



HAL
open science

Structural basis for catalysis in a CDP-alcohol phosphotransferase.

Giuliano Sciara, Oliver B Clarke, David Tomasek, Brian Kloss, Shantelle Tabuso, Rushelle Byfield, Raphael Cohn, Surajit Banerjee, Kanagalaghatta R Rajashankar, Vesna Slavkovic, et al.

► **To cite this version:**

Giuliano Sciara, Oliver B Clarke, David Tomasek, Brian Kloss, Shantelle Tabuso, et al.. Structural basis for catalysis in a CDP-alcohol phosphotransferase.. Nature Communications, 2014, 5, pp.4068. 10.1038/ncomms5068 . hal-01594683

HAL Id: hal-01594683

<https://hal.science/hal-01594683>

Submitted on 26 Sep 2017

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - ShareAlike 4.0 International License



HHS Public Access

Author manuscript

Nat Commun. Author manuscript; available in PMC 2014 December 13.

Published in final edited form as:

Nat Commun. ; 5: 4068. doi:10.1038/ncomms5068.

Structural basis for catalysis in a CDP-alcohol phosphotransferase

Giuliano Sciara^{1,†}, Oliver B. Clarke^{2,†}, David Tomasek^{1,†}, Brian Kloss³, Shantelle Tabuso³, Rushelle Byfield¹, Raphael Cohn¹, Surajit Banerjee⁴, Kanagalaghatta R. Rajashankar⁴, Vesna Slavkovic⁵, Joseph H. Graziano⁵, Lawrence Shapiro², and Filippo Mancía¹

¹Department of Physiology and Cellular Biophysics, Columbia University, New York, NY 10032, USA

²Department of Biochemistry and Molecular Biophysics, Columbia University, New York, NY 10032, USA

³New York Consortium on Membrane Protein Structure, New York Structural Biology Center, 89 Convent Avenue, New York, NY 10027, USA

⁴Department of Chemistry and Chemical Biology, Cornell University, NE-CAT, Advanced Photon Source, Argonne, IL 60439, USA

⁵Department of Environmental Health Sciences, Mailman School of Public Health, Columbia University, New York, NY 10032

Abstract

The CDP-alcohol phosphotransferase (CDP-AP) family of integral membrane enzymes catalyzes the transfer of a substituted phosphate group from a CDP-linked donor to an alcohol-acceptor. This is an essential reaction for phospholipid biosynthesis across all kingdoms of life, and it is catalyzed solely by CDP-APs. Here we report the 2.0 Å resolution crystal structure of a representative CDP-AP from *Archaeoglobus fulgidus*. The enzyme (AF2299) is a homodimer, with each protomer consisting of six transmembrane helices and an N-terminal cytosolic domain. A polar cavity within the membrane accommodates the active site, lined with the residues from an absolutely conserved CDP-AP signature motif (D₁xxD₂G₁xxAR...G₂xxxD₃xxxD₄). Structures in the apo, CMP-bound, CDP-bound and CDP-glycerol-bound states define functional roles for each

Users may view, print, copy, and download text and data-mine the content in such documents, for the purposes of academic research, subject always to the full Conditions of use:http://www.nature.com/authors/editorial_policies/license.html#terms

^{*}To whom correspondence should be addressed: fm123@columbia.edu (F.M.).

[†]These authors made equal contributions.

Author contributions

G.S. and B.K. designed the constructs. B.K. and S.T. performed the initial expression screens. G.S., D.T., R.C. and R.B. purified the protein and obtained the initial crystals. D.T. produced diffraction quality crystals. O.B.C. and D.T. screened crystals and collected the diffraction data. S.B. and K.R.R. assisted with data collection and processing. O.B.C. determined the structure and refined the model. V.S. and J.H.G. performed the atomic absorption spectrophotometry experiments. G.S., O.B.C., L.S. and F.M. analyzed the data and wrote the manuscript. F.M. coordinated and supervised the project.

Accession codes: Coordinates and structure factors have been deposited in the Protein Data Base under accession codes 4O6M (CMP/apo), 4O6N (CDP), 4Q7C (CDP-glycerol)

Competing financial interests

The authors declare no competing financial interests.

of these eight conserved residues and allow us to propose a sequential, base-catalyzed mechanism universal for CDP-APs, in which the fourth aspartate (D₄) acts as the catalytic base.

A key step in the synthesis of most phospholipids in both eukaryotes¹ and prokaryotes^{2,3} involves the transfer of a substituted phosphate group from a CDP-linked donor to an acceptor alcohol to generate a phosphodiester linked product. This essential reaction for the formation of phospholipid bilayers is catalyzed exclusively by a family of membrane-embedded enzymes known as CDP-alcohol phosphotransferases (CDP-APs). CDP-APs facilitate the conjugation of a polar headgroup, such as choline, inositol or ethanolamine, to a diacylglycerol lipid tail, resulting in the formation of polar phospholipids such as phosphatidylcholine (PC), phosphatidylinositol (PI) and phosphatidylethanolamine (PE). Depending on the enzyme and on organism, the diacylglycerol tail or the polar headgroup may participate as either the CDP-linked donor substrate, or the acceptor alcohol (Supplementary Table 1). Furthermore, both substrates may be lipids, as in cardiolipin synthase⁴, or polar small molecules⁵, further exemplifying the broad substrate specificity of this diverse class of enzymes for which the only two required elements are a terminal hydroxyl on the acceptor and a CDP-linkage on the donor. CDP-APs exhibit broad variety in the identity and lipophilic versus hydrophilic nature of the acceptor and donor substrates they process. Irrespective of the identity of the substrates, they are all predicted to be integral membrane proteins and possess a universally conserved signature motif of 8 amino acids spaced over a small stretch of sequence (D₁xxD₂G₁xxAR...G₂xxxD₃xxxD₄; Fig. 1a–b). The absolute conservation of the signature motif in enzymes with such diverse substrate specificities suggests that it plays a fundamental role in catalyzing phosphotransfer, but the absence of any structural information has hampered progress in understanding the mechanism of this essential class of integral membrane enzymes. Here we report the structure of AF2299, a representative CDP-AP from *Archaeoglobus fulgidus*, revealing the transmembrane (TM) architecture of an integral membrane phosphotransferase, unveiling the role of each conserved residue in the signature motif in the reaction, and allowing us to postulate a universal mechanism of catalysis for this reaction.

RESULTS

Structure determination

To understand the structural basis for catalysis in CDP-APs we determined crystal structures of AF2299, a representative member of this family from *Archaeoglobus fulgidus*, in the apo form, with bound CDP, and with bound CMP. The crystals, grown from lipidic cubic phase, belong to space group P 2₁ with one homodimer per asymmetric unit, and they diffract X-rays to 1.9 Å (CMP/apo) and 2.1 Å (CDP) respectively. We report two structures, one with both active sites occupied by CDP (PDB accession number 4O6N), and the other with one active site bound by CMP and the other one in the apo state (PDB accession number 4O6M). Both structures were obtained with co-purified ligands. To shed light on the key elements for substrate specificity, we also determined the structure to 3.1 Å resolution of AF2299 co-crystallized with CDP-glycerol, the likely donor substrate for this enzyme.

Architecture of a CDP-alcohol phosphotransferase

AF2299 is a homodimer, with each protomer consisting of 2 distinct domains. An N-terminal cytoplasmic domain of unknown function has clear structural similarity to the cytidyltransferases⁶, but the active site residues are not conserved, suggesting that it may lack catalytic activity. The cytidyltransferase domain perhaps represents an evolutionary vestige of a bifunctional enzyme that performed the function of generating the CDP-donor substrate in addition to the phosphotransferase activity retained in AF2299. Such dual-function enzymes have been characterized in archaea⁷. The N-terminal cytidyltransferase domain is followed by a C-terminal TM domain consisting of 6 TM helices arranged as two inverted repeats of three helices (Fig. 1c–d and Fig. 2a). The dimer interface is symmetric, and involves TM 3 and TM 4 (Fig. 2a). All TM helices except TM5 extend beyond the predicted intracellular surface of the membrane by three to four turns (Fig. 2b). A curved juxtamembrane helix wraps around the outer surface of TM2 and TM5. A mobile loop (ordered in only one of the two subunits) provides a connection between the last helix of the cytidyltransferase domain and the juxtamembrane helix (Fig. 1d).

CDP binds within a cavity lined by the signature motif

The 6 TM helices of AF2299 surround a polar, positively charged cavity extending from the cytosol to a depth of 8 Å into the membrane (Fig. 3a and Supplementary Fig. 1). The CDP-AP signature motif lines one side of this polar cavity formed by the cytosolic ends of TM2 and TM3, and defines the active site (Fig. 3b–c). The two conserved segments of the motif reside on TM2 and TM3 respectively, and the residues that separate them constitute the loop joining the two TM helices (Fig. 3c). Of the four conserved aspartic acid residues in the signature motif, the first (D₁, Asp214) and the second (D₂, Asp217) on TM2 directly face the fourth (D₄, Asp239) and the third (D₃, Asp235) respectively on TM3 (Fig. 3b). The first conserved glycine (G₁, Gly218) provides the necessary flexibility to TM2 to allow the carbonyl of D₁ to reorient away from the helix axis towards the active site. In the structures with bound CDP or CMP, the ligand is wedged in a cleft between TM2 and TM3, and backed by TM1 (Fig. 3c and Supplementary Fig. 2). The conserved alanine (Ala221) and G₁ interact with one face of the planar cytosine ring whilst the second conserved glycine (G₂, Gly231) interacts with the opposite face. The N-terminal end of TM1 also contributes to nucleotide binding, primarily through interactions of a small segment that includes a conserved threonine (Thr178) which forms part of a hydrogen bonding network stabilizing the polar edge of the cytosine ring (Supplementary Fig. 3 a–b). The invariant arginine (Arg222) interacts with the α-phosphate of CDP, and this interaction is retained in the CMP bound state (Supplementary Fig. 3 a–b). Finally, it may be worth mentioning that Ser228, another residue in this network, forms a water-mediated hydrogen bond to N3 in the pyrimidine ring. This residue is typically either a serine or threonine in CDP-APs (Fig. 1a), and a well-characterized mutation (T60P) of the equivalent residue in the *E. coli* gene PgsA, which encodes a phosphatidylglycerophosphate (PGP) synthase, results in a defect in phosphatidylglycerol synthesis^{8,9}.

Divalent cation binding sites

All CDP-APs require a divalent cation for activity, most often Mg^{2+} (ref.¹⁰), Mn^{2+} (ref.¹¹), or Co^{2+} (ref.¹²). In the structure of AF2299, a bound metal ion mediates the interaction of the pyrophosphate moiety of CDP with the sidechains of D₂ and D₃, and with the backbone carbonyl of D₁ (Fig. 3b). In the CDP-bound and apo structures only, a secondary metal binding site is located 4 Å away from the first, coordinated by the side chains of D₁, D₃ and D₄ (Supplementary Fig. 3 c). We could not observe any anomalous signal derived from these ions, a strong indication that points to the exclusion of transition metals, leaving Ca^{2+} or Mg^{2+} as possibilities. Details of their coordination are most consistent with Ca^{2+} , due to both the pentagonal-bipyramid coordination geometry and bond distances in the range of 2.3–2.7 Å, which are considerably longer than would be expected for Mg^{2+} ions¹³. We confirmed the identity of the divalent cations as Ca^{2+} by flame atomic absorption spectrophotometry on purified protein. Indeed, by this method we could measure a signal from Ca above background and equimolar to the concentration of AF2299 (Methods and Supplementary Table 2), whilst we could not detect the presence of any protein-associated Mg.

Two substrate-binding pockets are located within the polar cavity

Deeper in the active site of the CMP and apo structures, a tetrahedral oxyanion, probably sulfate given its presence in the crystallization solution, is bound by Arg240 and Arg304 (Fig. 4a and Supplementary Figure 3 d). The sulfate also forms a hydrogen bond with the amide nitrogen of Arg301, which lies at the N-terminal end of the short TM5, consistent with a role for the helix dipole in stabilization of anion binding¹⁴. The oxyanion-binding site is formed by residues that are conserved in a manner that appears to correlate with the identity of the acceptor. In the CDP structure, for which tartrate rather than sulfate was present in the crystallization solution, a tartrate ion is found in the same position occupied by the sulfate in the CMP and apo structures (Supplementary Fig. 4). Sequence analysis reveals that Arg239 is conserved in all enzymes that take either inositol or inositol phosphate as the acceptor and that a short motif (R₃₀₁xxR₃₀₄, in AF2299 numbering) is specifically conserved in the subset of these enzymes that utilize inositol-1-phosphate (Fig. 4b–c). We hypothesize that Arg240 forms hydrogen bonds with hydroxyls 2/3 or 5/6 of the inositol, fixing the orientation of the ring in a catalytically competent orientation, with the 4-OH positioned in-line with the pyrophosphate moiety of the CDP-donor. The data that we present here are insufficient to assign the identity of the acceptor alcohol for AF2299, however they do allow us to identify with reasonable certainty the location of the acceptor binding-site, which we would expect to be conserved throughout the family of CDP-APs.

Another apparent binding pocket extends into the wall of the polar cavity, perpendicular to the CDP pyrophosphate (Fig. 4b). This site is lined by residues specifically conserved in the archaeal family of CDP-APs that utilize donors with small, polar substituents, such as CDP-inositol and CDP-glycerol, to generate soluble phosphodiester linked osmolytes that apparently help their host organisms endure thermal and osmotic stress¹⁵. This family includes the di-myoinositol-1,3-phosphate synthase (DIPPS) subfamily of CDP-APs, which are the best-characterized enzymes of this class¹⁶. Based on this observation, we suggest that this pocket may accommodate the donor substituent, and that AF2299 may catalyze the

reaction of inositol-1-phosphate with a small molecule, possibly CDP-glycerol or CDP-inositol⁵. Indeed, co-crystallization with CDP-glycerol reveals additional density in this pocket consistent with binding of the glycerol moiety (Fig 5). This is in agreement with the observation that the AF2299 sequence is most closely related to those of the DIPPS family (Fig. 4c and Fig. 6). Given our structure with bound CDP-glycerol and the fact that an *A. fulgidus* DIPPS enzyme (AF0263) has already been identified⁷, the most plausible hypothesis is that AF2299 catalyses the condensation of CDP-glycerol with *L-myo*-inositol-1-phosphate, to yield the glyceryl phosphodiester of inositol-1-phosphate, although experimental confirmation of this conjecture awaits further studies.

DISCUSSION

The structure of AF2299 reveals that CDP-APs adopt what is, to the best of our knowledge, a novel fold, with six TM helices surrounding a polar cavity. The active site of AF2299 is located near the surface of the membrane with direct access to the cytosol, consistent with the function of many CDP-APs in binding one soluble and one membrane-embedded substrate. It is tempting to speculate that the gap in the wall of the polar cavity created by the shorter TM5 helix may facilitate access of CDP-linked lipids to the active site, in those enzymes that utilize CDP-archaeol or CDP-diacylglycerol as donor substrates. Access of CDP-linked lipids by such a route would not be possible in AF2299 from our current structures, as the N-terminal end of the juxtamembrane helix occludes this hypothetical pathway, but only a slight alteration in the length or conformation of this helix would be necessary to facilitate lipid binding. The juxtamembrane helix is less conserved than the other helices – it is present in the mycobacterial PIP-synthases, but not in bacterial PGP synthases such as *E. coli* PgsA or in the eukaryotic PI synthases. Sequence alignment and secondary structure prediction for other CDP-alcohol phosphotransferases suggests that at least the PI-synthases share the same fold as the TM domain of AF2299, with an amphipathic juxtamembrane helix followed by 6 TM helices, of which the fifth is considerably shorter than the other five (Supplementary Fig. 5).

Interestingly, enzymes that utilize a lipid acceptor, such as the eukaryotic phosphatidylcholine synthases, which take CDP-choline as a donor and diacylglycerol as an acceptor, have a somewhat different architecture, characterized by the presence of an additional 3–4 TM helices at the C-terminus, perhaps in order to accommodate the bulkier lipid acceptor.

The membrane localization of CDP-APs such as AF2299 and the DIP synthases, which condense two soluble substrates, is intriguing and we suggest that it may have an evolutionary explanation. If the ancestral proto-CDP-AP from which all current CDP-APs are derived was an integral membrane protein with hydrophobic substrates, it is difficult to envision how the enzyme architecture could be modified during evolution to become a soluble protein, but it is not unlikely that slight modifications around the active site could result in entry of and affinity for a pair of soluble substrates.

The structures of AF2299 in the apo, CMP-bound, and CDP-bound forms presented here provide us with the opportunity to dissect the structural basis for catalysis in a CDP-AP. It

has been suggested that reactions catalyzed by CDP-APs proceed by a sequential mechanism involving a nucleophilic attack of the acceptor on the β -phosphorous of the CDP-donor^{17,18}, but the roles of residues in the signature motif have remained unclear, although it has been previously suggested that either D₃ or D₄ may act as a catalytic base¹⁹. The AF2299 structures reveal that D₁, D₂, and D₃ coordinate the pyro-phosphate of CDP via the primary divalent cation. D₄, in contrast, based on its position approximately equidistant between the acceptor and donor binding sites (Supplementary Fig. 4), is likely to play a key role in catalysis. In *S. cerevisiae* choline phosphotransferase, mutation of the equivalent residue results in complete abolition of enzymatic activity¹⁹. Likewise, mutation of D₄ in an aminoalcoholphosphotransferase from *Brassica napus* (rapeseed) results in loss of detectable phosphotransferase activity²⁰. A more extensive alanine mutagenesis scan in *Rhizobium meliloti* phosphatidylcholine synthase yielded the same conclusions²¹. Furthermore, D₄ does not interact with the CDP either directly or via the bridging divalent cation. In summary, D₄ is absolutely conserved, essential for activity in enzymes from two different kingdoms (*Plantae* and *Fungi*), but does not appear to participate in substrate binding, and is located between the binding sites of the acceptor and donor substrates. These observations support a direct role for D₄ in the mechanism of phosphotransfer in CDP-APs. We propose that this residue acts as a catalytic base, abstracting a proton from the terminal hydroxyl of the acceptor alcohol, facilitating nucleophilic attack by the activated acceptor on the β -phosphorus of the CDP (Fig. 7a). Following either an associative or dissociative mechanism (Fig. 7b) this yields CMP and the phosphodiester product (Fig. 7c). We cannot unambiguously determine whether this reaction occurs by an associative mechanism via a pentacoordinate, trigonal bipyramidal transition state or by a dissociative mechanism via a trigonal planar intermediate. A plausible associative mechanism is depicted in Figure 7. Although the pKa of an isolated aspartate is too low to act as an efficient catalytic base under physiological conditions, substantial elevation of the pKa of catalytic aspartates has been previously described²², particularly in the presence of neighboring negative charges, as is observed here with the clustering of the 4 conserved aspartates. A base-catalyzed mechanism is supported by the observation that many CDP-APs have alkaline pH optima^{23,24}. Consistent with the presence of a general acid-base catalyst, rabbit phosphatidylinositol (PI) synthase has an alkaline pH optimum for the forward synthase reaction, whilst displaying a mildly acidic optimum for the reverse activity²⁵.

The role of the primary divalent cation appears to be to prime the pyro-phosphate for catalysis via an in-line mechanism of substrate transfer, in addition to withdrawing charge from the pyrophosphate, making it more electrophilic and increasing susceptibility to nucleophilic attack. The function of the secondary cation is less obvious. It may contribute to orienting the carboxyl of D₄ appropriately to deprotonate the acceptor. Secondary divalent cation binding sites have also been described for other phosphotransferases, such as cAMP-dependent protein kinase²⁶. Strikingly, this mechanism appears to have many common features with that of other phosphotransferases including kinases, which have a completely unrelated fold, but a very similar configuration of active site residues and substrates²⁷.

A remarkable feature of the CDP-AP family is the chemical diversity of both acceptor and donor substrates. In the active site of the AF2299 structure, in addition to the nucleotide-

binding site, we observe two pockets that are likely to host the donor substituent and the acceptor alcohol (Fig. 4b). This hypothesis is confirmed for the donor group by our structure showing CDP-glycerol bound to AF2299 (Fig. 5). Given the absolute conservation of residues involved in catalysis and nucleotide binding, the position of the donor and acceptor substrates will most probably also remain constant, so as to allow the reaction to occur. Therefore, the structures reported here are likely to provide a framework in which to investigate the molecular elements underlying substrate specificity.

AF2299 appears likely to take inositol-1-phosphate as the acceptor alcohol, an observation that is of particular significance when considered in the context of mycobacterial PI-synthesis, which proceeds via a somewhat different pathway to that in other organisms. PI-synthases from eukaryotes catalyze the direct formation of phosphatidylinositol from CDP-diacylglycerol (CDP-DAG) and myo-inositol, and they cannot process inositol phosphate²⁸. In contrast, mycobacterial phosphatidylinositol is formed in two steps. In the first step, a PI-synthase-like CDP-AP catalyzes the conjugation of CDP-DAG with inositol-1-phosphate. Importantly, this enzyme does not recognize inositol, instead specifically recognizing inositol-1-phosphate²⁹. This reaction generates phosphatidylinositol-phosphate, which is then processed by a phosphatase of unknown identity to yield phosphatidylinositol. Phosphatidylinositol biosynthesis is a particularly attractive target for the development of anti-tuberculosis therapeutics, as all downstream steps in the synthesis of lipoarabinomannan, an important component of the cell wall, are contingent upon the attachment of mannan chains to a phosphatidylinositol anchor^{30,31}. The orthogonal substrate specificities of mycobacterial and eukaryotic PI-synthases suggest that this may be a feasible avenue for therapeutic intervention. The transmembrane domain of AF2299 shares only 20% identity with *M. tuberculosis* PI synthase, but the high degree of conservation of specific elements present in AF2299, MT-PI synthases and the DIP synthases (Supplementary Fig. 6), suggests that AF2299 may provide a useful framework for the design of inositol-phosphate analogs for the inhibition of the mycobacterial PI-synthases. Previous development of inositol-1-phosphate analogs has focused on replacement of the 1-phosphate with an analogous non-hydrolyzable moiety such as phosphonate, but these modifications resulted in inhibitors with significantly lower affinity than the cognate substrate³². We would predict that the 1-phosphate interacts with two arginine residues (Arg300 and Arg303 in AF2299), and likely makes a significant contribution to binding affinity, suggesting that alterations at this position are likely to result in reduced affinity. A more productive avenue for exploration may involve replacement of the terminal hydroxyl with a less reactive substituent.

The best-characterized role of CDP-APs is in the biosynthesis of phospholipids. In the scheme of this reaction, the lipidic substrate can participate either as a donor substituent, as an acceptor alcohol, or two lipids may act as donor and acceptor respectively. The prokaryotic PGP-synthases and eukaryotic PI-synthases process a lipophilic donor (CDP-DAG) and a polar acceptor, respectively glycerol-3-phosphate³³ and myo-inositol³⁴. By contrast, the eukaryotic choline and ethanolamine phosphotransferases that participate in the Kennedy pathway of zwitterionic phospholipid biosynthesis utilize CDP-choline/CDP-ethanolamine as donors, and diacylglycerol as the acceptor alcohol^{35,36}. Eukaryotic

cardiolipin synthases⁴ represent a third class in which CDP-DAG condenses with phosphatidyl-glycerol. While representing the first reported structure for any CDP-AP, our current model of AF2299 does not allow us to draw firm conclusions regarding the binding mode of lipid substrates. However, attachment of a bulky, hydrophobic lipid substituent to the β phosphate of CDP would appear to restrict the number of possibilities for the overall disposition of the alkyl chains to only two, of which only one seems plausible. In one orientation, the alkyl chain would project into the polar cavity of the active site towards the acceptor. The more likely scenario, in agreement with the structure of bound CDP-glycerol, has the glycerol moiety of a lipid positioned in the secondary pocket, with the alkyl chains of the lipid exiting into the bilayer between TM2 and TM5.

METHODS

Target identification and cloning

A member of the CDP-alcohol phosphatidyltransferase class-I family from *Archaeoglobus fulgidus* DSM 4304 (AF2299, UniProt accession #O27985) was identified as a promising candidate for crystallization experiments by the NYCOMPS (New York Consortium on Membrane Protein Structure) high throughput screening pipeline^{39,40}. The gene encoding AF2299 was amplified from *Archaeoglobus fulgidus* DSM 4304 genomic DNA using (5'-tacttccaatccaatgccATGAGGTTAGCGTACGTTAAAAAC-3') as the forward primer and (5'-ttatccactccaatgCTAGTTCGTGTAACCTTCTGAGTG-3') as the reverse primer for PCR, where upper case letters indicate gene-specific sequences and lower case letters indicate sequences incorporated into the PCR product that are required for ligation independent cloning. The gene was cloned into the bacterial expression vector pMCSG7-10xHis, introducing an N-terminal decahistidine tag and a TEV protease cleavage site.

Protein expression and purification

AF2299 was overexpressed in *E. coli* strain BL21 (DE3) pLysS. Bacterial cultures were grown at 37°C in 2xYT medium supplemented with 100 μ g/mL ampicillin and 50 μ g/mL chloramphenicol. Once the OD₆₀₀ reached 1.0, the temperature was reduced to 22°C, and cells were induced with 0.2 mM isopropyl β -D-1-thiogalactopyranoside (IPTG) after 15 minutes. After overnight induction at 22°C, cells were harvested by centrifugation and stored at -80°C. Cells expressing selenomethionine-substituted protein were grown with a kit containing a minimal medium of M9 salts (M9 SeMET High-Yield Growth Media Kit, Shanghai Medicilon Inc.). 150 mg selenomethionine was added per liter of culture in addition to selenomethionine provided in the kit. The rest of the protocol was the same as for native protein expression.

For large-scale purification of both native and selenomethionine-substituted protein, frozen cell pellets were resuspended in lysis buffer containing 20 mM Na-HEPES pH 7.5, 200 mM NaCl, 1 mM EDTA, 1 mg/ml lysozyme, 20 mM MgSO₄, 10 μ g/mL DNase I, 10 μ g/mL RNase A, 1 mM TCEP, 1 mM PMSF, and Complete Mini EDTA-free protease inhibitor cocktail (Roche) as described in instructions. Cells were lysed with an Emulsiflex C3 homogenizer (Avestin). Lysate was solubilized for 1.5 hours with 1% (w/v) n-decyl- β -D-maltopyranoside (DM, Anagrade, Affymetrix) in a volume of approximately 40 mL per cell

pellet from 800 mL culture (~6 grams). After solubilization, insoluble material was pelleted by ultracentrifugation at 134,000 x g for 30 min at 4°C. Protein was purified by metal-affinity chromatography (Ni-NTA, Qiagen) followed by size-exclusion chromatography (SEC) on a Superose 12 column (GE Healthcare) in a buffer of 20 mM Na-HEPES pH 7.0, 200 mM NaCl, 0.2% (w/v) DM, and 1 mM tris(2-carboxyethyl)phosphine hydrochloride (TCEP-HCl). Approximately 1.5 mg of purified protein could be obtained from an 800 mL bacterial culture. Selenomethionine-substituted protein was purified using the same protocol.

Crystallization

Protein from peak fractions from size-exclusion chromatography was concentrated to 30–35 mg/ml (approximated by $A_{280\text{nm}}$) for crystallization using a centrifugal concentrator (Millipore) with a 100 kDa MWCO. Selenomethionine-substituted protein was concentrated to 25–30 mg/ml. Crystals were grown at room temperature (20–22°C) in lipidic cubic phase. Concentrated protein was mixed with monoolein (Sigma) in a 1:1.5 (w/w) ratio of protein:monoolein. A Mosquito LCP (TTP Labtech) robot was used to dispense a typical volume of 50–100 nL of protein/monoolein mixture onto a 96-well glass plate, which was covered with 750–1000 nL precipitant solution and sealed with a glass cover slip. Glass sandwich plates were stored in a 22°C incubator. Largest crystals grew to approximately $10 \times 30 \times 100 \mu\text{m}^3$ in 1 to 2 days in (a) 14–20% (v/v) PEG 550 MME, 0.1 M Na-HEPES or Tris-HCl pH 7.0–8.0, 0.2 M lithium sulfate (CMP/apo structure) (b) 25–30% (v/v) PEG 400, 0.1 M Na-HEPES pH 7.0–7.5, 0.2 M potassium sodium tartrate (CDP structure) or (c) 15% PEG 550 MME, 0.1 M Na-HEPES pH 7.0, 0.2 M potassium sodium tartrate, 2.5 mM CDP-glycerol, 1 mM TCEP (CDP-glycerol structure). Crystals of selenomethionine-substituted protein grew in similar conditions. A tungsten carbide glass-cutter (Hampton Research) was used to cut and remove the glass cover slip, and crystals were harvested using 20–100 μm MicroLoops and MicroMounts (MiTeGen). Crystals were flash-cooled directly in liquid nitrogen without additional cryoprotection.

Diffraction data collection and processing

Diffraction data were collected on NECAT beamlines 24-ID-C and 24-ID-E at the Advanced Photon Source (Argonne, IL). Crystals were centered by diffraction, rastering across a grid in one orientation, and then along the orthogonal axis intersecting the position where the best diffraction was obtained in the first orientation. All diffraction data were indexed, integrated scaled using XDS and XSCALE^{41,42}. Data sets collected above the Se K-edge from two selenomethionine-substituted crystals were used for structure solution by anomalous diffraction methods. Selenium sites were located using SHELXD^{43,44}, and refined in SHARP⁴⁵. Density modification, including solvent flattening, histogram matching and two-fold non-crystallographic symmetry averaging were performed using DM to give an initial experimentally phased electron density map⁴⁶.

Model building and refinement

An initial model was built using BUCCANEER⁴⁷. The model was manually completed using COOT and refined using the PHENIX crystallographic software package, alternating between cycles of manual building in COOT⁴⁸ and refinement in PHENIX⁴⁹. Local torsion angle NCS restraints were employed throughout refinement. Riding hydrogens were

included in the last round of refinement. The quality of the model was analyzed using the validation module of the PHENIX package, which incorporates Molprobrity clash score, density correlation and rotamer analysis⁵⁰. Data collection, refinement and validation statistics for both structures are provided in Table 1. Protein structure figures were prepared using UCSF Chimera⁵¹. Within the crystal, the protein packs as stacks of two-dimensional sheets (Supplementary Fig. 7). Within each layer, adjacent dimers are arranged in an up-down orientation, such that the N-terminal domains of adjacent layers in the stack are interdigitated. A stereo figure of representative electron density for the CMP/apo structure and a C-alpha trace depicting the fold is given in Supplementary Figure 8.

Ca and Mg measurements

Ca and Mg measurements were conducted using a dual flame/graphite furnace Atomic Absorption Spectrometer AAnalyst 800, manufactured by Perkin Elmer. Data were collected and processed with WinLab32 software. Ca and Mg concentrations were determined at wavelengths of 422.7nm, and 285.2nm respectively, using an air-acetylene flame and standard nebulizer with flow spoiler. Detection limits were 0.02mg/L, and 0.003mg/L for Ca and Mg, respectively. Each run included calibration with 3 standards, as well as quality control samples with known concentrations of Ca and Mg. Calibration standards were made in 1% (v/v) HCl. All other samples were diluted in 1% (v/v) HCl. Duplicate samples of AF2299 at a concentration of 40mg/L were analyzed, as well as a control sample of protein-free buffer. Raw data and derived results are reported in Supplementary Table 2.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Crystallographic data for this study were measured at the NE-CAT beamlines 24ID-C and E (supported by NIH-NIGMS grant P41 GM103403) at the Advanced Photon Source. This work was supported by an NIH-NIGMS initiative to the New York Consortium on Membrane Protein Structure (NYCOMPS; U54 GM095315). Flame atomic absorption spectrophotometry experiments were performed in a dedicated core facility in the Mailman School of Public Health of Columbia University (supported by NIH-NIEHS P30 ES 09089). GS was supported by the European Union Seventh Framework Programme (FP7/2007-2013) under grant agreement no. 255508. The authors would like to thank Wayne A. Hendrickson for his leadership of NYCOMPS, Helena Santos and George Carman for their invaluable advice, and Leora Hamberger for her assistance running the Mancia Lab.

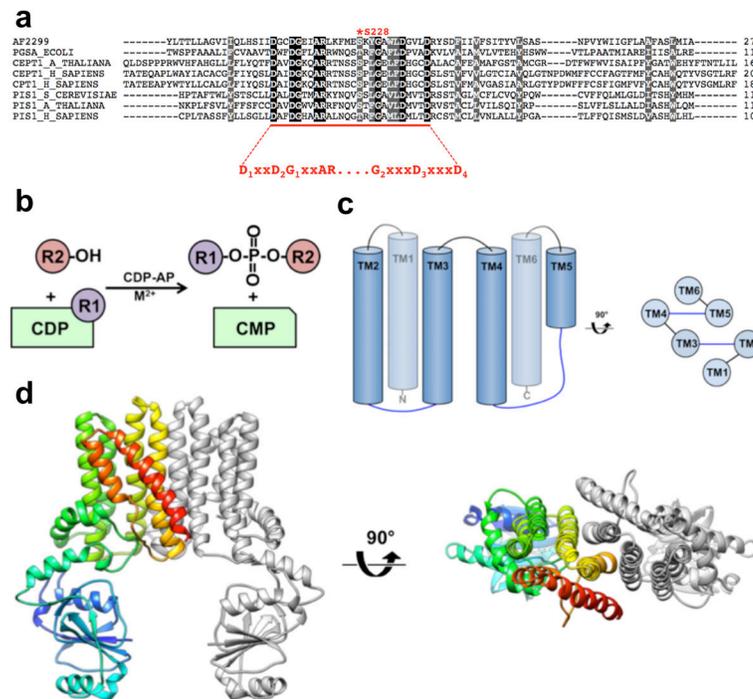
References

1. Vance JE, Vance DE. Phospholipid biosynthesis in mammalian cells. *Biochem Cell Biol.* 2004; 82:113–128. [PubMed: 15052332]
2. Cronan JE. Bacterial membrane lipids: where do we stand? *Annu Rev Microbiol.* 2003; 57:203–224. [PubMed: 14527277]
3. Jackson M, Crick DC, Brennan PJ. Phosphatidylinositol is an essential phospholipid of mycobacteria. *J Biol Chem.* 2000; 275:30092–30099. [PubMed: 10889206]
4. Houtkoper RH, et al. Identification and characterization of human cardiolipin synthase. *FEBS Letters.* 2006; 580:3059–3064. [PubMed: 16678169]
5. Borges N, et al. Biosynthetic pathways of inositol and glycerol phosphodiester used by the hyperthermophile *Archaeoglobus fulgidus* in stress adaptation. *J Bacteriol.* 2006; 188:8128–8135. [PubMed: 17028285]

6. Brito JA, Borges N, Vonnrhein C, Santos H, Archer M. Crystal structure of Archaeoglobus fulgidus CTP:inositol-1-phosphate cytidyltransferase, a key enzyme for di-myo-inositol-phosphate synthesis in (hyper)thermophiles. *J Bacteriol.* 2011; 193:2177–2185. [PubMed: 21378188]
7. Rodrigues MV, et al. Bifunctional CTP:inositol-1-phosphate cytidyltransferase/CDP-inositol:inositol-1-phosphate transferase, the key enzyme for di-myo-inositol-phosphate synthesis in several (hyper)thermophiles. *J Bacteriol.* 2007; 189:5405–5412. [PubMed: 17526717]
8. Miyazaki C, Kuroda M, Ohta A, Shibuya I. Genetic manipulation of membrane phospholipid composition in Escherichia coli: pgsA mutants defective in phosphatidylglycerol synthesis. *Proc Natl Acad Sci USA.* 1985; 82:7530–7534. [PubMed: 2999767]
9. Usui M, Sembongi H, Matsuzaki H, Matsumoto K, Shibuya I. Primary structures of the wild-type and mutant alleles encoding the phosphatidylglycerophosphate synthase of Escherichia coli. *J Bacteriol.* 1994; 176:3389–3392. [PubMed: 8195097]
10. Woodard DS, Lee TC, Snyder F. The final step in the de novo biosynthesis of platelet-activating factor. Properties of a unique CDP-choline:1-alkyl-2-acetyl-sn-glycerol choline-phosphotransferase in microsomes from the renal inner medulla of rats. *J Biol Chem.* 1987; 262:2520–2527. [PubMed: 3029085]
11. de Rudder KE, Sohlenkamp C, Geiger O. Plant-exuded choline is used for rhizobial membrane lipid biosynthesis by phosphatidylcholine synthase. *J Biol Chem.* 1999; 274:20011–20016. [PubMed: 10391951]
12. Sandoval-Calderon M, Geiger O, Guan Z, Barona-Gomez F, Sohlenkamp C. A eukaryote-like cardiolipin synthase is present in Streptomyces coelicolor and in most actinobacteria. *J Biol Chem.* 2009; 284:17383–17390. [PubMed: 19439403]
13. Zheng H, Chruszcz M, Lasota P, Lebioda L, Minor W. Data mining of metal ion environments present in protein structures. *J Inorg Biochem.* 2008; 102:1765–1776. [PubMed: 18614239]
14. Copley RR, Barton GJ. A structural analysis of phosphate and sulphate binding sites in proteins. Estimation of propensities for binding and conservation of phosphate binding sites. *J Mol Biol.* 1994; 242:321–329. [PubMed: 7932692]
15. Martins LO, Santos H. Accumulation of Mannosylglycerate and Di-myo-Inositol-Phosphate by Pyrococcus furiosus in Response to Salinity and Temperature. *Applied and Environmental Microbiology.* 1995; 61:3299–3303. [PubMed: 16535119]
16. Rodionov DA, et al. Genomic identification and in vitro reconstitution of a complete biosynthetic pathway for the osmolyte di-myo-inositol-phosphate. *Proc Natl Acad Sci USA.* 2007; 104:4279–4284. [PubMed: 17360515]
17. Pontoni G, et al. Studies on enzyme-substrate interactions of cholinephosphotransferase from rat liver. *Biochim Biophys Acta.* 1985; 836:222–232. [PubMed: 2992601]
18. Schlame M, Zhao M, Rua D, Haldar D, Greenberg ML. Kinetic analysis of cardiolipin synthase: a membrane enzyme with two glycerophospholipid substrates. *Lipids.* 1995; 30:633–640. [PubMed: 7564918]
19. Williams JG, McMaster CR. Scanning alanine mutagenesis of the CDP-alcohol phosphotransferase motif of Saccharomyces cerevisiae cholinephosphotransferase. *J Biol Chem.* 1998; 273:13482–13487. [PubMed: 9593682]
20. Qi Q, Huang YF, Cutler AJ, Abrams SR, Taylor DC. Molecular and biochemical characterization of an aminoalcoholphosphotransferase (AAPT1) from Brassica napus: effects of low temperature and abscisic acid treatments on AAPT expression in Arabidopsis plants and effects of over-expression of BnAAPT1 in transgenic Arabidopsis. *Planta.* 2003; 217:547–558. [PubMed: 12739150]
21. Solis-Oviedo RL, Martinez-Morales F, Geiger O, Sohlenkamp C. Functional and topological analysis of phosphatidylcholine synthase from Sinorhizobium meliloti. *Biochim Biophys Acta.* 2012; 1821:573–581. [PubMed: 22333179]
22. Harris TK, Turner GJ. Structural basis of perturbed pKa values of catalytic groups in enzyme active sites. *IUBMB Life.* 2002; 53:85–98. [PubMed: 12049200]
23. Nowicki M, Muller F, Frentzen M. Cardiolipin synthase of Arabidopsis thaliana. *FEBS Letters.* 2005; 579:2161–2165. [PubMed: 15811335]

24. Antonsson BE. Purification and characterization of phosphatidylinositol synthase from human placenta. *Biochem J.* 1994; 297:517–522. [PubMed: 8110188]
25. Bleasdale JE, Wallis P, MacDonald PC, Johnston JM. Characterization of the forward and reverse reactions catalyzed by CDP-diacylglycerol:inositol transferase in rabbit lung tissue. *Biochim Biophys Acta.* 1979; 575:135–147. [PubMed: 41587]
26. Zheng J, et al. 2.2 A refined crystal structure of the catalytic subunit of cAMP-dependent protein kinase complexed with MnATP and a peptide inhibitor. *Acta Crystallogr Sect F Struct Biol Cryst Commun.* 1993; 49:362–365.
27. Matte A, Tari LW, Delbaere LT. How do kinases transfer phosphoryl groups? *Structure/Folding and Design.* 1998; 6:413–419. [PubMed: 9562560]
28. Morii H, Ogawa M, Fukuda K, Taniguchi H. Ubiquitous distribution of phosphatidylinositol phosphate synthase and archaetidylinositol phosphate synthase in Bacteria and Archaea, which contain inositol phospholipid. *Biochemical and Biophysical Research Communications.* 2013.1016/j.bbrc.2013.11.054
29. Morii H, Ogawa M, Fukuda K, Taniguchi H, Koga Y. A revised biosynthetic pathway for phosphatidylinositol in Mycobacteria. *J Biochem.* 2010; 148:593–602. [PubMed: 20798167]
30. Berg S, Kaur D, Jackson M, Brennan PJ. The glycosyltransferases of Mycobacterium tuberculosis--roles in the synthesis of arabinogalactan, lipoarabinomannan, and other glycoconjugates. *Glycobiology.* 2007; 17:35R–56R. [PubMed: 17261566]
31. Jackson M, Crick DC, Brennan PJ. Phosphatidylinositol is an essential phospholipid of mycobacteria. *J Biol Chem.* 2000; 275:30092–30099. [PubMed: 10889206]
32. Morii H, et al. Studies of inositol 1-phosphate analogues as inhibitors of the phosphatidylinositol phosphate synthase in mycobacteria. *J Biochem.* 2013; 153:257–266. [PubMed: 23225597]
33. Hagio M, et al. Direct evidence for requirement of phosphatidylglycerol in photosystem II of photosynthesis. *Plant Physiology.* 2000; 124:795–804. [PubMed: 11027727]
34. Nikawa J, Yamashita S. Phosphatidylinositol synthase from yeast. *Biochim Biophys Acta.* 1997; 1348:173–178. [PubMed: 9370330]
35. WEISS SB, SMITH SW, Kennedy EP. The enzymatic formation of lecithin from cytidine diphosphate choline and D-1,2-diglyceride. *J Biol Chem.* 1958; 231:53–64. [PubMed: 13538947]
36. Gibellini F, Smith TK. The Kennedy pathway--De novo synthesis of phosphatidylethanolamine and phosphatidylcholine. *IUBMB Life.* 2010; 62:414–428. [PubMed: 20503434]
37. Kyte J, Doolittle RF. A simple method for displaying the hydropathic character of a protein. *J Mol Biol.* 1982; 157:105–132. [PubMed: 7108955]
38. Lomize MA, Pogozheva ID, Joo H, Mosberg HI, Lomize AL. OPM database and PPM web server: resources for positioning of proteins in membranes. *Nucleic Acids Research.* 2011; 40:D370–D376. [PubMed: 21890895]
39. Nishijima M, Raetz CR. Membrane lipid biogenesis in Escherichia coli: identification of genetic loci for phosphatidylglycerophosphate synthetase and construction of mutants lacking phosphatidylglycerol. *J Biol Chem.* 1979; 254:7837–7844. [PubMed: 381294]
40. Mancía F, Love J. High throughput platforms for structural genomics of integral membrane proteins. *Current Opinion in Structural Biology.* 2011; 21:517–522. [PubMed: 21807498]
41. Williams JG, McMaster CR. Scanning alanine mutagenesis of the CDP-alcohol phosphotransferase motif of Saccharomyces cerevisiae cholinephosphotransferase. *J Biol Chem.* 1998; 273:13482–13487. [PubMed: 9593682]
42. Kabsch W. XDS. *Acta Crystallogr D Biol Crystallogr.* 2010; 66:125–132. [PubMed: 20124692]
43. Sheldrick GM. Experimental phasing with SHELXC/D/E: combining chain tracing with density modification. *Acta Crystallogr D Biol Crystallogr.* 2010; 66:479–485. [PubMed: 20383001]
44. Pape T, Schneider TR. HKL2MAP: a graphical user interface for macromolecular phasing with SHELX programs. *J Appl Cryst.* 2004; 37:843–844. 1–2.10.1107/S0021889804018047
45. Vonrhein C, Blanc E, Roversi P, Bricogne G. Automated structure solution with autoSHARP. *Methods Mol Biol.* 2007; 364:215–230. [PubMed: 17172768]
46. Cowtan K, Main P. Miscellaneous algorithms for density modification. *Acta Crystallogr Sect F Struct Biol Cryst Commun.* 1998; 54:487–493.

47. Cowtan K. The Buccaneer software for automated model building. 1. Tracing protein chains. *Acta Crystallogr Sect F Struct Biol Cryst Commun.* 2006; 62:1002–1011.
48. Emsley P, Lohkamp B, Scott WG, Cowtan K. Features and development of Coot. *Acta Crystallogr D Biol Crystallogr.* 2010; 66:486–501. [PubMed: 20383002]
49. Adams PD, et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr.* 2010; 66:213–221. [PubMed: 20124702]
50. Chen VB, et al. MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr D Biol Crystallogr.* 2010; 66:12–21. [PubMed: 20057044]
51. Pettersen EF, et al. UCSF Chimera--a visualization system for exploratory research and analysis. *J Comput Chem.* 2004; 25:1605–1612. [PubMed: 15264254]
52. Pei J, Kim BH, Grishin NV. PROMALS3D: a tool for multiple protein sequence and structure alignments. *Nucleic Acids Research.* 2008; 36:2295–2300. [PubMed: 18287115]
53. Larkin MA, et al. Clustal W and Clustal X version 2.0. *Bioinformatics.* 2007; 23:2947–2948. [PubMed: 17846036]
54. Sievers F, et al. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol Syst Biol.* 2011; 7:539. [PubMed: 21988835]
55. Makarenkov V. T-REX: reconstructing and visualizing phylogenetic trees and reticulation networks. *Bioinformatics.* 2001; 17:664–668. [PubMed: 11448889]

**Fig. 1.**

The structure of AF2299 reveals the architecture of a CDP-alcohol phosphotransferase. **(a)** The CDP-AP signature motif is absolutely conserved across all kingdoms of life. Alignment of eight CDP-APs of divergent sequence shows that the CDP-AP signature motif ($D_1xxD_2G_1xxAR\dots G_2xxxD_3xxxD_4$) is absolutely conserved, and is present in AF2299. Alignment performed using PROMALS3D⁵². Sequences, from top, (as species - protein; Uniprot ID) are: *A. fulgidus* - AF2299; O27985, *E. coli* - PGP synthase PgsA; P0ABF8, *A. thaliana* – choline/ethanolamine phosphotransferase; O82567, *H. sapiens* -choline/ethanolamine phosphotransferase; Q9Y6K0, *H. sapiens* -choline phosphotransferase; Q8WUD6, *S. cerevisiae* – phosphatidylinositol synthase; P06197, *A. thaliana* – phosphatidylinositol synthase; Q8LBA6, *H. sapiens* – phosphatidylinositol synthase; O14735. **(b)** CDP-alcohol phosphotransferases catalyse the transfer of a phosphate group, with donor substituent R1 (where R1 may represent choline, ethanolamine, or diacylglycerol, amongst others) to an alcohol, with substituent R2 (where R2 may represent inositol, inositol-1-phosphate, glycerol-3-phosphate choline or diacylglycerol, for example). **(c)** Each subunit of AF2299 possesses six TM helices arranged in two inverted repeats of three. TM5 is noticeably shorter than the rest. Intracellular loops are represented by blue lines and extracellular loops by black lines. The juxtamembrane helix and N-terminal domain are not shown. **(d)** AF2299 is a dimer of two subunits each with six TM helices and an N-terminal domain of the cytidyltransferase fold. The protein is depicted with one subunit in gray and the other in rainbow colored ribbon representation, from blue (N-terminus) to red (C-terminus), viewed (on the left) in the plane of the membrane, and (on the right) from the extracellular side of the membrane down the dimer interface.

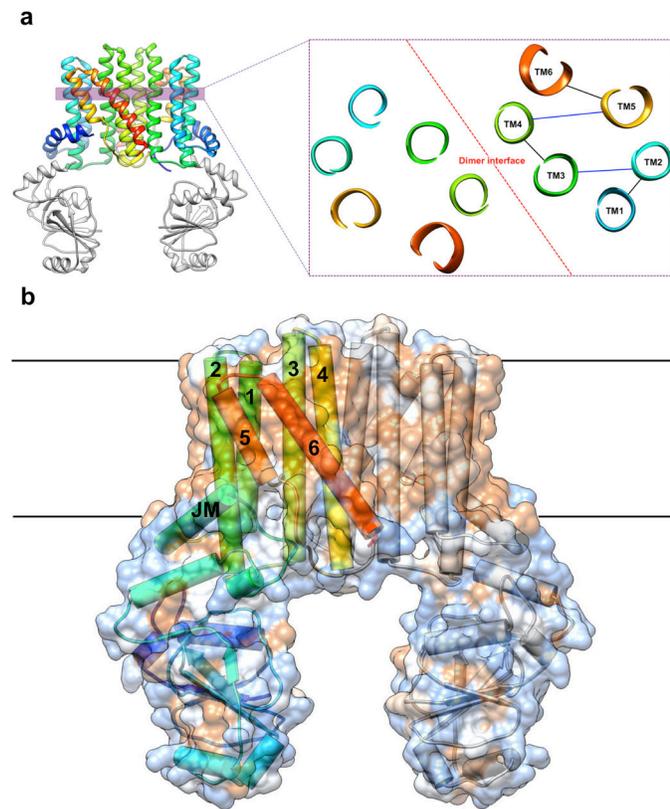


Fig. 2. Architecture of the AF2299 dimer. **(a)** Two inverted repeats of three helices (TM1–3 and TM4–6) surround a central cavity in each monomer. The dimer interface is formed by TM 3 and 4, and buries an area of 1963 \AA^2 . Here, on the left, a dimer is represented with the N-terminal domains in gray and the TM subunits in rainbow ribbon representation, from blue (N-terminal) to red (C-terminal). A 4 \AA thick slab (purple shading) is withdrawn from the protein, and on the right, viewed from the extracellular side of the membrane down the dimer interface (red dashed line). The helices are labeled TM1–TM6, and connections between them are represented either in black (for extracellular loops) or blue (for intracellular loops). **(b)** All TM helices except TM5 extend beyond the intracellular border of the membrane. Here, the TM helices of AF2299 are represented as cylinders, and the secondary structure elements of the protomer on the left are colored in rainbow representation, from blue (N-terminus) to red (N-terminus). The TM helices are labeled from 1 to 6. The dimer is represented inside a translucent molecular surface colored by the Kyte-Doolittle hydrophobicity scale from -4.5 (most polar, light blue) to 4.5 (most hydrophobic, orange). The approximate boundaries of the membrane are represented by black lines.

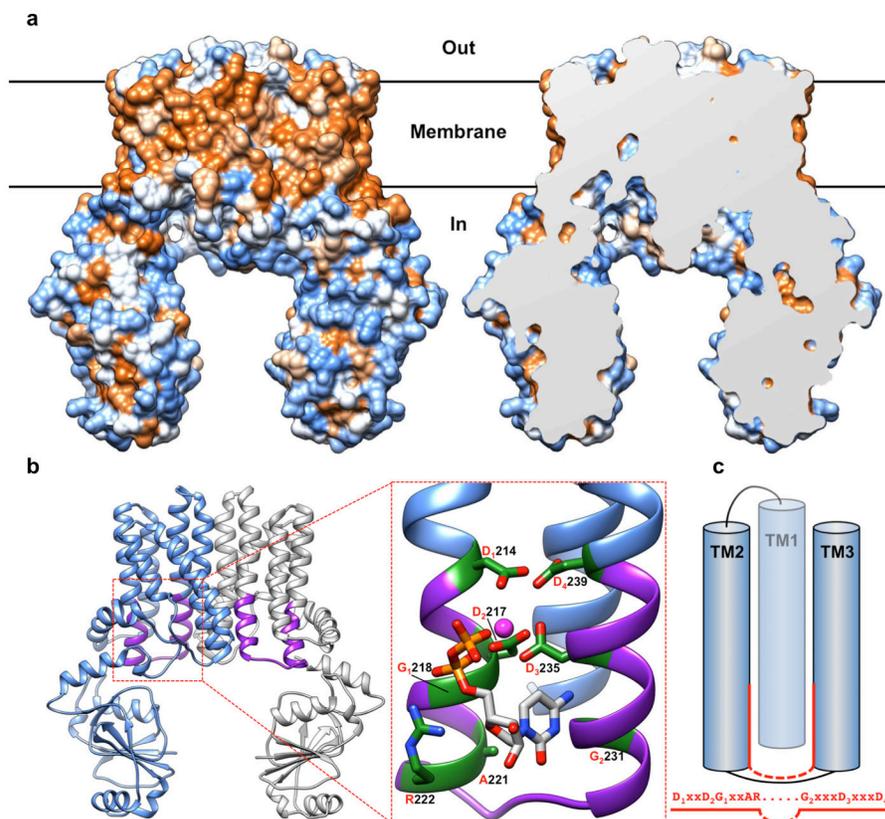


Fig. 3.

The signature motif of the CDP-alcohol phosphotransferase family is located around a cleft between TM helices 2 and 3. **(a)** Each subunit possesses a polar cavity near the surface of the membrane. Here, the molecular surface of AF2299 is colored by Kyte-Doolittle hydrophobicity³⁷ from -4.5 (most polar, light blue) to 4.5 (most hydrophobic, orange). On the right, the front half of the dimer has been removed to expose the polar cavities in each subunit that penetrate 8 \AA beyond the intracellular border of the membrane. The approximate borders of the membrane, as calculated by the PPM server³⁸, are indicated by black lines. **(b)** The CDP-alcohol phosphotransferase signature motif spans the cytosolic ends of TM2 and TM3 (highlighted here in purple; the conserved residues of the signature motif are shown in green stick representation in a zoom-in panel on the right, as is the CDP (in gray stick) and the primary divalent cation (as a pink sphere). **(c)** TM 2 and 3 (blue cylinders) form a groove, backed by TM1, lined by the residues of the signature motif (solid red line; the non-conserved segment between the two helices is indicated by a dashed red line).

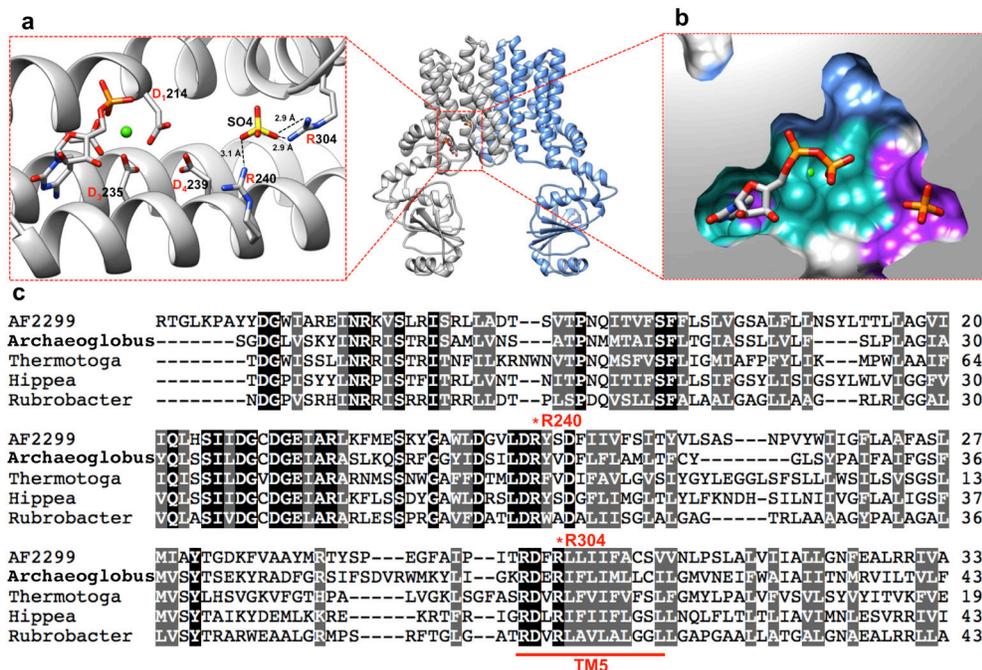


Fig. 4.

Two binding pockets within the cavity form putative binding sites for the donor substituent and the acceptor alcohol. **(a)** The putative catalytic aspartate, D239, lies approximately equidistant between the bound CMP and a bound sulfate ion, located between two arginine residues, R240 and R304. **(b)** Family-specific conservation patterns delineate three regions within the cavity. Residues conserved in all members of the CDP-alcohol phosphotransferase family are colored in cyan, those conserved only in enzymes that utilize inositol-1-phosphate as an acceptor are colored in purple, and those that are conserved in all enzymes known to bind CDP-inositol or CDP-glycerol in light blue. The bound sulfate and CDP are shown in stick representation. **(c)** An alignment of AF2299 with archaeal di-inositol phosphate synthases (DIPPS) shows several highly conserved regions outside the signature motif; notably, the two arginines that coordinate the bound sulfate, R240 and R304 (indicated in red), are absolutely conserved.

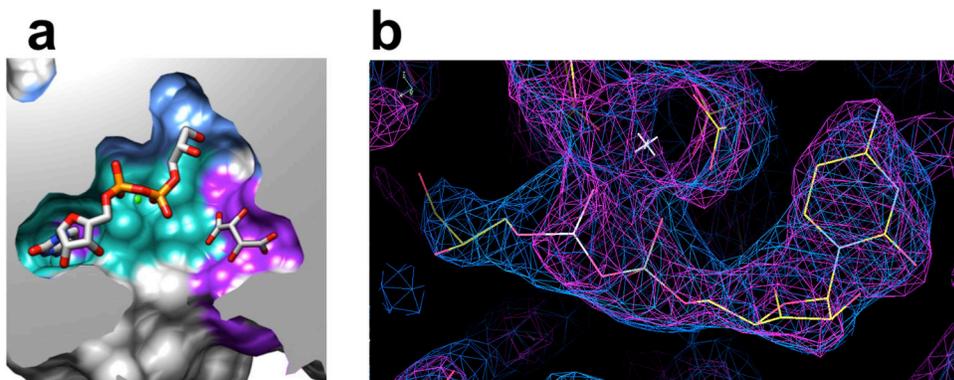


Fig. 5.

(a) A co-crystal structure with CDP-glycerol reveals that the glycerol moiety binds in the putative donor-substituent binding pocket (Colors as for Fig. 4). **(b)** $2F_o-F_c$ omit maps (calculated prior to ligand placement) contoured at 1σ for the CDP and CDP-glycerol structures are represented as purple and blue mesh, respectively. The CDP-glycerol ligand from the final refined structure is depicted in stick representation.

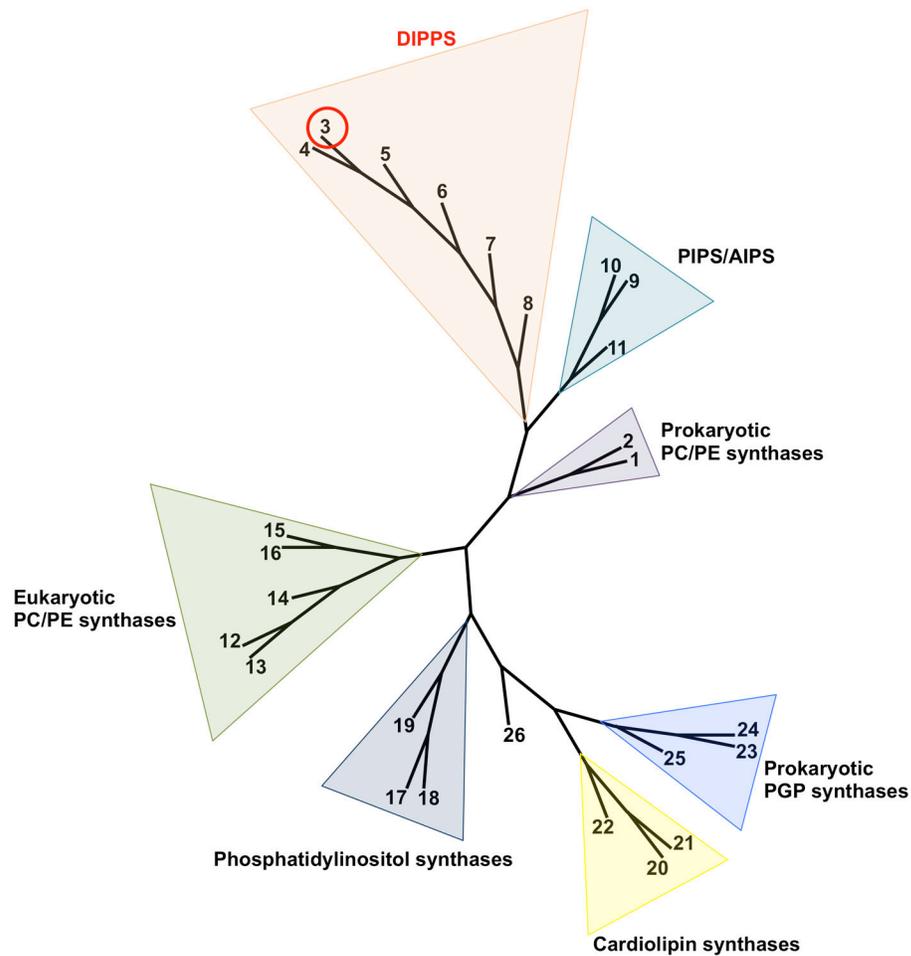


Fig. 6.

AF2299 is most closely related in sequence to the di-inositol-phosphate synthases. AF2299 (3, highlighted by a red circle), is most similar in sequence to the previously characterized di-inositol phosphate synthases (4–8). Here, an unrooted phylogenetic tree, generated using CLUSTALW-PHYLOGENY⁵³ from a multiple-alignment of 26 sequences (performed using CLUSTAL-OMEGA⁵⁴), was visualized in radial representation using T-REX⁵⁵, and manually edited for clarity. The sequences cluster into distinct families of related function (shaded triangles), and AF2299 is most closely related to those that have been functionally characterized as di-inositol phosphate synthases (orange shaded triangle). The species and Uniprot IDs of the sequences used to generate the tree are as follows: 1: Q9KJY8; *Rhizobium meliloti*. 2: Q98MN3; *Rhizobium loti*. 3: O27985; *Archeoglobus fulgidus* (**AF2299**). 4: O29976; *Archeoglobus fulgidus*. 5: Q5JDA9; *Thermococcus kodakaraensis*. 6: Q8U1Z6; *Pyrococcus furiosus*. 7: O67379; *Aquafex aeolicus*. 8: Q1AWQ0; *Rubribacter xylanophilus*. 9: O27726; *Methanothermobacter thermautotrophicus*. 10: Q9F7Y9; *Mycobacterium smegmatis*. 11: Q7D6W6; *Mycobacterium tuberculosis*. 12: Q8WUD6; *Homo sapiens*. 13: Q5ZKD1; *Gallus gallus*. 14: Q9Y6K0; *Homo sapiens*. 15: P17898; *Saccharomyces cerevisiae*. 16: O82567; *Arabidopsis thaliana*. 17: Q8LBA6; *Arabidopsis thaliana*. 18: P06197; *Saccharomyces cerevisiae*. 19: O14735; *Homo sapiens*. 20: Q8MZC4;

Drosophila melanogaster. 21: Q9UJA2; *Homo sapiens*. 22: O01916; *Caenorhabditis elegans*. 23: Q68XS5; *Rickettsia typhi*. 24: P0ABF8; *Escherichia coli*. 25: P44528; *Haemophilus influenza*. 26: P63753; *Mycobacterium tuberculosis*.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

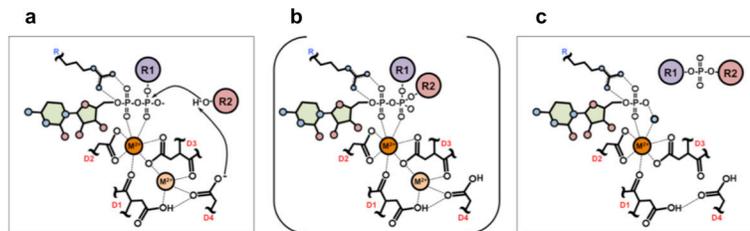


Fig. 7. CDP-alcohol phosphotransferases effect phosphotransfer by a base-catalysed mechanism. **(a)** The fourth aspartate in the signature motif, D4, deprotonates the acceptor alcohol, (R2, red), activating it for nucleophilic attack upon the beta-phosphorus of the CDP (the pyrimidine ring of which is shown in pale green), to which the donor substituent (R1, purple) **(b)** Here, the transition state of the reaction is shown assuming an associative mechanism proceeding via a penta-coordinate transition state, to yield the product, **(c)** a phosphodiester linked conjugate of the donor, R1, and the acceptor R2. A water molecule (aqua) replaces the beta-phosphorus in coordinating the primary divalent cation.

Table 1

Data collection and refinement statistics

	CMP/Apo (4O6M)	CDP (4O6N)	CDP-Glycerol (4Q7C)	SeMet
Data collection				
Space group	P 2 ₁	P 2 ₁	P 2 ₁	P 2 ₁
Cell dimensions				
<i>a</i> , <i>b</i> , <i>c</i> (Å)	46.36, 91.15, 107.02	46.33, 90.67, 106.44	46.42, 90.85, 106.61	46.35, 91.11, 106.92
β (°)	92.5	92.07	92.2	91.88
Wavelength (Å)	1.024	0.979	0.979	0.979
Resolution (Å)	46.33-1.90(1.97-1.90)*	46.30-2.10(2.18-2.10)*	69.13-3.10(3.32-3.10)*	69.33-2.60(2.72-2.60)*
<i>R</i> _{merge}	0.071 (1.72)	0.124 (1.62)	0.183 (0.780)	0.174 (1.018)
<i>I</i> /σ <i>I</i>	11.73 (0.72)	7.43 (0.81)	8.1 (2.0)	10.8 (2.2)
Completeness (%)	99.29 (99.15)	98.08 (98.08)	99.8 (99.8)	99.8 (100)
Redundancy	3.8 (3.8)	3.4 (3.5)	4.2 (4.2)	7.4 (7.6)
Resolution (<i>I</i> /σ <i>I</i> =2)	2.15	2.35	3.21	2.60
Refinement				
Resolution (Å)	46.33-1.90(1.97-1.90)	46.30-2.10(2.18-2.10)	69.13-3.10(3.21-3.10)	-
No. unique reflections	69560 (6905)	50389 (5000)	16102 (1562)	-
<i>R</i> _{work} / <i>R</i> _{free}	0.198 (0.343)/0.224 (0.360)	0.226 (0.351)/0.262 (0.379)	0.224 (0.313)/0.274 (0.347)	-
No. non-hydrogen atoms	5630	5584	5412	-
Protein	5257	5282	5272	-
Ligand/ion	209	158	140	-
Water	164	144	0	-
<i>B</i> -factors	47.3	46.8	41.1	-
Protein	47.0	46.7	41.3	-
Ligand/ion	55.3	47.6	33.6	-
Water	47.2	47.8	-	-
R.m.s. deviations				-
Bond lengths (Å)	0.005	0.003	0.004	-
Bond angles (°)	0.86	0.62	0.77	-

* Values in parentheses are for highest-resolution shell