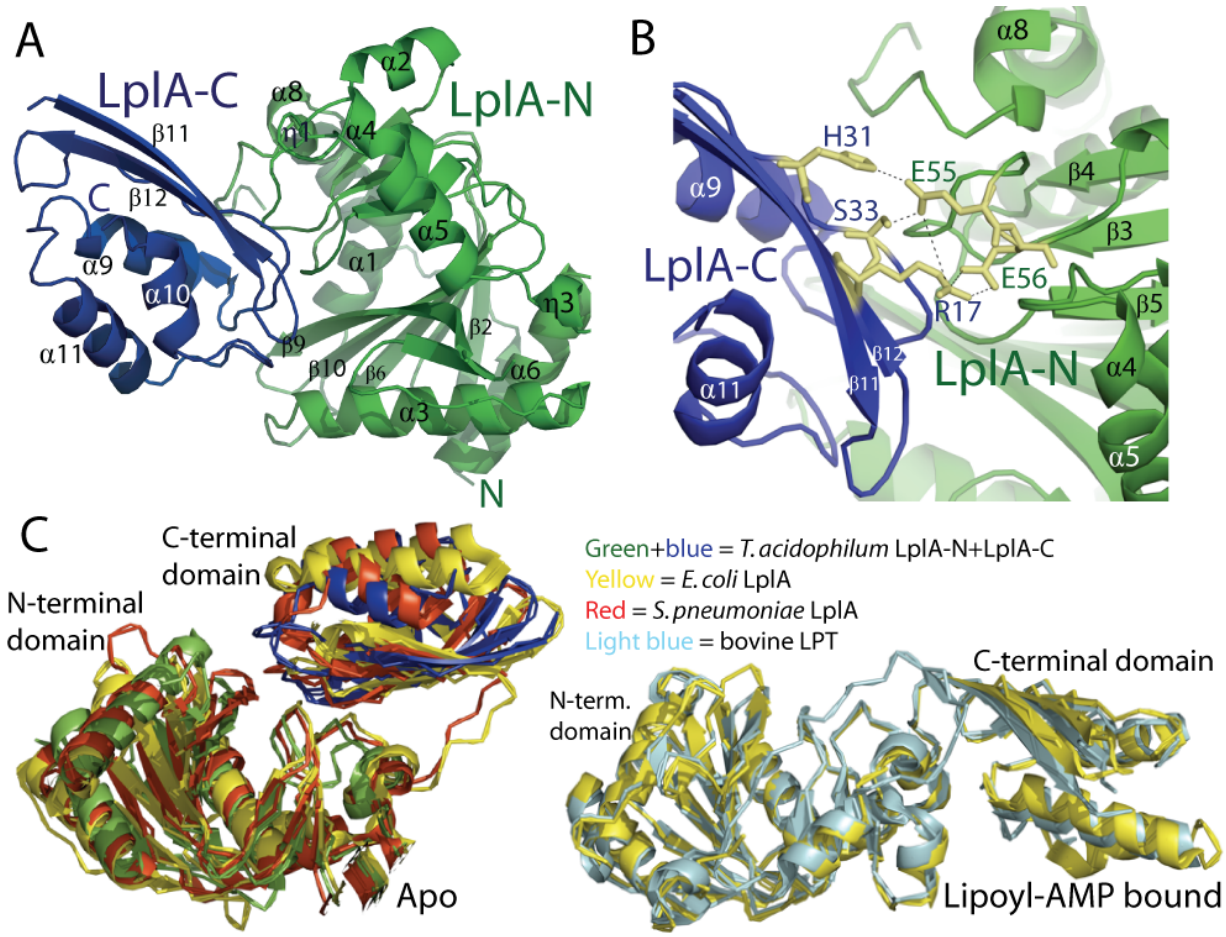

Figure 2 Structure of the *T. acidophilum* LplA-N+LplA-C complex. (**A**) LplA-N (green) and LplA-C (blue). Secondary structure elements, and the LplA-N N-terminus and LplA-C C-terminus, are indicated. (**B**) Some of the ionic interaction and hydrogen bond network interactions (indicated by dashed lines) between LplA-N (green) and LplA-C (blue) with side chains in pale yellow (different orientation of the complex to that in A). (**C**) Overlays of unliganded/apo *T. acidophilum* LplA-N+LplA-C (PDB: 3R07), *E. coli* LplA (PDB: 1X2G) and *S. pneumoniae* LplA (PDB: 1VQZ) (left) and of lipoyl-AMP bound *E. coli* LplA (PDB: 3A7R) and bovine LPT (PDB: 3A7U) (right). The N-terminal domain orientation is the same in both overlays.

A. *Crystallographic data*	
Resolution (Å)	2.7
Space group	$P3_121$
Unit cell parameters	
a = b (Å)	118.57
c (Å)	72.92
α = β (°)	90
γ (°)	120
B. *Merging statistics*	
No. of reflections	15063
Average redundancy	7.1 (7.1)
I/σI	37.1 (6.1)
Completeness (%)	96.0 (98.5)
R-merge (%)	7.7 (43.2)
C. *Refinement statistics*	
RMSD bond length (Å)	0.015
RMSD bond angle (°)	1.547
Residues in disallowed regions (%)	0
Mean B value ($Å^2$)	51.561
R-factor (%)	19.8
R-free (%)	25.4

Table 1 Data collection and structural refinement statistics. Numbers in parentheses represent statistics at the highest resolution (2.7-2.8 Å).

Organism	**Ligands**	**PDB code**	**Dali search with 2ARS; Z-score (RMSD, Å)**	**Dali search with *T.acidophilum* LplA-N+LplA-C; Z-score (RMSD, Å)**
T. acidophilum	None	2ARS	45.3	39.1 (0.6)
	None	2C7I	43.2 (0.7)	38.2 (0.7)
	Lipoic acid	2C8M	43.2 (0.7)	
	ATP	2ARU	44.12 (0.3)	38.9 (1.0)
	Lipoyl-AMP	2ART	44.12 (0.5)	38.9 (1.0)
S. pneumoniae	None	1VQZ	26.7 (2.5)	LplA-N: 26.9 (2.8) LplA-C: 11.4 (1.5)
E. coli	None	1X2G		LplA-N: 25.9 (2.8) LplA-C: 8.8 (2.3)
	Lipoic acid	1X2H		LplA-N: 25.4 (2.9) LplA-C: 8.7 (2.3)
	Lipoyl-AMP	3A7R		LplA-N: 28.5 (2.2) LplA-C: 9.7 (2.1)
	Octanoyl-AMP, ApoH	3A7A		LplA-N: 28.5 (2.3) LplA-C: 9.7 (2.1)
Bos taurus		2E5A		LplA-N: 27 (2.1) LplA-C: 7.7 (2.5)

Table 2 Summary of Dali similarity searches**.** The values as determined for *T. acidophilum* LplA-N and LplA-C are listed individually and have been obtained from pairwise Dali comparisons. 2ARS was compared only with the most similar non-*T. acidophilum* LplA, which is *S. pneumoniae* LplA (PDB: 1VQZ).

The LplA-N+LplA-C interface has a buried surface area of 805 $Å^2$ compared to 993 $Å^2$ between the N- and C-terminal domains of the closest single protein homologue, *S. pneumoniae* LplA [3, 18]. The LplA-N+LplA-C interface involves 50 residues and includes 12 hydrogen bonds plus five salt bridges involving three pairs of residues [26]. These include a five-residue network that forms salt bridges ($Glu_{LplA\text{-}N}56$-$Arg_{LplA\text{-}C}17$, $Glu_{LplA\text{-}N}55$-$His_{LplA\text{-}C}31$) and four hydrogen bonds ($Glu_{LplA\text{-}N}55$-$His_{LplA\text{-}C}31$, $Glu_{LplA\text{-}N}55$-$Ser_{LplA\text{-}C}33$, $Glu_{LplA\text{-}N}56$-$Arg_{LplA\text{-}C}17$) (Figure 2B). In the

corresponding location, *S. pneumoniae* LplA has an inter-domain three-residue (Arg45-Asp284-His46) network with two inter-domain ionic interactions involving Asp284, and *E. coli* LplA has no obvious ionic interaction. In addition, the LplA-N+LplA-C interface has a substantial hydrophobic component with about 25 hydrophobic residues contributing to the interface.

LplA-C-induced conformational change of LplA-N: "capping" loop and adenylate binding loop. LplA-N undergoes a substantial local structural rearrangement upon binding LplA-C. In isolated LplA-N (i.e. without LplA-C), β8 consists of residues 138-141 and is connected to β9 (residues 144-154) by a short β-turn, while strands β7 and β8 are connected by a long loop consisting of residues 124-137 (orange and labelled as the "capping loop" in Figure 3). Interestingly, this loop makes several contacts with lipoyl-AMP in 2ART, and may play a role in ensuring that isolated LplA-N is catalytically inert. This region is reorganized in the LplA-N+LplA-C structure such that residues 128-131 form β8, and a short turn comprising residues 125-127 connects β7 to β8, while β8 is connected to β9 by a disordered loop comprising residues 132-142, for most of which electron density is not observed (Figure 3). This structural shift seems to be facilitated by the similarity of the two motifs that alternate as β8 - residues 128-131 are DVSI, while residues 138-141 are DIMA. It is noteworthy that in *E. coli* LplA, β8 is a fixed motif, connected to β7 and β9 by short loops on either side.

The lipoate binding loop adopts the same conformation with and without LplA-C (Figure 3), whereas in LplA-N+LplA-C a substantial portion of the region corresponding to the adenylate binding loop strikingly forms contiguous α-helices (α4 and α5; residues 181-183 and 184-194). In other LplA structures, this region is either a loop or is largely disordered such that electron density is absent. In several structures of isolated LplA-N, for example, much of the adenylate binding loop region is disordered (e.g. 2ARS, 2C7I – both unliganded, 2C8M – lipoic acid bound, 2ART – lipoyl-AMP bound), although it is noteworthy that following an electron density gap in the structures 2ARS, 2ART and 2C8M, there is a nascent α-helix that overlaps with part of LplA-N+LplA-C α5 (e.g. region shown in purple in Figure 3). In the structure of unliganded *E. coli* LplA (1X2G), the adenylate binding loop occupies a similar position to LplA-N+LplA-C helices α4 and α5, whereas there is again missing electron density in unliganded bovine LPT. The adenylate binding loops of lipoyl-AMP bound *E. coli* LplA and bovine LPT overlap closely with each other and are shifted towards the active site relative to the adenylate binding loops of the respective unliganded enzymes.

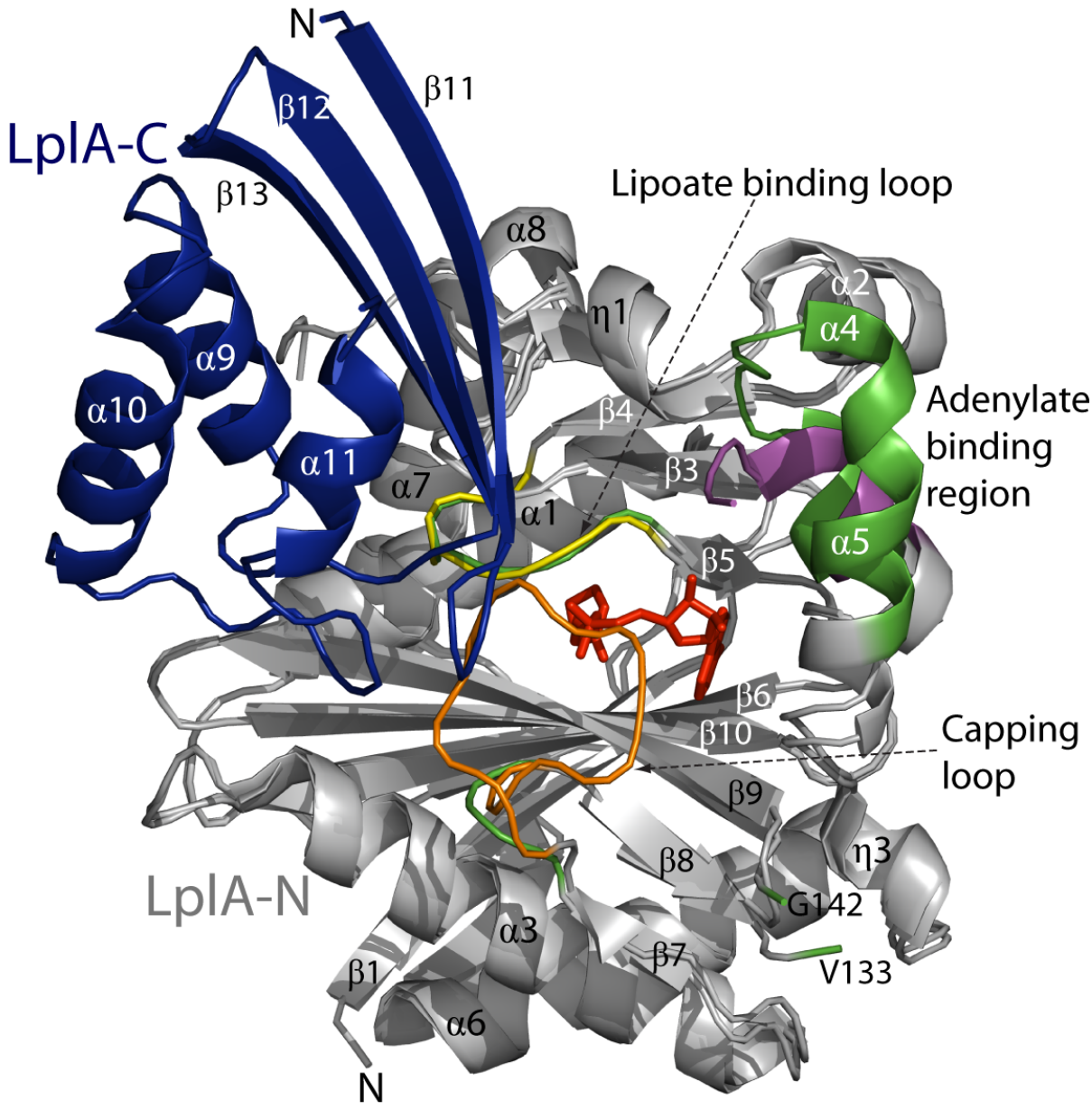

Figure 3 LplA-C-induced conformational change of LplA-N. The structure of the LplA-N+LplA-C complex (3R07) is shown with LplA-N predominantly in grey and LplA-C in blue. The structure of isolated (i.e. without LplA-C) LplA-N (2ART) is superimposed on the LplA-N+LplA-C complex and is also shown predominantly in grey. Structural features of 3R07 are highlighted in green and structural features of 2ART are highlighted in yellow, orange and purple. Lipoyl-AMP from 2ART is shown in red. Secondary structure elements of 3R07 are labelled. The lipoate binding loop conformations are almost identical in 3R07 (green) and 2ART (yellow). In the absence of LplA-C, residues 124-137 form a long loop (orange – labelled as "capping loop"); the corresponding loop (green) is much shorter in the LplA-N+LplA-C complex. Instead, the subsequent loop is much longer in the LplA-N+LplA-C complex than in isolated LplA-N (electron density is lacking between V133 and G142 in the LplA-N+LplA-C complex). Residues identified by structure-based alignment to form the adenylate binding region are shown in green (3R07) and purple (2ART), although in 2ART much of the adenylate binding region lacks defined electron density.

LplA-N+LplA-C interaction with lipoyl-AMP and E2lipD. We used NMR titrations to investigate LplA-N+LplA-C interactions, monitoring the ^{1}H-^{15}N HSQC spectrum of ^{15}N-labelled LplA-C in complex with unlabelled LplA-N. In one titration (Titration 1), lipoic acid, ATP and Mg^{2+}

were added together to LplA-N+^{15}N-LplA-C, then E2lipD was added (Figure 4A and 4B). Upon addition of lipoic acid, ATP and Mg^{2+} (and presumably therefore upon formation of the lipoyl-AMP-bound form of LplA-N+LplA-C), fifteen of the seventy-six distinct backbone amide NH peaks in the ^{1}H-^{15}N HSQC spectrum of LplA-N+^{15}N-LplA-C were significantly broadened (intermediate timescale exchange), five exhibited slow exchange (two peaks observed per backbone amide NH), four underwent a chemical shift change, four exhibited both chemical shift change and broadening, and at least one peak increased significantly in intensity (Figure 4A). Upon subsequent titration of E2lipD into the LplA-N+^{15}N-LplA-C NMR sample, most of the peaks that had been perturbed (Figure 4A) reverted to a state the same as or close to that observed prior to addition of lipoic acid, ATP and Mg^{2+} (Figure 4B), consistent with lipoyl transfer to E2lipD and hence consumption of substrate.

In a separate titration (Titration 2), unlabelled E2lipD was first titrated into a LplA-N+^{15}N-LplA-C NMR sample to a final molar ratio of 1:1 LplA-N+LplA-C to E2lipD, producing essentially no change in the LplA-N+^{15}N-LplA-C ^{1}H-^{15}N HSQC spectrum. Significant perturbation of the ^{1}H-^{15}N HSQC was observed, however, when lipoic acid, ATP and Mg^{2+} were then added (Figure 4C): at least thirty peaks underwent chemical shift change, five peaks were broadened, three showed both broadening and chemical shift change, one showed chemical shift change plus increased intensity, and two new peaks appeared in the vicinity of an initially absent peak at around 8.8 ppm, 116.3 ppm that is generally weak/absent in LplA-N+^{15}N-LplA-C ^{1}H-^{15}N HSQC spectra. Notably, at least 75% of the peaks perturbed in Titration 1 were also perturbed in Titration 2. When a ^{1}H-^{15}N HSQC spectrum was recorded the next day on the same sample, some peaks had reverted towards a pre-lipoic acid/ATP/Mg^{2+} position/intensity, but some had not (Figure 4D).

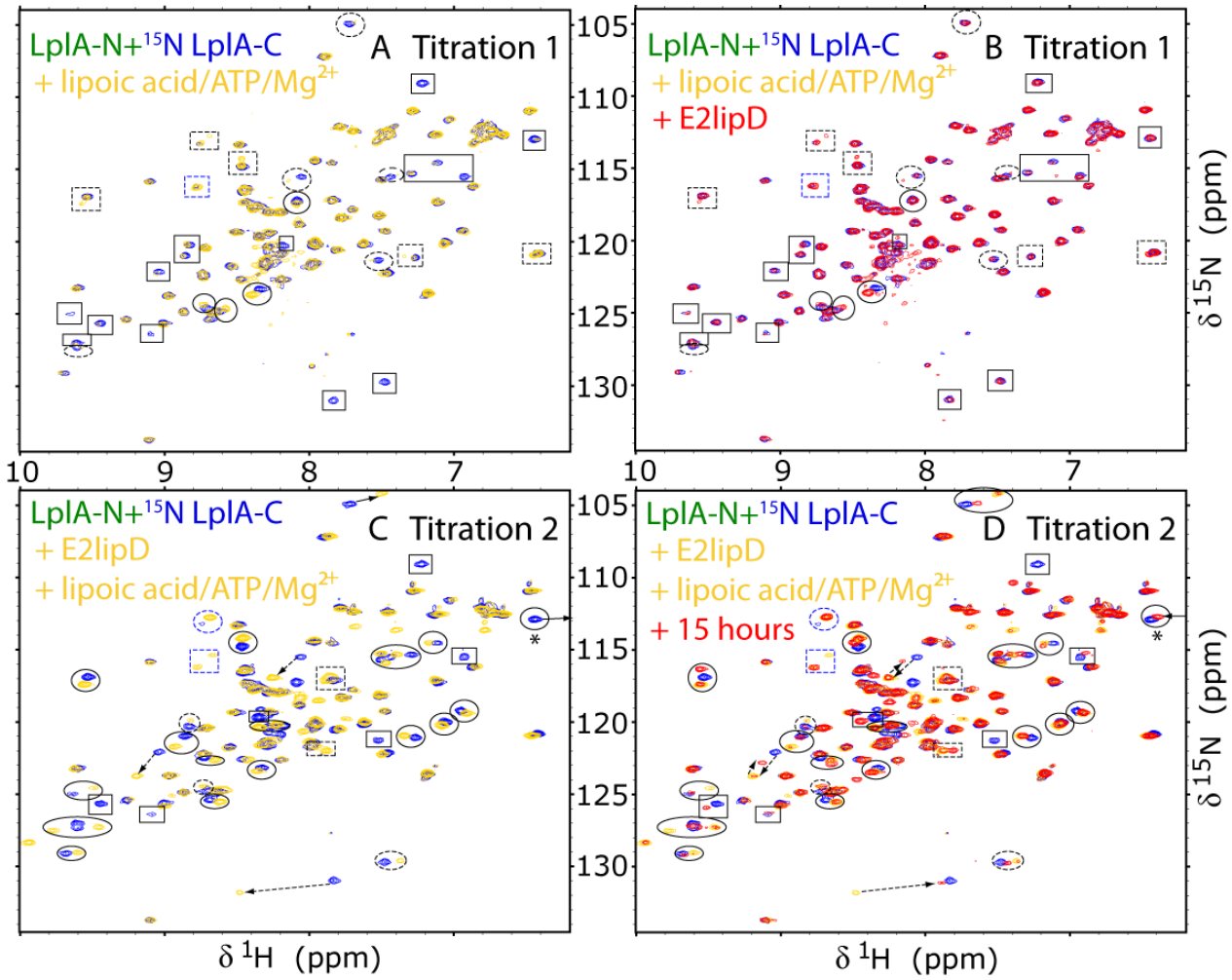

Figure 4 ^{1}H-^{15}N HSQC spectra of LplA-N+^{15}N-LplA-C: titration with lipoic acid, ATP and Mg^{2+}, and with E2lipD. The results of two titrations are shown. In (**A**) and (**B**), lipoic acid, ATP, and Mg^{2+} were added before E2lipD. In (**C**) and (**D**), E2lipD was added before lipoic acid, ATP and Mg^{2+}. (**A)** Lipoic acid, ATP and Mg^{2+} were added together in step-wise fashion to an NMR sample containing LplA-N+LplA-C (unlabelled LplA-N and uniformly ^{15}N-labelled LplA-C) in 20 mM Tris, pH 7.5, 150 mM NaCl. A ^{1}H-^{15}N HSQC spectrum was recorded at each titration point. The molar ratio of LplA-N+LplA-C complex to lipoic acid at each titration point was 1:0, 1:0.25, 1:0.50, 1:1.25. The initial spectrum is shown in blue, the final spectrum in yellow. LplA-C peaks perturbed upon addition of lipoic acid, ATP and Mg^{2+} are highlighted as follows: intermediate exchange (broadening) - solid rectangle; slow exchange (two peaks) - broken line rectangle; chemical shift change – solid ellipse; chemical shift change and broadening - broken line ellipse; increase in intensity – blue broken line rectangle. (**B**) E2lipD was then added to the same NMR sample as (A) with molar ratios of LplA-N+LplA-C to E2lipD of 1:0.25, 1:0.50, 1:1.25. The final spectrum is shown in red with the same peaks highlighted as in (A). Most of the peaks perturbed in (A) reverted to the pre-lipoic acid/ATP/Mg^{2+} state, which is again represented by blue peaks. (**C**) In the second titration, E2lipD was added in four steps (molar ratio of LplA-N+LplA-C to E2lipD at each titration point was 1:0, 1:0.25, 1:0.50, 1:0.75, and 1:1) with essentially no change in the LplA-C ^{1}H-^{15}N HSQC spectrum (not shown). Lipoic acid (2 mM), ATP (2.5 mM) and Mg^{2+} (1 mM) were then added together. The spectrum after E2lipD addition is shown in blue, and the spectrum after lipoic acid, ATP and Mg^{2+} addition is shown in yellow. Perturbed LplA-C peaks are highlighted using the same scheme as in (A), except that some of the larger chemical shift changes are indicated by an arrow (dotted line arrow where the connection between the shifted peak and the original peak is tentative), newly appearing peaks are indicated by a blue dashed rectangle, and chemical shift change plus increased intensity is indicated by a blue dashed ellipse. The peak at around 6.4 ppm, 113 ppm labelled with an asterisk moved to 6.1 ppm, 113 ppm. Peaks subject to smaller chemical shift changes have not been highlighted. (**D**) ^{1}H-^{15}N HSQC spectrum of the same sample as in (C), recorded after leaving the sample overnight at 4 °C.

Substrate promiscuity and recognition of E2 lipoyl domains by LplA-N+LplA-C. As its bipartite nature may affect substrate recognition and specificity, we tested *T. acidophilum* LplA-N+LplA-C activity with different acceptor domains and substrates, including LA, OA and octanoyl-AMP, with *E. coli* LplA serving as a positive control. In gel shift lipoylation assays [18] with LA, co-expressed LplA-N+LplA-C and a 1:1 mixture of individually expressed and purified LplA-N and LplA-C showed equal activity. LplA-N+LplA-C showed activity with both OA and octanoyl-AMP, in agreement with previous findings that LplAs can catalyse the formation and transfer of octanoyl-AMP [38]. As judged by mass spectrometry, E2lipD modification efficiency by LplA-N+LplA-C with OA substrate was 20-30% of that with LA substrate.

E2lipD cross-reactivity was analysed next. E2lipD residues both N-terminal (-) and C-terminal (+) of the target lysine are important for efficient lipoylation [39]. Previous large-scale sequence alignments identified a highly conserved Asp(-1), hydrophobic residues at +1, +5 and -4, Glu/Asp enrichment at -3 and +4, and Ser/Ala at +7 [40]. Glu(-3) and Gly(-16) are involved in LplA recognition. Sequence alignment with *E. coli* E2lipDs and *E. coli* H protein shows that Gly(-16) is conserved in *T. acidophilum* E2lipD. Glu(-3) is conserved in *E. coli* E2lipDs and H protein, whereas *T. acidophilum* E2lipD has Met(-3); residues other than Glu at -3 reduce lipoylation efficiency in *E. coli* [41]. Correspondingly, *E. coli* LplA lipoylated *T. acidophilum* E2lipD with about 50% efficiency relative to *E. coli* E2lipD, and *T. acidophilum* LplA-N+LplA-C lipoylated *E. coli* E2lipD with about 15-20% efficiency relative to *T. acidophilum* E2lipD.

NMR analysis of *T. acidophilum* E2lipD structure and lipoylation. In order to facilitate modelling studies of the complete *T. acidophilum* lipoylation system (see below), *T. acidophilum* E2lipD structure was determined by NMR. *T. acidophilum* E2lipD is similar overall to other lipoyl domains (DaliLite Z-score of 6.8 and r.m.s.d. over all Cα atoms of 2.7 Å versus *E. coli* E2lipD (PDB:1QJO; 27% sequence identity)). In an NMR titration to monitor E2lipD lipoylation, *T. acidophilum* E2lipD ^{1}H-^{15}N HSQC did not change upon step-wise addition of catalytic quantities of LplA-N+LplA-C (final molar ratio 100 E2lipD: 1 LplA-N+LplA-C), nor upon addition of LA (final molar ratio approximately 1 E2lipD: 2 LA) (Figure 5). Upon subsequent addition of ATP (final molar ratio approximately 1 E2lipD: 2 ATP), however, several E2lipD backbone amide peaks underwent chemical shift perturbation (Figures 5A and 5B); the largest chemical shift

perturbations were observed for E2lipD residues 42-44 (Lys $_{E2lipD}$42 is the lipoylation target residue), followed by residues 9-10; the Thr$_{E2lipD}$40 peak was broadened. When mapped onto the E2lipD structure, the pattern of largest chemical shift perturbations (plus broadening for residue 40) indicates that lipoylation induces a localized conformational change in E2lipD (Figure 5C).

Total number of NOE restraints intraresidue sequential/med. range (*i* to *i*+1— 4) long range	648 146 270 232
Number of dihedral angle restraints	76
Number of hydrogen bond restraints	12
Rmsd for backbone atoms[a]	0.48 Å
Rmsd for non-hydrogen atoms[a]	1.11 Å
Average number of NOE violations >0.5Å (per structure)	0
Average number of dihedral angle violations >1^{o} (per structure)	4
Ramachandran plot statistics[b] Most favoured (%) Additional allowed (%) Generously allowed (%) Disallowed (%)	 81.4 15.2 3.2 0.1

Table 3 Structural statistics for the ensemble of NMR-derived structures of E2lipD. [a] The rmsd from the mean structure calculated over residues 2-5, 16-20, 27-29, 37-39, 44-46, 53-58, 64-66, 73-76.[b] Calculated with PROCHECK-NMR.

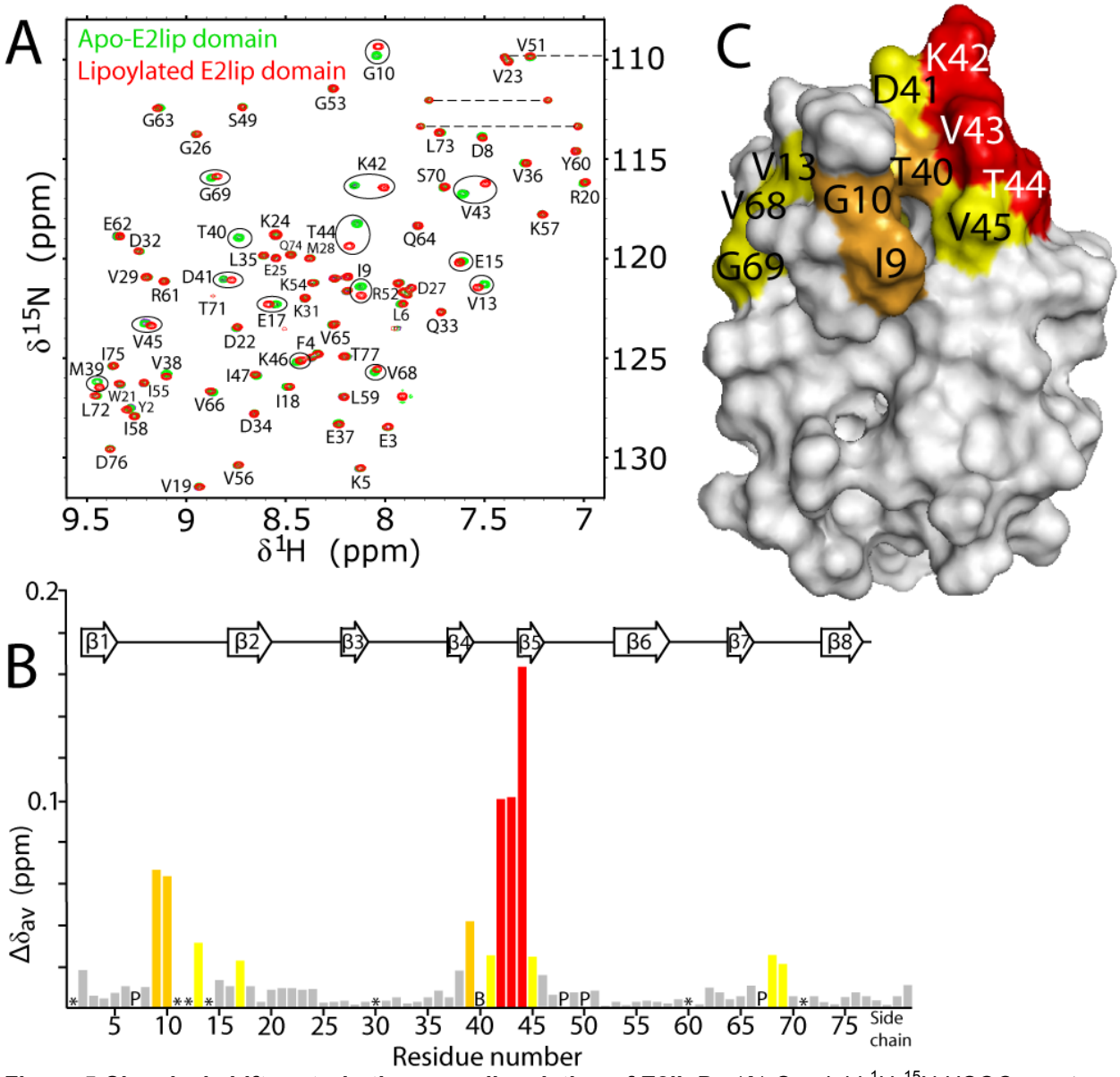

Figure 5 Chemical shift perturbation upon lipoylation of E2lipD. (**A**) Overlaid ^{1}H-^{15}N HSQC spectra of uniformly ^{15}N-labelled E2lipD from titration of ^{15}N-labelled 1.1 mM E2lipD with catalytic quantities of LplA-N+LplA-C in four steps (molar ratio of E2lipD to LplA-N+LplA-C of 1:0.0025, 1:0.005, 1:0.0075 and 1:0.01), followed by two additions of lipoic acid to a final concentration of 2.25 mM; the final spectrum from the titration to this point is shown in green (no change from original spectrum). ATP was then added to a concentration of 2.25 mM; the subsequent ^{1}H-^{15}N HSQC spectrum is shown in red. Peaks are labelled with amino acid assignments (assignments could not be made for E11, G12, T14, E30, Y60 and T71 due to lack of peaks in 3D spectra; unlabelled backbone NH peaks are assumed to arise from these residues or from the N-terminal 6His-tag residues; the G16 peak (^{1}H and ^{15}N chemical shifts of 8.22 ppm and 104.75 ppm, respectively) is omitted to allow greater overall clarity. Peaks showing the largest chemical shift changes upon addition of ATP are highlighted with an ellipse drawn around the pre-ATP (green) and post-ATP (red) peak positions (the peak due to T40 was broadened out of the spectrum upon E2lipD lipoylation). (**B**) Lipoylation-induced chemical shift changes (calculated using $\Delta\delta_{av}$ (ppm) = $[(\Delta\delta^{2}HN + \Delta\delta^{2}N/25)/2]^{1/2}$) plotted as a function of E2lipD residue number. ∗ indicates an unassigned residue, P indicates a proline residue (proline does not produce a signal in ^{1}H-^{15}N HSQC spectra), and B indicates a peak that was broadened out of the spectrum upon E2lipD lipoylation. Red = $\Delta\delta_{av}$ (ppm) ≥ 0.1 ppm, orange = $\Delta\delta_{av}$ (ppm) ≥ 0.04 ppm, and yellow = $\Delta\delta_{av}$ (ppm) ≥ 0.02 ppm. E2lipD secondary structure elements (β-strands) are indicated by arrows above the histogram. (**C**) E2lipD surface with most of the residues perturbed upon lipoylation highlighted using the same colour scheme as in (B).

Model of *T. acidophilum* LplA-N+LplA-C:E2lipD complex. In *E. coli* LplA, lipoate adenylation causes conformational changes, including 180º rotation of the C-terminal domain (Figures 2C and 6A), that prime the system for lipoyl transfer [13]. In *T. acidophilum* apo-LplA-N+LplA-C, LplA-C forms a canopy over the active site in a similar manner to the C-terminal domain of other LplAs, obstructing access to lipoate of $Lys_{E2lipD}42$, the lipoate acceptor residue (Figure 6A). Hence, a substantial change to the apo-LplA-N+LplA-C structure is required for lipoylation to occur. Using *E. coli* LplA:octanoyl-AMP:ApoH protein complex crystal structure as a template ([13]; Figure 6A), we have modelled a possible end-point of such a change with E2lipD incorporated (Figure 6B). The respective lipoate acceptor residues (*T. acidophilum* $Lys_{E2lipD}42$ and *E. coli* $Lys_{ApoH}64$) are in similar positions (Figure 4B). In our model, E2lipD bridges LplA-N and LplA-C, and LplA-C has undergone a rotation of approximately 120º. Thirty-five LplA-N residues, 8 LplA-C residues, and 34 E2lipD residues are involved in the interface (24).

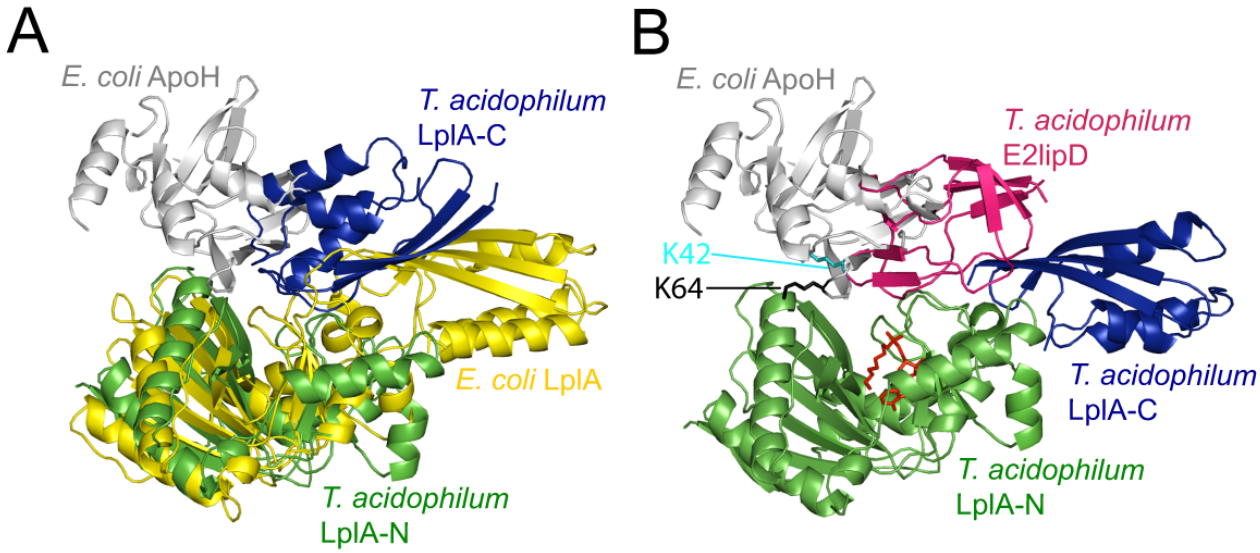

Figure 6 Lipoate protein ligase complexes with E2lipD or ApoH protein. (**A**) Superposition of *E. coli* octanoyl-5'-AMP-bound LplA (yellow) in complex with *E. coli* ApoH (grey) (PDB: 3A7A) and *T. acidophilum* LplA-N (green)+LplA-C (blue) (PDB: 3R07). (**B**) Comparison of the acceptor lysine residue positions in the *E. coli* LplA:ApoH complex and in the *T. acidophilum* LplA-N+LplA-C:E2lipD complex in which LplA-C has undergone a change in position and orientation relative to LplA-N+LplA-C. *E. coli* ApoH is positioned as in the LplA:ApoH complex but *E. coli* LplA has been omitted for clarity. The lipoyl acceptor residue of *E. coli* ApoH, $Lys_{ApoH}64$, is shown in black. *T. acidophilum* E2lipD is in magenta with its lipoyl acceptor residue, $Lys_{E2lipD}42$, in cyan. Octanoyl-5'-AMP (red) is positioned as in the *E. coli* LplA:ApoH complex.

Discussion

Biochemical data from our laboratory and elsewhere indicate that LplA-C is essential for lipoylation of E2 in *T. acidophilum* [17, 18], although LplA-N by itself can catalyse lipoate adenylation to form lipoyl-AMP [16]; it is still not clear, however, whether LplA-C enhances catalysis of lipoate adenylation, although it is known to be essential at least for lipoate transfer. Our genomic profiling, moreover, indicates that bipartite LplA-N+LplA-C is the evolutionarily predominant lipoylation system in the archaea. Despite these observations, structural and mechanistic information on LplA-C function has been lacking. The structures of *T. acidophilum* LplA-N+LplA-C and E2lipD presented here represent the first structural analysis of a complete archaeal lipoylation system and the first of a bipartite lipoate protein ligase. The structure of LplA-N was already known [3, 16], but the nature of LplA-N association with the functionally essential LplA-C was unknown [17, 18]. It was not clear, for example, whether LplA-C is always associated with LplA-N or only during lipoylation. We now know that LplA-C is probably not functional by itself as it is disordered and undergoes LplA-N-induced folding. The C-terminal domain of *E. coli* LplA, on the other hand, was found by limited proteolysis to be structurally stable [3]. We do not know, however, whether LplA-C retains its fold once it is released from LplA-N as we have been unable to establish a non-denaturing procedure to dissociate the LplA-N+LplA-C complex.

The current evidence indicates that the observed interface between LplA-N and LplA-C is a biological rather than crystal packing interface and that LplA-N and LplA-C exist permanently as a complex. This evidence includes the observations that isolated LplA-C is disordered, the LplA-N+LplA-C complex is stable to at least 50 °C (Figure 1), LplA-N and LplA-C associate strongly upon co-expression, there is an extensive buried hydrophobic surface between LplA-N and LplA-C, similar interfaces are observed in other LplA structures from several organisms, and LplA-N and LplA-C are functionally inter-dependent. There are thus mechanistic differences from the LipA-LipB system in which LipA and LipB operate sequentially, although we note that *E. coli* LipA and LipB have been found to form a tight non-covalent association with the E2 components of PDHC and OGDHC [4], presumably with resulting potential for greater processivity and for interaction between LipA and LipB themselves. The inherent robustness of *T. acidophilum* LplA, and presumably LplA from other thermophiles, makes these ligases attractive starting points for biotechnological and chemical biology applications such as have been demonstrated for *E. coli* LplA [42, 43].

The *T. acidophilum* LplA-N+LplA-C complex adopts the same fold and the same spatial arrangement of domains as the structurally characterised bacterial apo-LplAs (PDB: 1X2G, 2P0L, 1VQZ; Figure 2). LplA-N+LplA-C does, however, possess at least two distinctive local conformational features that could be functionally important. Firstly, in *E. coli* LplA and bovine LPT, the adenylate binding region is a loop, often at least partially disordered, whereas in LplA-N+LplA-C the equivalent region includes two contiguous α-helices (α4 and α5; Figure 3). We cannot say whether these helices persist at the optimum temperature (55 °C) for *T. acidophilum*, but their presence reduces the probability that the LplA-N adenylate binding region undergoes the same transition as the *E. coli* adenylate binding region upon lipoate adenylation which includes formation of a new β-strand anti-parallel to β13 of the C-terminal domain [13]. Secondly, LplA-C-induced conformational changes in LplA-N around strands β7 and β8 result in substantial shortening of a loop (that we label as the capping loop in Figure 3) that in isolated LplA-N partially occupies the space that the adenylate binding loop occupies in *E. coli* LplA. It is likely that the capping loop functions at least to repress catalytic activity of LpA-N in the absence of LplA-C. It would be interesting in future to establish whether replacement of the capping loop with the equivalent short loop from *E. coli* LplA shows gain of function effects in isolated LplA-N.

E. coli LplA undergoes significant structural changes upon lipoate adenylation, including reorientation of the C-terminal domain to produce a more stretched overall conformation (PDB: 3A7R). Bovine LPT adopts this stretched domain arrangement in both apo- and lipoyl-AMP-

bound forms (PDB: 2E5A, 3A7U). We have been unable to produce crystals of LplA-N+LplA-C complexes with lipoyl-AMP and with E2lipD that diffract to sufficient resolution to investigate structural changes in LplA-N+LplA-C. We have, however, studied LplA-N+LplA-C interactions with lipoic acid/ATP/Mg^{2+} and E2lipD using ^{1}H-^{15}N HSQC NMR spectra, which are highly sensitive to conformational changes and interactions. We did two titrations in which at each step we recorded ^{1}H-^{15}N HSQC spectra of ^{15}N-labelled LplA-C in complex with unlabelled LplA-N. In Titration 1, lipoic acid/ATP/Mg^{2+} were added before E2lipD such that lipoate adenylation and lipoate transfer are monitored separately. The observed NMR spectral changes (Figure 4A) are not immediately suggestive of a substantial LplA-C conformational change upon formation of the lipoyl-AMP intermediate, but a LplA-C positional change like the 180° rotation of the C-terminal domain observed in *E. coli* LplA [13] cannot be ruled out as the observation of exchange in about 25 peaks indicates that nearly a third of LplA-C residues sample more than one chemical environment.

In Titration 2, E2lipD was added before lipoic acid/ATP/Mg^{2+}. The lack of change upon E2lipD addition indicates that in apo-LplA-N+LplA-C, LplA-C does not interact with E2lipD and further, if there is any E2lipD interaction with LplA-N in apo-LplA-N+LplA-C, it does not occur in the vicinity of LplA-C. Chemical shift perturbations were observed in more than a third of LplA-C ^{1}H-^{15}N HSQC peaks when lipoic acid/ATP/Mg^{2+} were then added to the mixture of apo-LplA-N+LplA-C and E2lipD (Figure 4C). Thus a more substantial change in LplA-C occurs when E2lipD is already present at the time of adding lipoate adenylation ingredients. A reliable explanation of this observation would require extensive further investigation, but for now we note that lipoate adenylation and lipoate transfer are monitored simultaneously in Titration 2, rather than separately as in Titration 1. We note also that at least 75% of peaks perturbed in Titration 1 were also perturbed in Titration 2, indicating the involvement of substantially overlapping regions of LplA-C in any conformational/positional changes occurring during the two titrations.

Our NMR data, particularly from Titration 2, indicate that LplA-C does undergo significant conformational change at one or more stages of lipoylation. This is consistent with our structure-based hypothesis that rearrangement of the LplA-N+LplA-C complex, akin to that seen in *E. coli* LplA [13], is required to allow E2lipD access to the LplA-N active site and hence to allow lipoate transfer. In our model of a possible end-point of such a rearrangement, LplA-C has rotated through approximately 120º, compared to the approximately 180º rotation observed for *E. coli* LplA C-terminal domain upon lipoate adenylation [13]. One potential flaw in our model, in common with the *E. coli* LplA:octanoyl-AMP:ApoH protein complex crystal structure, is that the distance between the adenylated intermediate and the acceptor lysine ($Lys_{E2lipD}42$) is too great for initiation of lipoyl transfer. It remains to be seen whether, as Fujiwara et al suggest [13], this is rectified if a version of a LplA-E2lipD/H protein complex with “true substrates” can be crystallised.

Assuming that LplA-C becomes disordered if it is released from LplA-N, there is no evidence from our NMR data that LplA-C dissociates from LplA-N during lipoate adenylation or lipoate transfer; we detect only folded LplA-C species during catalysis of both steps, suggesting that LplA-C is not competed off LplA-N by incoming substrate and that LplA-C remains bound to LplA-N during any rearrangement of the complex. There remains the possibility, however, of minor populations of disordered LplA-C species at any one time that are not detected by the techniques used here. On the other hand, if LplA-C remains structured upon dissociation from LplA-N, then rearrangement of the LplA-N+LplA-C complex could clearly involve a simple release and rebind mechanism.

We also used NMR to monitor the effect of lipoylation on E2lipD. We believe that the observed chemical shift perturbations upon addition of the lipoylation ingredients (Figure 5) are more likely to result from lipoylation-induced localised conformational change in E2lipD than from non-covalent E2lipD interaction with LplA-N+LplA-C (present only in catalytic quantities) or substrate. Previous NMR analysis did not indicate any conformational change in E2lipD from *Bacillus stearothermophilus* PDHC upon lipoylation [44]. The reason for the difference is not obvious, but

it is clear that the results of this titration represent further evidence that the *T. acidophilum* LplA-N+LplA-C complex is functional.

We have previously described features of the genes encoding LplA-C and LplA-N, including the facts that their genes overlap by a single base pair, and a TATA box is readily identifiable upstream of *lpla-c*; however, no *cis*-regulatory sequence is observed in the proximity of *lpla-n*, suggesting that the genes are transcriptionally coupled. Given that the gene order is *lpla-c* then *lpla-n* [18], our structure-based observation that the LplA-C C-terminus and LplA-N N-terminus are located at opposite ends of the LplA-N+LplA-C complex confirms that LplA-N and LplA-C are made as separate polypeptides.

We have shown that, like other LplAs, *T. acidophilum* LplA-N+LplA-C can utilize OA and octanoyl-AMP as substrates, albeit less efficiently than LA. At first glance this may be unsurprising, but fatty acid synthesis, the source of the OA precursor, is thought to be absent in the archaea [45]. It has long been hypothesized that ancient enzyme promiscuity gave rise to the specialized enzyme activities known today [46]. Hence, the ability to utilize OA as well as LA may reflect an early evolutionary state. Notably, this promiscuity has a physiological advantage in *E. coli* where LipB mutants can still carry out lipoylation using LipA and LplA [38]. LplA is evolutionarily related to LipB and biotin protein ligase [14] but whether LplA may have served as an evolutionary platform for LipB or vice versa is uncertain.

Acknowledgements

We thank Prof. R. Perham (University of Cambridge, UK) for kindly providing TM202 and pET11c plasmids, and Geoff Kelly at the MRC Biomedical NMR Centre, Mill Hill for assistance with NMR data collection.

Funding

MGP gratefully acknowledges receipt of a UNESCO-L'Oréal FWIS Fellowship.

References

1 Perham, R. N. (2000) Swinging arms and swinging domains in multifunctional enzymes: catalytic machines for multistep reactions. Ann. Rev. Biochem. **69**, 961-1004
2 Cronan, J. E., Zhao, X. and Jiang, Y. F. (2005) Function, attachment and synthesis of lipoic acid in *Escherichia coli*. Adv. Microb. Physiol. **50**, 103-146
3 McManus, E., Luisi, B. F. and Perham, R. N. (2006) Structure of a putative lipoate protein ligase from *Thermoplasma acidophilum* and the mechanism of target selection for post-translational modification. J. Mol. Biol. **356**, 625-637
4 Hassan, B. H. and Cronan, J. E. (2011) Protein-protein interactions in assembly of lipoic acid on the 2-oxoacid dehydrogenases of aerobic metabolism. J. Biol. Chem. **286**, 8263-8276
5 Fujiwara, K., Toma, S., Okamura-Ikeda, K., Motokawa, Y., Nakagawa, A. and Taniguchi, H. (2005) Crystal structure of lipoate-protein ligase A from *Escherichia coli*: determination of the lipoic acid-binding site. J. Biol. Chem. **280**, 33645-33651
6 Morris, T. W., Reed, K. E. and Cronan, J. E. (1994) Identification of the gene encoding lipoate-protein ligase-A of *Escherichia coli*. Molecular cloning and characterization of the lplA gene and gene product. J. Biol. Chem. **269**, 16091-16100
7 Kang, S. G., Jeong, H. K., Lee, E. and Natarajan, S. (2007) Characterization of a lipoate-protein ligase A gene of rice (*Oryza sativa* L.). Gene **393**, 53-61

8 Fujiwara, K., Hosaka, H., Matsuda, M., Okamura-Ikeda, K., Motokawa, Y., Suzuki, M., Nakagawa, A. and Taniguchi, H. (2007) Crystal structure of bovine lipoyltransferase in complex with lipoyl-AMP. J. Mol. Biol. **371**, 222-234
9 Schonauer, M. S., Kastaniotis, A. J., Kursu, V. A. S., Hiltunen, J. K. and Dieckmann, C. L. (2009) Lipoic acid synthesis and attachment in yeast mitochondria. J. Biol. Chem. **284**, 23234-23242
10 Christensen, Q. H., Hagar, J. A., O'Riordan, M. X. D. and Cronan, J. E. (2011) A complex lipoate utilization pathway in *Listeria monocytogenes*. J. Biol. Chem. **286**, 31447-31456
11 Martin, N., Christensen, Q. H., Mansilla, M. C., Cronan, J. E. and de Mendoza, D. (2011) A novel two-gene requirement for the octanoyltransfer reaction of *Bacillus subtilis* lipoic acid biosynthesis. Mol. Microbiol. **80**, 335-349
12 Christensen, Q. H., Martin, N., Mansilla, M. C., de Mendoza, D. and Cronan, J. E. (2011) A novel amidotransferase required for lipoic acid cofactor assembly in *Bacillus subtilis*. Mol. Microbiol. **80**, 350-363
13 Fujiwara, K., Maita, N., Hosaka, H., Okamura-Ikeda, K., Nakagawa, A. and Taniguchi, H. (2010) Global conformational change associated with the two-step reaction catalyzed by *Escherichia coli* lipoate-protein ligase A. J. Biol. Chem. **285**, 9971-9980
14 Reche, P. A. (2000) Lipoylating and biotinylating enzymes contain a homologous catalytic module. Prot. Sci. **9**, 1922-1929
15 Wood, Z. A., Weaver, L. H., Brown, P. H., Beckett, D. and Matthews, B. W. (2006) Co-repressor induced order and biotin repressor dimerization: a case for divergent followed by convergent evolution. J. Mol. Biol. **357**, 509-523
16 Kim, D. J., Kim, K. H., Lee, H. H., Lee, S. J., Ha, J. Y., Yoon, H. J. and Suh, S. W. (2005) Crystal structure of lipoate-protein ligase A bound with the activated intermediate: insights into interaction with lipoyl domains. J. Biol. Chem. **280**, 38081-38089
17 Christensen, Q. H. and Cronan, J. E. (2009) The *Thermoplasma acidophilum* LplA-LplB complex defines a new class of bipartite lipoate-protein ligases. J. Biol. Chem. **284**, 21317-21326
18 Posner, M. G., Upadhyay, A., Bagby, S., Hough, D. W. and Danson, M. J. (2009) A unique lipoylation system in the Archaea: lipoylation in *Thermoplasma acidophilum* requires two proteins. FEBS J. **276**, 4012-4022
19 Otwinowski, Z. and Minor, W. (1997) Processing of X-ray diffraction data collected in oscillation mode. Meth. Enzymol. **276**, 307-326
20 Long, F., Vagin, A., Young, P. and Murshudov, G. N. (2008) BALBES: a molecular replacement pipeline. Acta Cryst. D Biol. Cryst. **64**, 125-132
21 Emsley, P. and Cowtan, K. (2004) Coot: model-building tools for molecular graphics. Acta Cryst. D Biol. Cryst. **60**, 2126-2132
22 Bailey, S. (1994) The CCP4 suite: programs for protein crystallography. Acta Cryst. D Biol. Cryst. **50**, 760-763
23 Davis, I. W., Leaver-Fay, A., Chen, V. B., Block, J. N., Kapral, G. J., Wang, X., Murray, L. W., Arendall, W. B., III, Snoeyink, J., Richardson, J. S. and Richardson, D. C. (2007) MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. Nucl. Acids Res. **35**, W375-W383
24 Laskowski, R. A., Macarthur, M. W., Moss, D. S. and Thornton, J. M. (1993) PROCHECK: a program to check the stereochemical quality of protein structures. J. App. Cryst. **26**, 283-291

25 Reynolds, C., Damerell, D. and Jones, S. (2009) ProtorP: a protein-protein interaction analysis server. Bioinf. **25**, 413-416
26 Krissinel, E. and Henrick, K. (2007) Inference of macromolecular assemblies from crystalline state. J. Mol. Biol. **372**, 774-797
27 Delaglio, F., Grzesiek, S., Vuister, G. W., Zhu, G., Pfeifer, J. and Bax, A. (1995) NMRPipe: a multidimensional spectral processing system based on UNIX pipes. J. Biomol. NMR **6**, 277-293
28 Vranken, W. F., Boucher, W., Stevens, T. J., Fogh, R. H., Pajon, A., Llinas, P., Ulrich, E. L., Markley, J. L., Ionides, J. and Laue, E. D. (2005) The CCPN data model for NMR spectroscopy: Development of a software pipeline. Prot. Struct. Func. Bioinf. **59**, 687-696
29 Wishart, D. S., Bigam, C. G., Yao, J., Abildgaard, F., Dyson, H. J., Oldfield, E., Markley, J. L. and Sykes, B. D. (1995) ^{1}H, ^{13}C and ^{15}N chemical shift referencing in biomolecular NMR. J. Biomol. NMR **6**, 135-140
30 Upadhyay, A., Burman, J. D., Clark, E. A., Leung, E., Isenman, D. E., van den Elsen, J. M. H. and Bagby, S. (2008) Structure-function analysis of the C3 binding region of *Staphylococcus aureus* immune subversion protein Sbi. J. Biol. Chem. **283**, 22113-22120
31 Grzesiek, S., Stahl, S. J., Wingfield, P. T. and Bax, A. (1996) The CD4 determinant for downregulation by HIV-1 Nef directly binds to Nef. Mapping of the Nef binding surface by NMR. Biochemistry **35**, 10256-10261
32 Bartoli, G., Bosco, M., Carlone, A., Dalpozzo, R., Marcantoni, E., Melchiorre, P. and Sambri, L. (2007) Reaction of dicarbonates with carboxylic acids catalyzed by weak Lewis acids: General method for the synthesis of anhydrides and esters. Synthesis, 3489-3496
33 Schall, O. F., Suzuki, I., Murray, C. L., Gordon, J. I. and Gokel, G. W. (1998) Characterization of acyl adenyl anhydrides: Differences in the hydrolytic rates of fatty acyl-AMP and aminoacyl-AMP derivatives. J. Org. Chem. **63**, 8661-8667
34 Comeau, S. R., Gatchell, D. W., Vajda, S. and Camacho, C. J. (2004) ClusPro: a fully automated algorithm for protein-protein docking. Nucl. Acids Res. **32**, W96-W99
35 Comeau, S. R., Gatchell, D. W., Vajda, S. and Camacho, C. J. (2004) ClusPro: An automated docking and discrimination method for the prediction of protein complexes. Bioinf. **20**, 45-50
36 Benkert, P., Kuenzli, M. and Schwede, T. (2009) QMEAN server for protein model quality estimation. Nucl. Acids Res. **37**, W510-W514
37 Holm, L., Kääriäinen, S., Rosenström, P. and Schenkel, A. (2008) Searching protein structure databases with DaliLite v.3. Bioinf. **24**, 2780-2781
38 Morris, T. W., Reed, K. E. and Cronan, J. E. (1995) Lipoic acid metabolism in *Escherichia coli*: the lplA and lipB genes define redundant pathways for ligation of lipoyl groups to apoprotein. J. Bact. **177**, 1-10
39 Gong, X. M., Peng, T., Yakhnin, A., Zolkiewski, M., Quinn, J., Yeaman, S. J. and Roche, T. E. (2000) Specificity determinants for the pyruvate dehydrogenase component reaction mapped with mutated and prosthetic group modified lipoyl domains. J. Biol. Chem. **275**, 13645-13653
40 Puthenveetil, S., Liu, D. S., White, K. A., Thompson, S. and Ting, A. Y. (2009) Yeast display evolution of a kinetically efficient 13-amino acid substrate for lipoic acid ligase. J. Am. Chem. Soc. **131**, 16430-16438

41 Fujiwara, K., OkamuraIkeda, K. and Motokawa, Y. (1996) Lipoylation of acyltransferase components of alpha-ketoacid dehydrogenase complexes. J. Biol. Chem. **271**, 12932-12936
42 Fernandez-Suarez, M., Baruah, H., Martinez-Hernandez, L., Xie, K. T., Baskin, J. M., Bertozzi, C. R. and Ting, A. Y. (2007) Redirecting lipoic acid ligase for cell surface protein labeling with small-molecule probes. Nat. Biotech. **25**, 1483-1487
43 Cohen, J. D., Zou, P. and Ting, A. Y. (2012) Site-Specific Protein Modification Using Lipoic Acid Ligase and Bis-Aryl Hydrazone Formation. Chembiochem **13**, 888-894
44 Dardel, F., Laue, E. D. and Perham, R. N. (1991) Sequence-specific ^{1}H-NMR assignments and secondary structure of the lipoyl domain of the *Bacillus stearothermophilus* pyruvate dehydrogenase multienzyme complex. Eur. J. Biochem. **201**, 203-209
45 Falb, M., Müller, K., Königsmaier, L., Oberwinkler, T., Horn, P., von Gronau, S., O., G., Pfeiffer, F., Bornberg-Bauer, E. and Oesterhelt, D. (2008) Metabolism of halophilic archaea. Extremophiles **12**, 177-196
46 Khersonsky, O. and Tawfik, D. S. (2010) Enzyme promiscuity: a mechanistic and evolutionary perspective. Ann. Rev. Biochem. **79**, 471-505