nature communications

Article

https://doi.org/10.1038/s41467-025-65968-6

Structural basis of undecaprenyl phosphate glycosylation leading to polymyxin resistance in Gram-negative bacteria

Received: 1 February 2025

Accepted: 24 October 2025

Published online: 09 December 2025

Check for updates

Khuram U. Ashraf[1,2], Mariana Bunoro-Batista[3,8], T. Bertie Ansell[4,6,8], Ankita Punetha[1,2,8], Stephannie Rosario-Garrido[1,2], Emre Firlar[5,7], Jason T. Kaelber[5], Phillip J. Stansfeld[3] & Vasileios I. Petrou[1,2]

In Gram-negative bacteria, the enzymatic modification of Lipid A with aminoarabinose (L-Ara4N) leads to resistance against polymyxin antibiotics and cationic antimicrobial peptides. ArnC, an integral membrane glycosyltransferase, attaches a formylated form of aminoarabinose to the lipid undecaprenyl phosphate, enabling its association with the bacterial inner membrane. Here, we present cryo-electron microscopy structures of ArnC from *S. enterica* in *apo* and nucleotide-bound conformations. These structures reveal a conformational transition that takes place upon binding of the partial donor substrate. Using coarse-grained and atomistic simulations, we provide insights into substrate coordination before and during catalysis, and we propose a catalytic mechanism that may operate on all similar metal-dependent polyprenyl phosphate glycosyltransferases. The reported structures provide a new target for drug design aiming to combat polymyxin resistance.

Antibiotic resistance is a growing global public health threat that endangers our ability to treat common infections[1]. In 2019, the Centers for Disease Control and Prevention (CDC) estimated that more than 2.8 million antibiotic resistant infections occur in the United States every year, leading to >35,000 deaths[2]. The situation is particularly critical with the emergence of multidrug resistant (MDR) Gram-negative (GN) strains of pathogens, including *Klebsiella pneumoniae*, *Acinetobacter baumanii*, *Pseudomonas aeruginosa*, *Salmonella enterica*, and *E. coli*, which are responsible for MDR nosocomial and non-nosocomial infections[3,4].

Polymyxins are currently the last line of defense against MDR Gram-negative bacterial infections[5–7]. Polymyxins are non-ribosomal lipopeptides that are polycationic at physiological pH[8]. They act by permeabilizing bacterial membranes[9], although alternative mechanisms like lipid exchange have been proposed[10,11]. In all cases, the necessary initial step is the binding of polymyxins to the lipopolysaccharide (LPS) of the outer membrane[12] (Supplementary Fig. 1). This molecular association is achieved through electrostatic interactions between the cationic amino groups of polymyxins and anionic phosphate groups of Lipid A, which is the amphipathic saccharolipid that anchors LPS to the membrane[12]. Resistance to polymyxins can emerge spontaneously in vitro[13–15] and has been observed in patients in case of suboptimal use[16,17]. The resistance is mainly acquired through enzymatic modifications of Lipid A that cap the glucosamine sugar phosphates, resulting in a reduction of the negative charge in bacterial outer membranes[18]. This mechanism also represents an evasion tactic of GN bacteria against natural antimicrobial peptides (AMPs), including those generated by the innate immune system[19,20]. Clinical and non-clinical isolates that exhibit resistance to polymyxins often carry chromosomal mutations in

[1]Department of Microbiology, Biochemistry, and Molecular Genetics, New Jersey Medical School, Rutgers Biomedical Health Sciences, Newark, NJ, USA. [2]Center for Immunity and Inflammation, New Jersey Medical School, Rutgers Biomedical Health Sciences, Newark, NJ, USA. [3]School of Life Sciences and Department of Chemistry, University of Warwick, Coventry, UK. [4]Department of Biochemistry, University of Oxford, Oxford, UK. [5]Rutgers Cryo-EM & Nanoimaging Facility, Rutgers, the State University of New Jersey, Piscataway, NJ, USA. [6]Present address: Division of Cryo-EM and Bioimaging, SLAC National Accelerator Laboratory, Menlo Park, CA, USA. [7]Present address: Bristol Myers Squibb, Discovery and Development Sciences, Lawrenceville, NJ, USA. [8]These authors contributed equally: Mariana Bunoro-Batista, T. Bertie Ansell, Ankita Punetha. e-mail: vasileios.petrou@rutgers.edu

control elements that allow upregulation of Lipid A-modifying enzymes[21].

In *E. coli* and *Salmonella enterica*, the most effective modification for reduction of negative membrane charge and development of polymyxin resistance is the attachment of an **aminoarabinose** sugar moiety (4-amino-4-deoxy-L-arabinose or L-Ara4N) to the 1 and 4′ phosphate groups of Lipid A[22,23]. This modification is also critical for development of colistin resistance in *P. aeruginosa*[24], and *K. pneumoniae*[25], both included in the ESKAPE pathogen list[26]. L-Ara4N is synthesized by the aminoarabinose biosynthetic pathway, a relay system of eight proteins that act in sequence to synthesize the sugar, associate it with the membrane and transport it to the periplasmic side of bacterial inner membranes[18]. The integral membrane glycosyltransferase ArnC enables the association of the L-Ara4N sugar with bacterial membranes by attaching it to undecaprenyl phosphate (UndP)[27]. ArnC (originally called PmrF) was first identified in *Salmonella* as the product of a gene that disrupted polymyxin resistance when randomly targeted via transposon mutagenesis[28]. It was later demonstrated that ArnC is the glycosyltransferase responsible for attaching the aminoarabinose sugar to the UndP lipid[27]. Only the formylated substrate UDP-L-Ara4FN, produced by the N-terminal domain of ArnA[29,30], can be converted into the glycolipid product (UndP-L-Ara4FN), whereas the unmodified substrate (UDP-L-Ara4N) cannot be processed[27] (Supplementary Fig. 1). ArnC was proposed to resemble dolichol-phosphate mannosyltransferases based on sequence analysis[28,31]. The product of ArnC is deformylated by ArnD[32,33], flipped to the outer leaflet of inner bacterial membranes by the heteromeric flippase ArnE/F[34], and finally, processed by the glycosyltransferase ArnT, which transfers the L-Ara4N sugar to Lipid A, enabling resistance to polymyxins and AMPs[22,23,35].

Glycosyltransferases are classified into families based on amino acid sequence similarity, with 138 distinct families identified to date[36,37]. Polyprenyl phosphate glycosyltransferases (Pren-P GTs) are a subclass of membrane-bound glycosyltransferases that catalyze the transfer of glycosyl groups from activated sugar donors to a polyprenyl lipid acceptor. This acceptor is then transported across the membrane for use in glycosylation reactions[38,39]. The polyprenyl lipid carrier used by Pren-P GTs varies between organisms, with dolichol phosphate (DolP) used in eukaryotes and archaea, and undecaprenyl phosphate (UndP) used in Gram-negative bacteria[40]. The structure of Pren-P GTs consists of two domains: a cytosolic GT-A-like catalytic domain and a transmembrane domain that anchors the enzyme to the lipid bilayer[39]. The GT-A fold consists of an open twisted β-sheet surrounded by α-helices on both sides, and is reminiscent of two adjoined Rossmann-like folds, typical of nucleotide-binding proteins[41,42]. GT-A enzymes possess a DXD signature in which carboxylates coordinate a divalent cation and/or a ribose[41,42]. Thus far, only two Pren-P GTs have been structurally characterized using X-ray crystallography. GtrB from *Synechocystis sp*. was solved in the presence of nucleotide donor substrate[43] (PDB code 5EKP), and the dolichol phosphate mannose synthase (DPMS) from *Pyrococcus furiosus* was crystallized in complex with nucleotide, complete donor substrate, and glycolipid product[44] (PDB codes 5MLZ, 5MM0, 5MM1). Despite the availability of crystal structures for two Pren-P GTs, the catalytic mechanism employed by Pren-P GTs remains poorly understood[39].

Here we report structures of the Pren-P GT ArnC from the Gram-negative bacterium *Salmonella enterica serovar Typhimurium LT2*, determined by single-particle cryo-electron microscopy (cryo-EM), in its *apo* and partial donor substrate-bound states. By combining structural information with molecular dynamics (MD) simulations, we provide a rationale for substrate binding and propose a hypothesis for the reaction mechanism. Our findings illuminate the structural basis for the catalytic activity of Pren-P GTs and pave the way for developing novel therapeutics targeting polymyxin resistance.

Results

Structure determination of ArnC from S. enterica

To identify a suitable candidate for structure determination of ArnC, we cloned *arnC* genes from the clinically-relevant species *S. enterica*, *K. pneumoniae*, and *E. coli* with either an N-terminal FLAG-10xHis-TEV cassette or a C-terminal TEV-10xHis cassette using pNYCOMPS vectors[45]. Cloned ArnC orthologs were expressed in an *E. coli* BL21 derivative strain (T7 express lysY/NEB) and extracted with n-Dodecyl-β-D-Maltoside (DDM). Expression was first evaluated in small scale after affinity purification with nickel-affinity resin (Supplementary Fig. 2a). ArnC from *Salmonella enterica* ($ArnC_{Se}$) was identified as the most promising candidate, due to the high yield of purified $ArnC_{Se}$ protein and good stability in detergent. After solubilization with DDM and initial affinity purification, $ArnC_{Se}$ was reconstituted into lipid-filled nanodiscs for structure determination by cryo-EM. Nanodiscs have been shown to enhance the stability of membrane proteins and provide a more native-like environment for structure determination[46,47]. Based on the resulting size-exclusion chromatography (SEC) profiles, POPC and POPG lipids produced monodisperse peaks in MSP1E3D1 nanodiscs (Supplementary Fig. 2c). POPG was chosen as the optimal lipid for the final $ArnC_{Se}$ sample (Supplementary Fig. 2d,e). $ArnC_{Se}$ was also extracted with Cubipol copolymer and cross-linked with glutaraldehyde to assess its oligomeric state under native membrane conditions (Supplementary Fig. 2b). Using this treatment, crosslinked species corresponding to a dimer and a tetramer were observed, with the tetrameric species increasing in intensity with increasing treatment.

For structure determination of the substrate-free (*apo*) state, $ArnC_{Se}$ embedded in nanodiscs was plunge-frozen in liquid ethane using UltrAuFoil gold grids as support foil to reduce specimen motion[48]. After initial screening, a dataset of 5552 micrographs was collected using a Talos Arctica cryo-electron microscope equipped with a Gatan K2 direct detector. Processing of this dataset using the workflow described in the methods section resulted in a 2.79 Å reconstruction from ~184 thousand particles after imposing C4 symmetry (Supplementary Fig. 3, Supplementary Fig. 11, and Supplementary Table 1). In addition, a second *apo* $ArnC_{Se}$ dataset of 23,259 micrographs was collected at the National Center for Cryo-EM Access and Training (NCCAT) in New York using a Titan Krios microscope equipped with a Gatan K3 direct detector and a duplicate grid from the same sample preparation. An initial set of over 4 million particles was cleaned up to a final set of ~490 thousand particles, leading to a 2.74 Å reconstruction after imposing C4 symmetry (Supplementary Fig. 4, Supplementary Fig. 11, and Supplementary Table 1). Reconstructions for the *apo* datasets in C1, without imposing any symmetry, still show clear fourfold symmetry in each density, thus reinforcing a tetrameric arrangement for ArnC (Supplementary Fig. 12a). A direct comparison of the two datasets is discussed in a later section. To generate the initial model, we built into the 2.79 Å substrate-free map reconstructed from the Arctica dataset starting from an AlphaFold generated model of the *E. coli* (K12 strain) ArnC (Uniprot: P77757) from the AlphaFold database[49]. The final model after refinement showed an RMSD of 9.53 Å across 312 Cα atom pairs with the current AlphaFold model of *E. coli* (K12 strain) ArnC in the AlphaFold database (AF-P77757-F1-model_v4) (Supplementary Fig. 5a). The atomic model of $ArnC_{Se}$ was also refined with the Krios microscope map leading to second model that has 0.67 Å RMSD across 1248 Cα atom pairs with the Arctica model. A superposition of the two apo $ArnC_{Se}$ models is shown in Supplementary Fig. 5b.

The structure of *apo* ArnC

The structure shows that ArnC is a tetramer (Fig. 1a, b), with each protomer consisting of an N-terminal cytosolic glycosyltransferase (GT) domain; two juxtamembrane (JM) helices (JM1 and JM2) located adjacent to the transmembrane domain (TMD); two transmembrane

https://doi.org/10.1038/s41467-025-65968-6

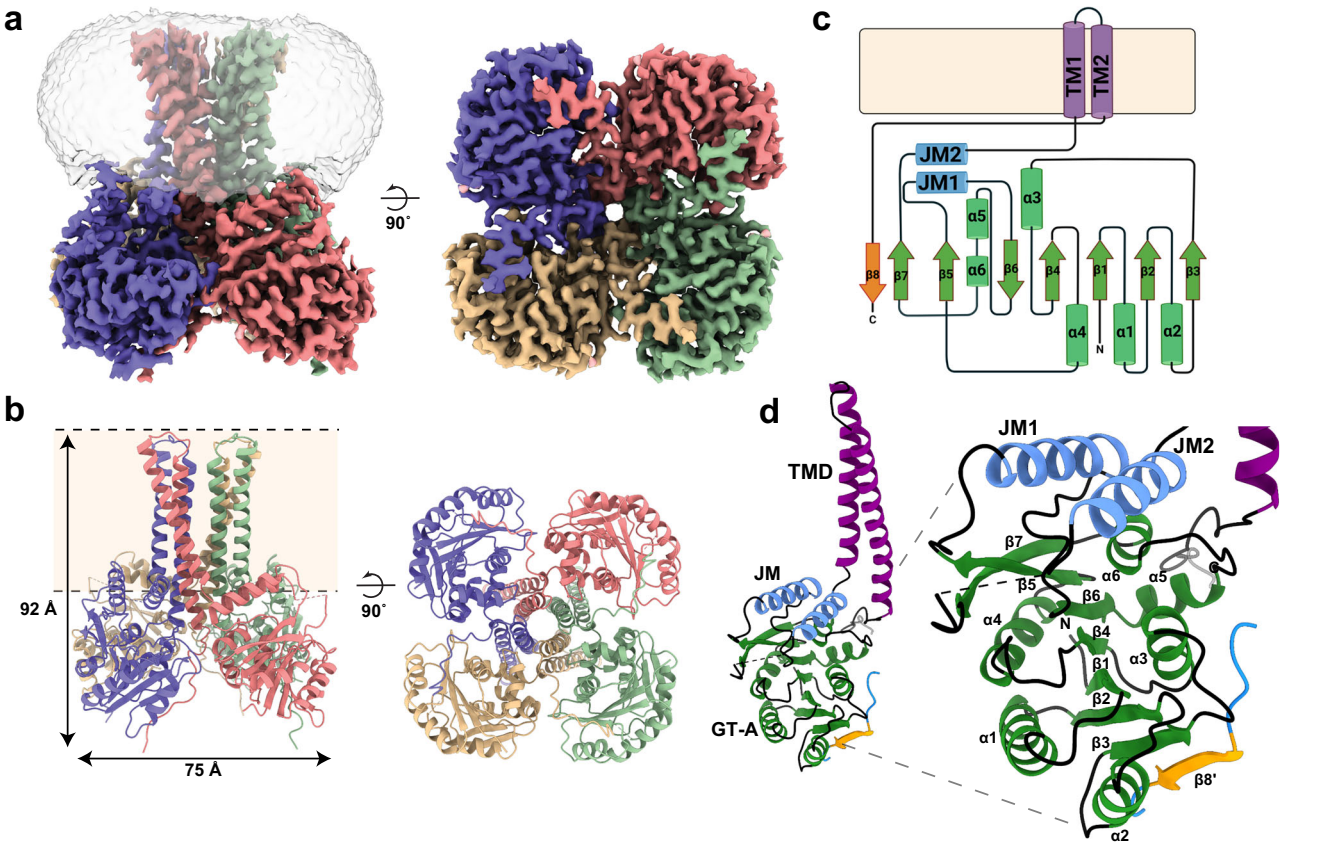

Fig. 1 | The cryo-EM structure of ArnC from *S. enterica*. **a** Cryo-EM density map of the ArnC tetramer from the Krios dataset in two orthogonal views. Coloring is per protomer, and the semi-transparent volume represents the nanodisc density observed in low contour levels of the reconstruction. **b** The cryo-EM structure of ArnC in ribbon representation. Approximate membrane boundaries are represented with dotted lines. The coloring is per protomer, and two orthogonal views are shown (side and bottom views). **c** Schematic representation showing the topology of each ArnC protomer, consisting of two transmembrane helices, linked to two juxtamembrane helices (JM1 and JM2), and a large soluble domain consisting of six α-helices and eight β-sheets in a classic GT-A like fold. The C-terminal β8 strand packs with β-strands of the adjacent protomer. Panel (c) was created in Biorender. **d** Ribbon representation of one ArnC protomer with the GT-A fold shown in green, and the two juxtamembrane helices in blue. β8′ strand from the adjacent protomer is shown. The numbering of helices and sheets is the same as in (c).

helices (TM1 and TM2), which are fully embedded within the nanodisc and anchor the intracellular domains to the membrane; and a short C-terminal β-hairpin (Fig. 1c, d). The two transmembrane helices create an eight-helix bundle within the TMD of the tetramer. The short β-hairpin at the C-terminus interacts with the adjacent protomer thus bridging two adjacent subunits within the tetramer. The two amphipathic JM helices are parallel to each other and orientated to expose their hydrophobic faces towards the membrane (Fig. 1d, Supplementary Fig. 6c).

The N-terminal glycosyltransferase GT domain features a classic GT-A fold composed of a seven-stranded β-sheet core flanked by α-helices on both sides. An eighth strand (β8) is contributed by the adjacent protomer and packs against the β3 strand. All eight strands are connected via hydrogen bonds, and two of them (β6 and β8) are antiparallel to the rest. Thus, the ArnC GT-A fold can be described as a Rossmann-like fold consisted of a mixed β-sheet of eight strands ordered in the sequence 83214657 (Fig. 1c, d). The conserved signature DXD motif, which is located on the loop between β4 and α4 in ArnC, is known to coordinate the catalytic metal ion and donor substrate in glycosyltransferases that carry a GT-A fold[41]. The ArnC GT-A fold can be further subdivided into a canonical four-stranded Rossmann fold, with four parallel β-strands (β1–β4), two α-helices (α1 and α2) on one side of the sheet, and one α-helix (α3) on the other side, within the N-terminal part of the GT-A fold; and a Rossman-like fold with a three-stranded mixed β-sheet (strands β5–β7) and one α-helix on either side (α4, α5/α6) within the C-terminal part of the GT-A fold (Fig. 1c, d).

The catalytic core of the protein between the JM helices and the GT-A domain forms a central cavity (cavity 1) that is lined with conserved charged and polar residues at the periplasmic surface, and hydrophobic residues within the core (Fig. 2, Supplementary Fig. 6). A cluster of highly conserved residues, which includes a number of conserved arginines and the GT-A DXD motif provide a net positive electrostatic charge within this region (Fig. 2c, Supplementary Fig. 6). No visible density was observed within this pocket, verifying that this structure is in its apo state; however amorphous lipid densities could be observed in the inter-protomer grooves of the TM domain, and in particular within a small cavity (cavity 2) formed at the top of the TM domain (Fig. 2a, b).

Comparison of 200 kV and 300 kV *apo* ArnC datasets

The *apo* ArnC datasets, collected at different sites with different electron microscopes, were refined to 2.74 Å and 2.79 Å, so we further explored resolution-limiting factors in the datasets collected on the Arctica and the Krios microscopes. To understand factors governing resolution from the two *apo* ArnC datasets, we first repeated the final 3D reconstruction with subsets of various sizes, randomly sampling particles present in the full datasets. At 180,000 particles (the largest subset contained in both datasets), ArnC was resolved to 2.81 Å at 300 kV and 2.80 Å at 200 kV. At all particle counts (from 2500 to

https://doi.org/10.1038/s41467-025-65968-6

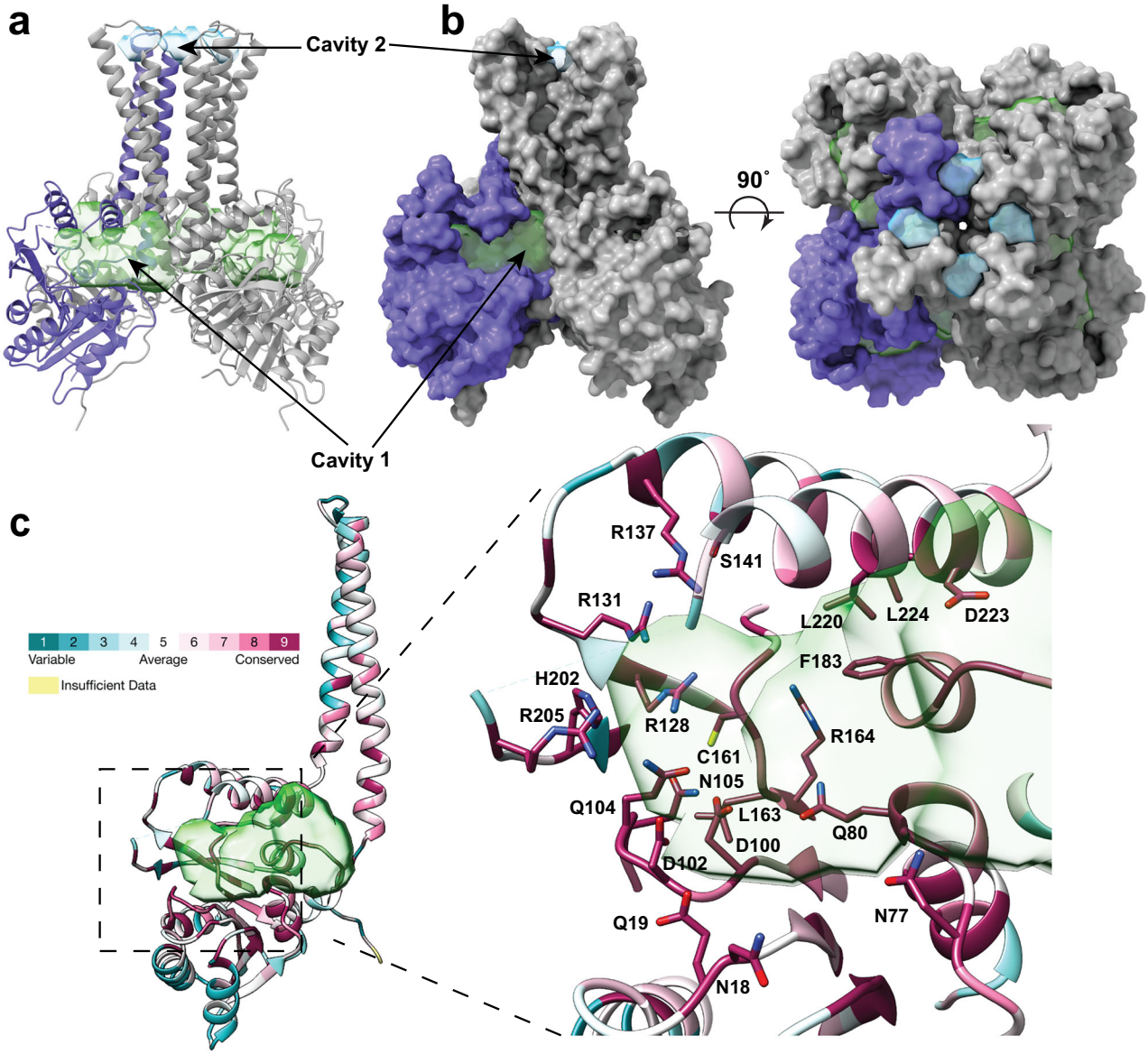

Fig. 2 | Structural features of ArnC. a Ribbon representation of ArnC showing the volumes of notable cavities. ArnC is colored gray with one protomer colored purple. The volumes are shown as semi-transparent surfaces, with cavity 1 colored green and cavity 2 blue. Volumes were calculated using the Voss Volume Voxelator (3V) server, using probes with 10 Å and 2 Å radii, corresponding to the outer and the inner probe, respectively. **b** Spacefill representation of ArnC showing the volumes of notable cavities in two orthogonal views. Coloring is the same as in (**a**). **c** Ribbon representation of one ArnC protomer colored by residue conservation from blue (low to no conservation) to purple (absolute conservation). The highly conserved residues are shown as sticks within the binding pocket of cavity 1.

180,000 particles), the difference in resolution (at FSC = 0.143) was small between datasets acquired at the two sites, averaging 0.04 Å better at 300 kV (Supplementary Fig. 7a,c,d). The most reasonable explanations for this difference include: effect of higher magnification (0.646 Å/px vs. 1.038 Å/px), effect of detector performance (Gatan K3 vs. Gatan K2 direct electron detector), effect of microscope model and voltage (300 kV Titan Krios vs. 200 kV Talos Arctica), effect of the experimenter/site, or other uncontrolled variation between experiments (e.g. intrabatch variation in grid quality, environmental interference, day-to-day variation in etc.). It is possible that an improvement at 300 kV or from another factor was masked by a countervailing variable: for instance, perhaps the higher dose rate used at 300 kV could have worsened data quality to approximately the same extent that the higher magnification improved data quality, resulting in similar-resolution structures. The difference of 0.04 Å resolution did not lead to meaningful differences in the biological interpretation of the macromolecular structure (Supplementary Fig. 5b).

In theory, the number of particles required to reach certain resolution is proportional to $e^{B/2d^2}$, where B is the temperature factor and d is the resolution[50], and B will be improved by superior microscope/detector quality. Varying particle number by reconstructing random subsets allows us to assess the conformance of the actual data to this prediction. In both datasets, resolution improves with increasing particle count in line with theory[50] and the B-factor plot is linear for all particle counts (Supplementary Fig. 7b). Holding particle alignment fixed, the spectral signal-to-noise ratio (SSNR) of the reconstruction is simply the product of the per-particle SSNR with the particle count[51]. Indeed, where the FSC > 0.143, this relationship holds at all particle counts tested (up to 490,000 particles, or 1.96×10^6 asymmetric units, for the Krios dataset); this is demonstrated in the "universal plot" of ln(SSNR/n) vs. spatial frequency[51] by the fact that the lines for each particle count are indistinguishable at resolutions >3 Å (Supplementary Fig. 7e, f). The SSNR has features in the same locations as the radially-averaged structure factor of the reconstructed map, but

convolved with a dampening factor representing resolution-dependent attenuation of information transfer. Extrapolating from the calculated B-factor, reaching 2.0 Å would take one millionfold more particles, whereas such resolution is readily attained under these conditions with 490,000 apoferritin particles. Although ArnC resolution did not reach a resolution cutoff and may continue to improve with particle count as $n \propto e^{B/2d^2}$, the B-factor (which was similar between microscopes/cameras/voltages) implies that resolution would be practically limited to ~2.5 Å unless biochemical or grid conditions were substantively different.

The B-factor and resolution are governed by sample-extrinsic factors such as the accelerating voltage of the imaging system and the detective quantum efficiency, as well as sample-intrinsic factors such as the flexibility of the molecules and the background scattering. In the case of ArnC, we show that (within these two high-quality imaging systems) extrinsic factors such as accelerating voltage had a negligible effect on resolution, but sample-intrinsic factors imposed definitive limits on resolution. This differs from the case of an 85 kDa CDK-activating kinase, where voltage effects were the most likely driver of a 0.3 Å resolution difference[52], and suggests that resolution gains with use of higher microscope voltage will be specimen-dependent.

Comparison with existing Pren-P GT structures

In addition to the structures reported here, the structures of two other Pren-P GTs have been solved previously by X-ray crystallography. In this section we provide a comparison between ArnC and the two existing Pren-P GT structures. GtrB is a glycosyltransferase that catalyzes the transfer of glucose, from UDP-Glucose to the undecaprenyl phosphate (UndP) lipid carrier[43]. The structure of GtrB from *Synechocystis* sp. PCC6803 [43] (PDB code 5EKP) has been determined in a nucleotide bound state at 3.2 Å resolution. Similar to ArnC, GtrB is a tetramer with two TM helices and a classic GT-A domain at the N-terminus (Supplementary Fig. 8a,b). ArnC and GtrB have overall similar topology. Yet, the two TM helices in GtrB assume different positions relative to the catalytic domain compared to ArnC (Supplementary Fig. 8c). The crystal structure shows that GtrB coordinates a UDP nucleotide within a GT-A domain pocket utilizing a magnesium metal to coordinate the phosphate groups of the donor. This pocket corresponds to cavity 1 in ArnC (Fig. 2). GtrB utilizes a DXD motif ($_{94}$DADLQD$_{99}$) to coordinate the magnesium metal using D96 and Q98, which is similar to the DXD motif of ArnC ($_{100}$DADLQN$_{105}$). In addition, the loop containing the DXD motif appears to be in a similar conformation in the two enzymes (Supplementary Fig. 8d,e).

Dolichyl phosphate mannose synthase (DPMS) catalyzes the transfer of mannose from the donor GDP-mannose to dolichol phosphate (DolP) to produce DolP-mannose (DolP-Man)[44]. The structure of DPMS from *Pyrococcus furiosus* (*Pf*DPMS) has been determined by X-ray crystallography in complex with nucleotide GDP, full donor substrate (GDP-Man), and glycolipid product (PDB code 5MLZ, 5MM0 and 5MM1)[44]. Unlike ArnC, *Pf*DPMS is a monomer in the crystal state (or perhaps a dimer in its physiological state), and has four TM helices, arranged as two α-helical hairpins (Supplementary Fig. 9a,b)[44]. Due to the differences of the TM helices, a structural alignment of *Pf*DPMS with an ArnC protomer is only meaningful for the GT-A domain (Supplementary Fig. 9c). When only the GT-A domain of ArnC and *Pf*DPMS are aligned, they superimpose well (14.7 Å RMSD across 388 Cα atom pairs). Based on the *Pf*DPMS structure in the product-bound state (PDB 5MM1), the JM helices enable the threading of the acceptor lipid DolP within the active site of the GT-A domain. A similar function for the JM helices of GtrB was hypothesized based on the GtrB crystal structure[43]. ArnC exhibits similar JM helices that may subserve the same role. This will be examined in a later section. Finally, like GtrB, PfDPMS utilizes a DXD motif ($_{89}$DADLQH$_{94}$) to coordinate a Mg^{2+} or Mn^{2+} metal and the donor nucleotide (Supplementary Fig. 9d,e).

The structure of UDP-bound ArnC

Given the similarity of ArnC to GtrB and *Pf*DPMS, and the presence of a DXD motif, ArnC likely coordinates its nucleotide-sugar donor substrate via a metal ion. Our first goal was to confirm metal binding and identify the optimal metal ion for structural studies of ArnC in a nucleotide-bound state. Thus, we used Microscale Thermophoresis (MST) to characterize the interaction between ArnC and the nucleotide UDP either in the presence of Mg^{2+} or Mn^{2+} metal ions, or in the absence of metal ions (Fig. 3a, b). MST is a biophysical technique that enables quantification of subtle changes in fluorescence in response to ligand binding, while the target protein is subjected to a thermal gradient[53]. UDP was titrated into ArnC in the presence of either 1 mM $MnCl_2$, 1 mM $MgCl_2$, 1 mM EDTA to chelate any endogenously-bound metals, or without any exogenous metal added (Fig. 3b). A clear difference was observed when titrating UDP in the presence of $MnCl_2$ compared to the other three conditions. Calculated Kd values based on the fitted curves were 8.5 μM in the presence of $MnCl_2$, 320 μM in the presence of $MgCl_2$, 8.5 mM in the presence of EDTA, and 591 μM without addition of exogenous metals. Only the curve fitted to the titration in the presence of $MnCl_2$ is a high confidence fit, as judged by the reduced Chi square value. An additional caveat for this experiment is that we were unable to strip the protein of endogenous metals before the MST measurements because extensive treatment with EDTA caused the protein to crash out of solution. Thus, Kd values presented here should only be considered in relative terms rather than absolute, since it is very likely that only a portion of the metal binding sites of ArnC were available to bind the exogenous metals added. Yet, this experiment shows clearly that Mn^{2+} enables higher affinity of ArnC towards the UDP nucleotide, suggesting that it maybe the preferred metal for the ArnC metal-binding site(s). Though, the metal ion used by ArnC in the cellular environment may depend on the relative abundance of metal ions near each metal-binding site.

To prepare UDP-bound ArnC for cryo-EM analysis, 2 mM UDP and 1 mM $MnCl_2$ were added to the purified protein prior to vitrification. 7925 movies of UDP-bound ArnC were collected using a 200 kV Talos Arctica microscope equipped with a Gatan K2 direct detector. Processing of this dataset using the workflow described in the methods section resulted in a 2.96 Å reconstruction from 216 thousand particles after imposing C4 symmetry (Supplementary Fig. 10, Supplementary Fig. 11, and Supplementary Table 1). Analysis of the UDP-bound ArnC map shows an overall fold similar to the *apo* ArnC structure (Fig. 3c,d). Within the binding cavity of the GT-A domain we observed clear non-protein density, which we modeled as UDP coordinated by the Mn^{2+} metal ion (Fig. 3e and Supplementary Fig. 11). A reconstruction in C1, without imposing any symmetry, still shows clear fourfold symmetry in each density, similar to the *apo* datasets (Supplementary Fig. 12b). Moreover, the density observed for the UDP nucleotide in C1, appears to be similar for all four protomers, thus confirming that we have captured a high-occupancy state for the UDP nucleotide (Supplementary Fig. 12c). The UDP nucleotide binds in a shallow cleft on the surface of the GT-A domain, adjacent to the DXD motif, with D102 and Q104 clearly coordinating the Mn^{2+} metal ion, which in turn coordinates the diphosphate of the UDP (Fig. 3e). The UDP molecule is also coordinated via hydrogen bonds with the sidechains of S210, R205, N77, D48 (Supplementary Fig. 13a), and backbone atoms of V16, A101 and K211 (not shown). A comparison of this binding mode with nucleotides bound to the other Pren-P GT structures shows that coordination and localization of the nucleotide within the GT-A domain is similar between ArnC, GtrB and *Pf*DPMS (Supplementary Fig. 13a).

All structures of Pren-P GTs to date have been determined in a nucleotide-bound state. Since we have now obtained structures of a true *apo* state of ArnC and a nucleotide-bound state, we further compared the structures of the two states by aligning them (Fig. 4a). After aligning the two structures on their transmembrane domain (aa

https://doi.org/10.1038/s41467-025-65968-6

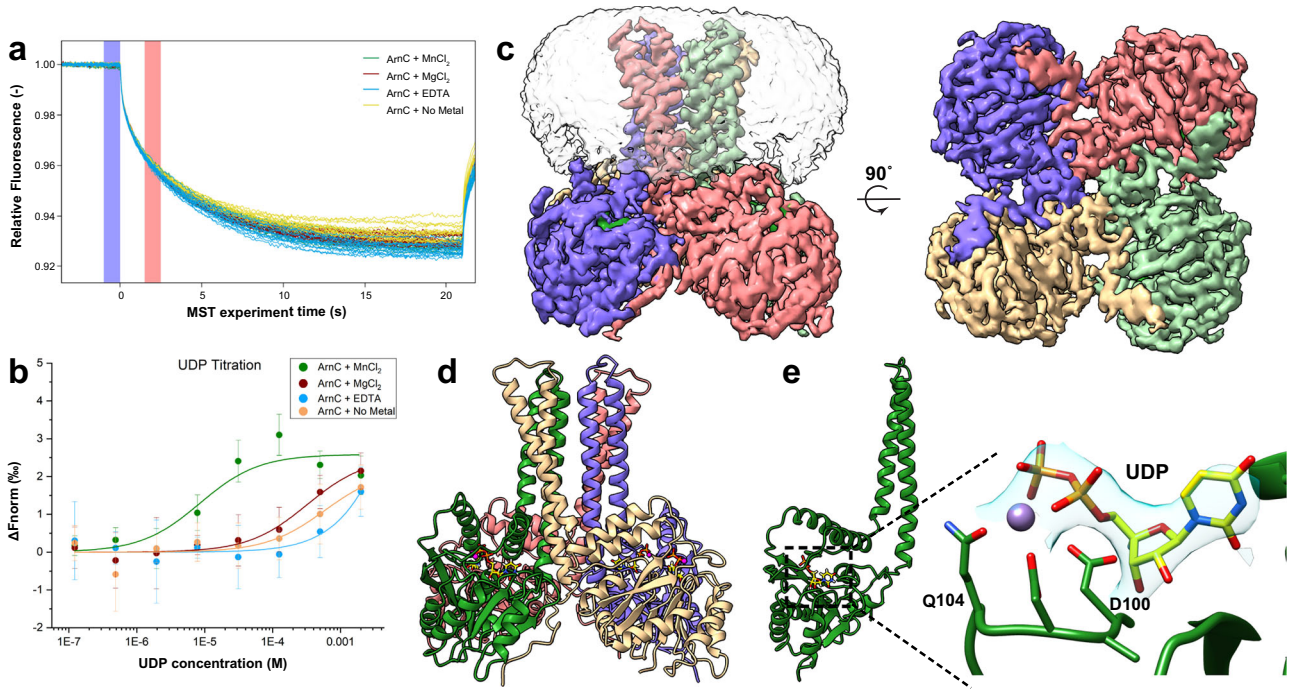

Fig. 3 | The cryo-EM structure of the UDP-bound state of ArnC. a Overlay of representative MST trace for all four experimental groups, showing the cold region (-1 to 0 s) and hot region (1.5 to 2.5 s) used for analysis of the thermophoresis traces. Traces are colored according to the inset; $MnCl_2$ (Green), $MgCl_2$ (Red), No added metal (Yellow) and EDTA (Blue). **b** Scatter plot showing the composite of three independent experiments for the titration of UDP under the four different experimental conditions. The Kd values calculated based on the fitted curves are: 8.5 ± 6.8 μM with 1 mM $MnCl_2$, 323.4 ± 106.1 μM with 1 mM $MgCL_2$, 8.6 ± 32.7 mM with 1 mM EDTA, and 591.5 ± 577 μM with no exogenous metal added. Kd values are listed with a 68% confidence interval. The presence of Mn^{2+} leads to a pronounced increase in the binding affinity of ArnC for the UDP ligand. Data points represent mean ± standard deviation (SD), $n = 6$ for each group. **c** Cryo-EM density map of the ArnC tetramer in the presence of Mn^{2+} and UDP in two orthogonal views. Coloring is per protomer, and the semi-transparent volume represents the nanodisc density observed in low contour levels of the reconstruction. **d** The cryo-EM structure of ArnC bound to Mn^{2+} and UDP. The coloring is per protomer, with UDP in yellow, the phosphates of UDP in orange, and the Mn^{2+} ion represented as a purple sphere. **e** A single protomer of ArnC bound to Mn^{2+} and UDP is shown. The magnified detail shows the characteristic $_{100}$DXD$_{102}$ motif found within the GT-A domain that facilitates coordination of the metal and UDP nucleotide. The cryo-EM density around UDP is shown as a semi-transparent surface. Coloring is the same as in (**d**).

230–305), we detected a significant conformational transition that occurs when UDP binds to its binding cleft (Fig. 4a, b). The overall fold of the GT-A domain remains largely unchanged between the two states, but in the UDP-bound state the GT-A domain comes closer to the JM helices (and the membrane). There is some movement at the JM helices (more in JM1 than JM2), but the major conformational changes occur in the GT-A domain, which moves like a "pendulum" upwards (Fig. 4b). The GT-A domain rotates upwards by 13.7° and different parts of the domain undergo a translation ranging from 2.7 to 7.7 Å. Moreover, the linker between strand β7 and JM2 becomes more ordered and can be fully modelled in the UDP-bound structure (aa 207–211 are missing from the *apo* structure). The conformational transition observed here is a typical clamshell motion to restrict the nucleotide within the binding cleft, and has been observed in cases of other domains binding to nucleotides (e.g. UDP-glucose 6-Dehydrogenase UGDH, UDP-MurNc-tripeptide ligase MurE, nucleoside triphosphate diphosphorylase NTPDase1)[54–56]. This transition and the concomitant binding of the nucleotide substrate likely represents the faster kinetic component of the enzymatic cycle compared to binding of the lipid acceptor UndP into the binding site.

During the final stages of preparation of this article, a complementary publication reported cryo-EM structures of ArnC from *S. enterica* in *apo* (2.75 Å) and UDP-bound conformations (3.8 Å)[57]. Notably, these structures (PDB IDs 9BHC and 9BHE) are also tetrameric and were determined in amphipol. The amphipol environment enabled reconstructions in C2 symmetry only, not the full C4 symmetry we have observed in our nanodisc-embedded structures. The overall monomeric fold of ArnC is quite similar in the two structures, though some differences exist at the top of portion of the transmembrane helices. At the quaternary level, we observe significant differences between *apo* structures in lipid nanodiscs vs amphipol, both in the overall fold and especially in the transmembrane domain: Overall: 6.194 Å RMSD *apo* Arctica structure with 9BHC (1208 Cα pairs), 6.228 Å RMSD *apo* Krios structure with 9BHC (1208 Cα pairs); TM: 6.493 Å RMSD *apo* Arctica structure with 9BHC (304 Cα pairs), 6.56 Å RMSD *apo* Krios structure with 9BHC (304 Cα pairs). Likely because of the absence of lipids and the C2 symmetry imposed in the amphipol structure, the transmembrane helices in the amphipol *apo* structure appear to be positioned in a rhomboid arrangement, unlike the square arrangement observed in our structure. These differences may be attributable to the more native-like membrane environment present in lipid nanodiscs. Interestingly, the UDP-bound structure in amphipol shows a square arrangement of transmembrane helices, similar to the one observed in the ArnC structure(s) in lipid nanodiscs. Significant differences between the UDP-bound structures are still observed, though not as pronounced as in the *apo* structures: Overall: 3.273 Å RMSD with 9BHE (1236 Cα pairs), TM: 3.302 Å RMSD with 9BHE (304 Cα pairs). The improved resolution of the UDP-bound structure in lipid nanodiscs has allowed us to more confidently model residues in the active site of each protomer.

Putative UndP-binding motif of ArnC

In the crystal structure of *Pf*DPMS bound to dolichol phosphate mannose (DolP-Man) (PDB 5MM1), the lipid DolP can be seen threading between the juxtamembrane helices to reach the active site of the GT-A domain[44] (Supplementary Fig. 17a). A similar function for the JM helices

https://doi.org/10.1038/s41467-025-65968-6

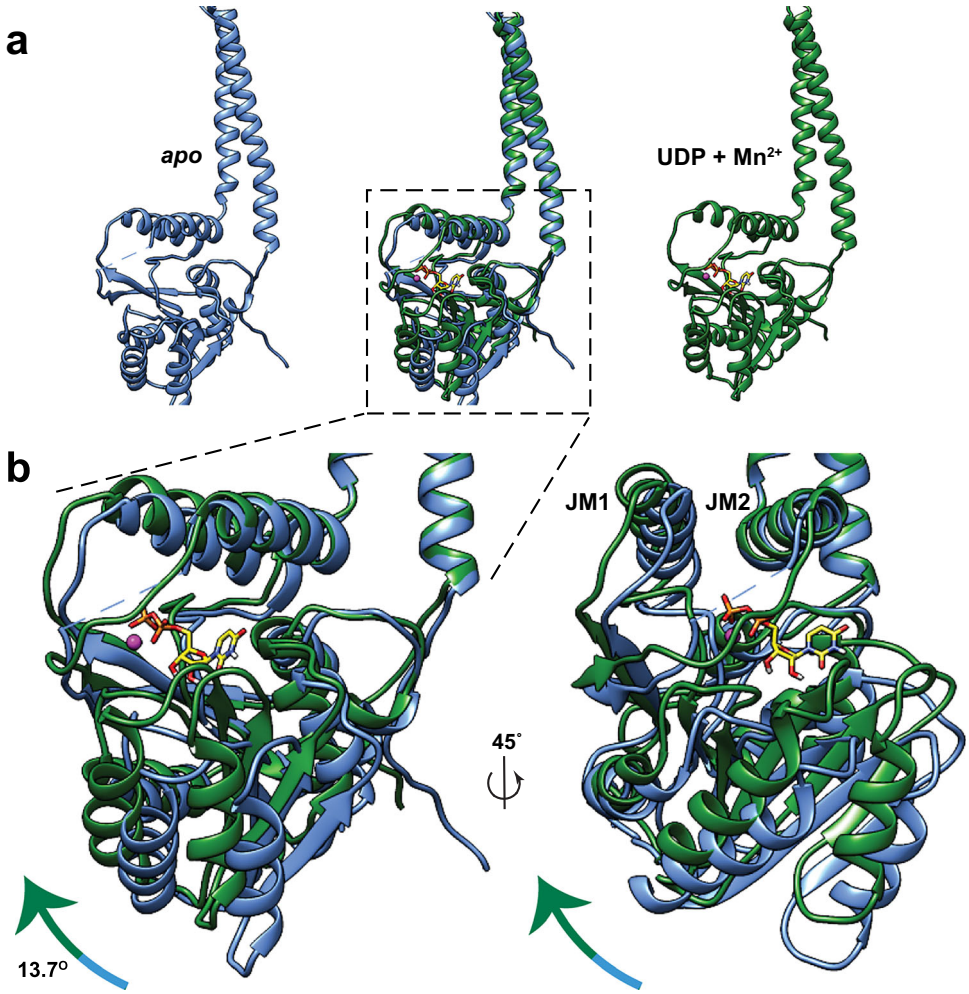

Fig. 4 | ArnC undergoes a conformational transition upon nucleotide binding. **a** Superposition of apo ArnC (blue) and UDP-bound ArnC (green) structures, with a single protomer from each structure aligned at the transmembrane helices (aa 230-305). **b** Detail from (**a**). Two views of the superposition rotated by 45° are presented. An upward movement is observed in the GT-A domain when the nucleotide is bound, with the domain rotating upward by 13.7°, and different parts of the domain showing a translation between 2.7-7.7 Å. The Mn^{2+} ion is shown as a pink sphere, and the UDP nucleotide is shown in stick representation and yellow color, with the phosphates shown in orange color.

of GtrB was hypothesized based on the GtrB crystal structure[43]. We propose that in ArnC the JM helices act in a similar manner to facilitate the entry of the acceptor substrate undecaprenyl phosphate (UndP) in the active site of ArnC. In addition, several non-protein densities likely corresponding to annular lipids were observed adjacent to TMD helices and especially within cavity 2 and at the base of the TMD close to the JM helices (Fig. 5a). To investigate these densities and the possible binding interface of UndP, we utilized LipIDens, a pipeline for molecular dynamics (MD) simulation-assisted interpretation of lipid and lipid-like densities in cryo-EM structures[58]. To assign the most likely identity of the additional lipid-like densities in the cryo-EM map we performed coarse-grained (CG) simulations of ArnC initiated from the apo conformation. Cardiolipin preferentially bound to a groove on the periplasmic TMD face with a residence time ~0.4-0.5 µs (Fig. 5a). The top half of the transmembrane domain contains only two positively-charged or polar residues, R258 and Q264, both located near cavity 2. This may account for the reduced residence time compared with high affinity cardiolipin sites across other prokaryotic membrane proteins[59]. Surprisingly, we observe spontaneous and stable binding of UndP (~ 8 µs) or a single tail of POPE (~ 10 µs) within the GT-A domain (Fig. 5a) whereby the phosphate group of UndP extends beyond the plane of the bilayer phosphate beads in the top ranked CG binding pose (Fig. 5b).

To further assess the binding of UndP within the GT-A domain, we performed an initial set of 100 ns atomistic MD simulations initiated from the top ranked CG pose. UndP was stably coordinated within the GT-A domain. The tail inserts between JM1 and JM2, forming hydrophobic interactions with small aliphatic sidechains (Ala, Leu, Ile). In contrast, the phosphate headgroup of UndP is coordinated by R131 and R137 (Fig. 5c). This binding pose is in broad agreement with that predicted for GtrB[43], and observed within the structure of *Pf*DPMS bound to DolP-Man[44]. Furthermore, we compared the position of UndP in the GT-A domain (as predicted from simulations) with the UDP binding pose in the UDP-bound ArnC cryo-EM structure (Supplementary Fig. 13b). Both UndP and UDP can be coordinated within the GT-A domain without coordination clashes. We also observed a rearrangement of a flexible loop (residues 201–208) that is stabilized by the presence of UndP compared to the *apo* conformation (Supplementary Fig. 13c). This rearrangement leads R128 and R205 to face towards the UDP binding site at the end of the simulations. R205 can be seen in the UDP-bound ArnC cryo-EM structure coordinating the UDP nucleotide (Supplementary Fig. 13a), and the flexible loop that "covers" the active site in each GT-A domain becomes fully ordered in the presence of the UDP nucleotide (aa 207-211 are only visible in the UDP-bound ArnC cryo-EM structure). Thus, the flexible loop appears to be stabilized by binding of either UndP or the UDP nucleotide, based on the

https://doi.org/10.1038/s41467-025-65968-6

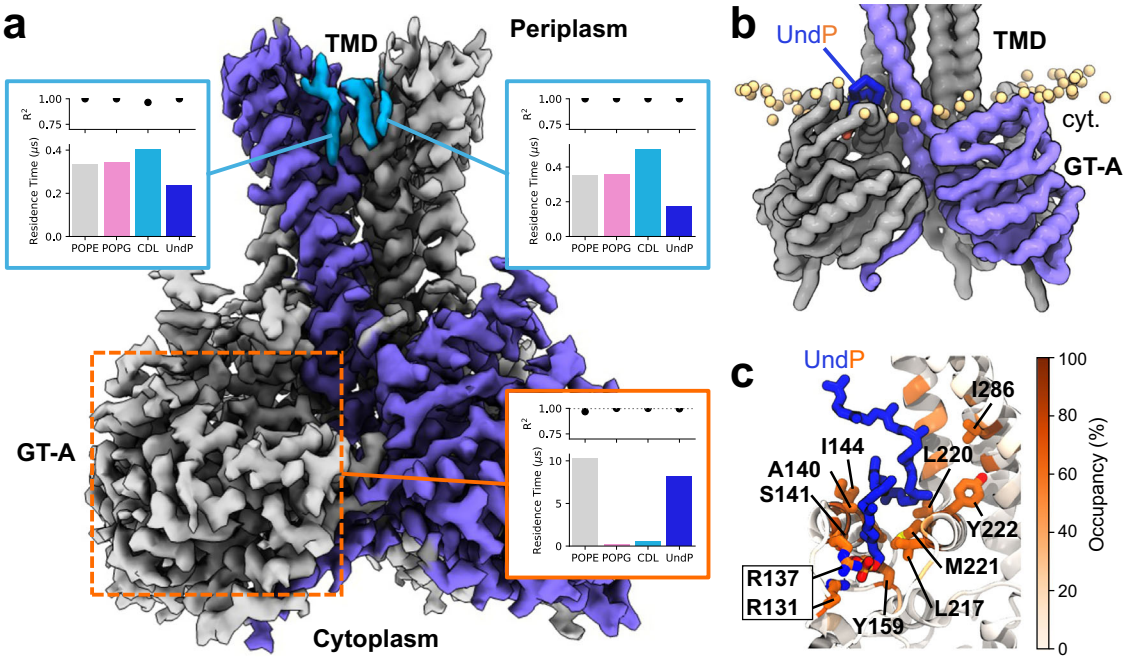

Fig. 5 | UndP and lipid interactions with ArnC. a Additional densities surrounding the apo ArnC TMD, colored according to preferential lipid interactions in coarse-grained (CG) MD simulations (10 x 15 μs) embedded within a bilayer contained POPE (64%), POPG (18%), cardiolipin (CDL) (9%) and UndP (9%). Plots indicate the relative residence times of lipids binding to the sites, as calculated using PyLipID. We also observe spontaneous binding of UndP within the GT-A domain in CG simulations (boxed in orange) and indicated in (**b**, **c**). **b** Top ranked CG binding pose for UndP within the GT-A domain. UndP is colored blue and the UndP phosphate bead is shown in orange with respect to the position of all lipid phosphate beads within the bilayer (yellow). **c** UndP pose at the end of atomistic MD simulations (3 × 100 ns) initiated by back-mapping from the top ranked CG UndP pose in (**b**). Residues with interaction occupancies >85% across all replicates are shown as sticks and labelled.

simulations and the UDP-bound cryo-EM structure. Residues R131, R137, and perhaps R128 appear to participate in coordination of UndP, while R205 coordinates the incoming nucleotide (together with the metal ion and the additional residues described in the previous section).

Molecular modelling and atomistic simulations of substrate-bound ArnC

To better describe the catalytically-relevant residues of ArnC, we proposed three models for the enzyme in complex with its full substrates, UndP and UDP-L-Ara4FN (UDPA) (Fig. 6a). For the first two models (State 1 and State 2), the lipid was positioned based on the binding observed in the CG and initial atomistic simulations, with the phosphate of UndP coordinated by R128 and R137. For State 1, the Ara4FN group of UDP-L-Ara4FN was placed in a groove between the JM2-β7 and JM1-β5 loops with the C1 atom within a hydrogen-bonding distance of the UndP phosphate group (State1—Fig. 6a). For the second model, the orientation of the Ara4FN group was based on the structure of PfDPMS in complex with GDP-mannose (PDB 5MM0) (State 2—Fig. 6a). To allow the reaction to occur, a third model was constructed in which the lipid was positioned deeper inside the active site, below the sugar, with the phosphate group within a hydrogen-bonding distance of the C1 atom of the Ara4FN group (State 3—Fig. 6a). Three independent unrestrained 500 ns MD simulations were performed for each system. UDP-L-Ara4N exhibits increased flexibility within the active site when initially placed in state 1 (Fig. 6b, c—bottom left panel and Supplementary Fig. 14). For all states the UDP portion of the UDP-L-Ara4N interacts with ArnC in a similar way, through several possible hydrogen bonds involving residues P15, Y17, E19, D48, A101 and R205. The diphosphate-binding site is formed at the canonical DXD motif, with hydrogen bonds between the diphosphate moiety of UDP and R205 also observed (Fig. 6d and Supplementary Fig. 15). While the UDP portion of UDP-L-Ara4FN behaves similarly for all models, the Ara4FN group is much more flexible within the active site for state 1. Interestingly, for this state, the sugar group rotates in the opposite direction of its initial configuration, adopting a binding pose similar to the one seen in the state 2 and 3 at the end of the simulation (Fig. 6b). For all states, hydrogen bonds between R137 and Ara4FN group were observed suggesting that R137 may also act in coordination of the sugar and stabilization of the substrate during catalysis. The high stability of UDPA within its binding site observed for states 2 and 3 suggest that the second proposed binding mode is more favorable.

To evaluate the viability of the proposed reaction models we measured the distance between UndP and UDP-L-Ara4N over the simulation (Fig. 6b top left panel). This distance was defined as the distance between the protonated oxygen of the phosphate group and the arabinose C1 atom. Even though for states 1 and 3 the UndP phosphate group was initially placed within 3.5 Å of the C1 atom, the higher distances sampled and the wider distance distribution for state 1 compared to state 3 suggest that the third model is most likely. The increased distance between UndP and UDPA observed in state 1 simulations is mainly due to a rearrangement in the lipid position, which is displaced toward the exit of the active site. In this position, the phosphate group retains its coordination by the R137 and R128 residues (Fig. 6c and Supplementary Fig. 14). While the phosphate group is more dynamic in state 1 (Fig. 6c—bottom right panel and Supplementary Fig. 14), for state 3, its coordination by Mg^{2+} and the R128 residue stabilizes the phosphate in the proximity of the Ara4FN group over the duration of the simulation. Moreover, for this state, the short distance between the UndP phosphate group and D100 (the first D in the DXD motif) suggests that this residue may act in the deprotonation of the phosphate group to allow the reaction to proceed (Fig. 6c—top right panel and Supplementary Fig. 14).

Even though the binding coarse-grained simulation and the atomistic simulations of state 2 suggest a somewhat different coordination for the UndP phosphate group, the CG simulations were

https://doi.org/10.1038/s41467-025-65968-6

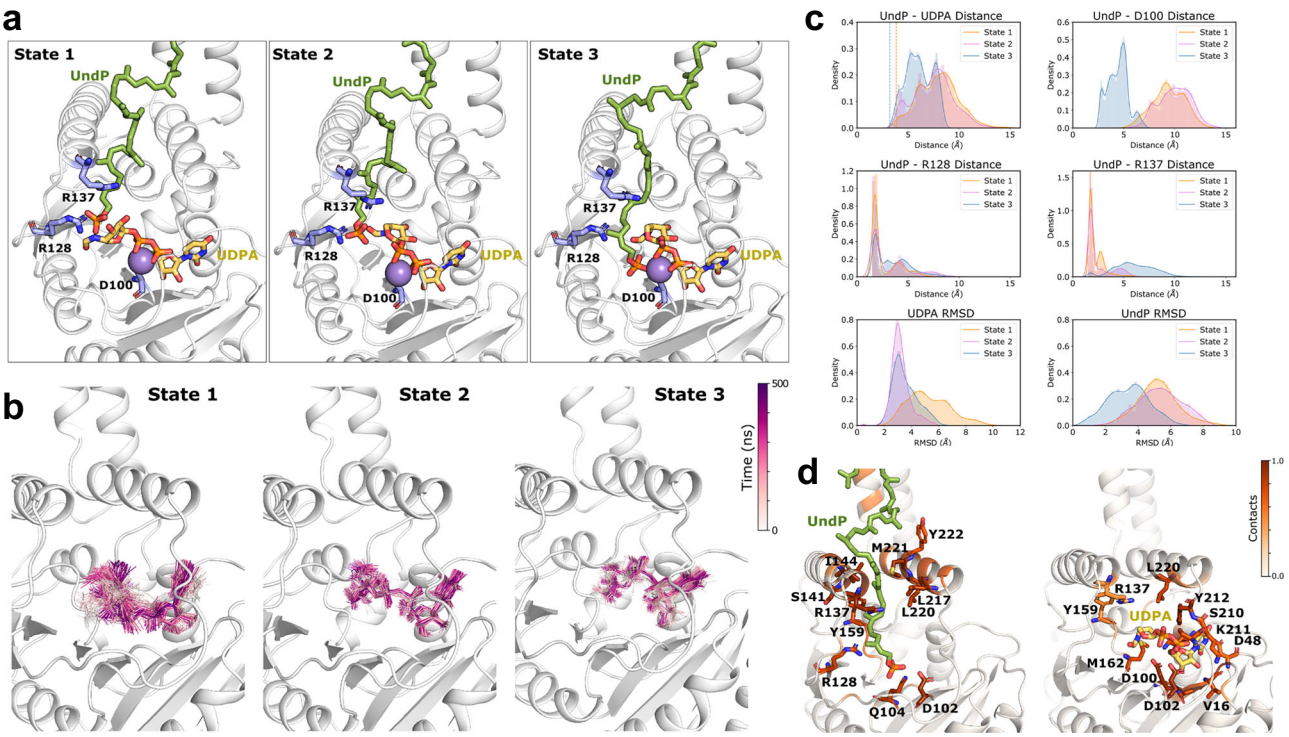

Fig. 6 | Atomistic simulations of ArnC with both substrates. a Models for the ArnC in complex with its substrates, UndP and UDP-L-Ara4FN: two conformations for UDP-L-Ara4FN and the UndP were modelled. In the first model the lipid was placed in the binding site predicted in the binding coarse-grained simulation, with its phosphate group coordinated by the conserved arginines R128 and R137. The Ara4FN group was positioned in an open groove within hydrogen-bonding distance from the phosphate of UndP (State 1). For the second model, the position of the sugar group was based on the structure of *Pf*DPMS in complex with GDP-mannose (PDB ID: 5MM0) while the coordinates for UndP were kept the same (State 2). For the third model, the lipid was positioned deeper inside the catalytic site, below the sugar (State 3). **b** Superposition of multiple frames showing the flexibility of UDP-L-Ara4FN within the ArnC active site for states 1,2 and 3. **c** UndP-UDPA distance distributions. The distance between UndP and UDPA was defined as the distance between the protonated oxygen of the phosphate group and the arabinose C1 atom (top left panel). UndP-D100 distance distributions (top right panel). UndP-R128 distance distributions (middle left panel). UndP-R137 distance distributions (middle right panel). The distances between UndP and D100, R128 or R137 were defined as the minimum distance between the phosphate group and each of the residues' sidechains. UDPA (bottom left panel) and UndP RMSD (bottom right). The RMSDs were calculated using backbone atoms of the protein as reference. To calculate the UndP RMSD only the phosphate group atoms and the first carbon were considered, due the high flexibility of the lipid tail. **d** Contacts between ArnC and UndP or UDPA for state 2. Coloring is the same as in (**a**).

conducted in the absence of UDPA and the divalent cation. Therefore, the CG and atomistic simulation are likely capturing different aspects of the catalytic cycle of the enzyme. The lipid would initially access the active site via the mechanism proposed by the CG simulation and move deeper inside the catalytic site when in the presence of UDPA and Mg^{2+}. Moreover, the similar behavior of UDPA observed for states 2 and 3 regardless of the position of the lipid, indicates that the presence of the lipid is not required for UDPA binding, which is also consistent with experimental results for UDP binding (Fig. 3b, Supplementary Figs. 11, 12c).

Mechanism of catalysis

The GT-A domain of ArnC is similar to the GT-A domain of Pren-P GTs GtrB and *Pf*DPMS, thus, comparison of the putative catalytic residues that operate in these three enzymes can help us better define the mechanism of catalysis for transfer of a sugar moiety from a nucleotide donor (UDP or GDP) to a lipid phosphate acceptor (UndP or DolP) in Pren-P GT enzymes. Based on the crystal structures of *Pf*DPMS bound to full donor substrate GDP-mannose (PDB 5MM0) and to product DolP-mannose (PDB 5MM1), it was proposed that dolichol phosphate is positioned within the active site by S135, R117 and R131 for catalysis[44] (Supplementary Figs. 16a, 17a). The acceptor phosphate of DolP, being in the dianionic form, is then suggested to perform a direct nucleophilic attack on the C1 atom of the mannose sugar[44]. All three of these residues are conserved in ArnC as S141, R128, and R137, which also has been seen interacting with the acceptor UndP in our CG and atomistic simulations (Figs. 5a, 6a, Supplementary Fig. 15). In addition, R205 in ArnC which is fully conserved in both other enzymes (R202 in *Pf*DPMS) participates in the coordination of the donor phosphate group (Supplementary Fig. 16a, b). The crystal structure of GtrB (PDB 5EKP) is missing densities for both β5-JM1 and β7-JM2 loops, has JM1 missing in some protomers, and missing sidechains for some critical residues, making comparison with ArnC more difficult. Despite that, R122 in GtrB likely corresponds to R117 in *Pf*DPMS and R128 in ArnC (Supplementary Fig. 16a), and K132 is the equivalent residue to R131 in *Pf*DPMS and R137 in ArnC (Supplementary Fig. 16b). R122 in GtrB has been proposed to coordinate the acceptor phosphate[43], similarly to the equivalent residues in *Pf*DMPS and ArnC. For GtrB, D157 was proposed to act as a catalytic acid that will donate a proton to the distal phosphate of the donor, after the acceptor phosphate of UndP in its dianionic form performs a direct nucleophilic attack on the C1 of the glucose sugar[43] (Supplementary Fig. 16a). The equivalent residue in *Pf*DPMS is a Gly and in ArnC a Met (Supplementary Fig. 16b). Thus, this mechanism is unlikely to operate in either *Pf*DPMS or ArnC.

In both *Pf*DPMS and GtrB, it has been proposed that the acceptor phosphate in its dianionic form performs a direct nucleophilic attack to the C1 of the sugar bound to the nucleotide donor to create the glycosidic bond of the product. Yet, phosphomonoesters have two pKas, with one expected to be in the physiological pH range[60]. The proportion of polyprenol phosphate lipids that are in their dianionic

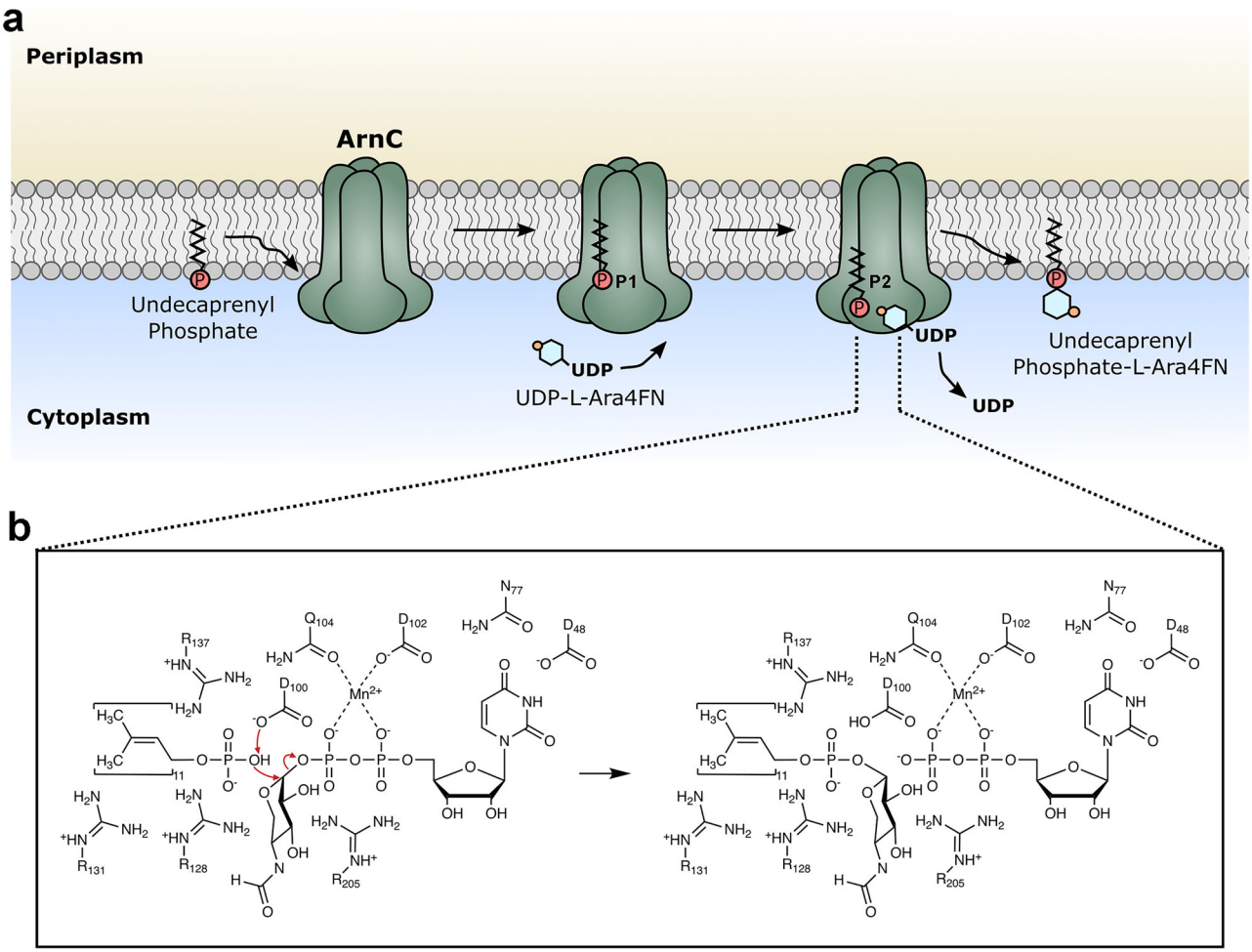

Fig. 7 | The catalytic mechanism of ArnC. a Schematic representation of the catalytic cycle of ArnC. The acceptor lipid undecaprenyl phosphate (UndP) enters the enzyme (by threading via the JM helices into the GT-A domain), and is coordinated in position 1 (P1), the "standby" position. Subsequently, the donor substrate UDP-L-Ara4FN binds into the same GT-A domain, and UndP moves to position 2 (P2), the "catalysis" position. The nucleophilic attack is rapidly accomplished and soluble UDP is released, while the new UndP-L-Ara4FN is released back into the inner bacterial membrane. **b** The proposed ArnC reaction mechanism involves a direct nucleophilic attack from the acceptor phosphate to the C1 of the L-Ara4FN ring. D100 acts as the catalytic base that abstracts a proton from the protonated oxygen of the UndP phosphate to enable the direct nucleophilic attack. The one step reaction enables the inversion of the anomeric configuration of the sugar. Several other residues that participate in coordination of the substrates are shown (D48, N77, D102, Q104, R128, R131, R137, R205).

form at physiological pH has not been determined as far as we know. The secondary pKa (pKa2) of phosphatidic acid (PA) and lysophosphatidic acid (LPA), which also carry a single phosphate connected to a glycerol and two or one acyl chain, respectively, has been measured to be between 6.8-7.5 for LPA and 6.9-7.9 for PA, and vary based on the lipid composition of the bilayer[60]. In monolayers, PA has been determined to have an even higher pKa2 value (10.5–11.5), which is also dependent on salt concentration[61]. Thus, it is very likely that at physiological pH a significant proportion of polyprenol phosphate acceptor lipids are in their monoanionic form, which would imply that simple coordination of UndP (or DolP) within the active site would not be sufficient for catalysis. We have observed that in all three cases of Pren-P GT enzymes, the first aspartate of the DXD motif (D89 in *Pf*DPMS, D94 in GtrB and D100 in ArnC) does not participate in the coordination of the catalytic metal ion and is available to act as a catalytic base. Our atomistic simulations have shown that the acceptor phosphate of UndP may be stably coordinated by D100, when the simulations start with both in proximity (Fig. 6c, Supplementary Fig. 17c). In addition, in *Pf*DPMS it has been shown that a D89A mutation leads to reduction of catalytic activity similar to the S135A mutation, while the second aspartate of the DXD motif when mutated (D91A), does not lead to as severe reduction of catalytic activity[44]. This argues that D89 may also operate as a catalytic base in *Pf*DPMS. In Fig. 7b we propose a catalytic mechanism for ArnC, which shows that a monoanionic form of UndP is coordinated in the active site via interactions with R128 and R137. D100 abstracts a proton from the acceptor phosphate and activates it. Subsequently, the acceptor phosphate performs an S_N2-like nucleophilic attack to the C1 atom of the L-Ara4FN sugar to create UndP-L-Ara4FN with inversion of the anomeric carbon. This reaction mechanism likely operates on all Pre-P GT enzymes, as the DXD motif is conserved throughout the family.

Finally, we propose an overall mechanistic scheme for the catalytic cycle of ArnC in Fig. 7a. In this scheme, the acceptor UndP threads through the JM helices of an ArnC protomer, and is coordinated in position 1 (P1), the "standby" position. Subsequently, a donor substrate UDP-L-Ara4FN binds to the same GT-A and triggers a conformational rearrangement in the flexible β7-JM2 loop, which allows UndP to move to position 2 (P2), the "catalysis position", for the reaction shown in Fig. 7B to take place. After that, the newly formed product will force UndP to backtrack into a "product position", which will facilitate release of the lipid back to the membrane. This is based on the observations that: i) UndP is coordinated stably within the GT-A domain in position 1 (Figs. 5c, 6a and Supplementary Fig. 17b) in the absence of a UDP nucleotide; (ii) In the presence of the nucleotide both

UndP and UDPA are less labile when UndP is located in position 2 (Fig. 6b, c and Supplementary Fig. 17c); (iii) Position 2 is optimal for catalysis to take place, as the acceptor phosphate is positioned in relative proximity of both the potential catalytic base D100 and the anomeric carbon of the L-Ara4FN sugar (Supplementary Fig. 17c, d); (iv) The "product position" for *Pf*DPMS shown in Supplementary Fig. 17a is different than position 1, and since the DolP-Man structure (PDM 5MM1) is a product-bound state, it likely represents a temporary coordinating position of the product. The unbinding of the UDP nucleotide after catalysis will likely trigger the product to be expelled back into the membrane via the JM helices. As a final note, there is no indication that the binding of UndP in position 1 always happens first, in fact the binding of the donor and the acceptor is likely independent. Yet, the threading of UndP via the JM helices is expected to be a slower process compared to binding of a diffusible ligand. Thus, in most cases the binding of UndP will happen first and be rapidly succeeded by binding of the donor substrate and catalysis.

Discussion

In this study we report the successful structure determination of the polyprenyl phosphate glycosyltransferase ArnC from the Gram-negative bacterium *S. enterica* embedded in lipid nanodiscs, in two conformations (*apo* and UDP-bound), using single-particle cryo-EM. We use MST to show that Mn^{2+} enables higher affinity for the partial donor substrate UDP, and by comparing the two conformations we show that binding of UDP on ArnC triggers a conformational rearrangement leading to a clamshell-like motion that brings the GT-A domain closer to the juxtamembrane helices of each protomer. We also perform a side-by-side comparison of two *apo* ArnC datasets collected using duplicate grids on a 200 kV Talos Arctica and a 300 kV Titan Krios microscope. At matched particle counts, datasets collected at 200 kV and 300 kV differed negligibly in resolution. No resolution cutoff was detected, as particle/resolution scaling conformed to theory throughout the range, but specimen-dependent limitations impose practical limits to resolution via the B-factor. Possible ways to improve the resolution of ArnC include alternative reconstruction algorithms, stabilizing or enlarging the complex, or physical improvements to the prepared grid. Next, we used CG simulations to show that the acceptor lipid UndP threads between the juxtamembrane helices of each ArnC protomer to reach the catalytic GT-A domain. Using additional atomistic simulations of ArnC with both the acceptor UndP and the full-donor substrate UDP-L-Ara4FN, we describe two different coordination positions for UndP within the GT-A domain (P1 and P2). We propose that the first position works as a "standby" position and the second as the "catalysis" position that enables the nucleophilic attack from the acceptor phosphate to the C1 carbon of the sugar. We also propose that the first aspartate of the DXD motif, which doesn't participate in coordination of the metal ion, functions as a catalytic base to abstract a proton from the UndP and activate it to perform the nucleophilic attack. The proposed catalytic mechanism likely operates similarly on all members of the Pren-P GT family to enable partially deprotonated UndP molecules to react with the donor substrate.

Overall, the findings of this study provide extensive characterization of the structure, conformational changes, and mechanistic basis of function for the Pren-P GT ArnC, the third member of the family to be structurally elucidated after GtrB and *Pf*DPMS. Moreover, the structures of ArnC add to the repertoire of structurally characterized proteins in the aminoarabinose biosynthetic pathway that can be targeted for rational drug design aimed at preventing or reversing resistance to polymyxin antibiotics.

Methods

Target cloning

arnC genes from the clinically relevant species *S. enterica* (WP_000458893), *K. pneumoniae* (WP_040165138), and *E. coli* (NP_416757) were codon optimized for expression in *E. coli* and synthesized with 30 bp overhangs (Genewiz) for direct assembly into expression vectors using NEBuilder® HiFi DNA Assembly Master Mix (NEB E2621). The genes were cloned in frame with either an N-terminal FLAG-10xHis-TEV cassette or a C-terminal TEV-10xHis cassette using pNYCOMPS vectors[45]. The sequences of the synthetic linear DNA fragments and the primers used to amplify the two pNYCOMPS vectors (N-term, C-term) with PCR are shown in Supplementary Table 2.

Small-scale expression and purification

The resulting ArnC constructs were transformed into T7 Express lysY E. coli competent cells (NEB, C3010I), an enhanced BL21 derivative strain, and grown overnight in 1 mL 2xYT medium supplemented with 50 μg/mL kanamycin at 37 °C with shaking (240 RPM). The next day, 20 mL of the same medium was inoculated with the starter culture at 1:100 ratio and grown at 37 °C with shaking (240 RPM), until OD600 1.0 was reached. Temperature was then decreased to 22 °C and protein expression was induced with 0.2 mM isopropyl β-D-1-thiogalactopyranoside (IPTG), and the culture was incubated overnight with shaking (240 RPM). Cells were harvested by centrifugation (4,000 g for 10 min) at 4 °C. For small scale purification of ArnC proteins, the cell pellets were resuspended in 700 μL lysis buffer containing 20 mM HEPES pH 7.5, 200 mM NaCl, 20 mM $MgSO_4$, 10 μg/mL DNase I, 8 μg/mL RNase A, 1 mM tris(2-carboxyethyl)phosphine hydrochloride (TCEP), 1 mM PMSF, and protease inhibitor cocktail set III EDTA-free (Calbiochem, 539134) at 1:1000 ratio. Cells were lysed using a probe sonicator (Microson) and the lysate was solubilized for 1.5 h with n-dodecyl-β-D-maltopyranoside (DDM; Anatrace, D310S) at a final concentration of 1% (w/v). Insoluble materials were removed by centrifugation at 18,000 g for 30 min at 4 °C. The supernatant was then subjected to metal-affinity chromatography using NiNTA agarose beads (Qiagen, 30230). Briefly, the supernatants were incubated with pre-equilibrated NiNTA agarose beads overnight (25 μL per pellet from a 20 mL culture). The beads were loaded on a 96-well filter plate and washed thrice with 100 μL of wash buffer containing 20 mM HEPES pH 7.5, 500 mM NaCl, 80 mM imidazole and 0.1% (w/v) DDM. The flow through at each step was removed by brief centrifugation (100 g for 30 s) at 4 °C. ArnC proteins were eluted with 60 μL elution buffer containing 20 mM HEPES pH 7.5, 200 mM NaCl, 300 mM imidazole, and 0.1% (w/v) DDM. the eluted proteins were run on 12% SDS-PAGE gel and stained with Imperial protein stain (Thermo Scientific 24617) for visualization (Supplementary Fig. 2a).

Extraction with Cubipol copolymer and cross-linking

A midscale pellet was resuspended in 15 mL of lysis buffer, followed by sonication for 12 mins. To extract ArnC from membrane Cubipol (Cube Biotech,18701) was added to a final concentration of 1%. The mixture was incubated for 4 hrs at 4 °C with gentle agitation. Ultracentrifugation at 4 °C (Ti70 rotor, 45 K, 20 mins) was performed to separate the supernatant, which was then bound to pre-equilibrated (0.5 mL) Ni-NTA beads overnight at 4 °C (supplemented with 40 mM Imidazole). The protein-bound beads were washed with 15 CV of wash buffer (20 mM HEPES pH 7.5, 500 mM NaCl, 80 mM Imidazole) and eluted with 5 CV elution buffer (20 mM HEPES pH 7.5, 200 mM NaCl, 300 mM Imidazole). To remove the imidazole, a PD-10 column was used. Final buffer used was 20 mM HEPES pH 7.5, 200 mM NaCl. The unconcentrated protein at 1.2 mg/mL was used for crosslinking experiment. ArnC purified in 1% Cubipol was cross linked with 3 mM glutaraldehyde for 2, 4 and 6 min in dark at room temperature. The reaction was quenched using 100 mM Tris 8.0 at specific time intervals followed by addition of SDS dye. The crosslinked samples were analyzed using 12% SDS-PAGE (Supplementary Fig. 2b).

Large-scale expression and purification in detergent

For large scale production of ArnC from Salmonella enterica ($ArnC_{Se}$), the construct encoding N-terminal FLAG-10xHis-TEV- $ArnC_{Se}$ was used

to transform T7 Express lysY E. coli competent cells (NEB, C3010I), and grown overnight in 2xYT medium supplemented with 50 µg/mL kanamycin at 37 °C with shaking (240 RPM). The next day, 200 mL of the same medium were inoculated with the starter culture at 1:100 ratio and left to grow at 37 °C with shaking (240 RPM), until OD600 reached 0.8-1.2. Temperature was then reduced to 22 °C, protein expression was induced with 0.2 mM IPTG, and the culture was incubated overnight with shaking (240 RPM). Cells were harvested by centrifugation (3700 RPM for 15 min) at 4 °C, washed once with phosphate buffered saline (PBS) and centrifuged again to produce a solid pellet that was stored at -80 °C until further use. For large-scale purification cell pellets were resuspended in lysis buffer containing 20 mM HEPES pH 7.5, 200 mM NaCl, 20 mM $MgSO_4$, 10 µg/mL Dnase I, 8 µg/mL Rnase A, 1 mM TCEP, 1 mM PMSF, and protease inhibitor cocktail set III EDTA-free (Calbiochem, 539134) at 1:1000 ratio. Cells were lysed with an Emulsiflex C3 high-pressure homogenizer (Avestin) and the lysate was solubilized for 2 h with DDM (Anatrace, D310S) added to a final concentration of 1% (w/v), in a volume of ~40 mL per cell pellet from 800 mL culture. Insoluble material was removed by ultracentrifugation at 34,000 RPM for 30 min at 4 °C and the protein was purified from the supernatant by metal-affinity chromatography using Ni-NTA agarose beads (Qiagen, 30230). The supernatant, after addition of 40 mM imidazole, was incubated with pre-equilibrated Ni-NTA agarose beads (0.7 mL per pellet from an 800 mL culture) overnight. The beads were then loaded on a column and washed with 10 column volumes of 20 mM HEPES pH 7.5, 500 mM NaCl, 75 mM imidazole and 0.03% (w/v) DDM. Protein was eluted with 4 column volumes of 20 mM HEPES pH 7.0, 150 mM NaCl, 300 mM Imidazole, and 0.03% (w/v) DDM. Imidazole was removed from the eluted protein by exchanging buffer to 20 mM HEPES pH 7.0, 200 mM NaCl, 0.03% (w/v) DDM (final protein buffer) using a PD-10 desalting column (Cytiva, 17085101).

Nanodisc incorporation

ArnC was incorporated into lipid nanodiscs with a 1:300:5 molar ratio of protein:1-palmitoyl-2-oleoyl-sn-glycero-3-phospho-(1'-rac-glycerol) (POPG):membrane scaffold protein 1E3D1 (MSP1E3D1)[62,63]. The same ratio was used in nanodisc incorporation trials with other lipids. This mixture was incubated at 4 °C for 2 h with gentle agitation. Reconstitution was initiated by removing detergent with the addition of Bio-beads SM-2 (Bio-Rad, 1523920) at 4 °C overnight with constant rotation. Subsequently, bio-beads were removed and the nanodisc reconstitution mixture was bound again to Ni-NTA resin at 4 °C for 2 h to remove empty nanodiscs. The resin was washed with 10 CV of wash buffer (20 mM HEPES pH 7.5, 150 mM NaCl, and 60 mM imidazole) followed by 4 CV of elution buffer (20 mM HEPES pH 7.0, 150 mM NaCl, and 300 mM imidazole). The sample was further purified by loading onto a Superdex 200 Increase 10/300 GL size-exclusion column (Cytiva, 28990944) in gel filtration buffer (20 mM HEPES pH 7.0 and 150 mM NaCl). Protein typically eluted as a sharp monodispersed peak, observed by monitoring absorbance at 280 nm (Supplementary Fig. 2c–e).

Microscale Thermophoresis (MST)

Prior to labeling, detergent-purified ArnC was diluted to 20 µM in buffer A (20 mM HEPES, pH 7.5, 150 mM NaCl, 0.01% DDM). The diluted ArnC was labeled using a Monolith Protein Labeling Kit RED-NHS 2nd Generation (Amine reactive) (NanoTemper Technologies, MO-L011). To assess the binding affinity of ArnC for UDP in the presence of different metal ions, UDP was titrated into ArnC across four experimental conditions: 1 mM $MnCl_2$, 1 mM $MgCl_2$, 1 mM EDTA to chelate any endogenous metal ions, or without adding any exogenous metal ions. ArnC was used at a final concentration of 50 nM. The UDP titration was prepared by serial dilution in each condition, starting from 2 mM to 122 nM. After incubation, the samples were transferred into premium coated capillaries (Nanotemper Technologies, MO-K025) and read in a Monolith NT.115 Nano-Blue/Red instrument at room temperature using 40% light-emitting-diode (LED) power and 40% MST power. Binding affinities were calculated using the thermophoresis with T jump evaluation strategy. All experiments were repeated three times in duplicate ($n = 6$) for each experimental group. Data analyses were performed using the NanoTemper analysis software.

Single-particle cryo-EM vitrification and data acquisition

Purified ArnC incorporated into MSP1E3D1/POPG nanodiscs was concentrated to 1.5 mg/ml using an Amicon 100 kDa concentrator (EMD Milipore, UFC510096). To generate UDP-bound ArnC, the purified protein was laced with 2 mM Uridine 5-Diphosphate Disodium (UDP) (Sigma-Aldrich, 94330) and 1 mM $MnCl_2$, prior to freezing. For each sample, 3 µl of sample was added to a glow-discharged R 1.2/1.3, 300 mesh holey gold grid (Quantifoil, UltrAuFoil) and blotted using filter paper for 3.5 s using a Vitrobot Mark IV (ThermoFisher Scientific) with a blot force of 3 and a wait time of 30 s, before plunging into liquid ethane for vitrification. The sample chamber was held at 4 °C with >90% humidity to minimize evaporation and sample degradation.

apo ArnC data collection on Talos Arctica

Micrograph movies in TIFF format were collected at the Rutgers Cryo-EM & Nanoimaging Facility (RCNF) using a 200 kV Talos Arctica electron microscope (FEI/ThermoFisher Scientific), equipped with a K2 Summit direct electron detector (Gatan) and a BioQuantum energy filter (Gatan) with slit width of 20 eV. Data were collected automatically in counting mode using EPU (FEI/ThermoFisher Scientific), a nominal magnification of 130,000×, a nominal pixel size of 1.038 Å/pixel, and a dose rate of 5.420 electrons/pixel/s. Movies were recorded at 200 ms/frame for 7 s (35 frames total), resulting in a total dose of 35.21 electrons/Å^2. Nominal defocus range was -0.5 to -2.5 µm. A total of 5552 movies were recorded from one grid over three days. Micrographs were gain-normalized and defect-corrected. At the end of processing a calibrated pixel size of 1.05 Å/pixel was determined for this dataset, representing an 1.16% error from the nominal pixel size value. The dataset has been deposited to EMPIAR with accession code EMPIAR-11924.

apo ArnC data collection on Titan Krios

Micrograph movies in TIFF format were collected at the National Center for Cryo-EM Access and Training (NCCAT) using the 300 kV Titan Krios #4 "Elizabeth" electron microscope (ThermoFisher Scientific), equipped with a K3 direct electron detector (Gatan) and a BioQuantum energy filter (Gatan) with slit width of 20 eV. Data were collected automatically in counting mode using image shift in Leginon, a nominal magnification of 130,000×, a nominal pixel size of 0.646 Å/pixel, and a dose rate of 19.969 electrons/pixel/s. Movies were recorded at 30 ms/frame for 1.2 s (40 frames total), resulting in a total dose of 57.42 electrons/Å^2. Nominal defocus range was -0.8 to -2.0 µm. A total of 23,259 movies were recorded from one grid over 3 days. Micrographs were not gain-normalized or defect corrected. A gain reference was recorded separately and is provided together with the dataset. At the end of processing a calibrated pixel size of 0.67 Å/pixel was determined for this dataset, representing an error of 3.72% from the nominal pixel size value. The dataset has been deposited to EMPIAR with accession code EMPIAR-12002.

UDP-bound ArnC data collection on Talos Arctica

Micrographs in TIFF format were collected at RCNF using a 200 kV Talos Arctica electron microscope (FEI/ThermoFisher Scientific), equipped with a K2 Summit direct electron detector (Gatan) and a BioQuantum energy filter (Gatan) with slit width of 20 eV. Data were collected automatically in counting mode using EPU (FEI/ThermoFisher Scientific), a nominal magnification of 165,000×, a nominal pixel

size of 0.818 Å/pixel, and a dose rate of 5.034 electrons/pixel/s. Movies were recorded at 100 ms/frame for 5.3 s (53 frames total), resulting in a total dose of 39.88 electrons/Å^2. The nominal defocus range used was -0.5 to -2.5 µm. A total of 7,925 micrographs were recorded from one grid over 2 days. Micrographs were gain-normalized and defect corrected. At the end of processing a calibrated pixel size of 0.82 Å/pixel was determined for this dataset, representing an 0.24% error from the nominal pixel size value. The dataset has been deposited to EMPIAR with accession code EMPIAR-11987.

Cryo-EM data processing

For the Arctica *apo* ArnC dataset: Motion correction and contrast transfer function (CTF) estimation were performed using patch motion correction and patch CTF estimation as implemented in cryoSPARC (v.3.3.2)[64]. The resulting aligned micrographs were manually curated for CTF fit resolution (< 6 Å), defocus (> 5500 Å), average intensity (-35<intensity<150), relative ice thickness (<3), and total full-frame motion distance (< 45 pix), resulting in 3402 exposures (~61%) being accepted. Blob picker (min diameter 60 Å, max diameter 150 Å) was used on 500 random micrographs to pick particles, and 50,752 particles were extracted using a 256 pixel box size and binned 4X. The particle stack was then subjected to 2D classification in cryoSPARC v.3.3.2. Eight 2D classes corresponding to 8.585 particles where selected and used for template picking of 600 random micrographs to pick particles. 78,806 particles were extracted using a 288 pixel box size binned 4X. The particle stack was again subjected to 2D classification and seven 2D classes corresponding to 24,123 particles were used for template picking the full 3402 accepted micrograph dataset. A total 1,002,867 particles were extracted using a 288 pixel box with 3X binning. This particle stack was subjected to 3 rounds of cleanup using ab initio reconstruction with 3 classes, resulting in a particle stack of 224,496 that was re-extracted with 288 pixel box size with no binning. This particle stack was subjected to a series of ab initio and heterogeneous refinement steps, resulting in a 3.42 Å map, which was then subjected to non-uniform refinement and local refinement in cryoSPARC resulting in a 2.73 Å reconstruction map from a final particle stack of 187,649 particles after imposing C4 symmetry. A calibrated pixel size of 1.05 Å/pix was imposed on the final reconstruction giving a final resolution of 2.79 Å. The processing workflow is summarized in Supplementary Fig. 3a.

For the Krios *apo* ArnC dataset: Motion correction and CTF estimation was performed using patch motion correction and patch CTF as implemented in cryoSPARC (v.3.3.2). The micrographs were separated to 4 groups of ~4–5 K micrographs each. The resulting aligned micrographs were manually curated for CTF fit resolution (< 5.9 Å), defocus (>5500 Å), average intensity (-7<intensity<4.4), relative ice thickness (<1.5), and total full-frame motion distance (<60 pix), resulting in 17,547 exposures (~75%) being accepted. Templates were imported from the *apo* Arctica dataset and used for template picking of each group of 4 K micrographs. Approx. 1 m particles per group was extracted using a 384-pixel box size binned 4X, which were then subjected to 2 separate rounds of ab initio cleanup, The resulting particle stack containing 525,744 particles was re-extracted using a 480 pixel box size (no binning). Heterogeneous refinement with 1 model and 3 decoys gave a particle stack of 506,322 particles. This particle stack was subjected to non-uniform refinement, manual micrograph curation, and then local refinement in cryoSPARC resulting in a 2.62 Å reconstruction map from a final particle stack of 490,807 particles after imposing C4 symmetry. A calibrated pixel size of 0.67 Å/pix was imposed on the final reconstruction giving a final resolution of 2.74 Å. The processing workflow is summarized in Supplementary Fig. 4a.

For the UDP-bound ArnC dataset: Motion correction and CTF estimation was performed using patch motion correction and patch CTF as implemented in cryoSPARC (v.3.3.2). The resulting aligned micrographs were manually curated for CTF fit resolution (<6.8 Å), defocus (>1400 Å), average intensity (-45<intensity<100), astigmatism<1000, relative ice thickness (<1.26), and total full-frame motion distance (<35 pix), resulting in 6761 exposures (~85%) being accepted. Templates were imported from the *apo* Arctica dataset and used for template picking. 2,382,682 particles were extracted using a 288 pixel box size binned 4X. This particle stack was then subjected to separate rounds of ab initio and 2D classification cleanup resulting in a particle stack of 347,008 particles that were reextracted using an unbinned 288-pixel box size. This particle stack was again subjected to a series of ab initio reconstructions with 2 classes, heterogeneous and homogeneous refinement, and manual micrograph curation, resulting in a 2.87 Å map from 216,104 particles after imposing C4 symmetry. A calibrated pixel size of 0.82 Å/pix was imposed on the final reconstruction giving a final resolution of 2.96 Å. The processing workflow is summarized in Supplementary Fig. 10a.

Pixel calibration

The calibrated pixel size for each dataset was determined using aa 1-120 of the GtrB X-ray crystal structure (PDB 5EKP) to fit the density of the GT-A domain in ArnC. Voxel size was varied systematically using UCSF Chimera and the correlation was monitored until it was maximized. The voxel size at the maximal correlation was selected as the calibrated pixel size and was imposed on the final reconstruction for each dataset.

Atomic model building and refinement

For modeling, an AlphaFold generated model of the E. coli (K12 strain) ArnC (Uniprot: P77757) from the AlphaFold database[49] was fitted in the *apo* (Arctica) 2.79 Å cryo-EM map using UCSF Chimera, and was manually mutated in Coot[65] to match the sequence of *S. enterica* ArnC. Subsequent model refinement and adjustment was performed in Coot[65], Phenix[66] and Isolde[67], iteratively. The *apo* (Arctica) model was then used to build models for the *apo* (Krios) map and UDP-bound cryo-EM maps. UDP was manually placed into the observed density using Coot[65] and refined in Phenix[66]. After refinement of each model in Isolde[67], a final refinement using phenix.real_space_refine was performed before finalizing each structure. The program phenix.douse was used to add waters automatically to the *apo* ArnC structure from the Krios map. Waters were manually curated in a second step before a final refinement using phenix.real_space_refine.

Resolution analysis

Random subsets of particles were reconstructed in cryoSPARC (v.3.3.2.). Orientational assignments, defoci, and other particle parameters from the final 3D refinement job were preserved; no refinement was performed. Only the 3D reconstruction step was repeated using the job type “Homogeneous Reconstruction Only.” Spectral signal-to-noise ratio was calculated as the Fourier shell correlation between the even and odd half-maps. Radial averaging of structure factors was performed using EMAN[68], and simulated maps were formed by approximating each atomic position as a sphere of density with Gaussian spatial decay.

Model analysis

A cavity search using the Solvent Extractor from Voss Volume Voxelator server[69] was performed using an outer-probe radius of 10 Å and inner-probe radius of 2 Å. Chimera[70] and ChimeraX[71] were used to visualize the structures and the resulting cavity volumes.

Coarse-grained MD simulations

Coarse-grained MD simulations of ArnC were established using the MemProtMD[72,73] Google Colab Notebook and LipIDens[58] for the apo ArnC conformation, including the modelling of missing residues/loops. ArnC was embedded in a bilayer composed of POPE (64%), POPG (18%), cardiolipin (CDL) (9%) and UndP (9%) and solvated with

MARTINI water[74] and 0.15 M NaCl in a 12 x 12 x 12 nm^3 box. The Martini 3.0 forcefield was used to describe all components[74]. Systems were energy minimized using a steepest-descent method and equilibrated in 2 x 1 ns steps. 10 × 15 µs CG simulations were run using parameters from LipIDens[58] and the GROMACS 2020 simulation software[75]. Temperature was maintained at 310 K using the V-rescale thermostat[76] (τt = 1.0 ps) and pressure was maintained at 1 bar using the Parrinello-Rahman barostat[77] (τp = 12.0 ps, compressibility = 3 x 10-4). The time-step was 20 fs. Coulombic interactions were described using the reaction-field method and a 1.1 nm cut-off. Van der Walls interactions were described with the potential-shift Verlet method and a 1.1 nm cut-off. Lipid interactions were analysed using the PyLipID toolkit[78] within the LipIDens pipeline[58]. A range of dual cut-off schemes were sampled and a 0.5/0.7 nm dual cut-off was selected based on the lipid interaction probability distribution and screening of e.g. interaction durations and binding site numbers across cut-off schemes. Lipid interaction residence times and binding sites were calculated using PyLipID[78]. Site screening and comparisons were obtained using LipIDens with POPE as the reference lipid against which to compare sites.

Atomistic MD simulations

The top ranked binding pose for UndP bound within the ArnC GT-A domain was selected for simulation. The LipIDens pipeline employed the CG2AT2 tool[79] to backmap the CG frame to atomistic resolution. The protein conformation was backmapped to that within the apo structure. The CHARMM-36 forcefield was used to describe all components[80]. Water was described using the TIP3P model[81]. Atomistic simulations were run for 3x100 ns using GROMACS 2020[75] and parameters from LipIDens[58]. Temperature was maintained at 310 K using the Nosé-Hoover thermostat[82] (τt = 0.5 ns) and pressure was maintained at 1 bar using the Parrinello-Rahman barostat[77] (τp = 2.0 ps, compressibility = 4.5 x 10-5). Van der Walls interactions were switched from 1.0 nm to 1.2 nm with the force-switch modifier. Coulombic interactions were modelled with the Particle-Mesh Ewald method[83] and a 1.2 nm cut-off. The LINCS algorithm[84] was used to constrain bonds to their equilibrium values and a dispersion correction was not applied. Per residue interaction occupancies with the bound UndP were calculated using a 0.3 nm cut-off. PyMol, ChimeraX and VMD were used for visualisation[71,85,86].

Molecular modelling of substrate-bound ArnC and atomistic simulations

The UDP-bound E. coli ArnC structure was used to build molecular models for the interactions with substrates, UndP and UDP-L-Ara4FN (UDPA), and Mg^{2+}. While the coordinates for UDP portion of UDPA were modelled as in the cryo-EM structure, two conformations for the Ara4FN group of UDP-L-Ara4FN and the UndP were proposed (Fig. 6A). The first model is based on the coordination of the phosphate of UndP by the conserved arginines R128 and R137 as suggested by the CG simulations. For this model, the Ara4FN sugar was placed in an open groove in a distance from the phosphate of UndP that would allow the reaction to proceed (Fig. 6A—State 1). A second model was generated using the same coordinates for UndP as in State 1, while the position of the Ara4FN group was based on the structure of *Pf*DPMS in complex with GDP-mannose (PDB ID: 5MM0)[44] (Fig. 6A—State 2). For this model, UndP and UDPA are positioned in a distance that would not allow the reaction to occur; therefore a third model was proposed in which UDPA is positioned as in the previous model and lipid is placed further down in the active site with the phosphate group of UndP below the sugar (Fig. 6A—State 3).

Atomistic simulations of substrates bound to ArnC were performed using the CHARMM36m force field[80]. As parameters for Mn^{2+} are not readily available for the CHARMM36m force field, so Mg^{2+} was used for the atomistic simulations. Parameters for UndP and UDP were taken from the CHARMM36m force field. CHARMM-GUI was used to prepare the CHARMM36m parameters and coordinates for UndP-Ara4FN and UDP-Ara4FN.

The MemProtMD pipeline[72,73] was used to run an initial 1 µs CG MD simulation to permit the assembly and equilibration of a POPG:POPE bilayer around ArnC without its substrate. The input protein was aligned to *xy* plane using MEMEMBED[87] and then converted to a CG representation using the Martini 3 force field[74]. In brief, the ArnC model was placed at the center of a periodic box and a POPG-POPE bilayer at 1:4 molar ratio was built around the protein. The final snapshot was converted back to atomic details using CG2AT[79] and the substrates coordinates, as described above, were added to the system. NaCl was added in a concentration of 150 mM to render the systems neutral. The complete systems were further equilibrated for 1 ns maintaining the structure of the protein and ligands restrained. Three repeats of unrestrained 500 ns MD simulations were performed for each system. All simulations were performed in the isothermal-isobaric ensemble at 310 K and 1 bar using a timestep of 2 fs. Pressure was maintained at 1 bar with a semi-isotropic compressibility of 4 x 10-5 using the Parrinello-Rahman barostat[77]. Temperature was controlled using the velocity-rescale thermostat[76], with the solvent, lipids and protein coupled to an external bath. The long-range electrostatic interactions were computed with the Particle Mesh Ewald method[83], while a verlet cut-off method was used to compute the non-bonded interactions. All MD simulations were performed using GROMACS[75] 2022 and analyzed using GROMACS tools and MDAnalysis[88]. Images were generated using PyMOL[85].

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

Atomic coordinates for the cryo-EM structures generated in this study have been deposited to the Protein Data Bank (PDB) under accession numbers: 8VXH (*apo*—Arctica), 9B77 (*apo*—Krios), and 9ASC (UDP—Arctica). Cryo-EM maps corresponding to the final sharpened map and half-maps for each dataset have been deposited to the Electron Microscopy Data Bank (EMDB) under accession numbers: EMD-43617 (*apo*—Arctica), EMD-44302 (*apo*—Krios), and EMD-43812 (UDP—Arctica). Raw electron micrograph movies, aligned and dose-weighted micrographs, and the final particle stacks used in reconstructions have been deposited to the Electron Microscopy Public Image Archive (EMPIAR) under accession numbers: EMPIAR-11924 (*apo*— Arctica), EMPIAR-12002 (*apo*—Krios), and EMPIAR-11987 (UDP—Arctica). Additional atomic models used in this study are publicly available under accession codes 5EKP, 5MLZ, 5MM0, 5MM1, 9BHC, 9BHE. The MD simulations data generated in this study are available as supplementary dataset on Zenodo (https://doi.org/10.5281/zenodo.17045635).

References

1. World Health Organization. *Antimicrobial Resistance: Global Report On Surveillance*. http://www.who.int/drugresistance/documents/surveillancereport/en (2014).
2. Centers for Disease Control and Prevention (U.S.). *Antibiotic Resistance Threats in the United States*. https://doi.org/10.15620/cdc:82532 (2019).
3. Curcio, D. Multidrug-resistant Gram-negative bacterial infections: are you ready for the challenge? *Curr. Clin. Pharm.* **9**, 27–38 (2014).
4. Viale, P., Giannella, M., Tedeschi, S. & Lewis, R. Treatment of MDR-Gram negative infections in the 21st century: a never ending threat for clinicians. *Curr. Opin. Pharm.* **24**, 30–37 (2015).
5. Velkov, T. & Roberts, K. D. Discovery of novel polymyxin-like antibiotics. *Adv. Exp. Med. Biol.* **1145**, 343–362 (2019).
6. Li, J. et al. Colistin: the re-emerging antibiotic for multidrug-resistant Gram-negative bacterial infections. *Lancet Infect. Dis.* **6**, 589–601 (2006).

7. Bergen, P. J. et al. Pharmacokinetics and pharmacodynamics of 'old' polymyxins: what is new. *Diagn. Microbiol. Infect. Dis.* **74**, 213–223 (2012).
8. Vaara, M. Polymyxins and their potential next generation as therapeutic antibiotics. *Front. Microbiol.* **10**, 1689 (2019).
9. Velkov, T., Thompson, P. E., Nation, R. L. & Li, J. Structure—activity relationships of polymyxin antibiotics. *J. Med. Chem.* **53**, 1898–1916 (2010).
10. Cajal, Y., Rogers, J., Berg, O. G. & Jain, M. K. Intermembrane molecular contacts by polymyxin B mediate exchange of phospholipids. *Biochemistry* **35**, 299–308 (1996).
11. Clausell, A. et al. Gram-negative outer and inner membrane models: insertion of cyclic cationic lipopeptides. *J. Phys. Chem. B* **111**, 551–563 (2007).
12. Mares, J., Kumaran, S., Gobbo, M. & Zerbe, O. Interactions of lipopolysaccharide and polymyxin studied by NMR spectroscopy. *J. Biol. Chem.* **284**, 11498–11506 (2009).
13. Tan, C. H., Li, J. & Nation, R. L. Activity of colistin against heteroresistant Acinetobacter baumannii and emergence of resistance in an in vitro pharmacokinetic/pharmacodynamic model. *Antimicrob. Agents Chemother.* **51**, 3413–3415 (2007).
14. Bergen, P. J. et al. Comparison of once-, twice- and thrice- daily dosing of colistin on antibacterial effect and emergence of resistance: studies with Pseudomonas aeruginosa in an in vitro pharmacodynamic model. *J. Antimicrob. Chemother.* **61**, 636–642 (2008).
15. Poudyal, A. et al. In vitro pharmacodynamics of colistin against multidrug-resistant Klebsiella pneumoniae. *J. Antimicrob. Chemother.* **62**, 1311–1318 (2008).
16. Talbot, G. H. et al. Bad bugs need drugs: an update on the development pipeline from the Antimicrobial availability task force of the infectious diseases society of America. *Clin. Infect. Dis.* **42**, 657–668 (2006).
17. Matthaiou, D. K. et al. Risk factors associated with the isolation of colistin-resistant gram-negative bacteria: a matched case-control study. *Crit. Care Med.* **36**, 807–811 (2008).
18. Raetz, C. R. H., Reynolds, C. M., Trent, M. S. & Bishop, R. E. Lipid A modification systems in gram-negative bacteria. *Annu Rev. Biochem* **76**, 295–329 (2007).
19. Brogden, K. A. Antimicrobial peptides: pore formers or metabolic inhibitors in bacteria?. *Nat. Rev. Micro* **3**, 238–250 (2005).
20. Needham, B. D. & Trent, M. S. Fortifying the barrier: the impact of lipid A remodelling on bacterial pathogenesis. *Nat. Rev. Microbiol.* **11**, 467–481 (2013).
21. Olaitan, A. O., Morand, S. & Rolain, J.-M. Mechanisms of polymyxin resistance: acquired and intrinsic resistance in bacteria. *Front Microbiol* **5**, 643 (2014).
22. Trent, M. S., Ribeiro, A. A., Lin, S., Cotter, R. J. & Raetz, C. R. H. An inner membrane enzyme in Salmonella and Escherichia coli that transfers 4-amino-4-deoxy-L-arabinose to lipid A: Induction in polymyxin-resistant mutants and role of a novel lipid-linked donor. *J. Biol. Chem.* **276**, 43122–43131 (2001).
23. Trent, M. S. et al. Accumulation of a polyisoprene-linked amino sugar in polymyxin-resistant Salmonella typhimurium and Escherichia coli: structural characterization and transfer to lipid A in the periplasm. *J. Biol. Chem.* **276**, 43132–43144 (2001).
24. Lo Sciuto, A. & Imperi, F. Aminoarabinosylation of lipid A is critical for the development of colistin resistance in pseudomonas aeruginosa. *Antimicrob. Agents Chemother.* **62**, e01820–17 (2018).
25. Helander, I. M. et al. Characterization of lipopolysaccharides of polymyxin-resistant and polymyxin-sensitive Klebsiella pneumoniae O3. *Eur. J. Biochem.* **237**, 272–278 (1996).
26. De Oliveira, D. M. P. et al. Antimicrobial Resistance in ESKAPE Pathogens. *Clin. Microbiol. Rev.* **33**, e00181-19 (2020).
27. Breazeale, S. D., Ribeiro, A. A., McClerren, A. L. & Raetz, C. R. H. A formyltransferase required for polymyxin resistance in escherichia coli and the modification of lipid a with 4-amino-4-deoxy-L-arabinose. Identification and function of UDP-4-deoxy-4-formamido-L-arabinose. *J. Biol. Chem.* **280**, 14154–14167 (2005).
28. Gunn, J. S. et al. PmrA–PmrB-regulated genes necessary for 4-aminoarabinose lipid A modification and polymyxin resistance. *Mol. Microbiol.* **27**, 1171–1182 (1998).
29. Gatzeva-Topalova, P. Z., May, A. P. & Sousa, M. C. Crystal structure and mechanism of the Escherichia coli ArnA (PmrI) transformylase domain. An enzyme for lipid A modification with 4-amino-4-deoxy-L-arabinose and polymyxin resistance. *Biochemistry* **44**, 5328–5338 (2005).
30. Genthe, N. A., Thoden, J. B. & Holden, H. M. Structure of the Escherichia coli ArnA N-formyltransferase domain in complex with N5-formyltetrahydrofolate and UDP-Ara4N. *Protein Sci.* **25**, 1555–1562 (2016).
31. Zhou, Z., Lin, S., Cotter, R. J. & Raetz, C. R. Lipid A modifications characteristic of Salmonella typhimurium are induced by NH4VO3 in Escherichia coli K12. Detection of 4-amino-4-deoxy-L-arabinose, phosphoethanolamine and palmitate. *J. Biol. Chem.* **274**, 18503–18514 (1999).
32. Adak, T. et al. ArnD is a deformylase involved in polymyxin resistance. *Chem. Commun.* **56**, 6830–6833 (2020).
33. Muñoz-Escudero, D. et al. Structure and function of ArnD. A deformylase essential for lipid A modification with 4-Amino-4-deoxy-l-arabinose and polymyxin resistance. *Biochemistry* **62**, 2970–2981 (2023).
34. Yan, A., Guan, Z. & Raetz, C. R. H. An undecaprenyl phosphate-aminoarabinose flippase required for polymyxin resistance in Escherichia coli. *J. Biol. Chem.* **282**, 36077–36089 (2007).
35. Petrou, V. I. et al. Structures of aminoarabinose transferase ArnT suggest a molecular basis for lipid A glycosylation. *Science* **351**, 608–612 (2016).
36. Coutinho, P. M., Deleury, E., Davies, G. J. & Henrissat, B. An evolving hierarchical family classification for glycosyltransferases. *J. Mol. Biol.* **328**, 307–317 (2003).
37. Drula, E. et al. The carbohydrate-active enzyme database: functions and literature. *Nucleic Acids Res.* **50**, D571–D577 (2022).
38. Lukose, V., Walvoort, M. T. & Imperiali, B. Bacterial phosphoglycosyl transferases: initiators of glycan biosynthesis at the membrane interface. *Glycobiology* **27**, 820–833 (2017).
39. Allen, K. N. & Imperiali, B. Structural and mechanistic themes in glycoconjugate biosynthesis at membrane interfaces. *Curr. Opin. Struct. Biol.* **59**, 81–90 (2019).
40. Hartley, M. D. & Imperiali, B. At the membrane frontier: A prospectus on the remarkable evolutionary conservation of polyprenols and polyprenyl-phosphates. *Arch. Biochem. Biophys.* **517**, 83–97 (2012).
41. Lairson, L. L., Henrissat, B., Davies, G. J. & Withers, S. G. Glycosyltransferases: structures, functions, and mechanisms. *Annu. Rev. Biochem.* **77**, 521–555 (2008).
42. Gloster, T. M. Advances in understanding glycosyltransferases from a structural perspective. *Curr. Opin. Struct. Biol.* **28**, 131–141 (2014).
43. Ardiccioni, C. et al. Structure of the polyisoprenyl-phosphate glycosyltransferase GtrB and insights into the mechanism of catalysis. *Nat. Commun.* **7**, 10175 (2016).
44. Gandini, R., Reichenbach, T., Tan, T.-C. & Divne, C. Structural basis for dolichylphosphate mannose biosynthesis. *Nat. Commun.* **8**, 120 (2017).
45. Bruni, R. & Kloss, B. High-throughput cloning and expression of integral membrane proteins in Escherichia coli. *Curr. Protoc. Protein Sci.* **74**, 29.6.1–29.6.34 (2013).
46. Efremov, R. G., Gatsogiannis, C. & Raunser, S. Lipid nanodiscs as a tool for high-resolution structure determination of membrane

proteins by single-particle cryo-EM. In *Methods in Enzymology* (ed. Ziegler, C.) 1–30 (Academic Press, 2017).
47. Notti, R. Q. & Walz, T. Native-like environments afford novel mechanistic insights into membrane proteins. *Trends Biochem Sci.* **47**, 561–569 (2022).
48. Russo, C. J. & Passmore, L. A. Ultrastable gold substrates for electron cryomicroscopy. *Science* **346**, 1377–1380 (2014).
49. Varadi, M. et al. AlphaFold Protein Structure Database: massively expanding the structural coverage of protein-sequence space with high-accuracy models. *Nucleic Acids Res.* **50**, D439–D444 (2022).
50. Rosenthal, P. B. & Henderson, R. Optimal determination of particle orientation, absolute hand, and contrast loss in single-particle electron cryomicroscopy. *J. Mol. Biol.* **333**, 721–745 (2003).
51. Baldwin, P. R. & Lyumkis, D. Non-uniformity of projection distributions attenuates resolution in Cryo-EM. *Prog. Biophys. Mol. Biol.* **150**, 160–183 (2020).
52. Cushing, V. I. et al. High-resolution cryo-EM of the human CDK-activating kinase for structure-based drug design. *Nat. Commun.* **15**, 2265 (2024).
53. López-Méndez, B., Uebel, S., Lundgren, L. P. & Sedivy, A. Microscale thermophoresis and additional effects measured in nanotemper monolith instruments. *Eur. Biophys. J.* **50**, 653–660 (2021).
54. Jung, K. H. et al. Wide-open conformation of UDP-MurNc-tripeptide ligase revealed by the substrate-free structure of MurE from Acinetobacter baumannii. *FEBS Lett.* **595**, 275–283 (2021).
55. Egger, S., Chaikuad, A., Kavanagh, K. L., Oppermann, U. & Nidetzky, B. Structure and mechanism of human UDP-glucose 6-dehydrogenase*. *J. Biol. Chem.* **286**, 23877–23887 (2011).
56. Zebisch, M., Krauss, M., Schäfer, P. & Sträter, N. Crystallographic evidence for a domain motion in rat nucleoside Triphosphate Diphosphohydrolase (NTPDase) 1. *J. Mol. Biol.* **415**, 288–306 (2012).
57. Patel, D. H. et al. Cryo-EM SPR structures of Salmonella typhimurium ArnC; the key enzyme in lipid-A modification conferring polymyxin resistance. *Protein Sci.* **34**, e70037 (2025).
58. Ansell, T. B. et al. LipIDens: Simulation assisted interpretation of lipid densities in cryo-EM structures of membrane proteins. *Nat. Commun.* **14**, 7774 (2023).
59. Corey, R. A. et al. Identification and assessment of cardiolipin interactions with E. coli inner membrane proteins. *Sci. Adv.* **7**, eabh2217 (2021).
60. Kooijman, E. E. et al. What makes the bioactive lipids phosphatidic acid and lysophosphatidic acid so special? *Biochemistry* **44**, 17007–17015 (2005).
61. Zhang, T. et al. Effect of pH and salt on surface pka of phosphatidic acid monolayers. *Langmuir* **34**, 530–539 (2018).
62. Bayburt, T. H., Grinkova, Y. V. & Sligar, S. G. Self-assembly of discoidal phospholipid bilayer nanoparticles with membrane scaffold proteins. *Nano Lett.* **2**, 853–856 (2002).
63. Denisov, I. G., Grinkova, Y. V., Lazarides, A. A. & Sligar, S. G. Directed self-assembly of monodisperse phospholipid bilayer nanodiscs with controlled size. *J. Am. Chem. Soc.* **126**, 3477–3487 (2004).
64. Punjani, A., Rubinstein, J. L., Fleet, D. J. & Brubaker, M. A. cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. *Nat. Methods* **14**, 290–296 (2017).
65. Casañal, A., Lohkamp, B. & Emsley, P. Current developments in coot for macromolecular model building of electron cryo-microscopy and crystallographic data. *Protein Sci.* **29**, 1055–1064 (2020).
66. Liebschner, D. et al. Macromolecular structure determination using X-rays, neutrons and electrons: recent developments in Phenix. *Acta Cryst. D.* **75**, 861–877 (2019).
67. Croll, T. I. ISOLDE: a physically realistic environment for model building into low-resolution electron-density maps. *Acta Cryst. D.* **74**, 519–530 (2018).
68. Tang, G. et al. EMAN2: An extensible image processing suite for electron microscopy. *J. Struct. Biol.* **157**, 38–46 (2007).
69. Voss, N. R. & Gerstein, M. 3V: cavity, channel and cleft volume calculator and extractor. *Nucleic Acids Res.* **38**, W555–W562 (2010).
70. Pettersen, E. F. et al. UCSF Chimera—A visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).
71. Goddard, T. D. et al. UCSF ChimeraX: meeting modern challenges in visualization and analysis. *Protein Sci.* **27**, 14–25 (2018).
72. Stansfeld, P. J. et al. MemProtMD: automated insertion of membrane protein structures into explicit lipid membranes. *Structure* **23**, 1350–1361 (2015).
73. Newport, T. D., Sansom, M. S. P. & Stansfeld, P. J. The MemProtMD database: a resource for membrane-embedded protein structures and their lipid interactions. *Nucleic Acids Res.* **47**, D390–D397 (2019).
74. Souza, P. C. T. et al. Martini 3: a general purpose force field for coarse-grained molecular dynamics. *Nat. Methods* **18**, 382–388 (2021).
75. Abraham, M. J. et al. GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **1–2**, 19–25 (2015).
76. Bussi, G., Donadio, D. & Parrinello, M. Canonical sampling through velocity rescaling. *J. Chem. Phys.* **126**, 014101 (2007).
77. Parrinello, M. & Rahman, A. Polymorphic transitions in single crystals: a new molecular dynamics method. *J. Appl. Phys.* **52**, 7182–7190 (1981).
78. Song, W. et al. PyLipID: A python package for analysis of protein–lipid interactions from molecular dynamics simulations. *J. Chem. Theory Comput.* **18**, 1188–1201 (2022).
79. Vickery, O. N. & Stansfeld, P. J. CG2AT2: An enhanced fragment-based approach for serial multi-scale molecular dynamics simulations. *J. Chem. Theory Comput.* **17**, 6472–6482 (2021).
80. Huang, J. & MacKerell, A. D. Jr. CHARMM36 all-atom additive protein force field: Validation based on comparison to NMR data. *J. Comput. Chem.* **34**, 2135–2145 (2013).
81. Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W. & Klein, M. L. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **79**, 926–935 (1983).
82. Hoover, W. G. Canonical dynamics: equilibrium phase-space distributions. *Phys. Rev. A* **31**, 1695–1697 (1985).
83. Darden, T., York, D. & Pedersen, L. Particle mesh Ewald: an N·log(N) method for Ewald sums in large systems. *J. Chem. Phys.* **98**, 10089–10092 (1993).
84. Hess, B., Bekker, H., Berendsen, H. J. C. & Fraaije, J. G. E. M. LINCS: A linear constraint solver for molecular simulations. *J. Comput. Chem.* **18**, 1463–1472 (1997).
85. DeLano, W. L. Pymol: An open-source molecular graphics tool. *CCP4 Newsl. Protein Crystallogr* **40**, 82–92 (2002).
86. Humphrey, W., Dalke, A. & Schulten, K. V. M. D. Visual molecular dynamics. *J. Mol. Graph* **14**, 33–38 (1996).
87. Nugent, T. & Jones, D. T. Membrane protein orientation and refinement using a knowledge-based statistical potential. *BMC Bioinforma.* **14**, 276 (2013).
88. Michaud-Agrawal, N., Denning, E. J., Woolf, T. B. & Beckstein, O. MDAnalysis: A toolkit for the analysis of molecular dynamics simulations. *J. Comput. Chem.* **32**, 2319–2327 (2011).

Acknowledgements

This project was supported by NIH grants R00GM123228 and R35GM150831 to VIP. PJS acknowledges the NIH (R01AI174416 - PI: M. Stephen Trent), Wellcome (208361/Z/17/Z), MRC, BBSRC, and the Howard Dalton Centre for funding. PJS and CMB acknowledge Sulis at HPC Midlands + , which was funded by the EPSRC on grant EP/T022108/1, and the University of Warwick Scientific Computing Research Technology Platform for computational access. This project made use of time on ARCHER2 and JADE2 granted via the UK High-End Computing Consortium for Biomolecular Simulation, HECBioSim (http://www.hecbiosim.ac.uk), supported by EPSRC (grant no. EP/R029407/1). TBA

was supported by Wellcome (102164/Z/13/Z) and is currently supported by Schmidt Science Fellows, in partnership with the Rhodes Trust. The Krios dataset was collected at the National Center for Cryo-EM Access and Training (NCCAT) and the Simons Electron Microscopy Center located at the New York Structural Biology Center, supported by the NIH Common Fund Transformative High Resolution Cryo-Electron Microscopy program (U24 GM129539) and by grants from the Simons Foundation (SF349247) and NY State Assembly. We thank the staff at NCCAT that facilitated the collection of the *apo* ArnC dataset.

Author contributions

A.P. and V.I.P performed construct design and target cloning. A.P. performed small-scale protein purification and extraction with Cubipol. K.U.A and S.R.G. performed nanodisc incorporation. K.U.A. performed MST experiments. K.U.A prepared grids, and K.U.A., E.F., and J.T.K. collected cryo-EM data. K.U.A, J.T.K., and V.I.P. processed and analyzed cryo-EM data. K.U.A and V.I.P. built atomic models and deposited data into databases. M.B.B., T.B.A., and P.J.S. performed CG and atomistic MD simulations. All authors contributed to experiment design. K.U.A., M.B.B., T.B.A., S.R.G., P.J.S., and V.I.P. created figures. K.U.A., M.B.B., T.B.A., P.J.S., and V.I.P. wrote the manuscript with edits and comments from all authors. V.I.P. provided oversight for the entire project.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41467-025-65968-6.

Correspondence and requests for materials should be addressed to Vasileios I. Petrou.

Peer review information *Nature Communications* thanks the anonymous reviewers for their contribution to the peer review of this work. A peer review file is available.