LETTER

doi:10.1038/nature13912

Structure and function of a single-chain, multi-domain long-chain acyl-CoA carboxylase

Timothy H. Tran[1], Yu-Shan Hsiao[2]†, Jeanyoung Jo[1], Chi-Yuan Chou[1]†, Lars E. P. Dietrich[1], Thomas Walz[2,3] & Liang Tong[1]

Biotin-dependent carboxylases are widely distributed in nature and have important functions in the metabolism of fatty acids, amino acids, carbohydrates, cholesterol and other compounds[1–6]. Defective mutations in several of these enzymes have been linked to serious metabolic diseases in humans, and acetyl-CoA carboxylase is a target for drug discovery in the treatment of diabetes, cancer and other diseases[7–9]. Here we report the identification and biochemical, structural and functional characterizations of a novel single-chain (120 kDa), multi-domain biotin-dependent carboxylase in bacteria. It has preference for long-chain acyl-CoA substrates, although it is also active towards short-chain and medium-chain acyl-CoAs, and we have named it long-chain acyl-CoA carboxylase. The holoenzyme is a homo-hexamer with molecular mass of 720 kDa. The 3.0 Å crystal structure of the long-chain acyl-CoA carboxylase holoenzyme from *Mycobacterium avium* subspecies *paratuberculosis* revealed an architecture that is strikingly different from those of related biotin-dependent carboxylases[10,11]. In addition, the domains of each monomer have no direct contact with each other. They are instead extensively swapped in the holoenzyme, such that one cycle of catalysis involves the participation of four monomers. Functional studies in *Pseudomonas aeruginosa* suggest that the enzyme is involved in the utilization of selected carbon and nitrogen sources.

The reactions catalysed by biotin-dependent carboxylases proceed in two steps and involve at least three different protein components (Extended Data Fig. 1). In the first step, a biotin carboxylase (BC) component catalyses the carboxylation of the biotin cofactor, which is covalently linked to the biotin carboxyl carrier protein (BCCP) component. In the second step, the carboxylated biotin translocates to the carboxyltransferase (CT) active site and transfers the carboxyl group to the substrate. In bacteria, acetyl-CoA carboxylase (ACC) has been well characterized as a multi-subunit enzyme, with a BC subunit, a BCCP subunit and two subunits (α and β) for the CT activity (Extended Data Fig. 1). In contrast, ACC is a large (~250 kDa), single-chain, multi-domain enzyme in most eukaryotes (Extended Data Fig. 1), with domains that are homologous to the bacterial subunits. Other members of this family include propionyl-CoA carboxylase (PCC)[10], 3-methylcrotonyl-CoA carboxylase (MCC)[11], pyruvate carboxylase (PC)[12,13] and urea carboxylase (UC)[14] (Extended Data Fig. 1).

By examining the sequence database, we identified a novel single-chain (~120 kDa), multi-domain biotin-dependent carboxylase in bacteria[1]. The enzyme contains a BC domain at the amino terminus, a BCCP domain near the middle, and a CT domain that is homologous to that of ACC and PCC (Fig. 1a and Extended Data Fig. 1). Homologues of this enzyme are found in a large number of Gram-negative and Gram-positive bacteria, such as *Rhodopseudomonas palustris*, *Mycobacterium avium* subspecies *paratuberculosis*[15] and the human pathogen *Pseudomonas aeruginosa* (Extended Data Fig. 2), with highly conserved sequences (Extended Data Fig. 3). These homologues are in fact mis-annotated as PC[16] or carbamoyl-phosphate synthase (CPS) in the database, as was noted in an earlier report[17], probably because they have roughly the same size as PC and CPS. The CT domain of PC has a completely different sequence and structure[12,13] (Extended Data Fig. 1), whereas CPS is not a biotin-dependent enzyme and does not have a BCCP domain. These single-chain enzymes are somewhat related to a family of acyl-CoA carboxylases that have been characterized in *Mycobacterium tuberculosis* and other actinomycetes[3], in which BC and BCCP are present in one subunit and CT is in a separate subunit (Extended Data Fig. 1; see below).

We overexpressed several of these single-chain enzymes in *Escherichia coli* and purified them to homogeneity. The proteins migrated at the same position on a gel filtration column as the 750 kDa $\alpha_6\beta_6$ holoenzymes of PCC[10] and MCC[11], suggesting that these enzymes are hexamers, with a molecular mass of ~720 kDa for the holoenzyme.

We characterized the catalytic activities of the enzyme from *R. palustris*, which is annotated as both PC and CPS in the database. Consistent with the sequence analysis was our finding that this enzyme is an acyl-CoA carboxylase, and we did not observe any PC activity for it. The enzyme is active towards all the acyl-CoAs that we tested, with chain lengths from C_2 to C_{16}, but it prefers long-chain substrates (Extended Data Table 1). The k_{cat} values for all the substrates are comparable, while the K_m for palmitoyl-CoA is ~350-fold lower than that for acetyl-CoA. The high sequence conservation among these proteins (Extended Data Fig. 3) suggests that they have similar activity profiles. We have therefore named them long-chain acyl-CoA carboxylases (LCCs). Broad-spectrum activity has been observed for a few of the acyl-CoA carboxylases in actinomycetes[3,18].

To define the holoenzyme architecture, we determined the crystal structure of *M. avium* subspecies *paratuberculosis* LCC (MapLCC) at 3.0 Å resolution (Fig. 1b–e); MapLCC shares 52% amino acid sequence identity with *R. palustris* LCC (Extended Data Fig. 3). The holoenzyme hexamer is situated on a crystallographic three-fold axis, and there is a dimer in the asymmetric unit. The atomic model has good agreement with the crystallographic data and the expected geometric parameters (Extended Data Table 2). Several segments of the protein have poor or no electron density and are not included in the atomic model. These include parts of the linkers from the BCCP to the BC and CT domains (Fig. 1b), although there is no ambiguity in assigning the BCCP domain to a specific monomer. A different assignment will result in gaps that are too large to be bridged by the missing residues.

The overall structures of the two MapLCC monomers in the asymmetric unit are similar. With their CT domains superposed, a difference of 3° is seen in the orientation of their BC domains (Fig. 1b). The BCCP domains have a larger difference, corresponding to a rotation of 15°, indicating some asymmetry in the holoenzyme hexamer. The BC, BCCP and CT domains of each monomer do not have any direct contact with each other.

The structure of the holoenzyme hexamer of MapLCC has the shape of an equilateral triangle, obeying 32 symmetry and with a length of ~180 Å for each side (Fig. 1c) and a thickness of ~65 Å (Fig. 1d, e). A hexamer of the CT domain forms the central core of the structure, with three CT domains in each layer and CT dimers being formed by one domain from each layer. Dimeric BC domains are located at the vertices

[1]Department of Biological Sciences, Columbia University, New York, New York 10027, USA. [2]Department of Cell Biology, Harvard Medical School, Boston, Massachusetts 02115, USA. [3]Howard Hughes Medical Institute, Harvard Medical School, Boston, Massachusetts 02115, USA. †Present addresses: Department of Otolaryngology, Massachusetts Eye and Ear Infirmary, Boston, Massachusetts 02114, USA (Y.-S.H.); Department of Life Sciences and Institute of Genome Sciences, National Yang-Ming University, Taipei 112, Taiwan (C.-Y.C.).

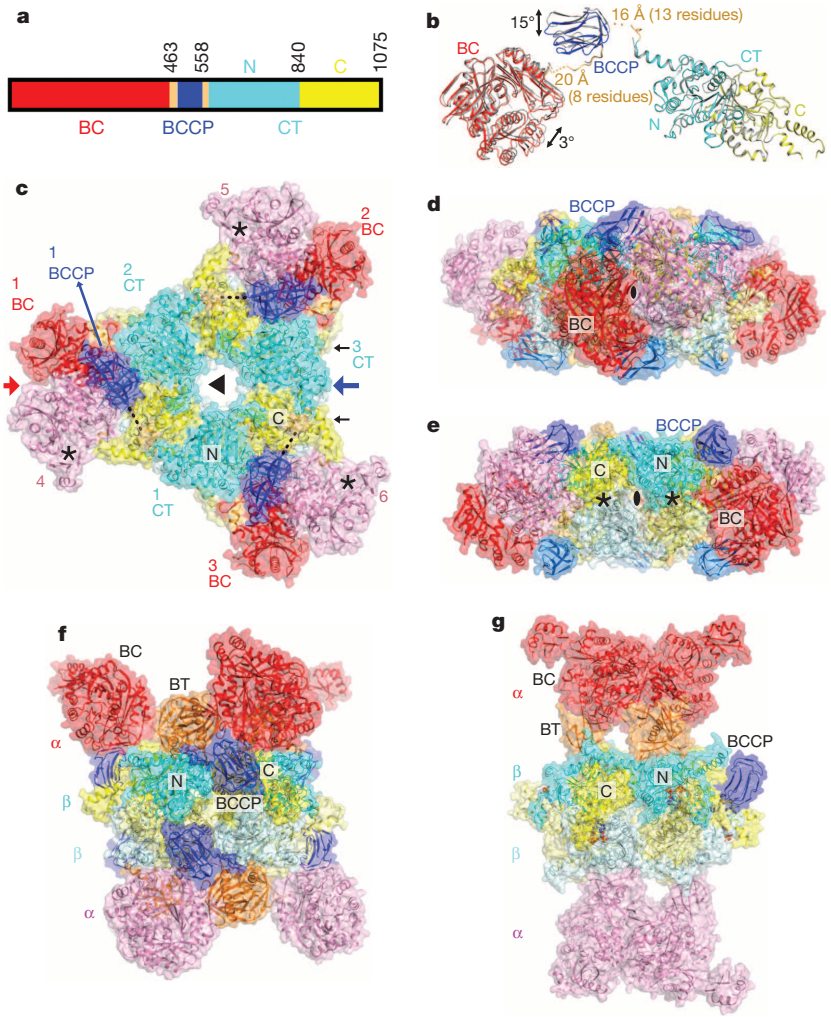


Figure 1 | Crystal structure of LCC from *M. avium* subspecies *paratuberculosis* (MapLCC). **a**, Domain organization of MapLCC. The domains are labelled and given different colours. **b**, Overlay of the structures of the two MapLCC monomers in the asymmetric unit (one in colour, the other in grey). Residues that are missing in the linkers from BCCP are indicated with dashed lines. **c**, Overall structure of the 720 kDa hexameric holoenzyme of MapLCC. The six monomers are labelled. The domains in the three monomers in the top layer (numbered 1, 2 and 3) are coloured as in **a**. The BC, BCCP, N and C CT domains in the three monomers in the bottom layer (numbered 4, 5 and 6) are coloured pink, pale blue, pale cyan and pale yellow, respectively. The disordered region of the BCCP–CT linker is indicated with a dashed line (black). The BC active sites are indicated with asterisks (black). The CT active sites are on the side of the CT domain core, indicated with black arrows. **d**, Structure of the MapLCC holoenzyme viewed down the BC domain dimer (red arrow in **c**). **e**, Structure of the MapLCC holoenzyme viewed down the blue arrow in **c**. The CT active sites are indicated with asterisks (black). **f**, Structure of the 750 kDa $\alpha_6\beta_6$ PCC holoenzyme[10]. The view is equivalent to that in **d**. **g**, Structure of the 750 kDa $\alpha_6\beta_6$ MCC holoenzyme[11]. The structure figures were produced with PyMOL (http://www.pymol.org).

of the triangle, contacting CT domain dimers. The BCCP domains are situated between the BC and CT domains but not in the active site of either domain. The lysine residue that would be biotinylated is located on the surface, more than 15 Å away from the nearest BC or CT domain, and its side chain is disordered. The BCCP domain is not biotinylated in this holoenzyme, even though it was expressed under conditions identical to those for *R. palustris* LCC, which was completely biotinylated.

The overall architecture of the MapLCC holoenzyme is strikingly different from those of PCC (Fig. 1f)[10] and MCC (Fig. 1g)[11], even though all three enzymes are ~750 kDa oligomers made up of homologous BC, CT and BCCP domains. The central CT domain core of MapLCC is similar to that of the β_6 hexamer core of PCC, with a root mean squared distance of 1.7 Å for 2,073 equivalent Cα atoms between them (Extended Data Fig. 4). In contrast, whereas the BC domains are located above and below the central core in PCC and MCC, they are positioned at the side of the central core in MapLCC. Moreover, the BC domain is a monomer in PCC and a weakly associated trimer in MCC, but it is a dimer in MapLCC. In fact, this dimer is similar to that for the BC subunit of *E. coli* ACC[19,20] (Extended Data Fig. 4) and the BC domain of PC[12,13].

A BT domain was identified in the structures of PCC (Fig. 1f) and MCC (Fig. 1g), located between the BC and BCCP domains in the primary sequence (Extended Data Fig. 1) and having an important function in mediating interactions between their α (BC) and β (CT) subunits[10,11]. This domain does not exist in LCC, because there are insufficient residues in the linker between BC and BCCP to form such a domain (Extended Data Fig. 1). However, the BC–BCCP linker and the BCCP–CT linker do participate in mediating interactions in the MapLCC holoenzyme, and there are also direct contacts between the BC and CT domains (see below).

The domains of the MapLCC monomers are swapped extensively in the holoenzyme hexamer. The CT domains of two monomers related by a BC domain dimer are located far from each other (Fig. 2a). Similarly, the BC domains of two monomers related by a CT domain dimer are also located far from each other (Fig. 2b). Interactions between these domains therefore occur only in the context of the holoenzyme, and the structure of the monomer alone is unlikely to be stable (Fig. 1b).

A total of ~8,200 Å^2 of the surface area of each MapLCC monomer is buried in the holoenzyme. The majority of this surface, 5,700 Å^2, is buried by the interface among the CT domains in the central core, and a 1,000 Å^2 surface is buried in the BC dimer interface (Fig. 3a, b). The remaining interfaces in the holoenzyme make smaller contributions to the surface area burial, 600 Å^2 in the BC–CT interface, 500 Å^2 for the linkers from BCCP, and 400 Å^2 from the BCCP domain. Nonetheless, all of these smaller interfaces together may be important for stabilizing the holoenzyme.

The primary BC–CT interface involves three monomers of MapLCC. For example, a β-hairpin structure in the BC domain of monomer 4 contacts the CT domains of monomer 2 (C domain) and monomer 6 (N domain; Fig. 3c). Similarly, the BC–BCCP linker just before the BCCP domain in monomer 1 contacts the CT domain of monomer 2 and the BC domain of monomer 4 (Fig. 3d). These interactions also demonstrate the extensive connections between the monomers in the holoenzyme. The interfaces involve van der Waals contacts, hydrogen bonding and ionic interactions.

We also carried out electron microscopy (EM) studies on the MapLCC holoenzyme. Negative-stain EM images showed monodisperse particles of similar sizes but variable shapes (Extended Data Fig. 5). Classification

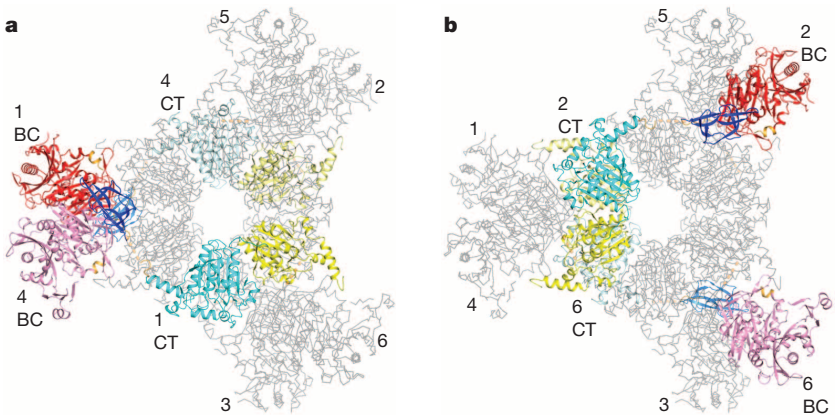


Figure 2 | Extensive domain swapping in the structure of the MapLCC holoenzyme. a, The CT domains of two monomers (in colour) related by a BC dimer are located far from each other (note that they are in opposite layers) and participate in dimerization with different monomers (in grey). The six monomers are labelled. **b**, The BC domains of two monomers related by a CT dimer are located far from each other.

in SPARX[21] of ~25,000 particles yielded 308 classes that represented 65% of the data set (Extended Data Fig. 5). Cross-correlation analysis identified class averages very similar to the top view (cross-correlation coefficient 0.856; Fig. 3e) and side view (cross-correlation coefficient 0.752; Fig. 3f) of the crystal structure, confirming the holoenzyme architecture seen in the crystal. At the same time, substantial variations in the three peripheral densities (corresponding to the BC domains) relative to the central core (the CT domains) are also observed (Extended Data Fig. 5). Each peripheral density can appear as a single or bilobed feature (Fig. 3g), probably representing different views of the BC dimer. The densities can also be located in different positions relative to the central core, breaking the alignment of the two-fold axes of the BC and CT dimers (Fig. 3h). Overall, the EM data indicate that the peripheral BC domains move as dimers and are flexibly tethered to the CT core (Supplementary Video 1), consistent with the crystal structure that shows that the BC–CT contact is relatively weak in MapLCC.

Each BC active site is located ~40 Å from a CT active site in the MapLCC holoenzyme (Fig. 4a). The BCCP domain, although not located in either active site, can access both of them, through conformational changes in its linkers. The crystal structure shows that the BCCP domain of monomer 1 visits the BC active site of monomer 4 and the CT active site at the interface of monomers 2 and 6 (Fig. 4a), suggesting that each cycle of catalysis requires the participation of four monomers. This again indicates the extensive communications between the monomers in this holoenzyme. The residues in both active sites are generally conserved with those in other biotin-dependent carboxylases, suggesting a similar catalytic mechanism[22] for LCC. The CT active site has a pocket that can accommodate short-chain and medium-chain acyl groups (Extended Data Fig. 6), and a conformational change is needed to bind long-chain substrates. This conformational flexibility may be important for the enzyme to adapt to and be active towards the broad collection of substrates.

In most of the other biotin-dependent carboxylases, BCCP is located at the end of a polypeptide chain, and there is therefore only one linker to the rest of the protein. The LCCs, and the eukaryotic ACCs (Extended Data Fig. 1), are distinct in that BCCP is located in the middle of those proteins. There are therefore two linkers from BCCP to the rest of the protein. The amino acid sequences of these two linkers in the LCC enzymes are not conserved (Extended Data Fig. 3). They are therefore likely to be flexible and allow the movement of BCCP during catalysis, which is consistent with the fact that a portion of both linkers is disordered in the structure (Fig. 1b). Residues in both linkers that are included in the current atomic model also have high *B* values (Extended Data Fig. 4). In addition, the B domain of BC and a loop near the carboxy terminus of the protein have high *B* values and are partly disordered.

The structure of MapLCC also has implications for the holoenzymes of other biotin-dependent carboxylases, especially the family of acyl-CoA carboxylases in *M. tuberculosis, S. coelicolor* and other actinomycetes[3]. These enzymes contain two subunits, with BC–BCCP in the α subunit and CT in the β subunit, and there are not enough residues in one of the α subunits for a BT domain (Extended Data Fig. 1). The holoenzyme is an $\alpha_6\beta_6$ dodecamer, and the structure of the β_6 hexamer[23] is similar to that of the CT domain hexamer in MapLCC. It is therefore likely that the holoenzymes of such two-subunit carboxylases (lacking the BT domain) share a similar architecture to that of MapLCC (Fig. 1c) rather than PCC (Fig. 1f). *M. tuberculosis* does not have an LCC homologue, indicating some differences between these mycobacterial species.

Eukaryotic ACCs are also single-chain, multi-domain enzymes (Extended Data Fig. 1), although there are substantial differences from LCC. The eukaryotic ACCs contain ~1,000 additional residues, including a unique central region of ~700 residues, and they are likely to carry a BT domain as well. The overall architecture of the eukaryotic ACCs is therefore likely to be different from that of MapLCC observed here.

We have begun to characterize the physiological functions of LCC, using *P. aeruginosa* (strain PA14) as the model organism. The *P. aeruginosa* enzyme (locus name PA14_46320, homologue of PA1400 in *P. aeruginosa* PAO1) shares 59% amino acid sequence identity with *R. palustris* LCC. This organism also carries the multi-subunit ACC, MCC and geranyl-CoA carboxylase (GCC)[24,25], although it lacks a homologue for PCC. The multi-subunit ACC is probably essential, in a similar manner to the *E. coli* enzyme, because no transposon insertions in its subunits are found in a PA14 transposon mutant library[26]. In contrast, transposon insertions are found in the loci encoding LCC, MCC and GCC, suggesting that they are not essential for growth under the conditions used to produce this library[26].

We produced a markerless deletion of LCC/PA14_46320, confirming that it is not essential for cell survival. The activity profiles of the mutant under ~2,000 conditions were characterized by using a phenotype microarray, which monitored its ability to reduce a tetrazolium dye[27]. The incubation conditions included different carbon or nitrogen sources, nutrient supplements, osmolytes, pH values and antibiotics. Although most of the conditions showed comparable profiles between the wild type and the LCC deletion mutant, phenotypic differences were observed for several conditions (Extended Data Fig. 7), and two of these were considered significant on the basis of the phenotype microarray analysis—using fumarate as the sole carbon source and the Met-Val dipeptide as the sole nitrogen source (Fig. 4b). It is not clear how this LCC enzyme is linked to the utilization of selected carbon and nitrogen sources in *P. aeruginosa*. The metabolism of both Met and Val is likely to produce propionyl-CoA, and the PCC activity of LCC may be important for its further degradation, as is the case in many other organisms. This is also supported by the fact that *P. aeruginosa* lacks a proper PCC enzyme.

A long-chain acyl-CoA carboxylase activity is needed for the biosynthesis of mycolic acid in *Mycobacterium* and other actinomycetes[3,18,28]. However, the LCC enzyme studied here is unlikely to be essential for this function, because its homologue is absent in *M. tuberculosis*. Moreover, mycolic acid is not known to be present in *P. aeruginosa*, *R. palustris* and other bacteria, suggesting that this LCC is likely to have different physiological functions.

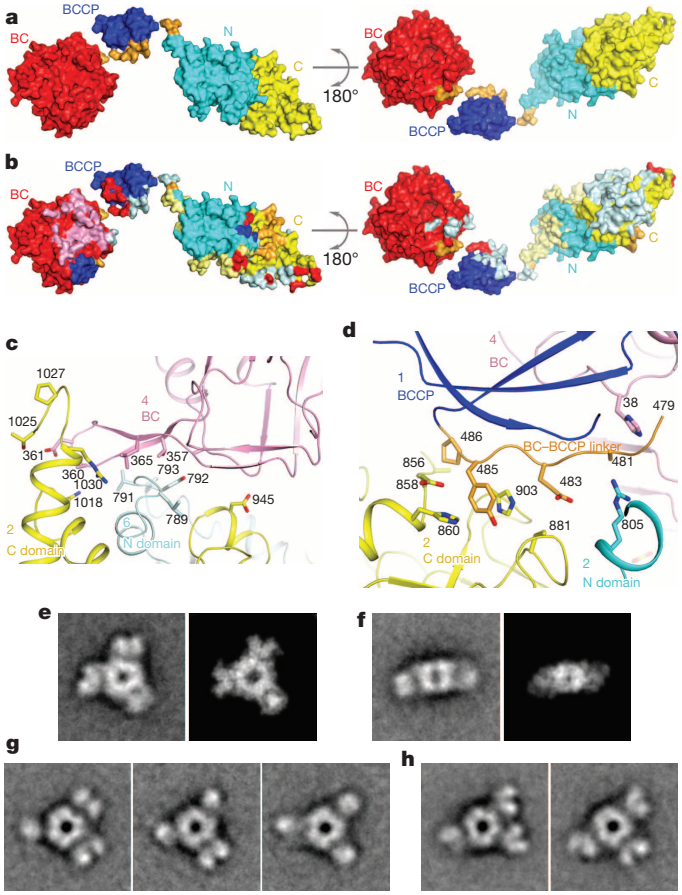


Figure 3 | Interactions between the central core and the rest of the holoenzyme. **a**, Two views of the molecular surface of a MapLCC monomer, coloured by the domains. **b**, Residues that mediate interactions in the MapLCC holoenzyme are indicated by the colours of the domains that they contact. For example, a pale cyan patch on the surface of the C domain of CT indicates residues that interact with the N domain of another monomer in the holoenzyme. The pink patch on BC indicates residues in the BC dimer interface, and the pale cyan patch indicates residues in the interface with CT. The major areas of contacts are between the CT domains and in the BC domain dimer. The views are identical to those in **a**. **c**, Residues in the interface between the BC and CT domains. The domains and the monomers they belong to are labelled. **d**, Residues in the interface between the BC–BCCP linker just before the BCCP domain and the rest of the holoenzyme. **e**, Class average representing the top view (left) most similar to the crystal structure and the corresponding projection from the crystal structure filtered to 30 Å resolution (right). **f**, Class average representing the side view (left) most similar to the crystal structure and the corresponding projection from the crystal structure filtered to 30 Å resolution (right). **g**, Three panels showing that each of the peripheral BC domains can appear as a single or bilobal density, probably representing different orientations of the BC dimer. **h**, Two panels illustrating that the BC domains can adopt different positions around the central CT core. The side length of the individual panels in **e**–**h** is 340 Å.

The striking differences between the architectures of the LCC, PCC and MCC holoenzymes are also linked to functional differences between them. Especially, the N and C domains of the β subunits of PCC and MCC are swapped relative to each other, and this is coupled to the different substrate specificity of the two enzymes. PCC carboxylates the α carbon of an acid (as a CoA ester), whereas MCC carboxylates the γ carbon of an α-β unsaturated acid (Extended Data Fig. 1). The structural differences therefore suggest that there are two lineages of the biotin-dependent carboxylases[11], one containing PCC, ACC and LCC, and the other containing MCC and possibly GCC. At the same time, PCC and LCC share propionyl-CoA carboxylase activity, indicating that similar biochemical activities can be supported by holoenzymes with markedly different architectures as well.

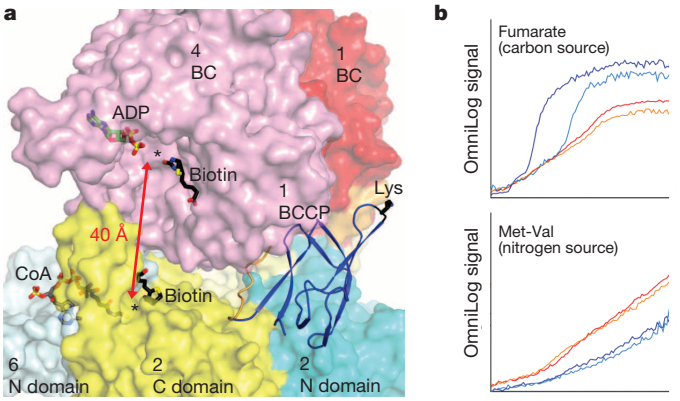


Figure 4 | Catalysis and function of LCC. **a**, The BC and CT active sites (black asterisks) of MapLCC are separated by ~40 Å (red arrow). Molecular surface of MapLCC is shown, coloured as in Fig. 1c. The domains are labelled by the monomer they belong to. The Lys side chain of BCCP to which biotin would be connected to is shown in black. The binding modes of ADP (green) and biotin (black) to the *E. coli* BC subunit are shown as stick models[20]. The binding mode of CoA (grey) to the CT domain of yeast ACC[29] and biotin (black) in the β (CT) subunit of PCC[10] are also shown. **b**, Phenotypic differences between wild-type and LCC knockout (ΔPA14_46320) *P. aeruginosa* strains, revealed by a colorimetric assay that monitors the reduction of a tetrazolium dye. Assays were performed twice in each medium for the wild-type (red and orange) and mutant (blue and cyan) strains. For each panel, the horizontal axis is time (24 h) and the vertical axis is OmniLog signal[27].

Overall, our studies have identified a new member of the biotin-dependent carboxylase family, and revealed a novel architecture for its holoenzyme. These observations are also relevant for other members of this family. Moreover, the differences between the architectures of LCC, PCC and MCC holoenzymes, despite their sharing domains with homologous structures, indicate that these domains can be arranged in remarkably different ways to form the various holoenzymes. This may have substantial implications for other multi-domain proteins, especially eukaryotic ACCs, and for protein structure and sequence conservation in general.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 4 May; accepted 2 October 2014.
Published online 10 November 2014.

1. Tong, L. Structure and function of biotin-dependent carboxylases. *Cell. Mol. Life Sci.* **70,** 863–891 (2013).
2. Waldrop, G. L., Holden, H. M. & St. Maurice, M. The enzymes of biotin dependent CO_2 metabolism: what structures reveal about their reaction mechanisms. *Protein Sci.* **21,** 1597–1619 (2012).
3. Gago, G., Diacovich, L., Arabolaza, A., Tsai, S.-C. & Gramajo, H. Fatty acid biosynthesis in actinomycetes. *FEMS Microbiol. Rev.* **35,** 475–497 (2011).
4. Jitrapakdee, S. *et al.* Structure, mechanism and regulation of pyruvate carboxylase. *Biochem. J.* **413,** 369–387 (2008).
5. Tong, L. Acetyl-coenzyme A carboxylase: crucial metabolic enzyme and attractive target for drug discovery. *Cell. Mol. Life Sci.* **62,** 1784–1803 (2005).
6. Cronan, J. E. Jr & Waldrop, G. L. Multi-subunit acetyl-CoA carboxylases. *Prog. Lipid Res.* **41,** 407–435 (2002).
7. Polyak, S. W., Abell, A. D., Wilce, M. C. J., Zhang, L. & Booker, G. W. Structure, function and selective inhibition of bacterial acetyl-CoA carboxylase. *Appl. Microbiol. Biotechnol.* **93,** 983–992 (2012).
8. Abramson, H. N. The lipogenesis pathway as a cancer target. *J. Med. Chem.* **54,** 5615–5638 (2011).
9. Wakil, S. J. & Abu-Elheiga, L. A. Fatty acid metabolism: target for metabolic syndrome. *J. Lipid Res.* **50,** S138–S143 (2009).
10. Huang, C. S. *et al.* Crystal structure of the $\alpha_6\beta_6$ holoenzyme of propionyl-coenzyme A carboxylase. *Nature* **466,** 1001–1005 (2010).
11. Huang, C. S., Ge, P., Zhou, Z. H. & Tong, L. An unanticipated architecture of the 750-kDa $\alpha_6\beta_6$ holoenzyme of 3-methylcrotonyl-CoA carboxylase. *Nature* **481,** 219–223 (2012).
12. St. Maurice, M. *et al.* Domain architecture of pyruvate carboxylase, a biotin-dependent multifunctional enzyme. *Science* **317,** 1076–1079 (2007).
13. Xiang, S. & Tong, L. Crystal structures of human and *Staphylococcus aureus* pyruvate carboxylase and molecular insights into the carboxyltransfer reaction. *Nature Struct. Mol. Biol.* **15,** 295–302 (2008).

14. Fan, C., Chou, C.-Y., Tong, L. & Xiang, S. Crystal structure of urea carboxylase provides insights into the carboxyltransfer reaction. *J. Biol. Chem.* **287,** 9389–9398 (2012).
15. Li, L. *et al.* The complete genome sequence of *Mycobacterium avium* subspecies *paratuberculosis*. *Proc. Natl Acad. Sci. USA* **102,** 12344–12349 (2005).
16. Stover, C. K. *et al.* Complete genome sequence of *Pseudomonas aeruginosa* PAO1, an opportunistic pathogen. *Nature* **406,** 959–964 (2000).
17. Lai, H., Kraszewski, J. L., Purwantini, E. & Mukhopadhyay, B. Identification of pyruvate carboxylase genes in *Pseudomonas aeruginosa* PAO1 and development of a *P.aeruginosa*-based overexpression system for a4- and $\alpha_4\beta_4$-type pyruvate carboxylases. *Appl. Environ. Microbiol.* **72,** 7785–7792 (2006).
18. Grande, R. *et al.* The two carboxylases of *Corynebacterium glutamicum* essential for fatty acid and mycolic acid synthesis. *J. Bacteriol.* **189,** 5257–5264 (2007).
19. Waldrop, G. L., Rayment, I. & Holden, H. M. Three-dimensional structure of the biotin carboxylase subunit of acetyl-CoA carboxylase. *Biochemistry* **33,** 10249–10256 (1994).
20. Chou, C.-Y., Yu, L. P. C. & Tong, L. Crystal structure of biotin carboxylase in complex with substrates and implications for its catalytic mechanism. *J. Biol. Chem.* **284,** 11690–11697 (2009).
21. Hohn, M. *et al.* SPARX, a new environment for Cryo-EM image processing. *J. Struct. Biol.* **157,** 47–55 (2007).
22. Knowles, J. R. The mechanism of biotin-dependent enzymes. *Annu. Rev. Biochem.* **58,** 195–221 (1989).
23. Lin, T. W. *et al.* Structure-based inhibitor design of AccD5, an essential acyl-CoA carboxylase carboxyltransferase domain of *Mycobacterium tuberculosis*. *Proc. Natl Acad. Sci. USA* **103,** 3072–3077 (2006).
24. Forster-Fromme, K. & Jendrossek, D. Catabolism of citronellol and related acyclic terpenoids in pseudomonads. *Appl. Microbiol. Biotechnol.* **87,** 859–869 (2010).
25. Aguilar, J. A. *et al.* Substrate specificity of the 3-methylcrotonyl coenzyme A (CoA) and geranyl-CoA carboxylases from *Pseudomonas aeruginosa*. *J. Bacteriol.* **190,** 4888–4893 (2008).
26. Liberati, N. T. *et al.* An ordered, nonredundant library of *Pseudomonas aeruginosa* strain PA14 transposon insertion mutants. *Proc. Natl Acad. Sci. USA* **103,** 2833–2838 (2006).
27. Shea, A., Wolcott, M., Daefler, S. & Rozak, D. A. Biolog phenotype microarrays. *Methods Mol. Biol.* **881,** 331–373 (2012).
28. Takayama, K., Wang, C. & Besra, G. S. Pathway to synthesis and processing of mycolic acids in *Mycobacterium tuberculosis*. *Clin. Microbiol. Rev.* **18,** 81–101 (2005).
29. Zhang, H., Yang, Z., Shen, Y. & Tong, L. Crystal structure of the carboxyltransferase domain of acetyl-coenzyme A carboxylase. *Science* **299,** 2064–2067 (2003).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank C. Huang for carrying out some initial studies in this project; A. Price-Whelan for discussions on *P. aeruginosa* physiology; R. Jackimowicz, N. Whalen and H. Robinson for access to the X29A beamline; Z. Li for EM support; P. Penczek for help with SPARX. The in-house instrument for X-ray diffraction was purchased with a National Institutes of Health (NIH) grant to L.T. (S10OD012018). This research is supported by grants from the NIH (R01DK067238 to L.T. and R01AI103369 to L.E.P.D.) and from the Protein Structure Initiative of the NIH (U54GM094597 to L.T.). The Orchestra High Performance Compute Cluster at Harvard Medical School is a shared facility partly supported by NIH grant NCRR 1S10RR028832-01. T.W. is an investigator with the Howard Hughes Medical Institute.

Author Contributions T.H.T. and C.-Y.C. performed cloning, protein expression, purification and crystallization experiments. T.H.T. and L.T. performed the crystallography experiments and calculation. Y.-S.H. and T.W. conducted the electron microscopy experiments. T.H.T. and C.-Y.C. performed the kinetic assays. J.J. and L.E.P.D. conducted the *P. aeruginosa* experiments. T.H.T., L.E.P.D., T.W. and L.T. wrote the paper.

Author Information Coordinates and structure factors have been deposited in Protein Data Bank under accession number 4RCN. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to L.T. (ltong@columbia.edu).

METHODS

Protein expression and purification. Full-length LCCs from several different bacterial organisms, including *R. palustris*, *M. avium* subspecies *paratuberculosis* and *P. aeruginosa*, were amplified from genomic DNA by PCR and cloned into pET28a, pET26b and/or pET24d vectors (Novagen). The plasmids were transformed into BL21Star (DE3) cells (Invitrogen). Protein expression was induced with the addition of 1 mM isopropyl β-D-thiogalactoside, and the cells were grown at 16 °C for 16–20 h.

To facilitate biotinylation, the recombinant enzyme was co-expressed with the *E. coli* biotin ligase BirA, and 15 mg l^{-1} biotin was added to the medium. An avidin shift assay of the purified enzymes showed that *R. palustris* LCC was completely biotinylated. However, purified MapLCC did not show any biotinylation, possibly indicating some degree of selectivity of the BirA enzyme. Expression of *P. aeruginosa* LCC did not produce any soluble protein and was not pursued further.

Cells were lysed by sonication in a buffer containing 20 mM Tris-HCl pH 8.0, 250 mM NaCl, 5% (v/v) glycerol, 10 mM 2-mercaptoethanol and 1 mM phenylmethylsulphonyl fluoride. Soluble enzyme was purified by Ni^{2+}-nitrilotriacetate (Qiagen), anion-exchange and gel-filtration (Sephacryl S-300; GE Healthcare) chromatography. The S-300 running buffer for MapLCC contained 25 mM HEPES pH 7.4, 250 mM NaCl and 2.5 mM dithiothreitol. The purified protein was concentrated to 6 mg ml^{-1}, and the solution was supplemented with 5% (v/v) glycerol before being flash-frozen in liquid nitrogen and stored at −80 °C.

The selenomethionyl MapLCC protein was produced in B834 (DE3) cells (Novagen) that were grown in defined LeMaster medium supplemented with selenomethionine[30]. The protein was purified with the same protocol as that for the native enzyme.

Protein crystallization. MapLCC was crystallized at 4 °C using the microbatch method under paraffin oil. The protein solution was mixed with a precipitant solution containing 0.1 M Bis-Tris propane pH 7.5–8.5 and 1.5–2.0 M ammonium sulphate. Crystals took 4–6 weeks to grow to full size, and larger crystals were obtained by microseeding. They were cryoprotected with reservoir solution supplemented with 12–15% (v/v) glycerol and flash-frozen in liquid nitrogen for data collection at 100 K. The C-terminal His tag on the protein was not removed before crystallization.

Data collection and structure determination. X-ray diffraction data for the native (wavelength 1.075 Å) and selenomethionyl (0.979 Å) crystals were collected with a Q315 charge-coupled device (Area Detector Systems Corporation) at the X29A beamline of the National Synchrotron Light Source. The diffraction images were processed with the HKL package[31]. The crystals belong to space group $P2_13$, with cell dimensions of $a = b = c = 220.9$ Å. There are two MapLCC monomers in the crystallographic asymmetric unit.

The structure of MapLCC was solved by a combination of molecular replacement and selenomethionyl SAD phasing. The orientation and position of the BC, CT and BCCP domains were located with the program Phaser[32]. The Se sites were located with the program SHELX[33], and SOLVE/RESOLVE was used for phasing the reflections and automated model building[34]. The atomic model was built with the program Coot[35]. The structure refinement was performed with the program CNS[36]. The crystallographic information is summarized in Extended Data Table 2.

We also obtained a second crystal form of MapLCC, with an entire hexamer in the asymmetric unit, and were able to collect an X-ray diffraction data set to 4.3 Å resolution (space group $P2_12_12_1$, $a = 102$ Å, $b = 292$ Å, $c = 314$ Å). The structure of this crystal form was readily solved by the molecular replacement method, and it revealed essentially the same holoenzyme architecture (data not shown).

Electron microscopy and image processing. Purified MapLCC was prepared by conventional negative staining with 0.75% (w/v) uranyl formate[37]. Images were collected with a Tecnai T12 electron microscope (FEI) equipped with an LaB_6 filament and operated at an acceleration voltage of 120 kV. Images were recorded using low-dose procedures on an UltraScan 895 4K x 4K charge-coupled device (CCD) camera (Gatan) using a defocus of −1.5 μm and a nominal magnification of ×52,000. The calibrated magnification was ×70,527, yielding a pixel size of 2.13 Å on the specimen level.

BOXER, the display program associated with the EMAN software package[38], was used to select 24,535 particles interactively from 270 CCD images, and the SPIDER software package[39] was used to window the particles into 160 × 160-pixel images. To perform iterative stable alignment and clustering (ISAC)[40] in SPARX[21], the size of the particle images was reduced to 64 × 64 pixels, and the particles were pre-aligned and centred. ISAC was run on the Orchestra High Performance Compute Cluster at Harvard Medical School (http://rc.hms.harvard.edu), specifying 200 images per group and a pixel error of 0.7. After 19 generations, 308 classes were obtained, accounting for 15,932 particles (65% of the entire data set) (Extended Data Fig. 5). Averages of these classes were calculated using the original 160 × 160-pixel images. To confirm that the ISAC averages were representative of the entire data set, the particles were also subjected to ten cycles of multi-reference alignment in SPIDER. Each round of multi-reference alignment was followed by K-means classification, specifying 300 output classes (Extended Data Fig. 5).

To compare the class averages with the crystal structure, the crystal structure was Fourier transformed, filtered to 30 Å with a Butterworth low-pass filter, and transformed back. Evenly spaced projections were calculated at 4° intervals and subjected to ten cycles of alignment with masked EM class averages. The class averages most similar to the top and side views of the crystal structure and their cross-correlation coefficients are presented in Fig. 3e, f.

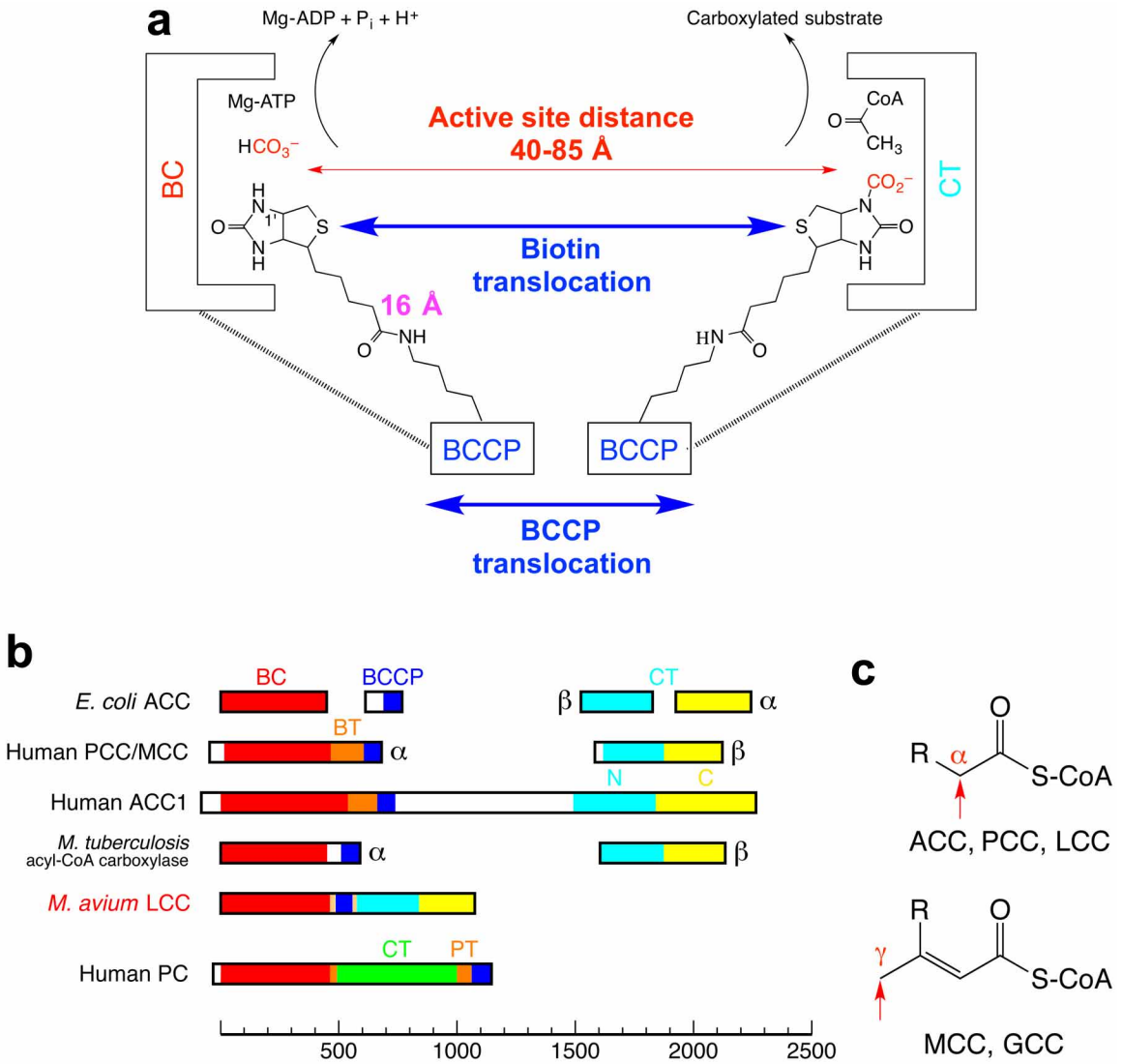
To visualize the structural heterogeneity of MapLCC in solution, 179 averages obtained by ISAC that showed three peripheral densities were selected (indicated by asterisks in Extended Data Fig. 5), ordered according to their correlation coefficients, and used to prepare Supplementary Video 1. This structural variability is probably the reason why it was not possible to calculate a three-dimensional map from cryo-EM images of vitrified MapLCC samples.

Enzymatic assays. The kinetic assays monitored the hydrolysis of ATP by *R. palustris* LCC in the presence of various acyl-CoA substrates, using coupling enzymes to convert the ADP product to NADH oxidation[41]. The reaction mixture contained 100 mM HEPES pH 7.5, 40 mM $KHCO_3$, 1.5 mM ATP, 0.4 mM NADH, 200 mM KCl, 10 mM $MgCl_2$, 0.5 mM phosphoenolpyruvate, 3.5/3.7 U of lactate dehydrogenase/pyruvate kinase (Sigma), 0.25 μM enzyme (except for MCC, which was at 1.2 μM) and various concentrations of acyl-CoA. The absorbance at 340 nm was monitored for 1.5 min. The initial velocities were fitted to the Michaelis–Menten equation using the program Origin (OriginLab).

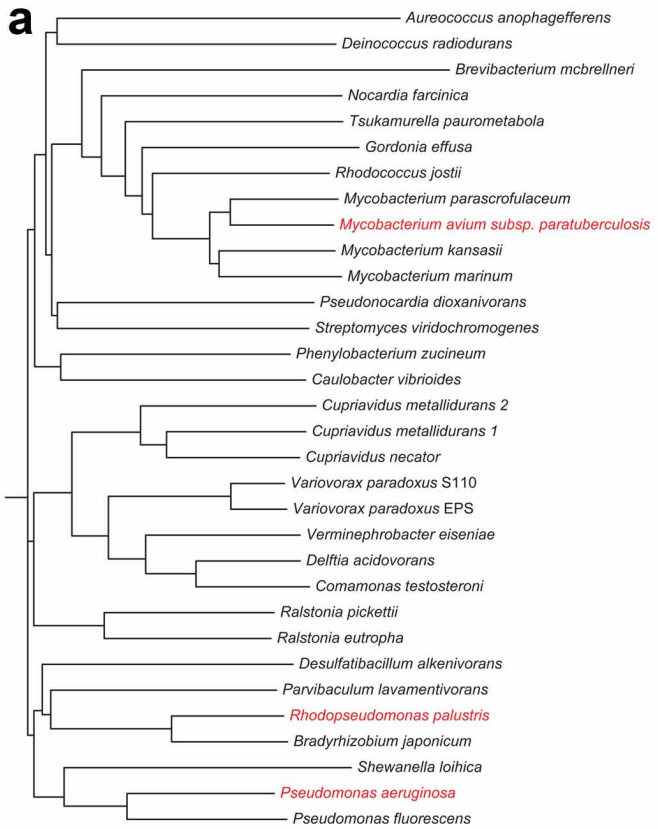
Construction of an LCC deletion mutant in *P. aeruginosa*. A markerless deletion was generated for the gene PA14_46320 in *P. aeruginosa* PA14, using previously described methods[42]. In brief, ~1-kilobase flanking regions for PA14_46320 were amplified with primers listed in Extended Data Table 3 and recombined into the allelic-replacement vector pMQ30 through gap repair cloning in the yeast strain InvSc1 (ref. 43). This plasmid was transformed into *E. coli* BW29427 and moved into PA14 using biparental conjugation. Luria–Bertani (LB) agar containing 100 μg ml^{-1} gentamicin was used to select for *P. aeruginosa* single recombinants. Markerless deletions in PA14_46320 (double recombinants) were then selected with the use of LB agar plates devoid of NaCl and containing 10% (w/v) sucrose as a counter-selection, and their genotypes were confirmed by PCR.

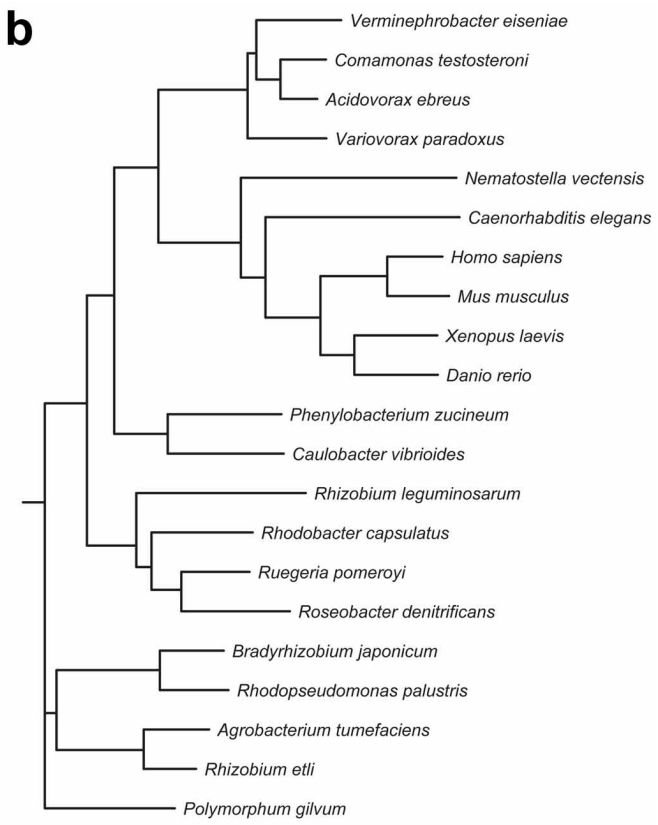
Phenotype microarrays. Phenotype microarray screening was performed by Biolog, Inc., as described[27].

30. Hendrickson, W. A., Horton, J. R. & LeMaster, D. M. Selenomethionyl proteins produced for analysis by multiwavelength anomalous diffraction (MAD): a vehicle for direct determination of three-dimensional structure. *EMBO J.* **9,** 1665–1672 (1990).
31. Otwinowski, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276,** 307–326 (1997).
32. McCoy, A. J. *et al.* Phaser crystallographic software. *J. Appl. Cryst.* **40,** 658–674 (2007).
33. Sheldrick, G. M. A short history of SHELX. *Acta Crystallogr. A* **64,** 112–122 (2008).
34. Terwilliger, T. C. SOLVE and RESOLVE: automated structure solution and density modification. *Methods Enzymol.* **374,** 22–37 (2003).
35. Emsley, P. & Cowtan, K. D. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D* **60,** 2126–2132 (2004).
36. Brunger, A. T. *et al.* Crystallography & NMR System: a new software suite for macromolecular structure determination. *Acta Crystallogr. D* **54,** 905–921 (1998).
37. Ohi, M., Li, Y., Cheng, Y. & Walz, T. Negative staining and image classification—powerful tools in modern electron microscopy. *Biol. Proced. Online* **6,** 23–34 (2004).
38. Ludtke, S. J., Baldwin, P. R. & Chiu, W. EMAN: semiautomated software for high-resolution single-particle reconstructions. *J. Struct. Biol.* **128,** 82–97 (1999).
39. Frank, J. *et al.* SPIDER and WEB: processing and visualization of images in 3D electron microscopy and related fields. *J. Struct. Biol.* **116,** 190–199 (1996).
40. Yang, Z., Fang, J., Chittuluru, J., Asturias, F. J. & Penczek, P. A. Iterative stable alignment and clustering of 2D transmission electron microscope images. *Structure* **20,** 237–247 (2012).
41. Blanchard, C. Z., Lee, Y. M., Frantom, P. A. & Waldrop, G. L. Mutations at four active site residues of biotin carboxylase abolish substrate-induced synergism by biotin. *Biochemistry* **38,** 3393–3400 (1999).
42. Recinos, D. A. *et al.* Redundant phenazine operons in *Pseudomonas aeruginosa* exhibit environment-dependent expression and differential roles in pathogenicity. *Proc. Natl Acad. Sci. USA* **109,** 19420–19425 (2012).
43. Shanks, R. M., Caiazza, N. C., Hinsa, S. M., Toutain, C. M. & O'Toole, G. A. *Saccharomyces cerevisiae*-based molecular tool kit for manipulation of genes from Gram-negative bacteria. *Appl. Environ. Microbiol.* **72,** 5027–5036 (2006).
44. Dereeper, A. *et al.* Phylogeny.fr: robust phylogenetic analysis for the non-specialist. *Nucleic Acids Res.* **36,** W465–W469 (2008).
45. Gouet, P., Courcelle, E., Stuart, D. I. & Metoz, F. ESPript: analysis of multiple sequence alignments in PostScript. *Bioinformatics* **15,** 305–308 (1999).

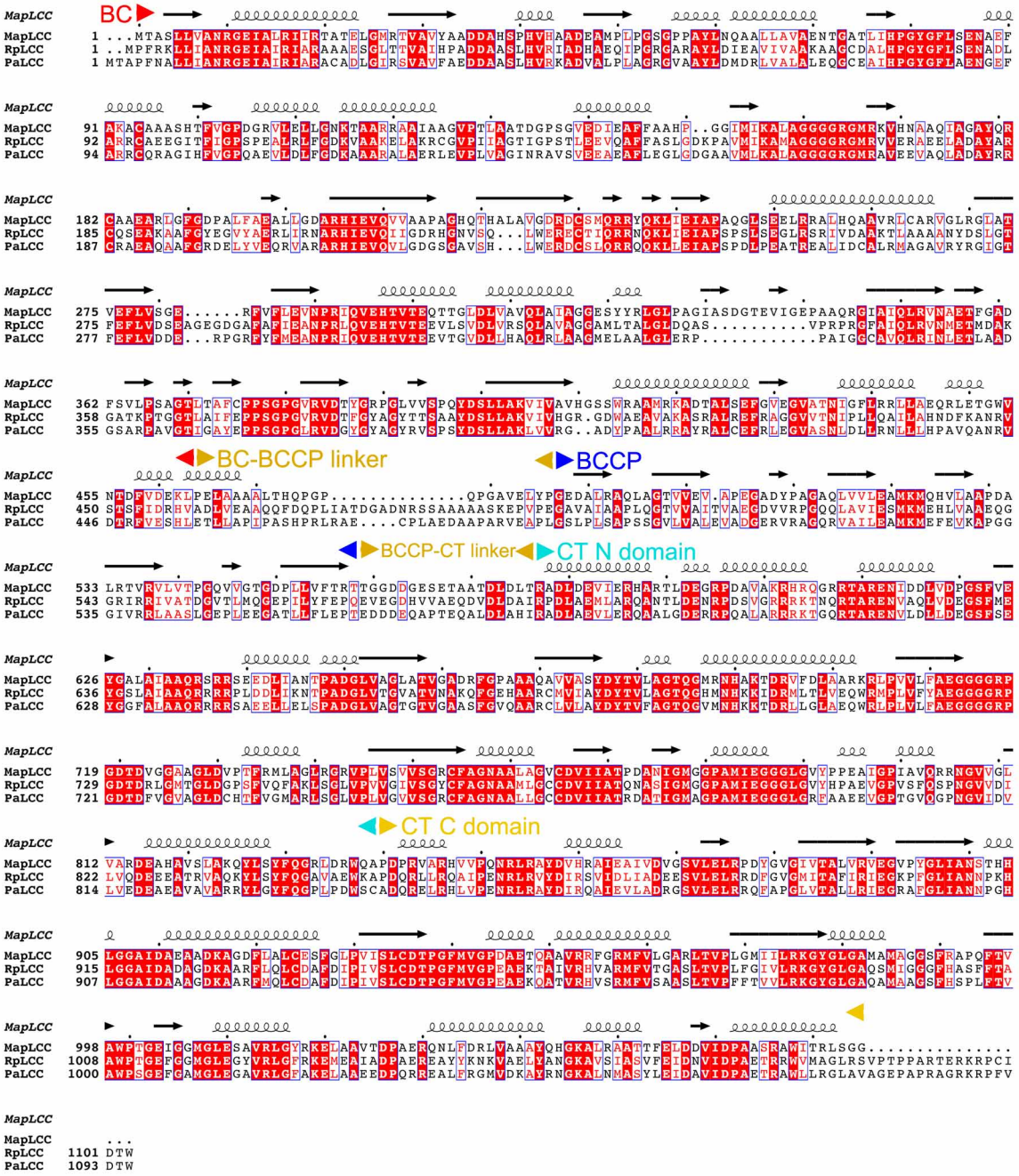


Extended Data Figure 1 | Domain organization of biotin-dependent carboxylases. **a**, Reactions catalysed by biotin-dependent carboxylases. Biotin is linked to the side chain of a Lys residue in BCCP, and this flexible arm has a maximum length of ~16 Å. The BCCP domain must also translocate to reach both active sites, separated by distances of 40–85 Å based on known holoenzyme structures (swinging domain model). **b**, Domain organizations of several representative biotin-dependent carboxylases. Homologous domains are given the same colours. The CT domain of PC has a completely different sequence and structure from those of ACC and PCC. The proteins are drawn to scale, and a scale bar is shown at the bottom. BT, BC–CT interaction domain; PT, PC tetramerization domain, also known as allosteric domain. **c**, Chemical structures of the substrates of ACC, PCC, LCC, MCC and GCC. The site of carboxylation is indicated with the red arrow.

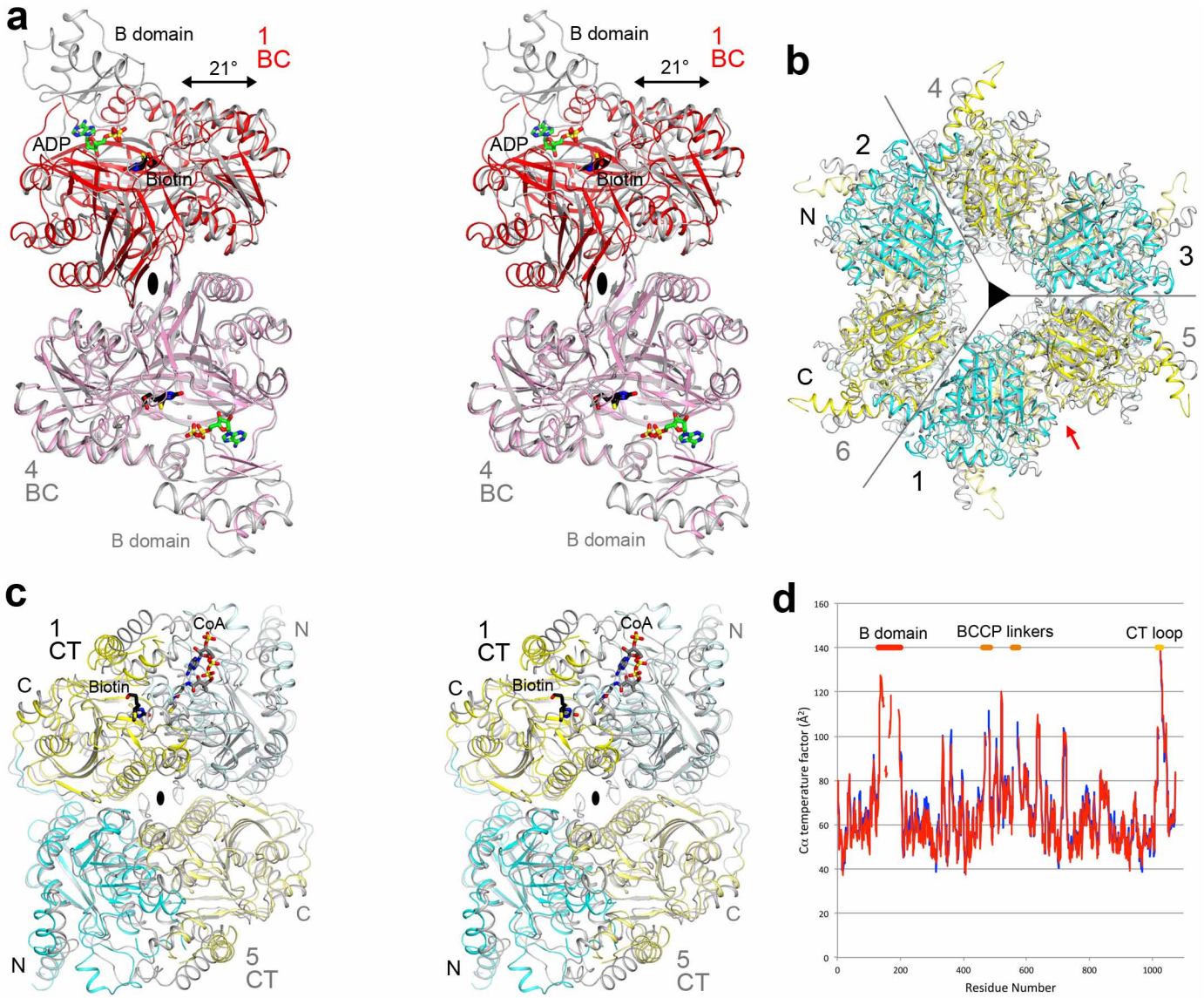




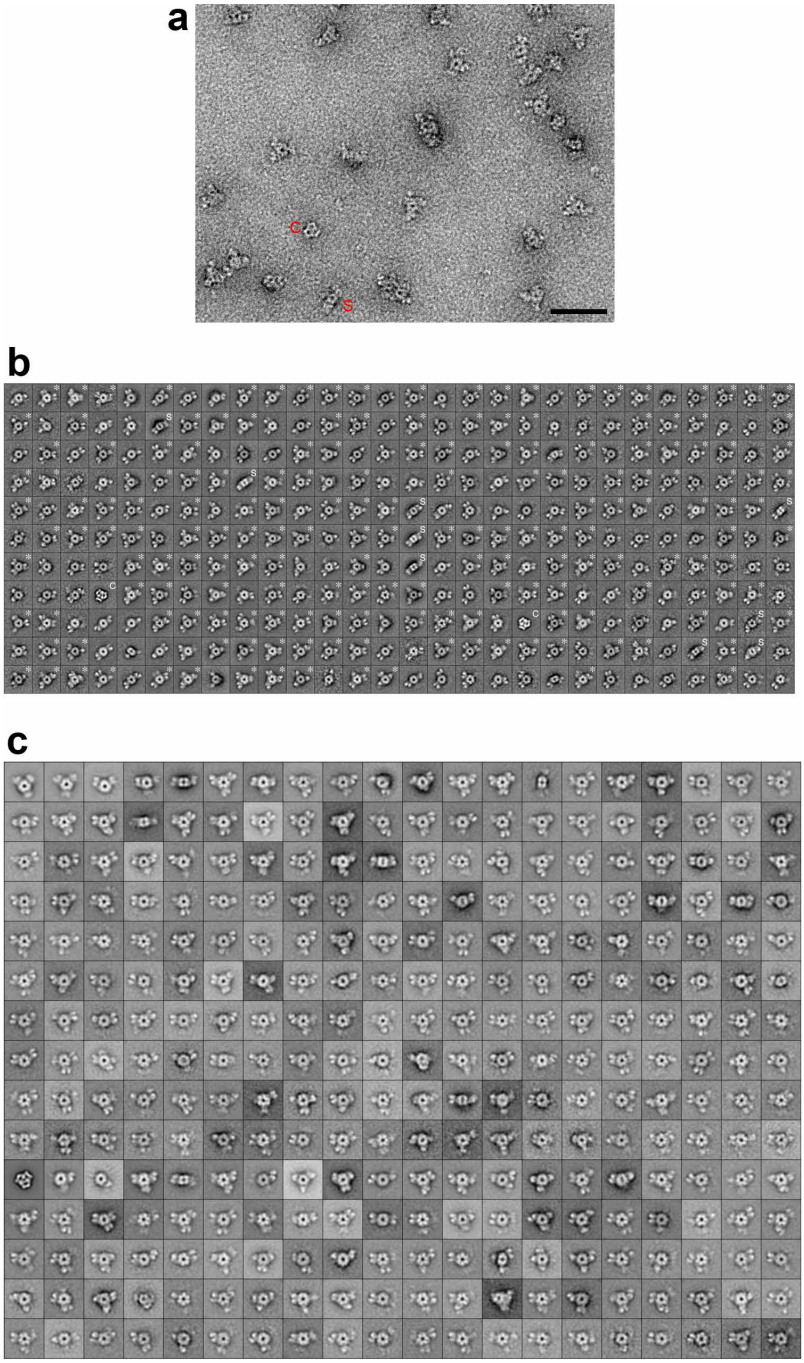
Extended Data Figure 2 | Phylogenetic trees for selected biotin-dependent carboxylases. **a**, Phylogenetic tree for LCC homologues in a collection of organisms. The three homologues studied in this paper are shown in red. **b**, Phylogenetic tree for PCC homologues in a collection of organisms, based on a sequence alignment of the β subunit. Modified from an output from the Phylogeny.fr server[44].



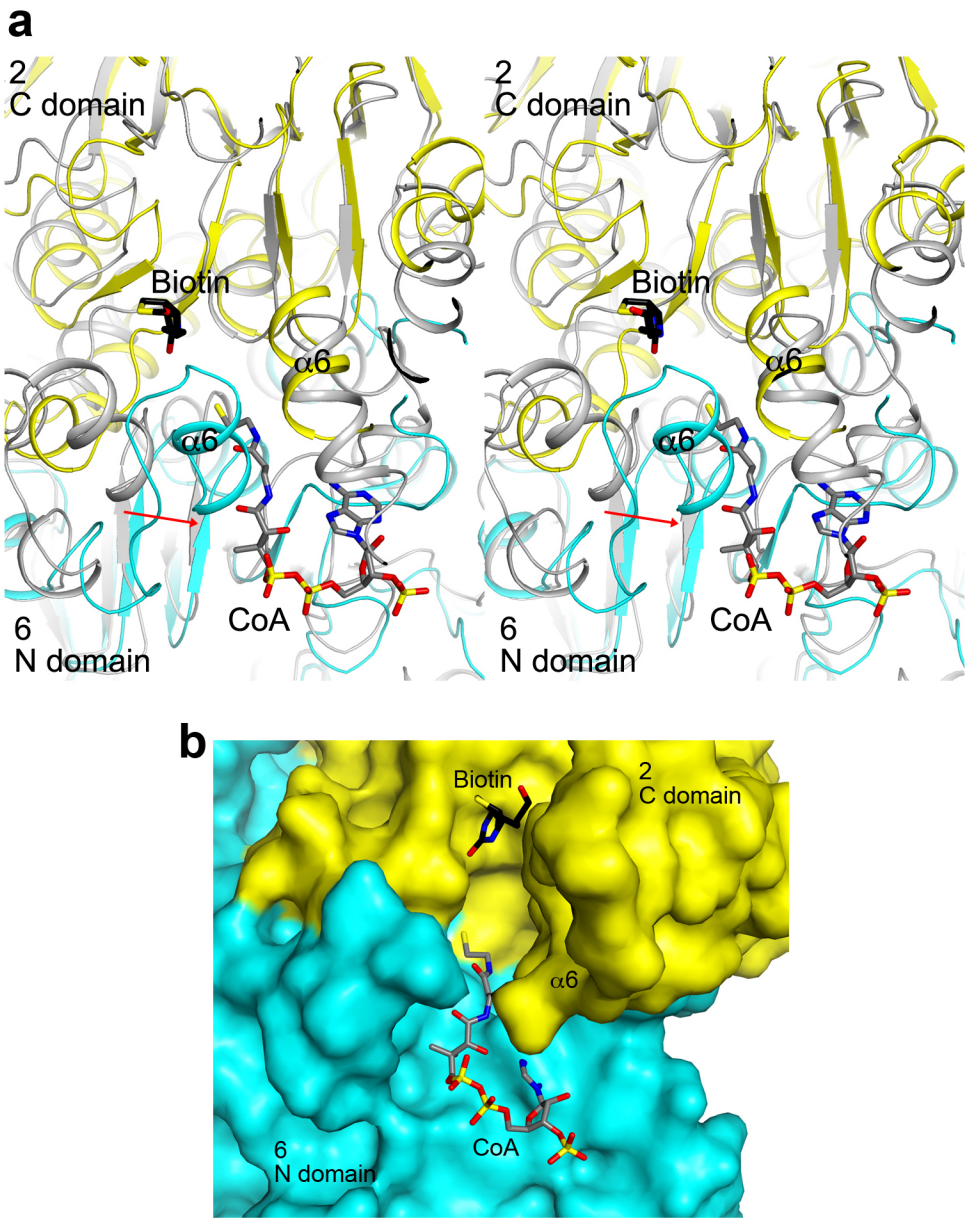
Extended Data Figure 3 | Sequence alignment of long-chain acyl-CoA carboxylases (LCCs) from *M. avium* subspecies *paratuberculosis* (MapLCC), *R. palustris* (RpLCC), and *P. aeruginosa* (PaLCC). The various domains in the proteins are labelled. The BCCP domain has two linkers to the rest of the protein. Modified from an output from ESPript[45].



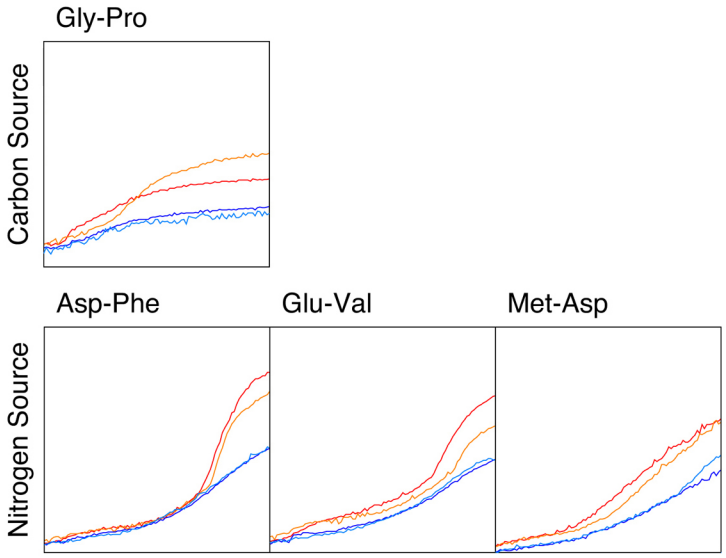
Extended Data Figure 4 | Structural comparisons of domains in LCC with related enzymes. **a**, Stereo drawing of the overlay of the structure of the BC domain dimer of MapLCC (in colour) with that of BC subunit dimer of *E. coli* ACC (in grey)[20]. The bound positions of biotin (black) and ADP (green) in the *E. coli* BC structure are also shown. The two-fold axis of the dimer is indicated by the black oval. With the two monomers at the bottom overlaid, a difference of 21° in the orientations of the two monomers at the top is observed. Most of the B domain of BC is ordered in one of the two monomers of MapLCC. In the other monomer, only weak electron density is observed for a few segments, and the B domain is not modelled. **b**, Overlay of the structures of the CT domain hexamer of MapLCC (in colour) and the β subunit of PCC (in grey)[10]. Each enzyme is highly conserved across species; the overlay should therefore be meaningful. **c**, Stereo drawing of the overlay of the CT domain dimer of MapLCC (in colour) and the β subunit of PCC (in grey). The view is down the red arrow in **b**. The bound position of biotin in the holoenzyme is shown in black. The position of CoA is modelled on that of CoA bound in the active site of the CT domain of yeast ACC[29]. **d**, Plot of the temperature factor value of each Cα atom in the two monomers (in red and blue). Several linker regions with high temperature factor values are indicated.



Extended Data Figure 5 | Electron microscopy studies of LCC.
a, Representative raw image of negatively stained MapLCC. 'S' marks a side view of the holoenzyme, and 'C' indicates a contaminant. Scale bar, 500 Å. **b**, The 308 class averages of negatively stained MapLCC obtained from 19 generations of the iterative stable alignment and clustering (ISAC) procedure[40] implemented in SPARX[21]. These class averages represent 65% (15,932 particles) of the entire data set (24,535 particles). Averages representing side views are marked with 'S', averages that were used to create Supplementary Video 1 are marked with an asterisk, and averages that represent a contaminant are marked with 'C'. The side length of the individual panels is 340 Å. **c**, The averages obtained by classifying all 24,535 particles of negatively stained MapLCC into 300 classes using K-means classification in SPIDER[39]. Averages are shown in rows, with the most populous class at the top left and the least populous class at the bottom right. The side length of the individual panels is 340 Å.



Extended Data Figure 6 | The CT active site of LCC. **a**, Stereo drawing of the overlay of the CT active site (cyan and yellow) of MapLCC with that of PCC (grey)[10]. The model of CoA was obtained from the structure of the complex with yeast ACC CT domain[29]. The α6 helix in the N domain of monomer 6 shows a more closed conformation (indicated by the red arrow) and clashes with the CoA model. There is also a clash with the adenine base of CoA. There may be a conformational change in this region of MapLCC for CoA binding. **b**, Molecular surface of the CT active-site region of MapLCC. The α6 helix in the N domain of monomer 6 was removed for a clearer view of the active site.



Extended Data Figure 7 | Phenotypic differences between wild-type and LCC knockout (ΔPA14_46320) *P. aeruginosa* strains, revealed by a colorimetric assay that monitors the reduction of a tetrazolium dye. The conditions were identified from a screen that sampled 1,920 different media (Biolog Inc.). Assays were performed twice in each medium for the wild-type (red and orange) and mutant (blue and cyan) strains. Shown are activity profiles for strains incubated with Gly-Pro as the sole carbon source (top panel) and Asp-Phe, Glu-Val and Met-Asp as the sole nitrogen source (bottom panels). For each panel, the horizontal axis is time (24 h), and the vertical axis is OmniLog signal[27].

Extended Data Table 1 | Kinetic parameters of *R. palustris* LCC towards various substrates

Substrate	K_m (mM)	k_{cat} (s^{-1})
Acetyl-CoA (C2)	2.0±0.3	0.77±0.05
Propionyl-CoA (C3)	2.2±0.5	0.54±0.07
Butyryl-CoA (C4)	0.82±0.16	0.35±0.03
Hexanoyl-CoA (C6)	0.20±0.04	0.30±0.01
Octanoyl-CoA (C8)	0.37±0.06	0.33±0.01
Decanoyl-CoA (C10)	0.033±0.005	0.28±0.01
Lauroyl-CoA (C12)	0.019±0.003	0.24±0.005
Myristoyl-CoA (C14)	0.026±0.002	0.45±0.01
Palmitoyl-CoA (C16)	0.0058±0.0005	0.16±0.003
3-methylcrotonyl-CoA	1.1±0.1	0.074±0.004

The errors are standard deviations from fitting one titration curve to the Michaelis–Menten equation.

Extended Data Table 2 | Data collection and refinement statistics

	MapLCC
Data collection	
Space group	$P2_13$
Cell dimensions	
a, *b*, *c* (Å)	220.9, 220.9, 220.9
α, β, γ (°)	90, 90, 90
Resolution (Å)	50-3.0 (3.1-3.0) *
R_{merge}	9.6 (44.6)
$I/\sigma I$	10.3 (1.9)
Completeness (%)	91 (72)
Redundancy	3.3 (2.1)
Refinement	
Resolution (Å)	50-3.0
No. reflections	64,953
R_{work}/ R_{free}	20.9 / 26.2
No. atoms	
Protein	14,632
Ligand/ion	0
Water	0
B-factors	
Protein	66.4
Ligand/ion	–
Water	–
R.m.s deviations	
Bond lengths (Å)	0.007
Bond angles (°)	1.4

Two crystals were used for data collection.
* The highest-resolution shell is shown in parenthesis.

Extended Data Table 3 | Primers used for making the LCC deletion mutant

Primer	Sequence (5' to 3')
ΔPA14_46320 flank 1F	CCAGGCAAATTCTGTTTTATCAGACCGCTTCTGCGTTCTGAT GCTGCCTGCTCTACATGCT
ΔPA14_46320 flank 1R	CCTTCAACGCCTTGCTGATCCAGCTACCTGGAGATCGAC
ΔPA14_46320 flank 2F	GTCGATCTCCAGGTAGCTGGATCAGCAAGGCGTTGAAGG
ΔPA14_46320 flank 2R	GGAATTGTGAGCGGATAACAATTTCACACAGGAAACAGCT GGCGCGACCAGTAGAGATT