Structural insights into the bacterial carbon-phosphorus lyase machinery

Paulina Seweryn[1], **Lan Bich Van**[1], **Morten Kjeldgaard**[1], **Christopher J. Russo**[2], **Lori A. Passmore**[2], **Bjarne Hove-Jensen**[1], **Bjarne Jochimsen**[1], and **Ditlev E. Brodersen**[1,3]

[1]Department of Molecular Biology and Genetics, Gustav Wieds Vej 10c, DK-8000 Aarhus C, Denmark

[2]Medical Research Council Laboratory of Molecular Biology, Francis Crick Avenue, Cambridge, CB2 0QH, United Kingdom

Summary

Phosphorous 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 organic phosphonate compounds, which require specialised enzymatic machinery for breaking the stable carbon-phosphorus (C-P) bond. Despite its importance, the details of how this machinery catabolises phosphonate remain unknown. Here we determine the crystal structure of the 240 kDa *Escherichia coli* C-P lyase core complex (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 likely couple organic 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 PhnJ via a conserved insertion domain. Our results provide a structural basis for understanding microbial phosphonate breakdown.

Phosphonate compounds that contain a stable carbon-phosphorus (C-P) bond are utilised as a source of phosphate by microorganisms in many natural environments where the low levels of free and organic phosphate limit growth[1]. The C-P lyase pathway, which converts phosphonate into 5-phosphoribosyl-α-1-diphosphate (PRPP) in an ATP-dependent fashion, is activated upon phosphate starvation in many bacterial species including *Escherichia coli*[2,3]. The enzymes of this pathway have a very broad substrate specificity enabling the

bacteria to utilise a wide range of compounds for growth including alkyl, amino-alkyl and aryl phosphonates[4-7].

In *E. coli*, the 14-cistron *phn* operon is required for phosphonate uptake and utilisation and encodes an ATP-binding cassette transporter (PhnC, PhnD, and PhnE), a regulatory protein (PhnF) and components required for enzymatic conversion of phosphonate into PRPP (PhnGHIJKLMNOP)[8-13]. PhnGHIJ have been shown to form a stable protein complex, which we term the *C-P lyase core complex*, which likely contains PhnG and PhnI at its centre[14,15]. The core complex stably associates with a fifth protein, PhnK, which resembles ABC cassette proteins, with unknown stoichiometry[14]. PhnJ contains an iron-sulphur cluster required for C-P bond cleavage via an S-adenosyl methionine (SAM)-dependent radical mechanism[16-18], while PhnI is a nucleosidase capable of deglycosylating ATP and GTP to ribose 5-triphosphate[18]. A reaction mechanism for the breakdown of phosphonate via the C-P lyase pathway was proposed where PhnI, supported by PhnG, PhnH, and PhnL, a protein not present in the core complex, catalyses the transfer of the phosphonate moiety to the ribose C-1′ position of ATP by displacing adenine, generating a ribose 5′-triphosphate alkyl phosphonate intermediate (Figure 1a and Extended Data Figure 1). Following pyrophosphate release by PhnM, PhnJ cleaves the C-P bond and PhnP/PhnI convert the resulting ribose cyclic phosphate into PRPP[17,18]. The C-P lyase core complex thus harbours two key activities of this pathway; coupling of the phosphonate to ATP (PhnG, PhnH, and PhnI) and C-P bond cleavage (PhnJ)[18]. PhnH is the only component of the C-P lyase core complex that has been structurally characterised and displays a fold related to the pyridoxal 5′-phosphate-dependent transferases. It forms a homodimer when expressed independently[19] and its role within the complex is unclear.

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 (MR-SAD) using a Ta_6Br_{12} cluster derivative and the PhnH crystal structure[19] 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 Figure 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 (Figure 1b and Extended Data Figure 3) and is complete except for a few residues located at the subunit termini. The structure includes four sulphate ions, four zinc ions, and 1,792 solvent molecules. Together, the eight polypeptides 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 (Figure 1c and 1d), consistent with its behaviour in solution[14].

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 between PhnJ and PhnH (Figure 1c). At the centre of the molecule, a compact PhnI homodimer forms a disc-like structure that serves as a central hub for attachment of the

other subunits (Figure 1c and 1d, green). The core domain of PhnI, which is the largest single domain in the structure, has a novel α+β fold comprised of a four-stranded, antiparallel β-sheet next to a four-helix bundle combining in a unique fold: the Beta-Barrel Domain (BBD, Figure 1c and Figure 2a). At both termini there are helical extensions of approximately 35 residues that grasp PhnJ and tether it to the complex via extensive interactions (Figure 3a). In turn, PhnJ attaches PhnH to the complex through packing of conserved α helices in both proteins (Figure 1c).

PhnJ has a compact α+β fold surrounded by two mini-domains, the Central Insertion Domain (CID) and C-terminal Mini Domain (CMD) (Figure 1b, 2b, and Extended Data Figure 3). Surprisingly, the core folds of PhnJ and PhnH are nearly identical (Cα rmsd 2.5 Å), despite very little sequence similarity (Extended Data Figure 4a). Moreover, the interactions in the PhnHJ heterodimer closely resemble those observed in the crystal structure of the isolated PhnH homodimer (Extended Data Figure 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 (Figure 2b). The CID is well conserved among PhnJ orthologues and contacts both of the central PhnI molecules (Figure 3b). Finally, the CMD is located at the C-terminus and consists of a small β-hairpin and a helix. It is stabilised by a zinc ion coordinated by four conserved cysteine residues, Cys241, Cys244, Cys266, and Cys272 (Figure 2b).

The PhnI monomers bind each other via an extensive, conserved surface interaction area comprising ~75% of the total PhnGHIJ dimerisation interface (Extended Data Figure 6). Each molecule of PhnI interacts with both copies of PhnG (Figure 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 (BBD, Figure 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 (Figure 2 and Extended Data Figure 4b). The PhnG β-hairpin and C-terminal helix form a molecular clamp that connect to a groove in PhnI forming an unusually long, combined β-barrel domain (80 Å, Figure 3d).

The iron-sulphur binding site

PhnJ belongs to the anaerobic radical SAM enzyme superfamily in which three conserved cysteine residues coordinate a cubane-like Fe_4S_4 cluster[20] that promotes formation of a free electron radical required for catalysis by reductive cleavage of SAM to a 5′-deoxyadenosyl radical (Ado-CH_2·) and L-methionine[21-23]. PhnJ does not contain the canonical CX_3CX_2C motif but rather a $CX_2CX_{21}C$ motif involving Cys241, Cys244, and Cys266, which are both necessary and sufficient for reconstitution of the iron-sulphur cluster *in vitro*[17,18]. In the structure, these cysteines are juxtaposed and coordinate a zinc ion (Figure 4a) in an arrangement that closely matches that expected for an Fe_4S_4 cluster-containing protein (Figure 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[24] (Extended Data Figure 5b).

According to the proposed reaction mechanism, the Ado-CH$_2\cdot$ radical is transferred to the universally conserved Gly32 of PhnJ generating a stable glycyl radical enzyme that supports multiple turnovers without further SAM consumption[17]. According to this scheme, transfer of the radical from Gly32 to the fourth conserved cysteine residue (Cys272) generates a thiyl radical capable of homolytic C-P bond cleavage of 5-phosphoribosyl 1-phosphonate (Extended Data Figure 1) through a thiophosphonate radical intermediate[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 (Figure 4c). A direct involvement of Gly32 in the reaction is therefore difficult to reconcile with the structure[20], however, it cannot be excluded that structural rearrangements could alter the position of the cluster relative to Gly32 to bring them into proximity. The CMD containing the cluster site has higher *B*-factors suggesting it is relatively loosely attached to the PhnJ core and perhaps could detach during the reaction (Figure 2b, Extended Data Figure 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 (Figure 4d). Two of the residues (PhnI His328 and His333) coordinate the zinc ion directly (2.4 Å), while the third (PhnJ 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 (Figure 4e). The cavity also contains a sulphate 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.

Structural 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 like observed in this case[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 utilise phosphonate (Extended Data Figure 7). A plasmid-borne copy of the wild-type *phnGHIJKLMNOP* allele was used to complement *E. coli* *phnHIJKLMNOP* under conditions where phosphonate was the sole phosphate source (Extended Data Figure 7)[26]. The wild type allele but none of the variants H333A (PhnI), H328A/H333A (PhnI), H108A (PhnJ), or C272A (PhnJ) were able to utilise phosphonate and 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)[14]. The function of PhnK is unclear but it contains the consensus elements of an ATP-binding cassette protein suggesting it might deliver nucleotide for the reaction (Extended Data Figure 8)[14]. Despite its ability to stably co-purify, we were unable to obtain crystals of a complex including PhnK. We therefore mapped the PhnK binding site on the complex using negative-stain electron microscopy by generating a 3D reconstruction of purified PhnGHIJK

(Figure 5a and Extended Data Figure 9). The crystal structure fits tightly within the resulting EM 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 (Figure 5a and 5b). The fold of PhnK can be roughly modelled using a homologous nucleotide-binding domain of an ABC transporter (PDB 4FWI)[27]. The EM map is consistent with a single PhnK binding unilaterally to the complex, breaking the two-fold symmetry (Figure 5c). Although the exact orientation of PhnK cannot be established at this resolution, we note that one side is highly conserved among orthologues suggesting it comprises the interaction surface (Figure 5c).

ABC modules often dimerise in a head-to-tail fashion, binding ATP between the Walker A/B motifs of one subunit and the ABC motif of the other[28]. PhnK contains a variant ABC motif (FSGGMQ versus LSGGQ), which could serve to bind the C-P lyase core particle (Extended Data Figure 8). The conserved CID domain protrudes into the PhnK-binding region, so to probe its importance we constructed C-P lyase complexes lacking residues 130-171 of PhnJ (Extended Data Figure 10). Purification PhnGHIJ ^{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 organisation 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, heterooctameric core 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[20].

The structure indicates the existence of a second active site at the interface of PhnI and PhnJ. Analysis of difference electron density from several independent data sets revealed consistent density next to the bound zinc ion, but we have been unable to identify the bound molecule. We also carried out co-crystallisation using a range of compounds including nucleotides and phosphonates, but no further substrate binding was observed. This suggests that the complex needs an Fe_4S_4 cluster or is not in the correct conformational state to bind substrate. We speculate that the His site is required for coupling phosphonate to ATP, which is known to depend on PhnI[18].

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 should focus on understanding the

requirements of the two reactions catalysed by the complex and definitively locating binding sites of substrates and reaction intermediates of these.

Methods

Protein expression and protein purification

The construction of **pHO572** (expressing phn*GHIJ*) and **pHO575** (expressing phn*GHIJK*) as well as gene expression in *E. coli* strain HO2735 (*(lac)X74*

(phnCDEFGHIJKLMNOP)33–30/F *lacI^q zzf*::Tn10) were described previously[14]. **pHO575** encodes a C-terminally six-histidine tagged version of PhnK while PhnGHIJ has no tag but still binds to Ni^{2+} NTA agarose beads. The PhnGHIJ ^{CID} and PhnGHIJ ^{CID}K constructs were created by site-directed mutagenesis using primers 5′-GTGCCAATCCCCGAG<u>GGCGGC</u>TATCCGGTGAAGGTA-3′ (delta-CID forward), and 5′-TACCTTCACCGGATA<u>GCCGCC</u>CTCGGGGATTGGCAC-3′, which result in replacement of residues 130-172 of PhnJ by two glycine residues (underlined in the primers). Cells were in all cases grown at 37°C in LB medium and gene expression achieved overnight at 18°C by induction using 0.5 mM isopropyl-β-D-thiogalactoside (IPTG). Cell pellets were resuspended in lysis buffer (50 mM Hepes, pH 7.5, 500 mM NaCl, 5 mM $MgCl_2$, 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 lysed cells were centrifuged at 16,000 rpm for 45 min to remove cell debris and bound to Ni^{2+} NTA agarose beads on a 5 ml pre-packed HisTrap HP column (GE Healthcare), pre-equilibrated with lysis buffer (PhnGHIJ) or lysis buffer plus 20 mM imidazole (PhnGHIJK). In all cases, the complexes were eluted by increasing the imidazole concentration to 250 mM. Following overnight dialysis at 4°C against buffer LS1 (50 mM Hepes, pH 7.5, 100 mM NaCl, 5 mM $MgCl_2$, and 5 mM 2-mercaptoethanol), the samples were applied to a 1 ml Source 15Q column (GE Healthcare), pre-equilibrated with buffer LS1 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 LS2 (50 mM Hepes, pH 7.5, 250 mM NaCl, 5 mM $MgCl_2$, and 5 mM 2-mercaptoethanol), washed, and eluted using a 250-400 mM NaCl gradient. Finally, the complexes were purified on a Superdex 200 10/300 GL size-exclusion column (GE Healthcare) equilibrated with buffer GF (50 mM Hepes pH 7.5, 300 mM NaCl, and 5 mM 2-mercaptoethanol). A 2 L culture typically yielded 4–8 mg of a purified protein complex. Purification from the PhnGHIJ ^{CID} and PhnGHIJ ^{CID}K constructs was stopped after the source Q column.

Crystallisation and structure determination of PhnGHIJ

Crystals of the PhnGHIJ complex were obtained using batch crystallisation at 4 °C with a reservoir solution containing 20% (w/v) PEG 10,000, 0.1 M Hepes, pH 7.5, 1 mM trisodium citrate dihydrate, 2.3% (w/v) 1,8-diaminooctane, and 5 mM 2-mercaptoethanol. Crystallisation drops contained 1 μl protein sample mixed with 0.7 μl reservoir solution and 0.3 μl of a micro-seed stock[30] obtained from early stage hits. Crystals suitable for data collection appeared within 2–3 days reaching maximum dimensions of $1.0 \times 0.2 \times 0.2$ mm. Crystals were collected, cryo-protected by gradual addition of glycerol to a final

concentration of 25% (v/v), and flash-frozen in a liquid nitrogen. For structure solution by MR-SAD, crystals were derivatised for 24 h with Ta_6Br_{12} (Proteros Biostructures)[31], which was added directly to the crystallisation drop as powder. Diffraction data were collected at 100K at the X06DA beamline at SLS, Villigen, Switzerland on a PILATUS 2M detector (native crystals) or a PILATUS 6M detector at the X06SA beamline (Ta_6Br_{12} data). Datasets were processed with xia2[32] and the structure was solved by MR-SAD method using Phaser and AutoSol via PHENIX[33]. The PhnH structure[19] (PDB 2FSU) was used as an MR search model in Phaser to locate the Ta_6Br_{12} sites by MR-SAD, whereby the partial MR solution allowed identification of the Ta_6Br_{12} cluster sites and served as a source of phase information in phenix.autosol. Initial phases were obtained after density modification using RESOLVE in the phenix.autosol pipeline and the resulting maps used to autobuild secondary structure elements using ARP/wARP from the CCP4 package[34]. The resulting partial model was then used as a search model for MR against the native data using Phaser and run through 25 cycles of backbone auto-tracing using the native dataset in SHELXE[35]. Finally, a near complete model of PhnGHIJ could be autobuilt using phenix.autobuild, RESOLVE, and Buccaneer[36]. Missing parts of the model were completed manually in COOT[37]. The model was iteratively improved and refined using phenix.refine[33] and validated using MolProbity[38]. The final model contains 1,958 amino acid residues, four bound sulphate ions, four zinc ions, and 1,792 molecules of ordered solvent. The model of complete except for 18 residues exclusively located at the termini of the subunits. Structure figures were made using PyMOL[39], Chimera[40], and ConSurf[29].

Electron microscopy

Purified samples of PhnGHIJK samples were applied to Quantifoil R2/2 holey carbon on copper grids (Quantifoil, Jena)[41], covered with an additional thin 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 44,000x magnification on a Tecnai T12 microscope equipped with a US4000 4K $\times$ 4K pixel CCD detector (Gatan) at 120 kV with defoci in the 0.8-2 μm range and using an electron dose of 20 electrons/$Å^2$. 10,137 single particles were manually picked from 105 micrographs using e2boxer from the EMAN2 software package[42] and contrast transfer function parameters were determined using CTFFIND[43]. Three iterations of 2D classification were performed in RELION[44] 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 16 Å by the 0.143 "gold standard" Fourier-shell correlation[45,46] and a resolution of 28 Å vs. the crystal structure at FSC = 0.5. The latter is likely 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° (p-value 0.01, κ=2.7)[47]. The FSC versus the crystal structure shows correlation between the crystal structure and the EM density at low resolution, after which

deviations due to structural differences between the C-P lyase core complex and the PhnGHIJK complex become apparent.

In vivo complementation

For the *in vivo* complementation studies, *E. coli* strain BW16711 (*phnHIJKLMNOP*) was transformed by the plasmid **pGY1**[26], conferring ampicillin resistance and encoding *phnGHIJKLMNOP* and analysed for its ability to grow on MOPS minimal plates[48] supplemented with 0.2% glucose, 100 μg/mL ampicillin, 0.1 mM IPTG and 0.2 mM of either methyl phosphonate, 2-aminoethyl phosphonate, or phosphate as positive control. The PhnI H328A variant, the H328A+H333A double variant, the PhnJ C272A variant as well as H108A variant, were introduced into **pGY1** by site-directed mutagenesis by PCR using the following primers, 5′-GCAGGCTTTGTCTC<u>GGC</u>CCTCAAACTCCCCCA-3′ (H328A forward), 5′-TGGGGGAGTTTGAGG<u>GCC</u>GAGACAAAGCCTGC-3′ (H328A reverse), 5′-GCAGGCTTTGTCTC<u>GGC</u>CCTCAAACTCCCC<u>GC</u>CTACGTCGATTTCCA-3′ (H328A +H333A forward), 5′-TGGAAATCGACGTAG<u>GC</u>GGGGAGTTTGAGG<u>GCC</u>GAGACAAAGCCTGC-3′ (H328A+H333A reverse), 5′-TCCGATACCGATTAT<u>GC</u>CCGCCAACAGAGCGA-3′ (C272A forward), 5′-TCGCTCTGTTGGCGG<u>GC</u>ATAATCGGTATCGGA-3′ (C272A reverse), 5′-<u>C</u>TTATCCAGACGCGT<u>GC</u>CCGCATCCCCGAAAC-3′ (H108A forward), and 5′-GTTTCGGGGATGCGG<u>GC</u>ACGCGTCTGGATAA<u>G</u>-3′ (H108A reverse), where changes relative to the Wt sequence are shown underlined. Template DNA was digested by the methylation-dependent endonuclease, DpnI, before transformation of 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 *phnGHIJKLMNOP* region of the **pGY1** vector to ensure that no other spontaneous mutations had been introduced that could prevent rescue of the BW16711 strain.
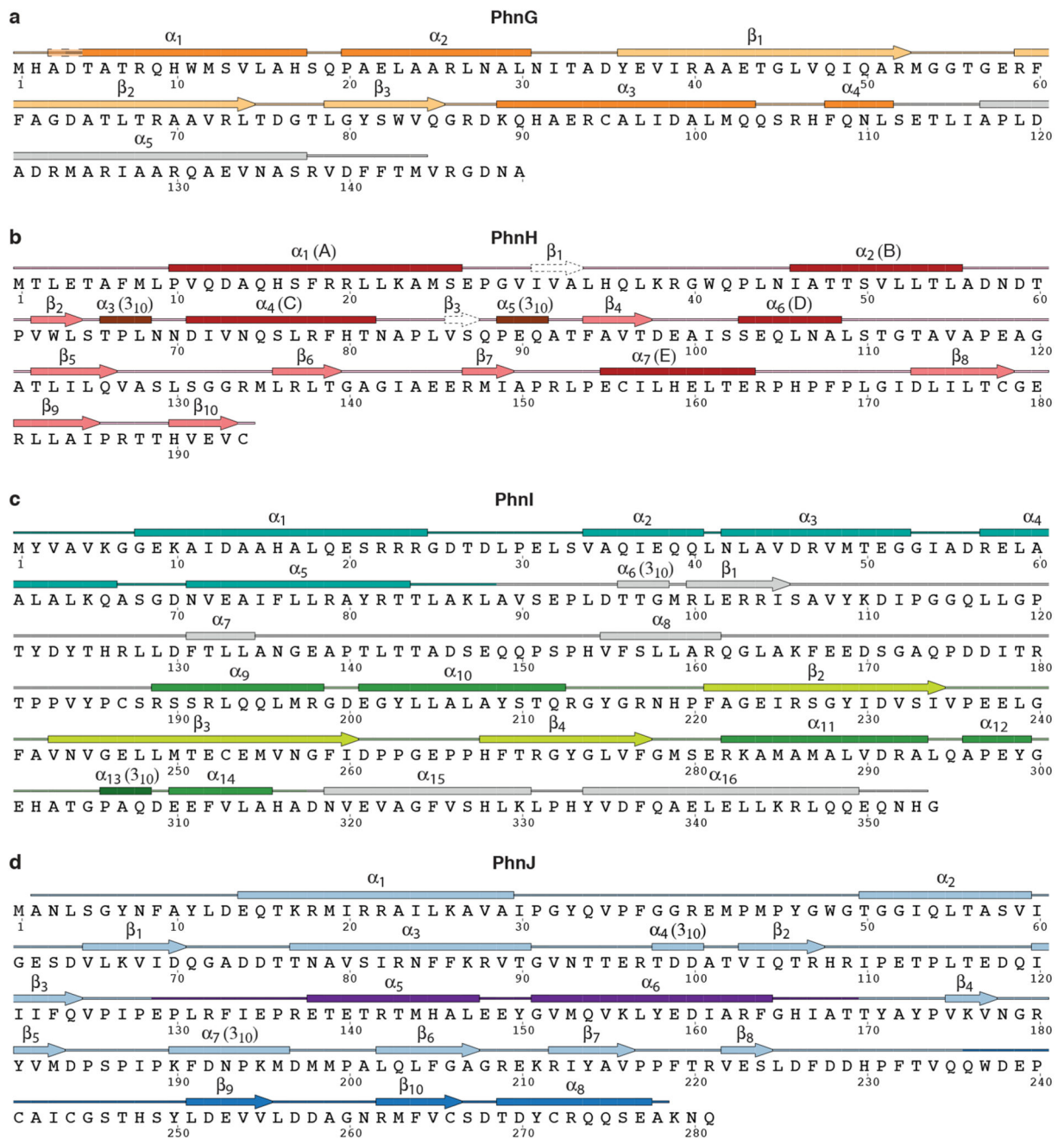
Extended Data
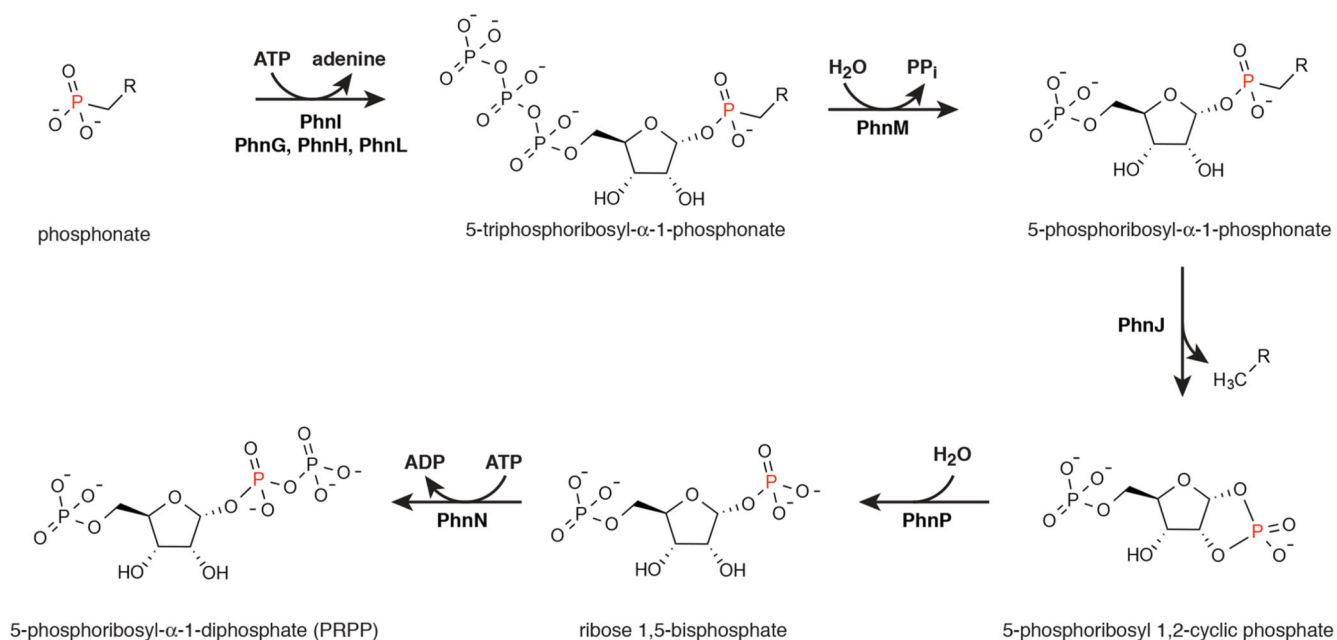
Extended Data Figure 1. The conversion of phosphonate to 5-phosphoribosyl-α-1-diphosphate (PRPP)

PhnI supported by PhnG, PhnH, and PhnL catalyses the transfer of the phosphonate moiety to the C-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 (phosphoribosyl cyclic phosphodiesterase) and PhnN (ribosyl bisphosphate phosphokinase) result in the formation of PRPP via ribose 1,5-bisphosphate.
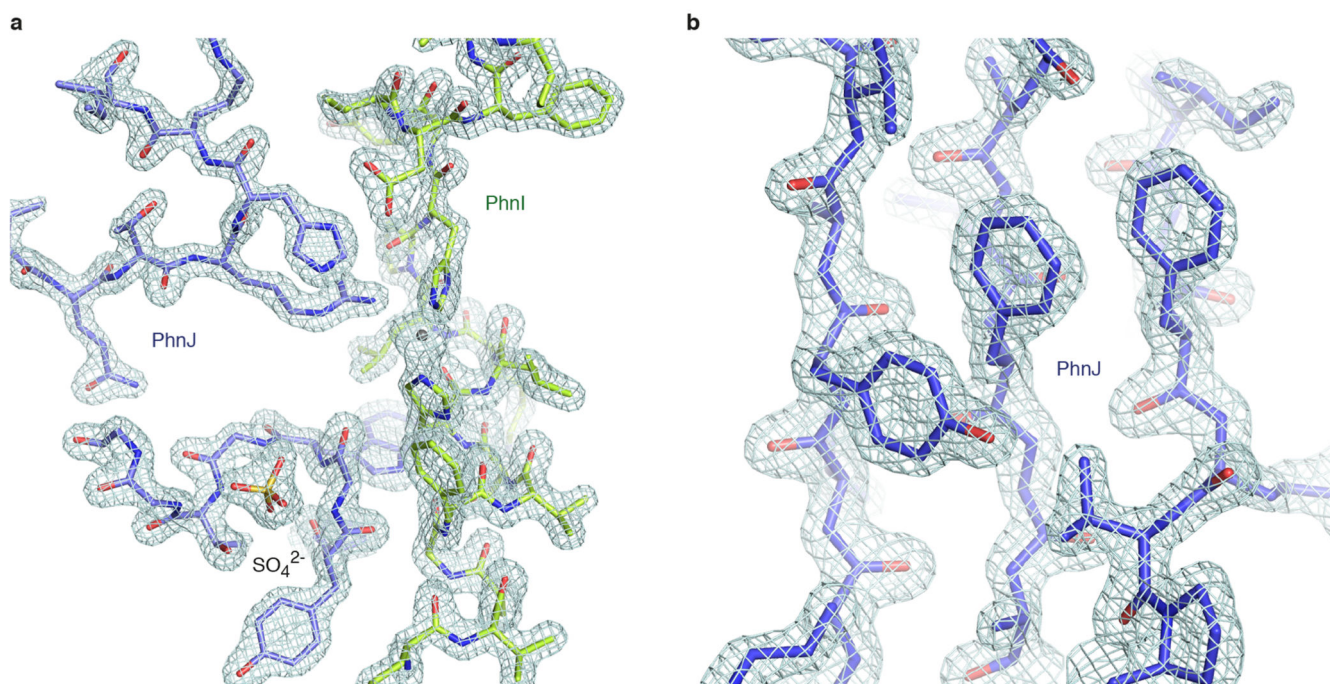
**Extended Data Figure 2. Representative examples of electron density**

a, The interface between PhnJ (blue, residues 45-52 and 104-110 including a bound sulphate ion) and PhnI (green, residues 321-341) showing $2F_o$-F_c electron density contoured at 2.0 rmsd. **b**, Close-up of the aromatic side chains in the central β-sheet of PhnJ (residues 118-126, 203-207, and 211-217), with the same contouring as above.

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 Figure 1. **a**, PhnG. The first α helix of is two residues longer in one of the two copies in the complex and indicated with a dashed box. The BBD is shown in yellow and orange colours. **b**, PhnH. The numbering of β-strands follows the convention from Adams *et al.*[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 NTD is shown with sea green and the BBD with green and light green colours.
d, PhnJ. The central insertion domain (CID) is shown in purple and the C-terminal mini
domain (CMD) in a darker blue colour. Figure produced using SecSeq (D. E. Brodersen,
unpublished, http://www.bioxray.au.dk/~deb/secseq).
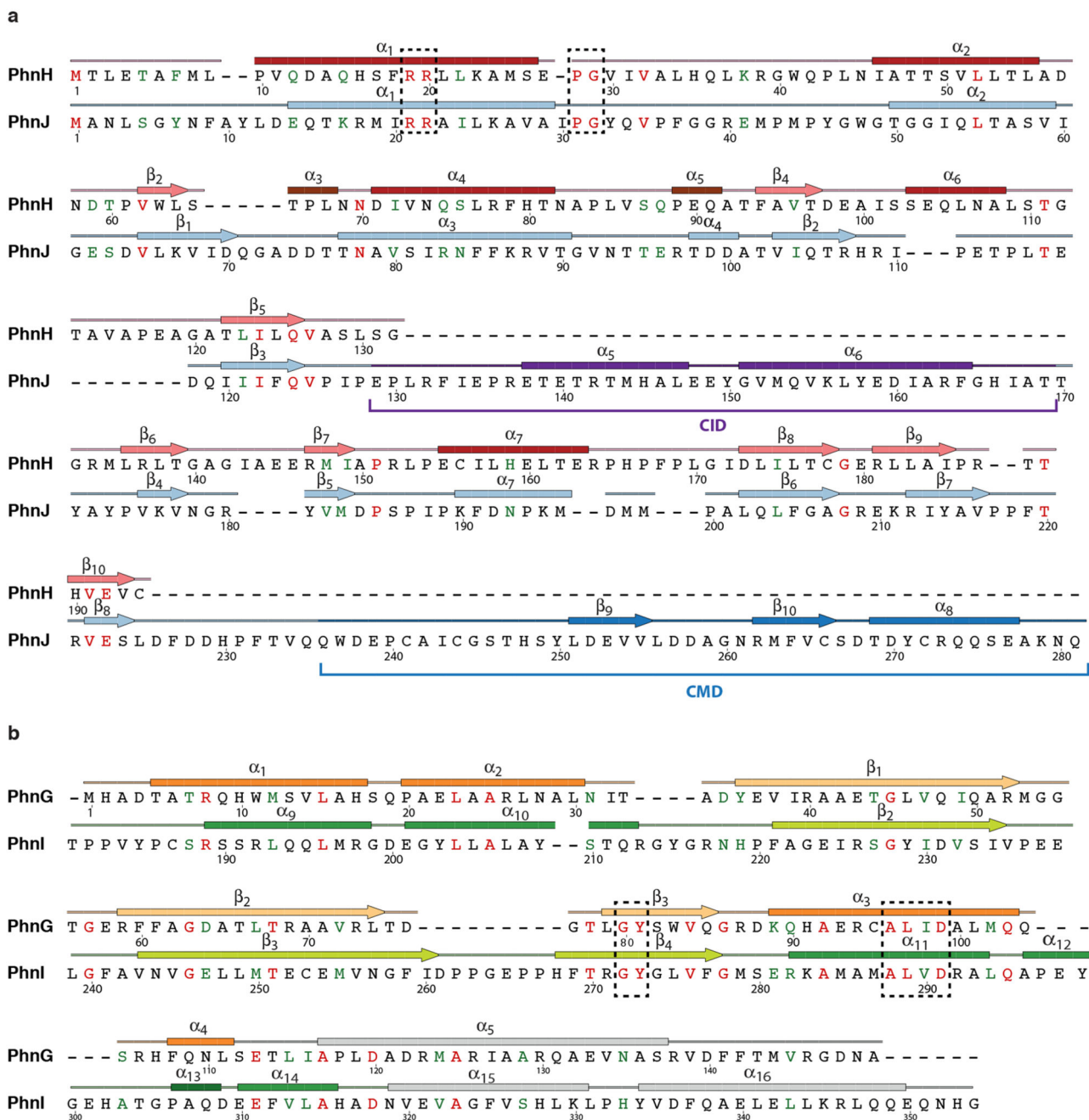

Extended Data Figure 4. Cross alignments
a, Alignment of the amino acid sequences of *E. coli* PhnH and PhnJ. Identical residues are
shown in red and conserved functionality in green. Secondary structure colours correspond

to Figure 1 and conserved regions are shown in dashed boxes. For PhnJ, the positions of the Central Insertion Domain (CID) and C-terminal Mini Domain (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).
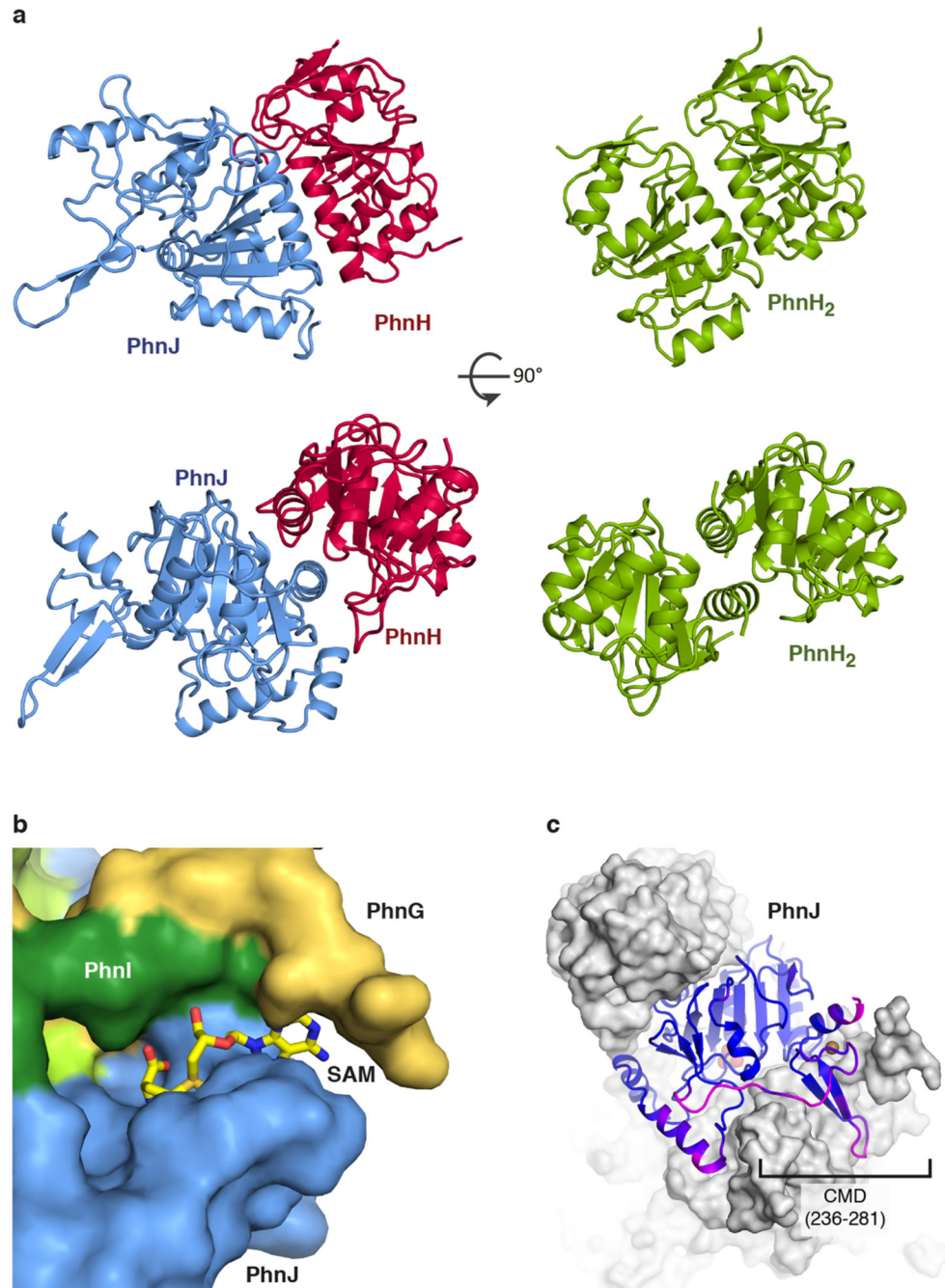

Extended Data Figure 5. The structure and function of PhnJ

a, Two perpendicular views of the PhnHJ heterodimer as observed within the C-P lyase core complex (blue and red, left) and aligned views of the $PhnH_2$ homodimer from the isolated crystal structure (PDB 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 4K37)[24] has been overlaid to visualise its putative placement in the pocket between PhnG and PhnJ. **c**, the C-P lyase core particle shown in surface representation with PhnJ and the Central Mini Domain (CMD) in cartoon coloured by B factor to show flexibility (B=25 $Å^2$, blue, B=45 $Å^2$, magenta). The zinc and sulphate ions are shown with spheres.
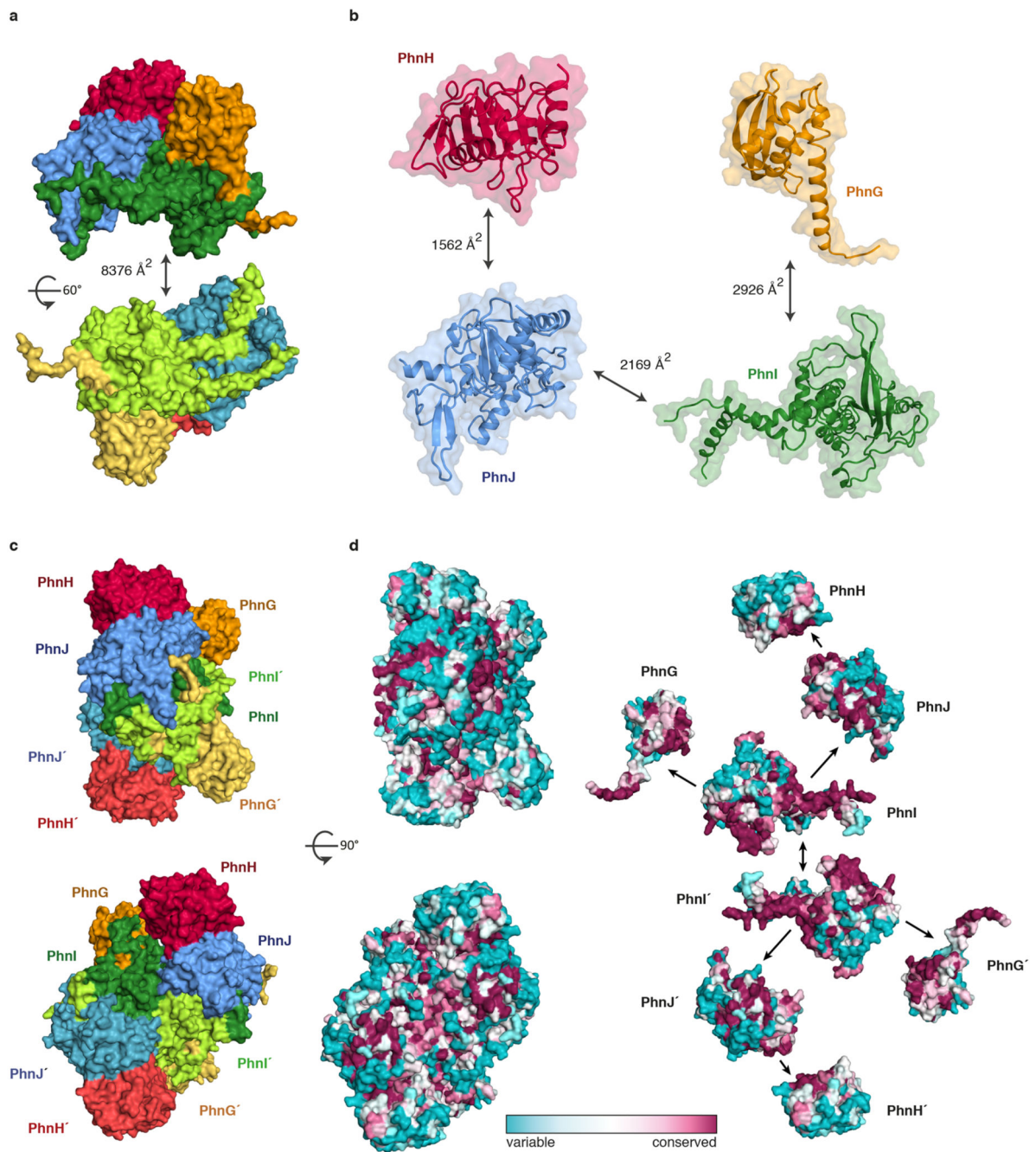
Extended Data Figure 6. Interaction areas within the C-P lyase core complex

a, Dimerisation interface between halves of the (PhnGHIJ)$_2$ complex. Colours as in Figure 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 Figure 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.
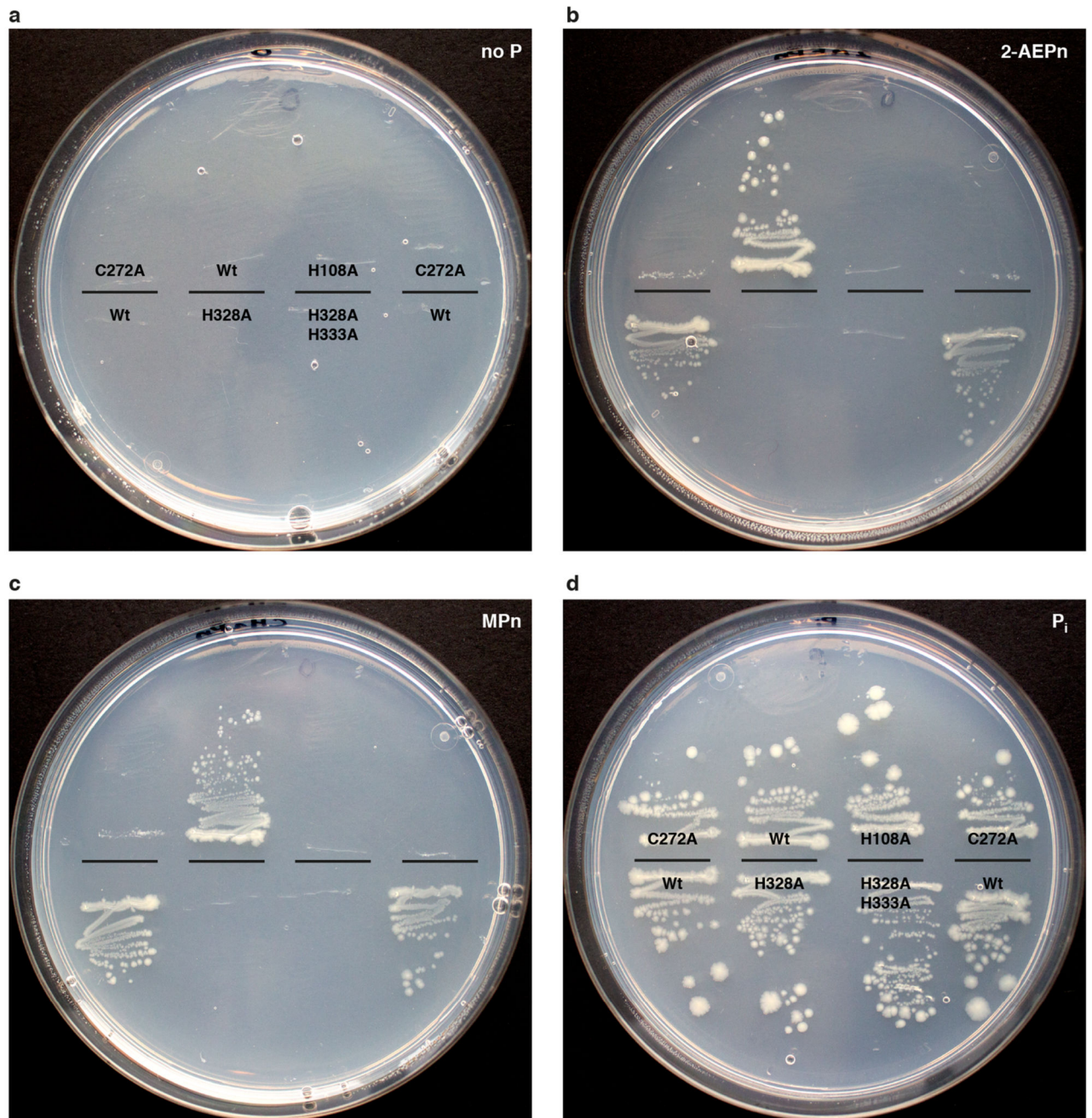
Extended Data Figure 7. *In vivo* **complementation of** *E. coli* *phnHIJKLMNOP*
E. coli strain BW16711 (*phnHIJKLMNOP*) complemented with a plasmid-borne copy of
either wild type (Wt) *phnGHIJKLMNOP* or variants thereof, including PhnJ C272A, PhnJ
H108A, PhnI H328A, or the PhnI H328A+H333A double variant. Growth is monitored on
minimal plates with either (**a**) no phosphorous source, (**b**) 2-aminoethyl phosphonate (2-
AEPn), (**c**) methyl phosphonate (MPn), or (**d**) phosphate. The data shown are representative
of three independent experiments.
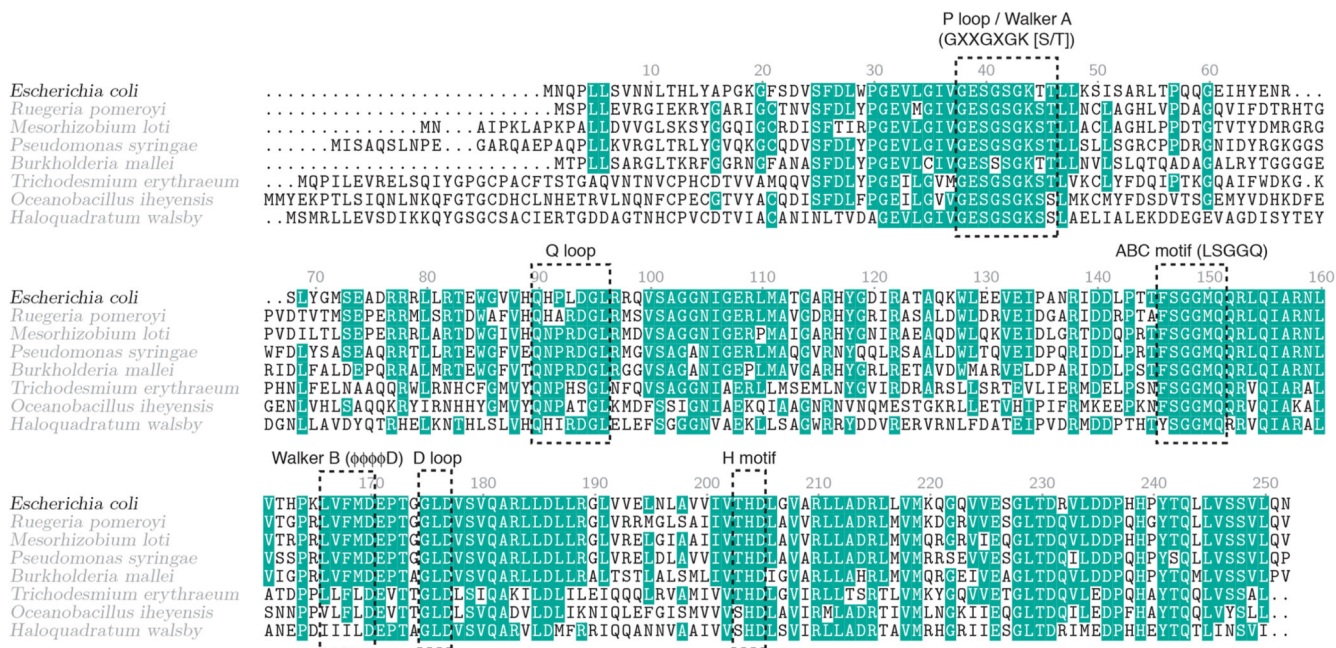
Extended Data Figure 8. PhnK sequence alignment

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.
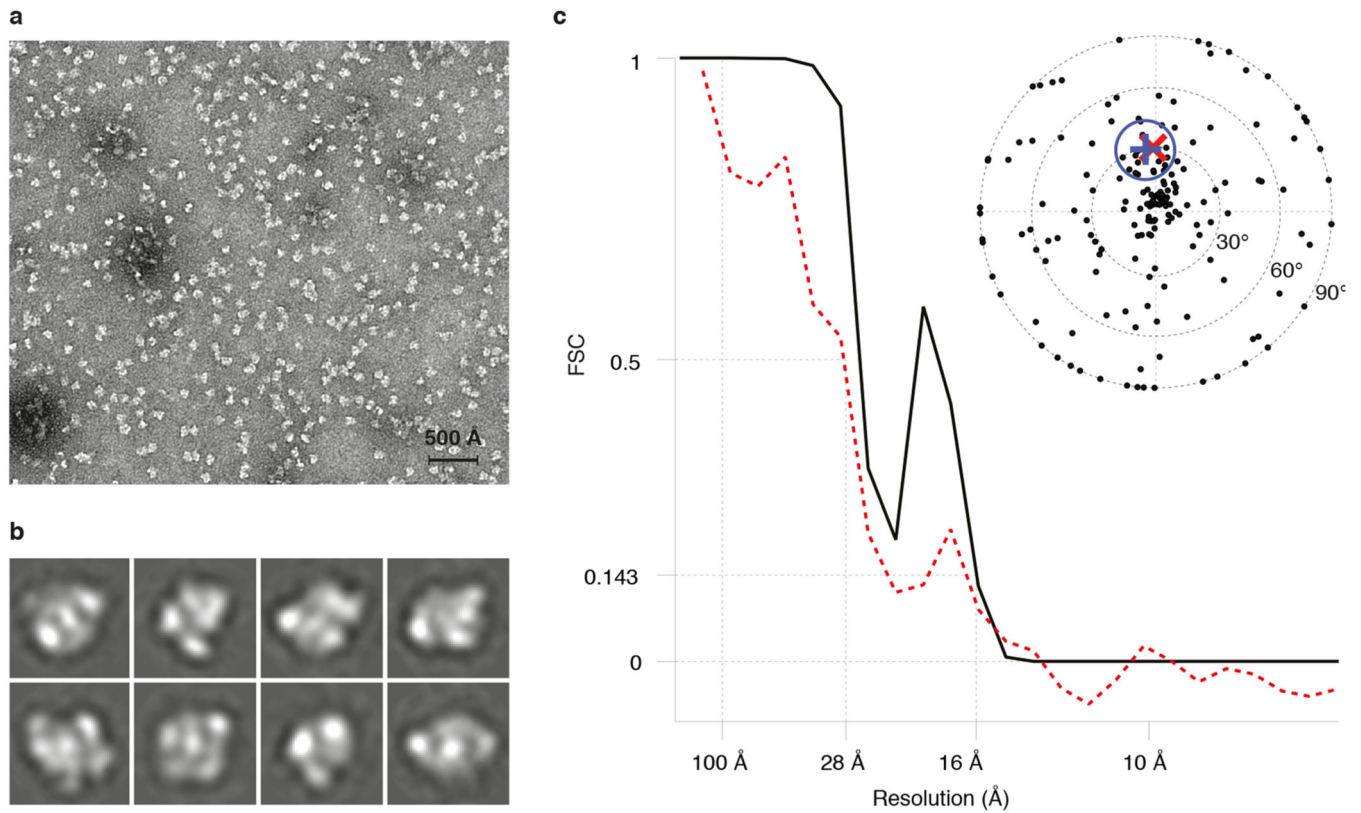
Extended Data Figure 9. Negative stain EM of PhnGHIJK

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 EM 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 EM density, which has FSC=0.5 at 28 Å. The inset figure shows an equal area projection plot for the EM tilt pair validation data set. The circle is an approximation of a 99% confidence interval that contains the representative direction (blue +), includes the true tilt direction (red cross), and excludes the untitled direction (origin).

a

b

c
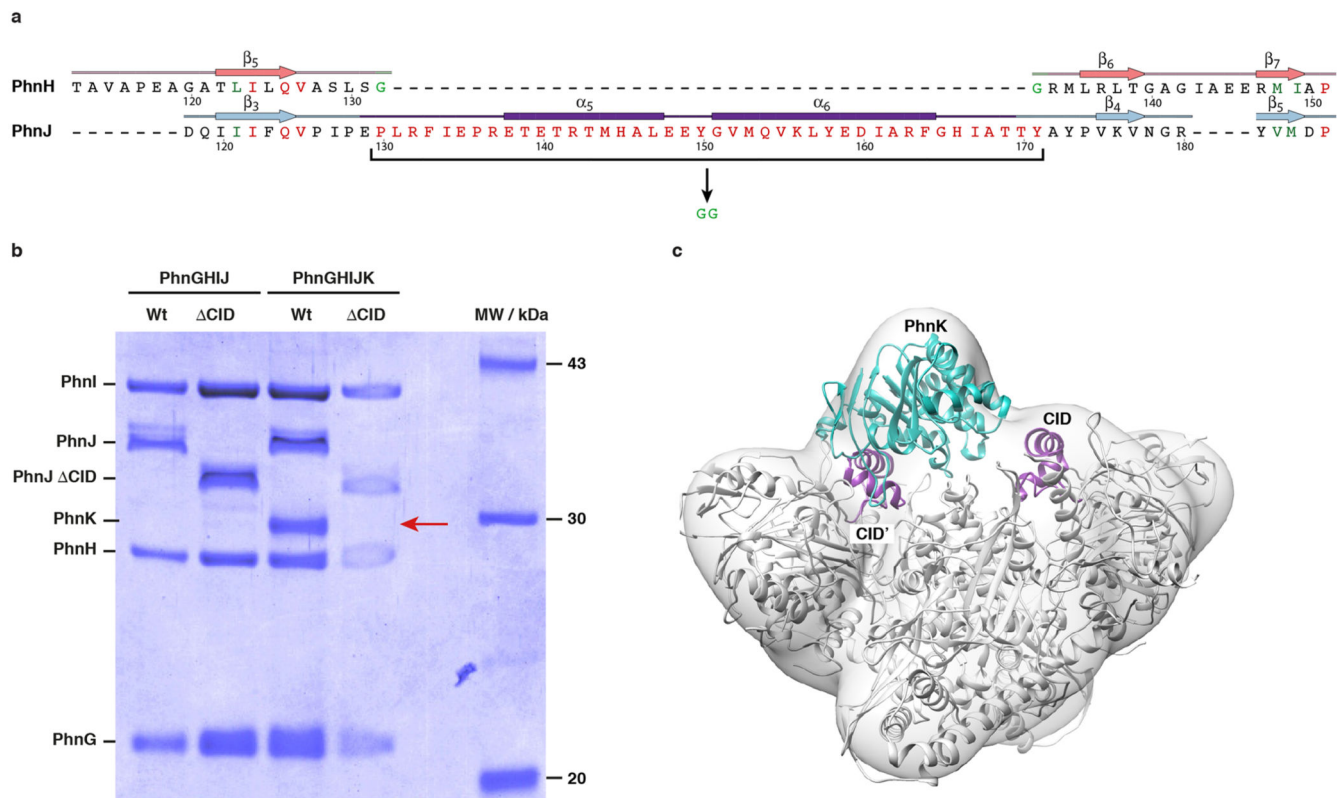
Extended Data Figure 10. Purification of the PhnGHIJ ^{CID} **and PhnGHIJ** ^{CID}**K complexes**
a, Sequence alignment of the CID-region of PhnJ with the corresponding part of the structural domain of PhnH, where strands 5 and 6 of the β-sheet are connected by a short Gly-Gly turn (green residues). In PhnJ ^{CID}, 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 central insertion domain on PhnJ. PhnK is missing from the complex purified from the PhnGHIJ ^{CID}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 EM 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	Ta$_6$Br$_{12}$ derivative
Data collection		
Space group	P2$_1$2$_1$2$_1$	P2$_1$2$_1$2$_1$
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)

	Native	Ta$_6$Br$_{12}$ derivative
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)
CC$_{½}$	99.9 (51.6)	99.7 (96.4)
I/σ_I	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_{work} / R_{free} (%)	14.9(17.6)	
No. of atoms		
Protein (non-hydrogen)	30,097 (15,203)	
SO$_4^{2-}$ / Zn^{2+}	24	
Water	1,792	
B-factors (Å^2)		
Protein	29.87	
SO$_4^{2-}$	40.28	
Zn^{2+}	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 parenthesis, except where otherwise stated.

[†] Statistics from MolProbity via Phenix[38].

Acknowledgements

References

1. Karl DM. Aquatic ecology. Phosphorus, the staff of life. Nature. 2000; 406:31–33. doi: 10.1038/35017686. [PubMed: 10894527]

2. Hsieh YJ, Wanner BL. Global regulation by the seven-component Pi signaling system. Curr Opin Microbiol. 2010; 13:198–203. doi:10.1016/j.mib.2010.01.014. [PubMed: 20171928]

3. Hove-Jensen B, Zechel DL, Jochimsen B. Utilization of glyphosate as phosphate source: biochemistry and genetics of bacterial carbon-phosphorus lyase. Microbiol Mol Biol Rev. 2014; 78:176–197. doi:10.1128/MMBR.00040-13. [PubMed: 24600043]

4. Imazu K, et al. Enhanced utilization of phosphonate and phosphite by *Klebsiella aerogenes*. Appl Environ Microbiol. 1998; 64:3754–3758. [PubMed: 9758795]

5. Quinn JP. Carbon phosphorus lyase activity - a novel mechanism of bacterial-resistance to the phosphonic acid antibiotics. Lett Appl Microbiol. 1989; 8:113–116. doi:Doi 10.1111/J.1472-765x. 1989.Tb00236.X.

6. Schowanek D, Verstraete W. Phosphonate utilization by bacterial cultures and enrichments from environmental samples. Appl Environ Microbiol. 1990; 56:895–903. [PubMed: 2339877]

7. Wackett LP, Shames SL, Venditti CP, Walsh CT. Bacterial carbon-phosphorus lyase: products, rates, and regulation of phosphonic and phosphinic acid metabolism. J Bacteriol. 1987; 169:710–717. [PubMed: 3804975]

8. White AK, Metcalf WW. Microbial metabolism of reduced phosphorus compounds. Annu Rev Microbiol. 2007; 61:379–400. doi:10.1146/annurev.micro.61.080706.093357. [PubMed: 18035609]

9. Chen CM, Ye QZ, Zhu ZM, Wanner BL, Walsh CT. Molecular biology of carbon-phosphorus bond cleavage. Cloning and sequencing of the phn (psiD) genes involved in alkylphosphonate uptake and C-P lyase activity in *Escherichia coli B*. J Biol Chem. 1990; 265:4461–4471. [PubMed: 2155230]

10. Hove-Jensen B, McSorley FR, Zechel DL. Physiological role of phnP-specified phosphoribosyl cyclic phosphodiesterase in catabolism of organophosphonic acids by the carbon-phosphorus lyase pathway. J Am Chem Soc. 2011; 133:3617–3624. doi:10.1021/ja1102713. [PubMed: 21341651]

11. Hove-Jensen B, Rosenkrantz TJ, Haldimann A, Wanner BL. *Escherichia coli* phnN, encoding ribose 1,5-bisphosphokinase activity (phosphoribosyl diphosphate forming): dual role in phosphonate degradation and NAD biosynthesis pathways. J Bacteriol. 2003; 185:2793–2801. [PubMed: 12700258]

12. Metcalf WW, Wanner BL. Evidence for a fourteen-gene, phnC to phnP locus for phosphonate metabolism in *Escherichia coli*. Gene. 1993; 129:27–32. [PubMed: 8335257]

13. Rizk SS, Cuneo MJ, Hellinga HW. Identification of cognate ligands for the *Escherichia coli* phnD protein product and engineering of a reagentless fluorescent biosensor for phosphonates. Protein Sci. 2006; 15:1745–1751. doi:10.1110/ps.062135206. [PubMed: 16751609]

14. Jochimsen B, et al. Five phosphonate operon gene products as components of a multi-subunit complex of the carbon-phosphorus lyase pathway. Proc Natl Acad Sci U S A. 2011; 108:11393–11398. doi:10.1073/pnas.1104922108. [PubMed: 21705661]

15. Ren Z, et al. Subunit interactions within the carbon-phosphorus lyase complex from *Escherichia coli*. Biochemistry. 2015 doi:10.1021/acs.biochem.5b00194.

16. Frost JW, Loo S, Cordeiro ML, Li D. Radical-based dephosphorylation and organophosphonate biodegradation. J Am Chem Soc. 1987; 109:2166–2171. doi:Doi 10.1021/Ja00241a039.

17. Kamat SS, Williams HJ, Dangott LJ, Chakrabarti M, Raushel FM. The catalytic mechanism for aerobic formation of methane by bacteria. Nature. 2013; 497:132–136. doi:10.1038/nature12061. [PubMed: 23615610]

18. Kamat SS, Williams HJ, Raushel FM. Intermediates in the transformation of phosphonates to phosphate by bacteria. Nature. 2011; 480:570–573. doi:10.1038/nature10622. [PubMed: 22089136]

19. Adams MA, et al. Crystal structure of PhnH: an essential component of carbon-phosphorus lyase in *Escherichia coli*. J Bacteriol. 2008; 190:1072–1083. doi:10.1128/JB.01274-07. [PubMed: 17993513]

20. Shisler KA, Broderick JB. Glycyl radical activating enzymes: structure, mechanism, and substrate interactions. Arch Biochem Biophys. 2014; 546:64–71. doi:10.1016/j.abb.2014.01.020. [PubMed: 24486374]

21. Booker SJ, Grove TL. Mechanistic and functional versatility of radical SAM enzymes. F1000 Biol Rep. 2010; 2:52. doi:10.3410/B2-52. [PubMed: 21152342]

22. Frey PA, Hegeman AD, Ruzicka FJ. The radical SAM superfamily. Crit Rev Biochem Mol Biol. 2008; 43:63–88. doi:10.1080/10409230701829169. [PubMed: 18307109]

23. Sofia HJ, Chen G, Hetzler BG, Reyes-Spindola JF, Miller NE. 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. 2001; 29:1097–1106. [PubMed: 11222759]

24. Goldman PJ, et al. X-ray structure of an AdoMet radical activase reveals an anaerobic solution for formylglycine posttranslational modification. Proc Natl Acad Sci U S A. 2013; 110:8519–8524. doi:10.1073/pnas.1302417110. [PubMed: 23650368]

25. McCall KA, Huang C, Fierke CA. Function and mechanism of zinc metalloenzymes. J Nutr. 2000; 130:1437S–1446S. [PubMed: 10801957]

26. Yakovleva GM, Kim SK, Wanner BL. Phosphate-independent expression of the carbon-phosphorus lyase activity of Escherichia coli. Appl Microbiol Biotechnol. 1998; 49:573–578. [PubMed: 9650256]

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. 2013; 69:256–265. doi:10.1107/S0907444912045180. [PubMed: 23385461]

28. Hollenstein K, Dawson RJ, Locher KP. Structure and mechanism of ABC transporter proteins. Curr Opin Struct Biol. 2007; 17:412–418. doi:10.1016/j.sbi.2007.07.003. [PubMed: 17723295]

29. Landau M, et al. ConSurf 2005: the projection of evolutionary conservation scores of residues on protein structures. Nucleic Acids Res. 2005; 33:W299–302. doi:10.1093/nar/gki370. [PubMed: 15980475]

30. D'Arcy A, Villard F, Marsh M. An automated microseed matrix-screening method for protein crystallization. Acta Crystallogr D. 2007; 63:550–554. doi:10.1107/S0907444907007652. [PubMed: 17372361]

31. Knablein J, et al. Ta6Br(2+)12, a tool for phase determination of large biological assemblies by X-ray crystallography. J Mol Biol. 1997; 270:1–7. [PubMed: 9231895]

32. Winter G. xia2: an expert system for macromolecular crystallography data reduction. J Appl Cryst. 2009; 43:186–190. doi:10.1107/s0021889809045701.

33. Adams PD, et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. Acta Crystallogr D. 2010; 66:213–221. doi:10.1107/S0907444909052925. [PubMed: 20124702]

34. Cohen SX, et al. Towards complete validated models in the next generation of ARP/wARP. Acta Crystallogra D. 2004; 60:2222–2229. doi:Doi 10.1107/S0907444904027556.

35. Thorn A, Sheldrick GM. Extending molecular-replacement solutions with SHELXE. Acta Crystallogr D. 2013; 69:2251–2256. doi:10.1107/S0907444913027534. [PubMed: 24189237]

36. Cowtan K. The Buccaneer software for automated model building. 1. Tracing protein chains. Acta Crystallogra D. 2006; 62:1002–1011. doi:Doi 10.1107/S0907444906022116.

37. Emsley P, Lohkamp B, Scott WG, Cowtan K. Features and development of Coot. Acta Crystallogr D. 2010; 66:486–501. doi:10.1107/S0907444910007493. [PubMed: 20383002]

38. Chen VB, et al. MolProbity: all-atom structure validation for macromolecular crystallography. Acta Crystallogra D. 2010; 66:12–21. doi:Doi 10.1107/S0907444909042073.

39. Schrodinger, LLC. The PyMOL molecular graphics system, version 1.3r1. 2010.

40. Pettersen EF, et al. UCSF Chimera--a visualization system for exploratory research and analysis. J Comput Chem. 2004; 25:1605–1612. doi:10.1002/jcc.20084. [PubMed: 15264254]

41. Ermantraut E, Wohlfart K, Tichelaar W. Perforated support foils with pre-defined hole size, shape and arrangement. Ultramicroscopy. 1998; 74:75–81. doi:Doi 10.1016/S0304-3991(98)00025-4.

42. Tang G, et al. EMAN2: an extensible image processing suite for electron microscopy. J Struct Biol. 2007; 157:38–46. doi:10.1016/j.jsb.2006.05.009. [PubMed: 16859925]

43. Mindell JA, Grigorieff N. Accurate determination of local defocus and specimen tilt in electron microscopy. J Struct Biol. 2003; 142:334–347. [PubMed: 12781660]

44. Scheres SH. RELION: implementation of a Bayesian approach to cryo-EM structure determination. J Struct Biol. 2012; 180:519–530. doi:10.1016/j.jsb.2012.09.006. [PubMed: 23000701]

45. Henderson R, et al. Outcome of the first electron microscopy validation task force meeting. Structure. 2012; 20:205–214. doi:10.1016/j.str.2011.12.014. [PubMed: 22325770]

46. Scheres SH, Chen S. Prevention of overfitting in cryo-EM structure determination. Nat Methods. 2012; 9:853–854. doi:10.1038/nmeth.2115. [PubMed: 22842542]

47. Russo CJ, Passmore LA. Robust evaluation of 3D electron cryomicroscopy data using tilt-pairs. J Struct Biol. 2014 doi:10.1016/j.jsb.2014.06.006.

48. Neidhardt FC, Bloch PL, Smith DF. Culture medium for enterobacteria. J Bacteriol. 1974; 119:736–747. [PubMed: 4604283]
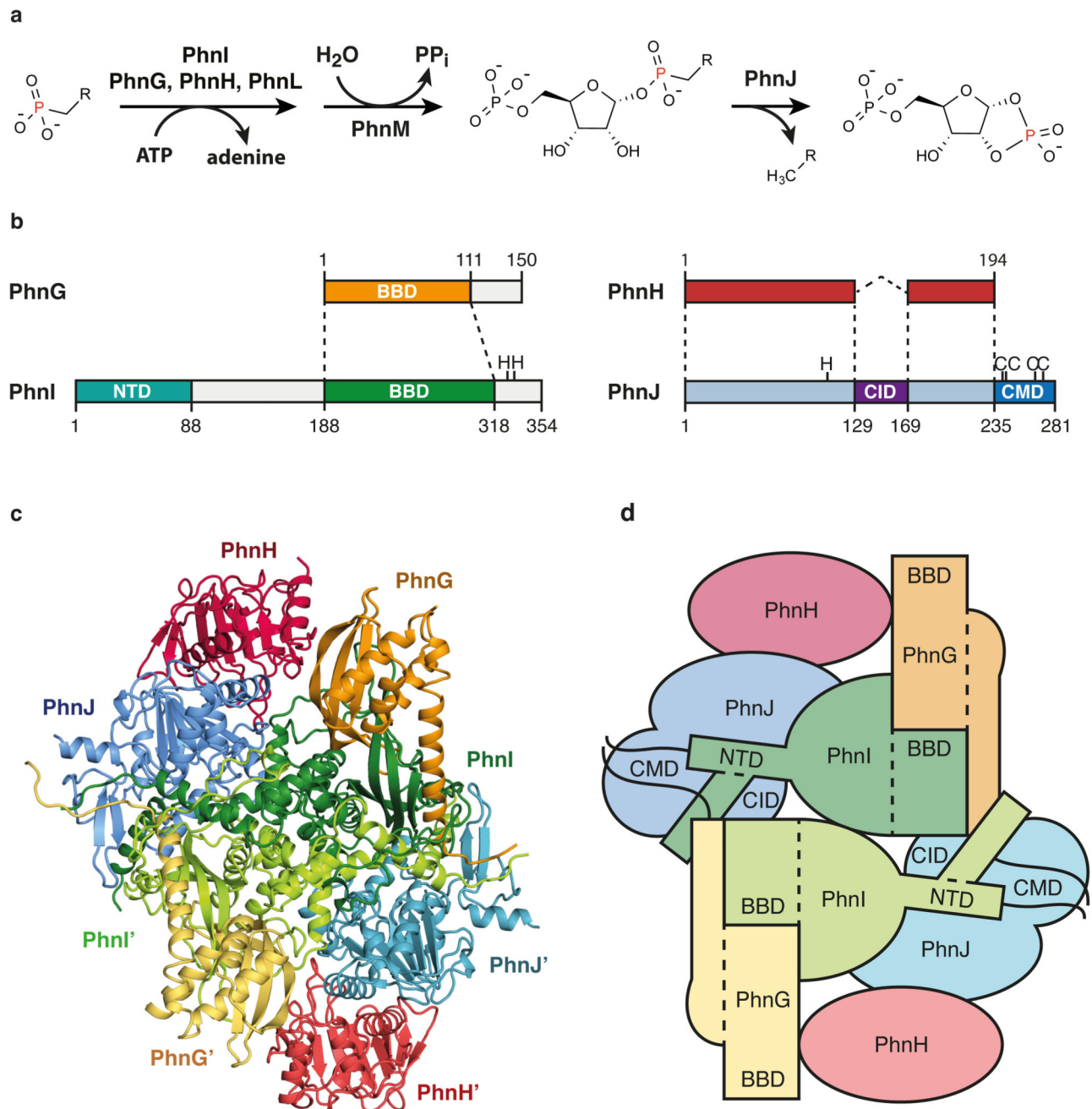
Figure 1. Overall architecture of the C-P lyase core complex

a, The C-P lyase core complex catalyses transfer of phosphonate to the 1′ position of ATP (PhnI assisted by PhnG, PhnH, and PhnL) and cleavage of the C-P bond (PhnJ). **b**, Overview of the four proteins with dashed lines showing conserved structural domains. Functional residues are shown with amino acid one-letter code. CID, central insertion domain, CMD, C-terminal mini domain; NTD, N-terminal domain, and BBD, β-barrel domain. **c**, Overall structure of the 240 kDa C-P lyase core complex. **d,** Schematic architecture of the complex with structural domains indicated as above.
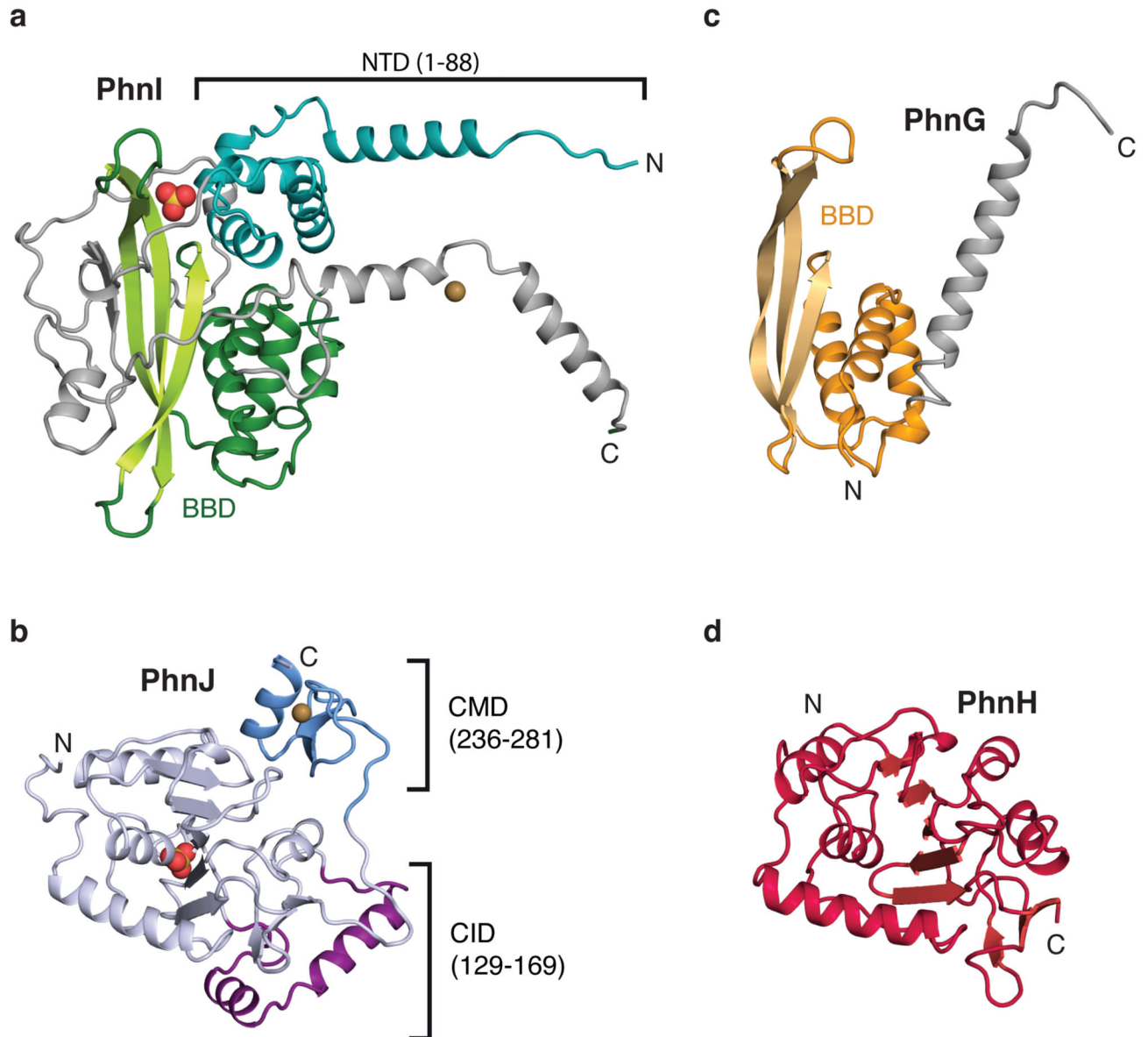
Figure 2. Details of subunit structures
Details of the individual protein structures in the complex, (**a**) PhnI, (**b**) PhnJ, (**c**) PhnG, and (**d**) PhnH, aligned to show their structural homologies and with domain colours as in Figure 1b. Ions are shown as spheres; sulphate, red and yellow; zinc; sand. CID, central insertion domain, CMD, C-terminal mini domain; NTD, N-terminal domain, and BBD, β-barrel domain.
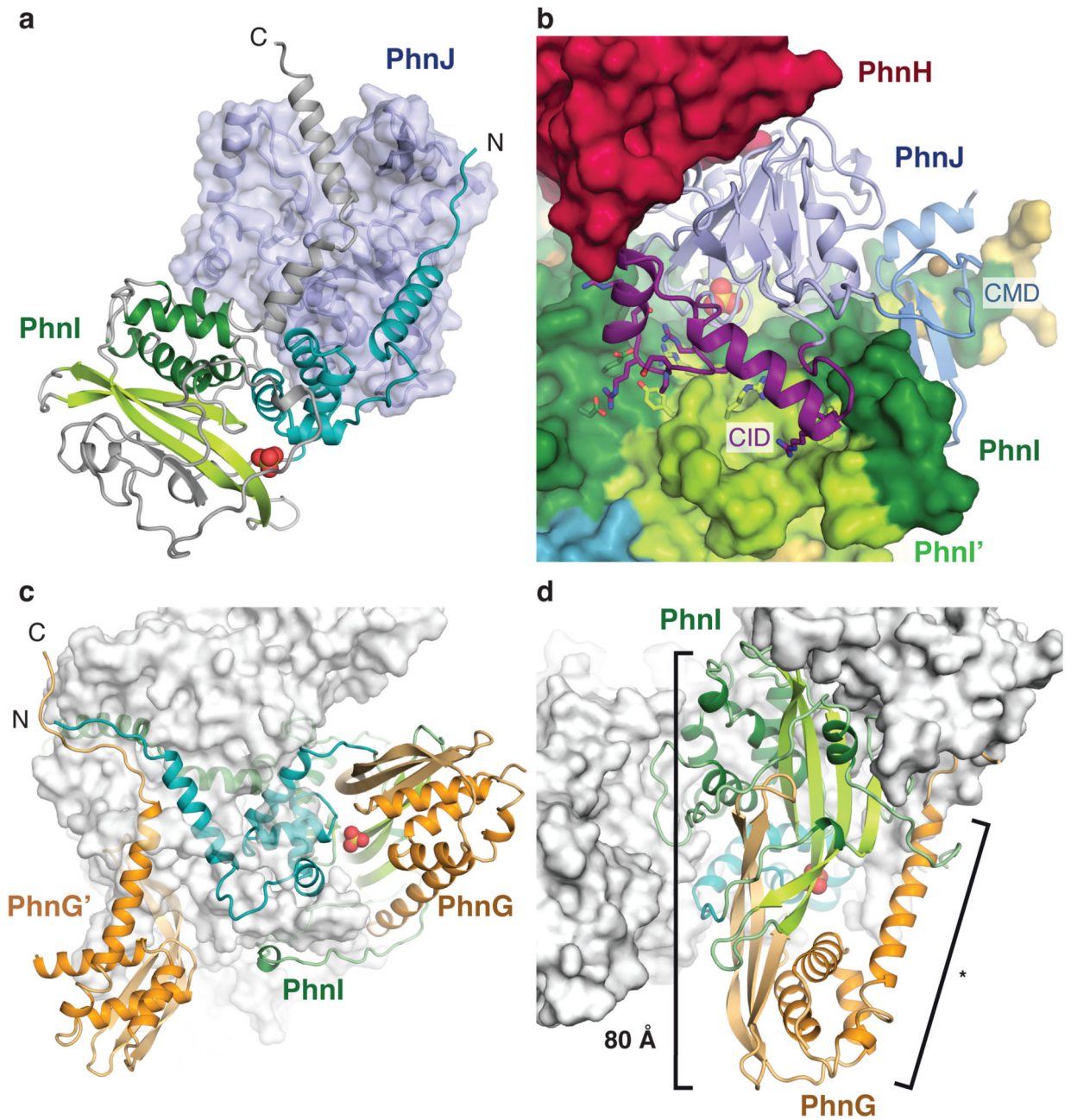
Figure 3. Subunit interactions within the C-P lyase core complex
a, The termini of PhnI (grey, C and cyan, N) grasp PhnJ (semi-transparent blue surface).
Domains in PhnI are coloured as in Figure 1 and the sulphate ion is shown as spheres. **b**,
The central insertion domain (CID, purple) contacts both copies of PhnI (green surface).
Residues involved in the interaction are shown with sticks. **c**, PhnI (green/cyan) interacts
with both copies of PhnG (orange). **d**, PhnG (orange) grabs both sides of PhnI (green) using
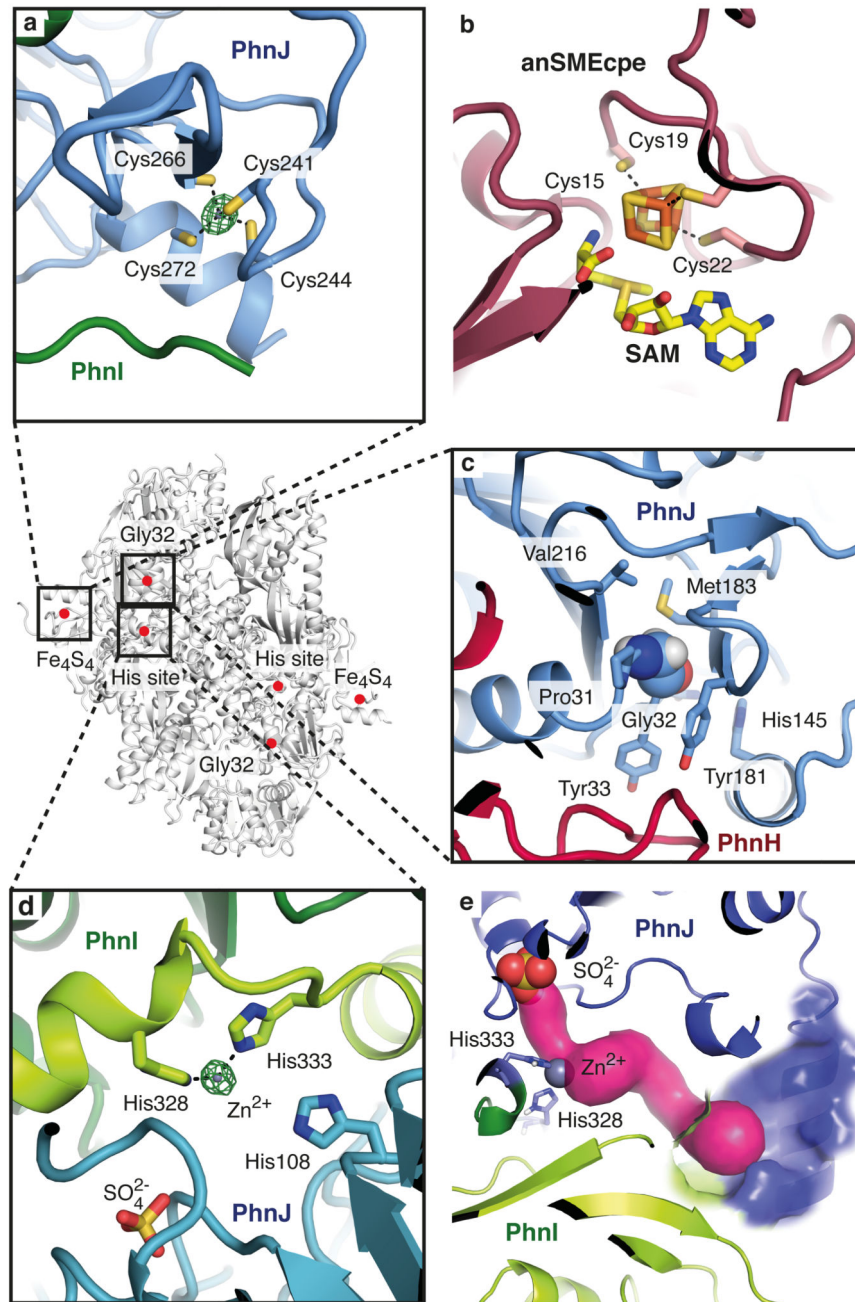the β-barrel domain and C-terminal helix (*). The combined β-barrel domain is 80 Å long.

Figure 4. Overview of the active sites of the C-P lyase core complex
a, Details of the iron-sulphur cluster site with a bound zinc ion (green = anomalous difference density at 6.0 σ). **b**, The Fe_4S_4 cluster found in *S*-adenosyl methionine activase (PDB 4K37)[24] shown in the same orientation as in **a**. **c**, The environment of PhnJ Gly32 (space fill). **d**, The zinc ion binding site at the interface between PhnI and PhnJ (His site) with the sulphate ion shown. **e**, A tunnel (magenta) leads from the sulphate (red and yellow) and zinc ions (grey) to the surface of the complex (blue area).
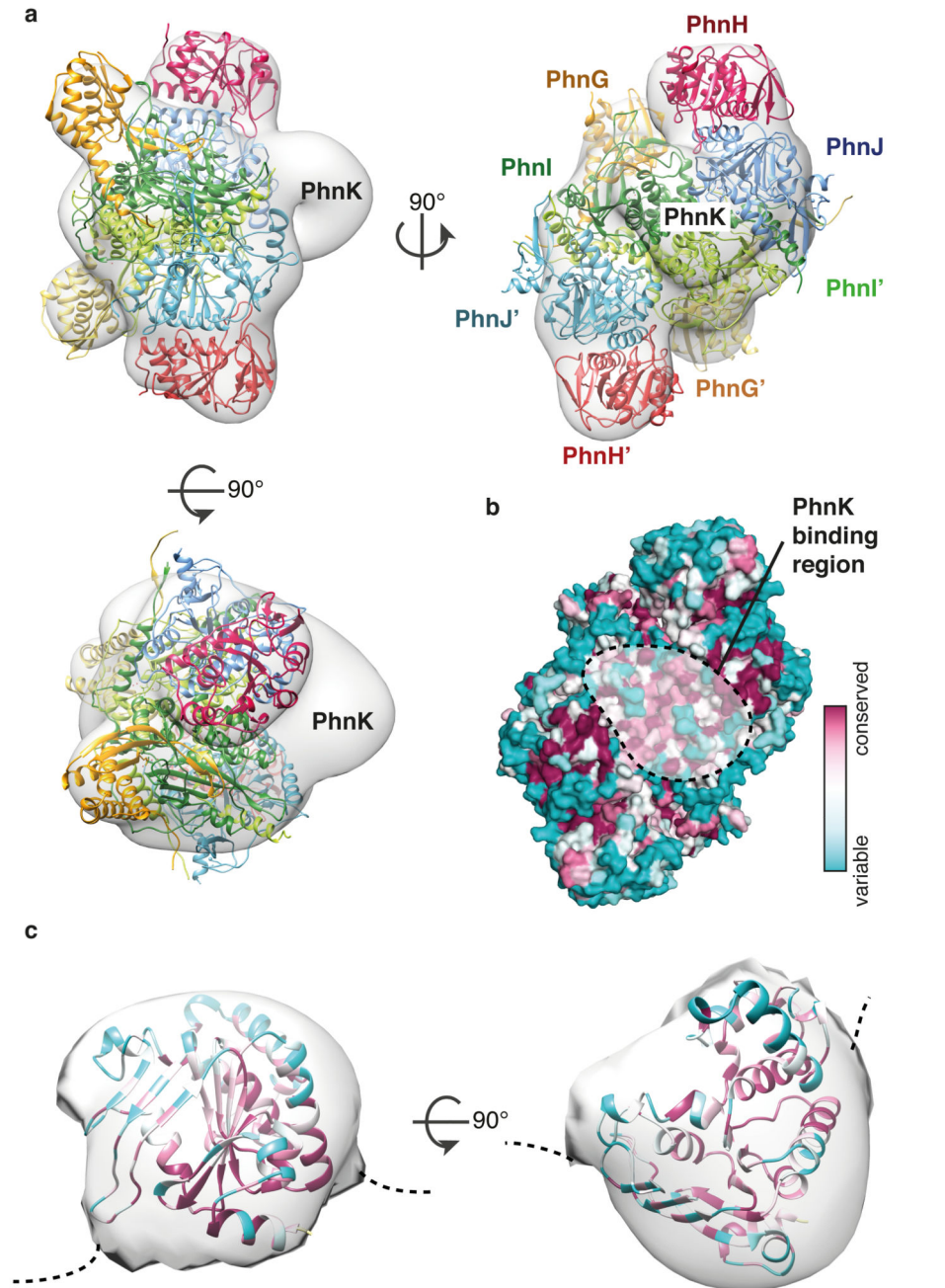
Figure 5. Mapping of PhnK using electron microscopy
a, Orthogonal views of the negative stain EM 3D reconstruction with the C-P lyase core complex crystal structure docked and the position of PhnK indicated. Colours are as in Figure 1c. **b**, Surface conservation of the C-P lyase core complex from variable (teal) to conserved (burgundy) with the PhnK binding region (dashed area) indicated. **c**, A single PhnK molecule (based on PDB ID 4FWI)[27] fitted into the EM density masked to remove the

C-P lyase core complex. The model is coloured by conservation using Consurf[29] at the same scale as in **b**.