Phosphorus is required for all life and microorganisms can extract it from their environment through several metabolic pathways. When phosphate is in limited supply, some bacteria are able to use phosphonate compounds, which require specialized enzymatic machinery to break the stable carbon–phosphorus (C–P) bond. Despite its importance, the details of how this machinery catabolizes phosphonates remain unknown. Here we determine the crystal structure of the 240-kilodalton *Escherichia coli* C–P lyase core complex (PhnG–PhnH–PhnI–PhnJ; PhnGHIJ), and show that it is a two-fold symmetric hetero-octamer comprising an intertwined network of subunits with unexpected self-homologies. It contains two potential active sites that probably couple phosphonate compounds to ATP and subsequently hydrolyse the C–P bond. We map the binding site of PhnK on the complex using electron microscopy, and show that it binds to a conserved insertion domain of PhnJ. Our results provide a structural basis for understanding microbial phosphonate breakdown.

Phosphonate compounds that contain a stable C–P bond are used as a source of phosphate by microorganisms in many natural environments where low levels of free and organic phosphate limit growth. The C–P lyase pathway, which converts phosphonate into 5-phosphoribosyl-α-d-1-diphosphate (PRPP) in an ATP-dependent fashion, is activated upon phosphate starvation in many bacterial species including *Escherichia coli*. The enzymes of this pathway have a very broad substrate specificity enabling the bacteria to utilize a wide range of compounds for growth including alkyl, amino-alkyl and aryl phosphonates.

In *E. coli*, the 14-cistron *phn* operon is required for phosphonate uptake and utilization and encodes an ATP-binding cassette transporter (PhnC, PhnD and PhnE), a regulatory protein (PhnF) and components required for enzymatic conversion into PRPP (PhnGHIJKLMNOP). PhnG, PhnH, PhnI and PhnJ have been shown to form a stable protein complex, which we term the C–P lyase core complex, probably with PhnG and PhnI at its centre. The core complex stably associates with a fifth protein, PhnK, which resembles ABC cassette proteins, with unknown stoichiometry. PhnK contains an iron–sulfur cluster required for C–P bond cleavage (PhnJ). PhnH is the only component of the C–P lyase core complex that has been structurally characterized and displays a fold related to the pyridoxal-5′-adenosyl methionine (SAM)-dependent radical mechanism.

The global architecture of C–P lyase

We purified the *E. coli* C–P lyase core complex and determined its crystal structure by molecular replacement in combination with single-wavelength anomalous dispersion using a T₄Br₂Cl₀₂⁻ cluster derivative and the PhnH crystal structure as a search model. The structure was refined using a native data set extending to 1.7 Å with resulting final *R* factors of 14.9% (*R*_work) and 17.6% (*R*_free) (Extended Data Table 1 and Extended Data Fig. 2). The structure consists of two copies of each of PhnG (16 kDa), PhnH (21 kDa), PhnI (39 kDa) and PhnJ (32 kDa), comprising a total of 1,958 amino acid residues in the asymmetric unit (Fig. 1b and Extended Data Fig. 3), and is complete except for a few residues located at the subunit termini. The structure includes four sulfate ions, four zircon ions and 1,792 solvent molecules. Together, the eight polyepitides form a compact and intertwined, two-fold symmetric hetero-octamer that can be described as (PhnGHIJ)₂, with a total molecular mass of 240 kDa (Fig. 1c, d), consistent with its behaviour in solution.

It forms a homodimer when expressed independently, and its role within the complex is unclear.

The C–P lyase core complex resembles the letter ‘H’ with rounded arms that are twisted approximately 45° in and out of the plane with respect to each other. The arms are composed on opposing sides by the two PhnG molecules and on the other sides by tight complexes of PhnH and PhnI (Fig. 1c). At the centre of the molecule, a compact PhnH homodimer forms a disc-like structure that serves as a central hub for attachment of the other subunits (Fig. 1c, d, green). The core domain of PhnH, which is the largest single domain in the structure, has a novel α + β fold comprised of a four-stranded, anti-parallel β-sheet next to a four-helix bundle combining in a unique fold: the β-barrel domain (Figs 1c and 2a). At both termini there are helical extensions of approximately 35 residues that grasp PhnI and tether it to the complex via extensive interactions (Fig. 3a). In turn, PhnH attaches PhnI to the complex through packing of conserved α-helices in both proteins (Fig. 1c).

---

1. Department of Molecular Biology and Genetics, Aarhus University, Gustav Wieds Vej 10c, DK-8000 Aarhus C, Denmark.
2. Medical Research Council Laboratory of Molecular Biology, Francis Crick Avenue, Cambridge CB2 0QH, UK.
PhnJ has a compact α + β fold surrounded by two mini-domains, the central insertion domain (CID) and the C-terminal mini domain (CMD) (Figs 1b, 2b and Extended Data Fig. 3). Surprisingly, the core folds of PhnJ and PhnH are nearly identical (Cα r.m.s.d. 2.5 Å), despite very little sequence similarity (Extended Data Fig. 4a). Moreover, the interactions in the PhnH–PhnJ heterodimer closely resemble those observed in the crystal structure of the isolated PhnH homodimer (Extended Data Fig. 5a)19. The CID is an insertion in PhnJ between β5 and β6 of the corresponding PhnH fold and consists of two α-helices and a short 3_10-helix (Fig. 2b). The CID is well conserved among PhnJ orthologues and contacts both of the central PhnI molecules (Fig. 3b). Finally, the CMD is located at the C terminus and consists of a small β-hairpin and a helix. It is stabilized by a zinc ion coordinated by four conserved cysteine residues: Cys241, Cys244, Cys266 and Cys272 (Fig. 2b).

The PhnI monomers bind each other via an extensive, conserved surface interaction area comprising ~75% of the total PhnGHJ complex dimerization interface (Extended Data Fig. 6). Each molecule of PhnI interacts with both copies of PhnG (Fig. 3c), the smallest protein in the complex displaying an elongated α + β fold with a four-stranded, antiparallel β-sheet against a four-helix bundle (β-barrel domain, Fig. 2c). Despite very little sequence similarity, the closest known structural homologue of PhnG is PhnI, with which it shares both the long β-hairpin and the helical bundle (Fig. 2 and Extended Data Fig. 4b). The PhnG β-hairpin and C-terminal helix form a molecular clamp that connects to a groove in PhnI, forming an unusually long, combined β-barrel domain (80 Å, Fig. 3d).

The iron–sulfur binding site
PhnJ belongs to the anaerobic radical SAM enzyme superfamily in which three conserved cysteine residues coordinate a cubane-like Fe₄S₄ cluster20 that promotes formation of a free electron radical required for catalysis by reductive cleavage of SAM to a 5’-deoxyadenosyl radical (Ado-CH₂) and L-methionine21–23. PhnJ does not contain the canonical CX₃CX₃C motif but rather a CX₂CX₂C motif involving Cys241, Cys244 and Cys266, which are both necessary and sufficient for reconstitution of

**Figure 1** | Overall architecture of the C–P lyase core complex. a, The C–P lyase core complex catalyses transfer of a phosphonate to the 1’ position of ATP (PhnJ assisted by PhnG, PhnH and PhnL) and cleavage of the C–P bond (PhnJ). b, Overview of the four proteins with dashed lines indicating conserved structural domains. Functional residues are shown with amino acid one-letter code. BBD, β-barrel domain; CID, central insertion domain; CMD, C-terminal mini domain; NTD, N-terminal domain. c, Overall structure of the 240 kDa C–P lyase core complex. d, Schematic architecture of the complex with structural domains indicated.

**Figure 2** | Details of subunit structures. a–d, Details of the individual protein structures in the complex: PhnJ (a), PhnH (b), PhnG (c) and PhnH (d), aligned to show their structural homologies and with domain colours as in Fig. 1b. Ions are shown as spheres: sulfate, red and yellow; zinc, pink. BBD, β-barrel domain; CID, central insertion domain; CMD, C-terminal mini domain; NTD, N-terminal domain.
Perhaps could detach during the reaction (Fig. 2b and Extended Data suggesting that it is relatively loosely attached to the PhnJ core and Cys272 is bond cleavage of 5-phosphoribosyl-1-phosphonate (Extended Data residue (Cys272) generates a thiyl radical capable of homolytic C–P transfer of the radical from Gly32 to the fourth conserved cysteine proximity. The CMD containing the cluster site has higher B the position of the cluster relative to Gly32 to bring them into prox–

The iron–sulfur cluster in vitro\textsuperscript{17,18}. In the structure, these cysteines are juxtaposed and coordinate a zinc ion (Fig. 4a) in an arrangement that closely matches that expected for an Fe\textsubscript{4}S\textsubscript{4} cluster-containing protein (Fig. 4b). Furthermore, super-positioning an S-adenosyl methionine activase structure on this region reveals a small groove on the surface of the C–P lyase core complex next to the cluster site that might accommodate SAM\textsuperscript{24} (Extended Data Fig. 5b).

According to the proposed reaction mechanism, the Ado-CH\textsubscript{2}\textsuperscript{+} radical is transferred to the universally conserved Gly32 of PhnJ, generating a stable glycyld radical enzyme that supports multiple turn–overs without further SAM consumption\textsuperscript{17}. According to this scheme, transfer of the radical from Gly32 to the fourth conserved cysteine residue (Cys272) generates a thyl radical capable of homolytic C–P bond cleavage of 5-phosphoribosyl-1-phosphonate (Extended Data Fig. 1) through a thiophosphonate radical intermediate\textsuperscript{17}. Cys272 is situated adjacent to the cluster site where it is the fourth ligand binding the zinc ion, while Gly32 is located more than 30 Å away, in the vicinity of PhnH (Fig. 4c). A direct involvement of Gly32 in the reaction\textsuperscript{20} is therefore difficult to reconcile with the structure; however, it cannot be excluded that structural rearrangements could alter the position of the cluster relative to bring them into prox–imity. The CMD containing the cluster site has higher B factors, suggesting that it is relatively loosely attached to the PhnJ core and perhaps could detach during the reaction (Fig. 2b and Extended Data Fig. 5c).

A second potential active site

At the interface of PhnI and PhnJ, three universally conserved histidine residues come together to form a second metal-ion-binding site. Analysis of the anomalous difference density confirms that this His

site also contains zinc (Fig. 4d). Two of the residues (PhnI His328 and His333) coordinate the zinc ion directly (2.4 Å), while the third (PhnI His108) is further away (4.5 Å). The three histidines are located in a cavity between PhnI and PhnJ that connects to the surface of the complex via a solvent-accessible tunnel (Fig. 4e). The cavity also contains a sulfate ion located 9.5 Å from the zinc, which may mimic a substrate phosphate or phosphonate. Finally, access to the cavity is defined by the PhnJ CID domain, which forms a lid-like domain.

Studies of zinc-binding proteins show that structural zinc sites usually have four protein ligands, while active-site zinc ions have a more open coordination sphere with 2–3 ligands similar to that observed in this case\textsuperscript{25}. To assess the functional importance of the His site, we used genetic complementation to determine whether mutation of the histidine residues affects the ability of E. coli to utilize phosphonates. A plasmid-born copy of the wild-type phnHJJKLMNOP allele was used to complement E. coli Apn–HJJKLMNOP under conditions where phosphonates were the sole phosphate source (Extended Data Fig. 7)\textsuperscript{26}. Unlike the wild type, none of the variants (PhnI(H333A), PhnI(H328A;H333A), PhnJ(H108A) and PhnJ(C272A)) could utilize phosphonates;
we therefore conclude that the His site is required for the activity of the C–P lyase core complex in vivo.

**PhnK binds via the PhnJ CID**

The C–P lyase core complex stably associates with a fifth component, PhnK (28 kDa). The function of PhnK is unclear but it contains the consensus elements of an ATP-binding cassette protein, suggesting that it might deliver nucleotides for the reaction (Extended Data Fig. 8). Despite its ability to stably co-purify, we were unable to obtain crystals of a complex including PhnK. We mapped the PhnK-binding site on the complex using negative-stain electron microscopy by generating a 3D reconstruction of purified PhnGHIJK (Fig. 5a and Extended Data Fig. 9). The crystal structure fits tightly within the resulting electron microscopy density map and reveals additional density in a groove close to the two-fold symmetry axis near two regions of highly conserved residues on PhnJ (Fig. 5a, b). The fold of PhnK can be roughly modelled using a homologous nucleotide-binding domain of an ABC transporter (Protein Data Bank accession code: 4FWI). The electron microscopy map is consistent with a single PhnK binding unilaterally to the complex, breaking the two-fold symmetry (Fig. 5c). Although the exact orientation of PhnK cannot be established at this resolution, we note that one side is highly conserved among orthologues, suggesting that it comprises the interaction surface (Fig. 5c).

ABC modules often dimerize in a head-to-tail fashion, binding ATP between the Walker A/B motifs of one subunit and the ABC motif of the other. PhnK contains a variant ABC motif (FSGGMQ versus LSGGQ), which could serve to bind the C–P lyase core complex (Extended Data Fig. 8). The conserved CID domain protrudes into the PhnK-binding region, so to probe its importance we constructed C–P lyase core complexes lacking residues 130–171 of PhnJ (Extended Data Fig. 10). Purification PhnGHJK(CID)K demonstrated that upon deletion of the CID, the C–P lyase core complex remains intact but PhnK is missing, thus indicating that the CID region of PhnJ is required for tethering PhnK to the core complex.

**Discussion**

In this paper, we delineate the organization and detailed molecular structure of a core complex involved in phosphonate catabolism in bacteria. We show that four of the proteins required for phosphonate breakdown assemble into a large, hetero-octameric complex with two-fold symmetry and that the symmetry is broken by binding of a fifth, ATP-binding subunit, PhnK. The structure is not immediately compatible with the direct involvement of Gly32 (PhnJ) in catalysis, but structural rearrangements may affect the location of this residue during the reaction. Many glycyl radical enzymes require separate activation enzymes that dissociate upon radical formation, a task that could also be maintained by a flexible internal domain.

The structure indicates the existence of a second active site at the interface of PhnI and PhnJ. Analysis of difference electron density maps from several independent data sets revealed a consistent density next to the bound zinc ion, but we have been unable to identify the bound molecule. We also carried out co-crystallization using a range of compounds including nucleotides and phosphonates, but no further substrate binding was observed. This suggests that the complex needs an Fe4S4 cluster, or is not in the correct conformational state to bind a substrate. We speculate that the His site is required for coupling a phosphonate to ATP, which is known to depend on PhnI.

Using electron microscopy we locate the binding region for PhnK on the C–P lyase core complex. While this does not reveal the role of PhnK in the reaction, we note that the region is close to the His site and it is therefore possible that structural changes in PhnK occurring upon ATP hydrolysis may affect substrate access. With the detailed architecture of the C–P lyase core complex thus delineated, future work will focus on understanding the requirements of the two reactions catalysed by the complex and definitively locating the binding sites of substrates and reaction intermediates.

**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 7 January; accepted 22 June 2015.
Published online 17 August 2015.

1. Karl, D. M. Aquatic ecology: Phosphorus, the staff of life. *Nature* **406**, 31–33 (2000).
2. Hsieh, Y. J. & Wanner, B. L. Global regulation by the seven-component Pi signaling system. *Curr. Opin. Microbiol.* **33**, 198–203 (2010).
3. Hove-Jensen, B., Zechel, D. L. & Jochimsen, B. Utilization of glyphosate as phosphate sources: biochemistry and genetics of bacterial carbon-phosphorus lyase. *Microbiol. Mol. Biol. Rev.* **8**, 176–197 (2014).
4. Imazu, K. et al. Enhanced utilization of phosphonate and phosphate by Klebsiella aerogenes. *Appl. Environ. Microbiol.* **64**, 3754–3758 (1998).
5. Quinn, J. P. Carbon phosphate lyase activity—a novel mechanism of bacterial resistance to the phosphonic acid antibi. *Lett. Appl. Microbiol.* **8**, 113–116 (1989).
6. Schowanek, D. & Verstraete, W. Phosphonate utilization by bacterial cultures and enrichments from environmental samples. *Appl. Environ. Microbiol.* **56**, 895–903 (1990).
18. Kamat, S. S., Williams, H. J. & Raushel, F. M. Intermediates in the transformation of phosphonates to phosphate by bacteria. *Nature* **525**, 710–717 (2015).

19. Adams, M. A. Physiological role of PheN-specified organophosphonate degradation in *Escherichia coli*. *J. Bacteriol.* **186**, 3461–3474 (2004).

20. Shisler, K. A. & Broderick, J. B. Glycyl radical activating enzymes: structure, mechanism, and substrate interactions. *Annu. Rev. Biochem.* **86**, 64–71 (2017).

21. Booker, S. J., Groth, T. L., Hall, B. L., Nusskern, D. R. & Ruzicka, F. J. Mechanistic and functional versatility of radical SAM enzymes. *Biochim. Biophys. Acta* **1797**, 977–987 (2010).

22. Frey, P. A., Hegeman, A. D. & Ruzicka, F. J. The radical SAM superfamily. *Crit. Rev. Biochem. Mol. Biol.* **43**, 63–88 (2008).

23. Sofía, H. J., Chen, G., Hetzler, B. G., Reyes-Spindola, J. F. & Miller, N. E. Radical SAM, a novel protein superfamily linking unresolved steps in familiar biosynthetic pathways with radical mechanisms: functional characterization using new analysis and information visualization methods. *Nucleic Acids Res.* **39**, 1097–1106 (2011).

24. Goldman, P. J. et al. X-ray structure of an AdoMet radical activase reveals an anaerobic solution for formylglycine posttranslational modification. *Proc. Natl Acad. Sci. USA* **110**, 8519–8524 (2013).

25. McColl, K. A., Huang, C. & Fierke, C. A. Function and mechanism of zinc metalloenzymes. *J. Nutr.* **130**, 14375–14465 (2000).

26. Yakovleva, G. M., Kim, S. K. & Wanner, B. L. Phosphate-independent expression of the carbon-phosphorus lyase activity of *Escherichia coli*. *Appl. Microbiol. Biotechnol.* **49**, 573–578 (1998).

27. Li, X. et al. Structure of the nucleotide-binding domain of a dipeptide ABC transporter reveals a novel iron-sulfur cluster-binding domain. *Acta Crystallogr. D Biol. Crystallogr.* **69**, 256–265 (2013).

28. Holleinstein, K., Dawson, R. J. & Locher, K. P. Structure and mechanism of ABC transporter proteins. *Curr. Opin. Struct. Biol.* **17**, 412–418 (2007).

29. Landau, M. et al. ConSurf 2005: the projection of evolutionary conservation scores of residues on protein structures. *Nucleic Acids Res.* **33**, W299–W302 (2005).

Acknowledgements We are thankful to T. L. Sørensen at Diamond, T. Weinert at SLS as well as beamline staff at ESRF and MAX-Lab for help during X-ray data acquisition and S. Chen, C. G. Savva, J. Grimmett and T. Darling at the MRC-LMB for technical assistance with electron microscopy. This work was supported by the European Research Council grant no. 261151 (L.A.P.), MRC grant MC_U105192715 (L.A.P.), and the Danish National Research Foundation ‘Centre for mRNP biogenesis and metabolism’ (D.E.B.).

Author Contributions P.S., L.A.P., B.H.J., B.J. and D.E.B. designed and carried out the experiments. P.S., M.K. and D.E.B. determined the crystal and EM structures while C.J.R. and L.A.P. carried out final refinement of the EM structure as well as EM structure validation. P.S., M.K., C.J.R., L.A.P., B.H.J., B.J. and D.E.B. wrote the manuscript.

Author Information Atomic coordinates and structure factors have been deposited in the Protein Data Bank (PDB) with accession code 4XB6. The EM density map has been deposited in the Electron Microscopy Data Bank (EMDB) with accession code EMDB-4183. 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 D.E.B. (deb@mbg.au.dk).
METHODS

No statistical methods were used to predetermine sample size. The experiments were not randomized and the investigators were not blinded to allocation during experiments and outcome assessment.

Protein expression and protein purification. The construction of pHO572 (expressing PhnGHIIK), and pHO575 (expressing PhnGHIIK) as well as gene expression in E. coli strain HB273 are described previously4,5. pHO575 encodes a C-terminally six-histidine tagged version of PhnK while PhnGHIIK has no tag but still binds to Ni2+ NTA agarose beads. The PhnGHIIK/(ACID) and PhnGHIIK/(ACID)K constructs were created by site-directed mutagenesis using primers 5′-TGGCAATCCCGGGGGGTCATCTGGAATGGA-3′ (ACID forward), and 5′-TACCTACCGGATATCCGCTGGAGGATTCCGAC-3′ (ACID reverse) which result in the replacement of residues 130–172 of Phn by two glycine residues (NZ1 lined in the proteins). Cells were in all cases grown at 37 °C in Luria Broth (LB) medium and gene expression achieved overnight at 18 °C by induction using 0.5 mM isopropyl-β-D-thiogalactoside (IPTG). Cells were pelleted by centrifugation at 16,000 rpm for 45 min at 4 °C and resuspended in lysis buffer (50 mM HEPES, pH 7.5, 500 mM NaCl, 5 mM MgCl2, 20% (v/v) glycerol, and 3 mM 2-mercaptoethanol) supplemented with complete Protease Inhibitor Cocktail tablets (Sigma) and lysed by high-pressure homogenization (EmulsiFlex-C5, Evesin) at 15,000 psi. The lysates were centrifuged at 16,000 rpm for 45 min to remove cell debris and bound to Ni2+ NTA agarose beads on a 5-ml pre-packed HisTrap HP column (GE Healthcare), pre-equilibrated with lysis buffer (PhnGHIIK) or lysis buffer plus 200 mM imidazole (PhnGHIIJK). In all cases, the complexes were eluted by increasing the imidazole concentration to 250 mM. Following overnight dialysis at 4 °C against buffer LSI (50 mM HEPES, pH 7.5, 750 mM NaCl, 5 mM MgCl2, and 5 mM 2-mercaptoethanol), the samples were applied to a 1 ml Source 15Q column (GE Healthcare), pre-equilibrated with buffer LSI and eluted using a linear gradient from 100–600 mM NaCl. The samples were then diluted to reach 250 mM NaCl and passed over a 1 ml Mono Q column (GE Healthcare) pre-equilibrated with buffer L52 (30 mM HEPES, pH 7.5, 250 mM NaCl, 5 mM MgCl2, and 5 mM 2-mercaptoethanol). Crystals suitable for data collection appeared within 2–3 days reaching maximum crystal size in 15–20 days. Data sets used for macromolecular structure solution were obtained using batch crystallization at 4 °C in Luria Broth (LB) medium and gene expression achieved overnight at 18 °C by induction using 0.5 mM isopropyl-β-D-thiogalactoside (IPTG). Cells were pelleted by centrifugation at 16,000 rpm for 45 min at 4 °C and resuspended in lysis buffer (50 mM HEPES, pH 7.5, 500 mM NaCl, 5 mM MgCl2, and 5 mM 2-mercaptoethanol), washed, and eluted using a 250–400 mM NaCl gradient finally.

Electron microscopy. Purified samples of PhnGHIIKJ were applied to Quantifoil R2/2 holey carbon on copper grids (Quantifoil, Jena) covered with a thin additional film of amorphous carbon, and rendered more hydrophilic with a 9:1 argon–oxygen plasma (Fischione Model 1070). The specimens were stained with 3% ammonium molybdate at pH 8 followed by 2% uranyl acetate. Micrographs were recorded at 44,000× magnification on a Tecnai T12 microscope equipped with a US4000 4K × 4K pixel CCD detector (Gatan) at 120 keV with defoci in the 0.8–2 μm range and using an electron dose of 20 electrons/A2. Leaflet particles were manually picked from 105 micrographs using eboxer from the EMAN2 software package28 and contrast transfer function parameters were determined using CTFFFIND3 (ref. 43). Three iterations of 2D classification were performed in REELION using 300 class averages to determine particles that did not align well with each other. These particles were removed from subsequent analysis. After 2D classification, the final set of 10,033 particles was used to calculate a 3D reconstruction in RELION without symmetry imposed. The initial model for the reconstruction was prepared by low-pass filtering a density map generated from the C–P lyase core complex crystal structure to 40 Å. The final model has a resolution of 1.6 Å by the 0.143 ‘gold standard’ Fourier–shell correlation4 and a resolution of 28 Å versus the crystal structure at FSC = 0.5. The latter is probably closer to the true resolution of the map as the granular nature of negative stain can introduce correlations in the half-maps that are not related to the protein structure. The map was validated using 419 tilt-pairs recorded using angles 0 ° and 30 ° (t′ < 0.11, t = 2.7)4. The FSC versus the crystal structure shows correlation between the crystal structure and the electron microscopy density at low resolution, after which deviations due to structural differences between the C–P lyase core complex and the PhnGHIIKJ complex become apparent.

In vivo complementation. For the in vivo complementation studies, E. coli strain BW16711 (pHnphnGHIIKLMNOP) was transformed by the plasmid pG17, conferring ampicillin resistance and encoding phnGHIIKLMNOP, and analysed for its ability to grow on MOPS minimal plates supplemented with 0.2% glucose, 100 μg ml−1 ampicillin, 0.1 mM IPTG and 0.2 mM of either methyl phosphytol, 2-aminophenylphosphate, or phosphate ion as a positive control. The Phn (H328A), variant, the Phn (H328A,H333A) double variant, the Phn(C272A) variant as well as the Phn(H108A) variant were introduced into pG17 by site-directed mutagenesis using PCR using the following primers, 5′-GAC-GTT-TTGC-GACTGG-CAT-GCC-CC-10,137 and 3′-CTGG-GAG-GAG-CAC-AC-134. The PCR product was ligated into pG17 linearized with XhoI and transformed into non-ligated DNA into NovaBlue Singles (Novagen) electrocompetent E. coli cells and selection on ampicillin plates. All mutations were confirmed by sequencing of the entire phnGHIIKLMNOP region of the pG17 vector to ensure that no other spontaneous mutations had been introduced that could prevent rescue of the BW16711 strain.

3. 3D-ARACY, A., Villard, F. & Marsh, M. An automated microtomed matrix-screening method for protein crystalization. Acta Crystallogr. D 63, 550–554 (2007).
4. Knäblein, J. et al. Ta6Br12, a tool for phase determination of large biological assemblies by X-ray crystallography. J. Mol. Biol. 270, 1–7 (1997).
5. Winter, G. xia2: an expert system for macromolecular crystallography data collection. J. Appl. Cryst. 43, 186–190 (2010).
6. Adams, P. D. et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. Acta Crystallogr. D 66, 213–221 (2010).
7. Cohen, S. K. et al. Towards complete validated models in the next generation of electron cryomicroscopy. Acta Crystallogr. F 66, 222–229 (2010).
8. Thorn, A. & Sheldrick, G. M. Extending molecular-replacement solutions with phnGHIIKLMNOP. Acta Crystallogr. D 66, 222–229 (2010).
9. Cowtan, K. The Buccaneer software for automated model building. 1. Tracing macromolecular structure solution. Acta Crystallogr. D 66, 698–501 (2010).
10. Chen, V. B. et al. MolProbity: all-atom structure validation for macromolecular crystallography. Acta Crystallogr. D 66, 12–21 (2010).
11. Schrödinger, LLC. The PyMOL molecular graphics system, version 1.3r1 (2010).
12. Poterssen, E. et al. UCSF Chimera—a visualization system for exploratory research and analysis. J. Comput. Chem. 25, 1605–1612 (2004).
13. Ermanbatur, E., Wohlfart, K. & Tichelaar, W. Perfumed support tools with predefined hole size, shape and arrangement. Ultramicroscopy 74, 75–81 (1998).
14. Tang, G. et al. EMAN2: an extensible image processing suite for electron microscopy. J. Struct. Biol. 157, 38–46 (2007).
43. Mindell, J. A. & Grigorieff, N. Accurate determination of local defocus and specimen tilt in electron microscopy. *J. Struct. Biol.* **142**, 334–347 (2003).

44. Scheres, S. H. RELION: implementation of a Bayesian approach to cryo-EM structure determination. *J. Struct. Biol.* **180**, 519–530 (2012).

45. Henderson, R. et al. Outcome of the first electron microscopy validation task force meeting. *Structure* **20**, 205–214 (2012).

46. Scheres, S. H. & Chen, S. Prevention of overfitting in cryo-EM structure determination. *Nature Methods* **9**, 853–854 (2012).

47. Russo, C. J. & Passmore, L. A. Robust evaluation of 3D electron cryomicroscopy data using tilt-pairs. *J. Struct. Biol.* **187**, 112–118 (2014).

48. Neidhardt, F. C., Bloch, P. L. & Smith, D. F. Culture medium for enterobacteria. *J. Bacteriol.* **119**, 736–747 (1974).
Extended Data Figure 1 | The conversion of a phosphonate to 5-phosphoribosyl-α-1-diphosphate (PRPP). PhnI supported by PhnG, PhnH and PhnL catalyses the transfer of the phosphonate moiety to the 1’ position of the ribose of ATP through displacement of adenine, generating a 5-triphosphoribosyl-α-1-phosphonate. Subsequent to the removal of pyrophosphate by PhnM yielding a 5-phosphoribosyl-α-1-phosphonate, PhnJ breaks the C–P bond of the ribose-coupled phosphonate liberating the alkyl moiety and generating 5-phosphoribosyl 1,2-cyclic phosphate. Finally, the combined activities of PhnP (a phosphoribosyl cyclic phosphodiesterase) and PhnN (a ribosyl bisphosphate phosphokinase) result in the formation of PRPP via ribose 1,5-bisphosphate.
Extended Data Figure 2 | Representative examples of electron density. 

a, The interface between PhnJ (blue, residues 45–52 and 104–110 including a bound sulfate ion) and PhnI (green, residues 321–341) showing $2F_o - F_c$ electron density contoured at 2.0 r.m.s.d. 
b, Close-up of the aromatic side chains in the central $\beta$-sheet of PhnJ (residues 118–126, 203–207 and 211–217), with the same contouring as a.
Extended Data Figure 3 | Sequences of the proteins of the C–P lyase core complex with secondary structure. Protein sequences are shown along with secondary structure assignment based on the crystal structure and colours as in Fig. 1. a, PhnG. The first α-helix is two residues longer in one of the two copies in the complex and indicated with a dashed box. The β-barrel domain is shown in yellow and orange colours. b, PhnH. The numbering of β-strands follows the convention from ref. 19. Helix names (A–E) used in that paper are shown in parentheses. β1 and β3 are not included in the figures in this paper as they only have two hydrogen bonds each. c, PhnI. The N-terminal domain is shown with sea green and the β-barrel domain with green and light-green colours. d, PhnJ. The central insertion domain is shown in purple and the C-terminal mini domain in a darker blue colour. Figure produced using SecSeq (D. E. Brodersen, unpublished, http://www.bioxray.au.dk/∼deb/secseq).
Extended Data Figure 4 | Cross alignments. **a**, Alignment of the amino acid sequences of *E. coli* PhnH and PhnJ. Identical residues are shown in red and conserved functionality in green. Secondary structure colours correspond to Fig. 1 and conserved regions are shown in dashed boxes. For PhnJ, the positions of the CID and CMD are indicated with brackets as well as with colours. **b**, Alignment of PhnG and PhnI (only partial sequence). Conserved regions are shown in dashed boxes. Figure produced using SecSeq (D. E. Brodersen, unpublished, http://www.bioxray.au.dk/–deb/secseq).
Extended Data Figure 5 | The structure and function of PhnJ. a, Two perpendicular views of the PhnH–PhnJ heterodimer as observed within the C–P lyase core complex (blue and red, left) and aligned views of the PhnH₂ homodimer from the isolated crystal structure (PDB ID: 2FSU; green, right)19. 
b, Surface view of the C–P lyase core complex with PhnG shown in yellow, PhnI in green, and PhnJ in blue. The position of SAM as modelled from an S-adenosyl methionine activase enzyme (PDB ID: 4K37)24 has been overlaid to visualize its putative placement in the pocket between PhnG and PhnJ. c, The C–P lyase core complex shown in surface representation with PhnJ and the CMD in cartoon, coloured by B factor to show flexibility (B = 25 Å², blue, B = 45 Å², magenta). The zinc and sulfate ions are shown with spheres.
Extended Data Figure 6 | Interaction areas within the C–P lyase core complex. a, Dimerization interface between halves of the (PhnGHIJ)₂ complex. Colours as in Fig. 1. b, Interaction areas between the individual subunits within each dimer half. c, Two perpendicular views of the C–P lyase core complex shown in surface representation with colours as in Fig. 1c. d, Left, overview of the surface conservation of the C–P lyase core complex shown as a colour gradient from teal (variable) to burgundy (conserved) as indicated. Right, conservation at the interaction interfaces between the individual subunits of the complex.
Extended Data Figure 7 | In vivo complementation of E. coli *AphnHIJKLMNOP*. *E. coli* strain BW16711 (*AphnHIJKLMNOP*) complemented with a plasmid-borne copy of either wild-type (Wt) *phnGHIJKLMNOP* or variants thereof, including PhnJ(C272A), PhnJ(H108A), PhnI(H328A), or the PhnI(H328A;H333A) double variant. Growth is monitored on minimal plates with either no phosphorus source (a), 2-aminoethyl phosphonate (2-AEPn) (b), methyl phosphonate (MPn) (c), or phosphate ion (d). The data shown are representative of three independent experiments.
Extended Data Figure 8 | PhnK sequence alignment. Alignment of the protein sequence of *E. coli* PhnK with homologous proteins from a wide range of microorganisms. Conserved residues are shown on a teal background, and residues mentioned in the main text and the location of the ABC cassette consensus motifs (Walker A, Q motif, ABC motif, Walker B, D loop, and H motif) are indicated.
Extended Data Figure 9 | Negative stain electron microscopy of PhnGHJK.

a. Raw micrograph representative of 100 images collected. Scale bar is 500 Å.
b. Selection of 2D reference-free class averages from a total of 300 classes showing the particle in various orientations. Each class is 172 Å wide.
c. Fourier-shell correlation (FSC) of the final electron microscopy density map as a function of resolution (black line) with $FSC = 0.143$ at 16 Å. The red line shows the correlation between the crystal structure and the electron microscopy density, which has $FSC = 0.5$ at 28 Å. The inset figure shows an equal area projection plot for the electron microscopy tilt pair validation data set. The circle is an approximation of a 99% confidence interval that contains the representative direction (blue plus), includes the true tilt direction (red cross), and excludes the untitled direction (origin).
Extended Data Figure 10 | Purification of the PhnGHIJ(ΔCID) and PhnGHIJ(ΔCID)K complexes. a, Sequence alignment of the CID region of PhnJ with the corresponding part of the structural domain of PhnH, where strands 5 and 6 of the β-sheet are connected by a short Gly–Gly turn (green residues). In PhnJ(ΔCID), the CID domain spanning residues 130–171 (red residues) are replaced by a similar dipeptide turn, which should maintain the overall domain fold. b, SDS–PAGE gel showing purified C–P lyase complexes (PhnGHIJ and PhnGHIJK) both with (Wt) and without (ΔCID) the CID on PhnJ. PhnK is missing from the complex purified from the PhnGHIJ(ΔCID)K construct (red arrow). The data shown are representative of three independent purifications. c, Overview of the C–P lyase core complex structure docked in the PhnGHIJK electron microscopy density with the location of the PhnJ CID domains (purple) and PhnK (cyan cartoon) indicated.
### Extended Data Table 1 | Data collection, phasing and refinement statistics

|                          | Native          | Ta₆Br₁₂ derivative |
|--------------------------|-----------------|-------------------|
| **Data collection**      |                 |                   |
| Space group              | P2₁2₁2₁         | P2₁2₁2₁           |
| Cell dimensions          |                 |                   |
| a, b, c (Å)              | 95.5, 133.7, 176.7 | 95.8, 143.1, 178.7 |
| α, β, γ (°)              | 90.0, 90.0, 90.0 | 90.0, 90.0, 90.0  |
| Resolution               | 58.91 - 1.70 Å (1.74- 1.70 Å)* | 48.24 - 3.50 (3.60 - 3.50 Å) |
| Wavelength               | 1.00004 Å       | 1.2552 Å          |
| Unique reflections       | 246,950 (17,488)| 59,846 (4,847)    |
| R-meas (%)               | 7.0 (114.7)     | 15.9 (48.0)       |
| CC₁/₂                    | 99.9 (51.6)     | 99.7 (96.4)       |
| I/σI                     | 17.7 (1.7)      | 17.0 (7.7)        |
| Completeness (%)         | 99.7 (99.1)     | 100.0 (100.0)     |
| Redundancy               | 5.6 (5.5)       | 26.6 (26.3)       |
| **Refinement**           |                 |                   |
| Resolution (Å)           | 58.4 - 1.7      |                   |
| No. of reflections       | 246,797         |                   |
| R<sub>work</sub> / R<sub>free</sub> (%) | 14.9 (17.6)   |                   |
| No. of atoms             |                 |                   |
| Protein (non-hydrogen)   | 30,097 (15,203) |                   |
| SO₄²⁻ / Zn²⁺             | 24              |                   |
| Water                    | 1,792           |                   |
| B-factors (Å²)           |                 |                   |
| Protein                  | 29.87           |                   |
| SO₄²⁻                    | 40.28           |                   |
| Zn²⁺                     | 25.97           |                   |
| Water                    | 21.15           |                   |
| R.m.s deviations         |                 |                   |
| Bond lengths (Å)         | 0.01            |                   |
| Bond angles (°)          | 1.17            |                   |
| **Ramachandran statistics** |               |                   |
| Favoured (%)             | 97.5            |                   |
| Allowed (%)              | 2.0             |                   |
| Outliers (%)             | 0.5             |                   |

* Highest resolution shell is shown in parentheses, except where otherwise stated.
† Statistics from MolProbity via Phenix**.