www.sciencemag.org/content/351/6273/608/suppl/DC1

Supplementary Materials for

Structures of aminoarabinose transferase ArnT suggest a molecular basis for lipid A glycosylation

Vasileios I. Petrou, Carmen M. Herrera, Kathryn M. Schultz, Oliver B. Clarke, Jérémie Vendome, David Tomasek, Surajit Banerjee, Kanagalaghatta R. Rajashankar, Meagan Belcher-Dufrisne, Brian Kloss, Edda Kloppmann, Burkhard Rost, Candice S. Klug, M. Stephen Trent, Lawrence Shapiro, Filippo Mancia*

*Corresponding author. E-mail: fm123@cumc.columbia.edu

Published 5 February 2016, *Science* **351**, 608 (2016)
DOI: 10.1126/science.aad1172

This PDF file includes

Materials and Methods

Target identification and cloning. Twelve target ArnT genes corresponding to a wide genomic background were identified based on a bioinformatics approach (*26*). Ligation independent cloning (LIC) was used to clone the targets into four different LIC-adapted expression vectors (pNYCOMPS-Nterm, pNYCOMPS-Cterm, pMCSG7, pMCSG28), all bearing protease-cleavable decahistidine tags for metal-affinity chromatography-based purification. All cloning and initial protein characterization were performed at the protein production core facility of the New York Consortium on Membrane Protein Structure (NYCOMPS) (*27*).

Protein expression and purification. ArnT from *Cupriavidus metallidurans* ($ArnT_{Cm}$) cloned in the pNYCOMPS-Nterm vector was transformed into BL21(DE3)pLysS *E. coli* competent cells, and grown overnight in 2xYT medium supplemented with 50 μg/mL kanamycin and 35 μg/mL chloramphenicol at 37°C with shaking (240 r.p.m.). The next day, 800 mL (large scale) or 8 mL (small scale to test expression) of the same medium were inoculated with the starter culture at 1:100 ratio, and left to grow at 37°C with shaking (240 r.p.m.), until OD_{600} reached 1.0 (approximately 3 hours). Temperature was then reduced to 22°C, protein expression was induced with 0.2 mM isopropyl β-D-1-thiogalactopyranoside (IPTG), and the culture was incubated overnight with shaking (240 r.p.m.). Cells were harvested by centrifugation (3,700 r.p.m. for 15 min) at 4°C, washed once with 1x phosphate buffered saline (PBS) and centrifuged again to produce a solid pellet that was stored at –80°C, until needed. For expression of selenomethionine-substituted protein, cells were grown in a minimal medium of M9 salts using a kit (M9 SeMet High-Yield Growth Media Kit, Shanghai Medicilon Inc.). An additional 150 mg selenomethionine were added per liter of medium above the amount provided by the kit. The rest of the growth conditions were identical as those described for the expression of native protein.

Small scale purification tests were performed as previously described (*11*). For large-scale purification of $ArnT_{Cm}$, 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 tris(2-carboxyethyl)phosphine hydrochloride (TCEP), 1 mM PMSF, and Complete Mini EDTA-free protease inhibitor cocktail (Roche) used according to the manufacturer's instructions. Cells were lysed using an Emulsiflex C3 homogenizer (Avestin) and the lysate was solubilized for 1.5-2 hours with *n*-dodecyl-β-D-maltopyranoside (DDM; Affymetrix) added to a final concentration of 1% (w/v), in a volume of approximately 40 mL per cell pellet from 800 mL culture (~6 grams cells). Insoluble material was removed by ultracentrifugation at 34,000 r.p.m. for 30 min at 4°C and the protein was purified from the supernatant by metal-affinity chromatography using Ni-NTA agarose beads (Qiagen). Briefly, the supernatant 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, 200 mM NaCl, 60 mM Imidazole and 0.1% (w/v) DDM. Protein was eluted with 4 column volumes of 20 mM HEPES pH 7.0, 200 mM NaCl, 300 mM Imidazole, and 0.05% (w/v) DDM. Imidazole was removed from the eluted protein by exchanging buffer to 20 mM HEPES pH 7.0, 200 mM NaCl, 0.05% (w/v) DDM, 1mM TCEP (Final protein buffer) using a PD-10 desalting column (GE Healthcare). Subsequently, the protein was subjected to TEV protease treatment (*28*) to cleave the decahistidine tag (~0.5 mg TEV protease added per pellet equivalent from 800 mL of initial bacterial culture) for 2-2.5 hours at 4°C. The sample was then re-passaged through the column containing Ni-NTA beads to remove the cleaved decahistidine tag, the TEV protease and

uncleaved protein. A protein aliquot was subjected to size-exclusion chromatography (SEC) using either a 24 mL bed volume Superdex 200 Increase 5/150 GL column (GE Healthcare) or a TSKgel G4000SW$_{XL}$, 7.8 mm x 30 cm column (TOSOH Biosciences) in Final protein buffer to assess quality of the purified sample. Protein typically eluted as a sharp monodispersed peak, observed by monitoring A_{280} (Fig. S1B). 0.7-1 mg of purified protein could typically be obtained from 800 mL of initial bacterial culture. Selenomethionine-substituted protein was purified using the same protocol as the native protein, and yields were comparable.

Crystallization. Lipidic cubic phase (LCP) (*12*) mixes were prepared at room temperature by mixing concentrated protein with host lipid at 1:1.5 (w/w) protein:lipid ratio using coupled syringes. Total purified ArnT$_{Cm}$ (not subjected to SEC) was concentrated using a centrifugal concentrator (Millipore) with a 100 kDa MW cutoff to 30 to 35 mg/mL and mixed with different host lipids: a) For the "apo" structure, native protein was mixed with a lipid mixture containing 0.8% decaprenyl phosphate (DecaP; Indofine Chemical Co., Inc.) and 99.2% monoolein (NuChek Prep), b) For the UndP-ArnT$_{Cm}$ structure, native protein was mixed with a lipid mixture containing 1% undecaprenyl phosphate (UndP; Indofine Chemical Co., Inc.) and 99% monoolein, and c) Selenomethionine-substituted protein was mixed with 100% monoolein. The host lipid mixtures were prepared by dissolving appropriate amounts of both lipids in a chloroform methanol mixture (1:1 chloroform:methanol for UndP, 74:23:3 chloroform:methanol: water for decaP) and mixing them until homogeneous. The mixture was then evaporated under nitrogen gas and desiccated overnight in a vacuum desiccator.

A LCP Mosquito robot (TTP Labtech) was used to dispense 50 nL of protein/lipid mixture and 750 nL of precipitant solution onto a 96-well glass plate, which was sealed with a glass cover slip and stored in a 22°C incubator. Crystals typically appeared after 1 or 2 days and grew to full size in about 1 week. Initial crystallization trials were set up using the MemMeso HT screen (Molecular Dimensions). Crystals grew in (a) 30% (v/v) PEG 400, 0.1 M MES pH 6.0, 0.1 M Sodium chloride, 0.1 M Magnesium chloride hexahydrate, (b) 30% (v/v) PEG 500 DME, 0.1 M HEPES pH 7.0, 0.1 M Sodium chloride, 0.1 M Magnesium chloride hexahydrate and (c) 30% (v/v) PEG 500 DME, 0.1 M HEPES pH 7.0, 0.1 M Sodium chloride, 0.1 M Calcium chloride dihydrate. These conditions were further optimized to yield best quality crystals. Selenomethionine-substituted protein crystals showed a preference towards PEG 400 as precipitant. A tungsten carbide glass-cutter (Hampton Research) was used to cut the glass cover slip, and crystals were flash-frozen directly in liquid nitrogen without additional cryoprotection.

Both ArnT$_{Cm}$ and UndP-ArnT$_{Cm}$ crystallized in space group P21 21 21 with four monomers in the asymmetric unit. Unit cell parameters are given in Table S1.

Data collection and structure determination. X-ray diffraction data were collected on the NE-CAT beamlines 24-ID-C and 24-ID-E at the Advanced Photon Source (Argonne National Laboratory, Argonne, IL). For structure determination, six datasets collected from selenomethionine-substituted protein crystals were scaled together using XDS/XSCALE (*29*). Twenty one selenium sites were located using these data employing SHELXD via HKL2MAP software (*30*, *31*). The sites were input to the program PHASER (*32*) for SAD phasing. PHASER was able to locate six additional sites. The resulting phases showed helical features in the electron density map. Initial phases were improved by placing a model derived from PDB entry 3WAJ in the electron density map using the program MOLREP (*33*). One copy of the molecule could be clearly positioned in the electron density map. PHASER was run once again using the

selenium sites and the positioned model together. The resulting phases were subjected to density modification using the program RESOLVE (*34*), yielding phases of great quality that allowed tracing the molecule in the electron density map.

Data were collected from five isomorphous crystals for the "apo" structure, and from one crystal for the UndP-bound structure. The datasets were indexed, integrated, scaled and merged using XDS (*29*) and AIMLESS (*35*). The initial model was rigid-body fitted in the final datasets using the PHENIX crystallographic software package (*36*). The models were manually corrected and completed using Coot (*37*), and refined using PHENIX, alternating between cycles of manual building in Coot and refinement in PHENIX. The electron density maps for both structures featured many partially ordered lipid molecules, which most likely correspond to monoolein. As no head groups or identifying features were discernable in the densities, all unidentified lipids were modeled as isolated alkyl chains. The UndP-bound dataset contained densities for the head and tail of UndP (Fig. 3 and Fig. S8)**.** The "apo" dataset also had a density inside cavity 3 corresponding to the two terminal prenyls of decaP or an endogenous polyprenyl ligand (since a similar density was observed in a true "apo" dataset, from crystals prepared using monoolein alone as a host lipid, data not shown). The SeMet dataset did not feature a similar density. UndP was built from a SMILES string and had geometrical restrains generated using the program eLBOW (in the PHENIX suite). It was manually fitted to the densities and position-optimized during regular refinement of the model with PHENIX. UndP in the bound dataset was assigned the residue code 5TR. Many ordered water molecules can be seen in the "apo" structure and some in the UndP dataset, and have been added to each model. The final ArnT$_{Cm}$ model has an R_{work}/R_{free} of 0.21/0.26 and the final UndP-ArnT$_{Cm}$ model an R_{work}/R_{free} of 0.22/0.26. All protein structure figures were prepared using UCSF Chimera (*38*).

The densities presented in Fig. 3 and Fig. S8 correspond to a feature-enhanced omit map (*39*) for UndP. To remove the possibility of bias, the UndP ligand was set to zero occupancy, and the coordinates for the resulting model were perturbed using phenix.dynamics. The model was then refined against the UndP-bound dataset and a feature-enhance map was calculated in PHENIX.

In silico molecular docking. Initial structures for the 4-amino-4-deoxy-**L**-arabinose-phospate, with either zero, one, two or three prenyl group(s) attached to it were built using MarvinSketch. These four initial structures were prepared for docking using Ligprep (version 2.9, Schrödinger Inc.), with the OPLS 2.1 force field. Possible ionization states of the phosphate and amine groups around pH 7 (±2) were calculated using the Epik option.

The structure of ArnT$_{Cm}$ in complex with UndP was prepared for docking as follows, using the Protein Preparation Wizard (Schrodinger Release 2014-3). First, the UndP ligand was removed from the structure. Second, missing side chains were built (including for missing residues 275-277 which are located around the binding site). Finally, hydrogens were added and a minimization was performed on them to optimize hydrogen bonding.

We performed *in silico* docking of the prepared ligand structures into the prepared ArnT$_{Cm}$ structure (UndP-bound conformation), using Glide (version 6.2, Schrödinger Inc.). Standard precision flexible docking procedure was used, allowing the generation and docking of multiple ligand conformations. The search space was defined as a cube of 36 Å edge length centered on the center of mass of residues Asp 55, Asp 158 and Tyr 59. As part of the standard Glide search algorithm, the diameter midpoint of each ligand had to stay within a smaller concentric cube (size set at 16 Å edge length). The potential for the non-polar part of the ligands was "softened"

by scaling down the van der Waals radii of atoms with a partial charge smaller than 0.15, by a factor of 0.85. Finally, poses with the best energy were kept and post-docking minimization was performed.

Number of poses obtained is as follows: No prenyl (only sugar-phosphate) 57, 1 prenyl group 37, 2 prenyl groups 48, and 3 prenyl groups 38. The poses obtained from the ligand containing 3 prenyl groups show the best match with experimentally-derived positions for the phosphate and first three prenyl groups of UndP, and are presented in Fig. S13.

Isolation of labeled lipid A. Overnight bacterial cultures were diluted to an OD_{600} of 0.05 and labeled with 2.5 μCi ml^{-1} $^{32}P_i$ (Perkin Elmer) in 5 mL LB broth. The cells were harvested at OD_{600} 1.0. Labeled ^{32}P-lipid A isolation, separation and visualization were performed as previously described (*9*).

Isolation of lipid A and mass spectrometry. A total of 500 mL bacterial culture was grown in LB medium at 37°C. When the culture reached OD_{600} 1.0, cells were harvested by centrifugation at 10,000 x *g* for 10 minutes. Lipid A was prepared by mild acidic hydrolysis (*40*). Lipid A species were separated by Thin Layer Chromatography (TLC) Silica Gel 60 (Merck KGaA) using the solvent system chloroform/pyridine/88% formic acid/water (50:50:16:5). Lipid A was purified as previously described (*7*) and analyzed using a MALDI-TOF/TOF (ABI 4700 Proteomic Analyzer) mass spectrometer in the negative-ion linear mode, also as previously described (*9*).

Determination of polymyxin B Minimum Inhibitory Concentration (**MIC**)**.** Overnight cultures were diluted 100-fold in fresh LB medium and grown until OD_{600} reached 0.5. To determine the MIC value, these cultures were diluted again to OD_{600} 0.05 in LB containing a range of the antimicrobial peptide concentrations. Cultures were incubated for 3 hours at 37°C and 200 r.p.m. The MIC was defined as the lowest concentration of polymyxin B that reduced the cell growth (A_{600}) by 50% compared to the growth in LB without polymyxin B. The table in Fig. S2 indicates the MIC ranges after 3 biological repetitions.

Mutagenesis. Mutations of $ArnT_{Se}$ (in pET21 vector) or $ArnT_{Cm}$ (in pWSK29 vector) were generated with whole-plasmid PCR using Phusion High-Fidelity PCR Master Mix with HF buffer (New England Biolabs) or QuikChange Lightning site-directed mutagenesis kit (Agilent Technologies) and custom primers. The "wild-type" $ArnT_{Se}$ used carries the background mutation P547A. This mutation has been shown to improve the yield of purified protein (*13*), and is present in every mutant examined here using the polymyxin B growth assay. All mutants shown for $ArnT_{Se}$ and $ArnT_{Cm}$ have been tested for total protein expression and all express to comparable levels to their respective wt ArnT (data not shown).

Polymyxin B growth assay. The *in vivo* polymyxin B growth assay used to assess function of $ArnT_{Se}$ mutants has been described previously (*13, 14*). Briefly, plasmids encoding mutant $ArnT_{Se}$ proteins were electroporated into BL21(DE3)ΔarnT *E. coli* cells, plated onto LB/ampicillin plates and grown overnight at 37°C. 10^5 cells from a single colony were inoculated into 2 mL LB/ampicillin and incubated overnight at 37°C with shaking in the presence and absence of 2 μg/mL polymyxin. Growth was continuously quantitated spectrophotometrically at 600 nm with a Varioscan microplate reader (Thermo Scientific).

Assays were carried out twice and in duplicate for each mutation. Empty vector (pET21) and P547A ArnT$_{Se}$ (wt) were included in all experiments as polymyxin-sensitive and polymyxin-resistant controls, respectively. All error bars in bar graphs represent standard deviation.

Energy-dispersive X-ray spectroscopy (**EDS**). EDS was carried out using a silicon drift detector (model X-123SDD, Amptek Inc, USA) at NE-CAT beamline 24-ID-C at the Advanced Photon Source (Argonne National Laboratory, Argonne, IL). The built-in multi-channel analyzer (MCA) of X-123SDD was calibrated with known fluorescent emission lines of several metals. The gain of the detector was set to 75%, corresponding to an energy range of 0-16.7 keV for the MCA. EDS experiments were carried out with incident X-ray energy of 12.66 keV, just above the K absorption edge of Selenium. First, an EDS spectrum was recorded for approximately 120 seconds with X-rays incident on a cryo-loop containing several crystals of ArnT$_{Cm}$ (grown in the absence of substrate lipids) (Fig. S7C). As a negative control experiment, EDS was also recorded for the same duration on a cryo-loop containing just the monoolein lipid and crystallization solution (Fig. S7D).

Fig. S1.

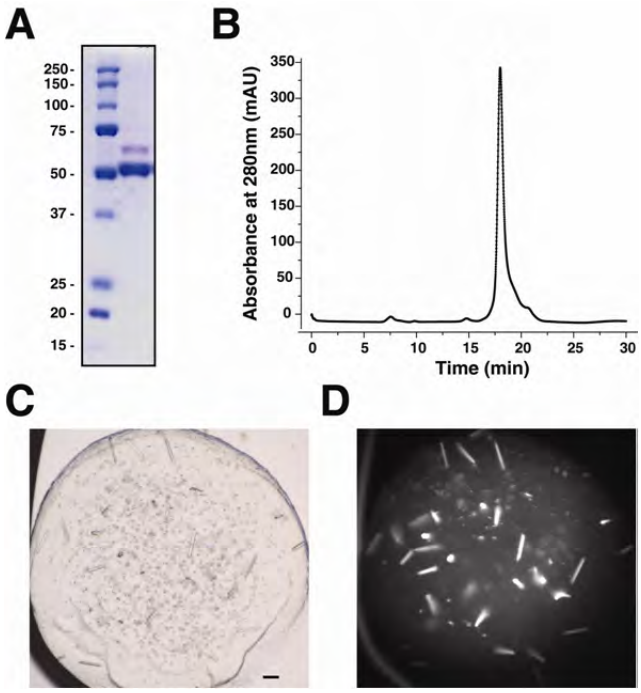

Fig. S1. Purification and crystallization of ArnT$_{Cm}$. (**A**) Representative coomassie-stained SDS-PAGE gel of purified ArnT$_{Cm}$. ArnT$_{Cm}$ (63 kDa) after treatment with TEV protease to cleave the N-terminal poly-histidine tag runs slightly above the 50 kDa marker. The higher band corresponds to uncleaved protein. (**B**) Representative elution profile from analytical size-exclusion chromatography (TSKgel G4000SW$_{XL}$, 7.8 mm x 30 cm column, TOSOH Biosciences, LLC) showing a monodispersed peak. (**C**) ArnT$_{Cm}$ rod-shaped crystals, reaching 100-200 μm in the longest dimension. The black scale bar corresponds to 100 μm. (**D**) UV exposure of the crystals from (C).

Fig. S2

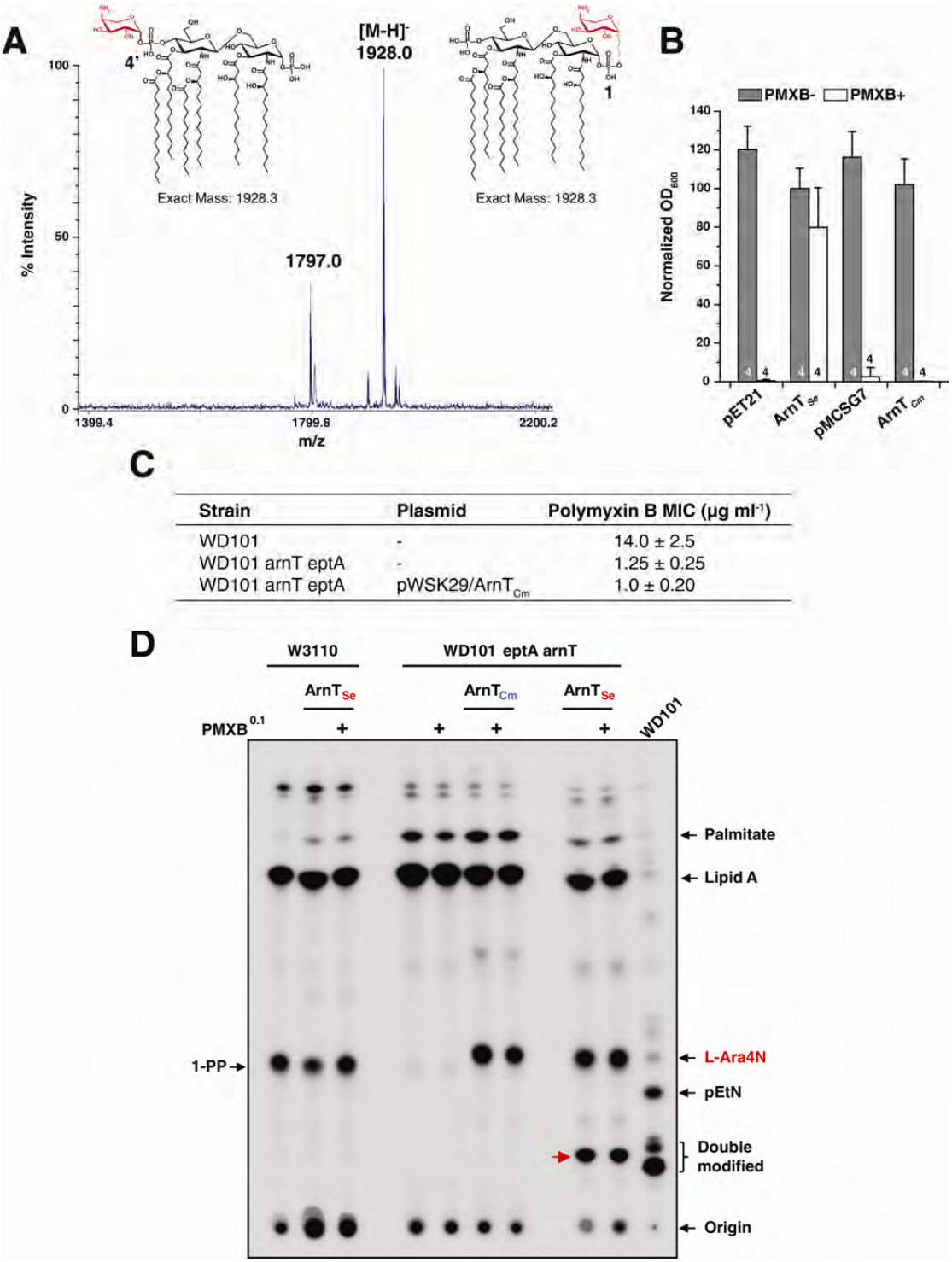

Fig. S2. Functional characterization of ArnT$_{Cm}$ (A) Negative ion MALDI-TOF spectrometry of lipid A purified from *E. coli* WD101 Δ*arnT*Δ*eptA* strain expressing ArnT$_{Cm}$. The major peak

at 1928.0 corresponds to lipid A modified with **L**-Ara4N added at either the 4' (shown on the left, in red) or 1 (shown on the right, in red) acceptor position. The secondary peak at 1797.0 corresponds to *bis*-phosphorylated hexa-acylated lipid A. (**B**) Bar graph showing the effect of introduction of $ArnT_{Se}$ and $ArnT_{Cm}$ into the BL21(DE3)Δ*arnT* knockout *E. coli* strain. Only $ArnT_{Se}$ is able to rescue polymyxin resistance. $ArnT_{Cm}$ was expressed here using the pMCSG7 vector, and not pNYCOMPS-Nterm used for crystallography-related protein production, because of matching antibiotic resistance of the former vector with pET21 used for $ArnT_{Se}$ expression. Data presented are mean + SD. *N* is shown for each data column. (**C**) Determination of minimal inhibitory concentration of polymyxin B (PMXB) in WD101 strain (PMXB-resistant) and Δ*arnT*Δ*eptA* double-mutant (PMXB-sensitive). Expression of $ArnT_{Cm}$ in Δ*arnT*Δ*eptA* does not rescue resistance to PMXB. (**D**) Expanded analysis of ^{32}P-labeled lipid A by thin-layer chromatography (TLC). Lipid A in W3110 strain carries only a pyrophosphate modification (1-PP) and is sensitive to polymyxin B (PMXB). WD101 produces double-modified lipid A with Ara4N and phosphoethanolamine (pEtN). In the absence of both *arnT* and *eptA* (a gene required for the addition of phosphoethanolamine to the 1 phosphate group of lipid A), $ArnT_{Se}$ adds **L**-Ara4N to both the 1 and 4' phosphates, leading to both single- and double-modified lipid A spots (double-modified **L**-Ara4N lipid A indicated by red arrow). In contrast, $ArnT_{Cm}$ only yields a single-modified lipid A species. Moreover, in a Δ*arnT* single knockout strain, $ArnT_{Cm}$ is unable to modify lipid A (data not shown). The conclusion is that $ArnT_{Cm}$ is stereospecific for the 1 phosphate group of lipid A. Pre-induction with PMXB (0.1 μg/mL) did not significantly change the amount and/or character of lipid A modification.

Fig. S3

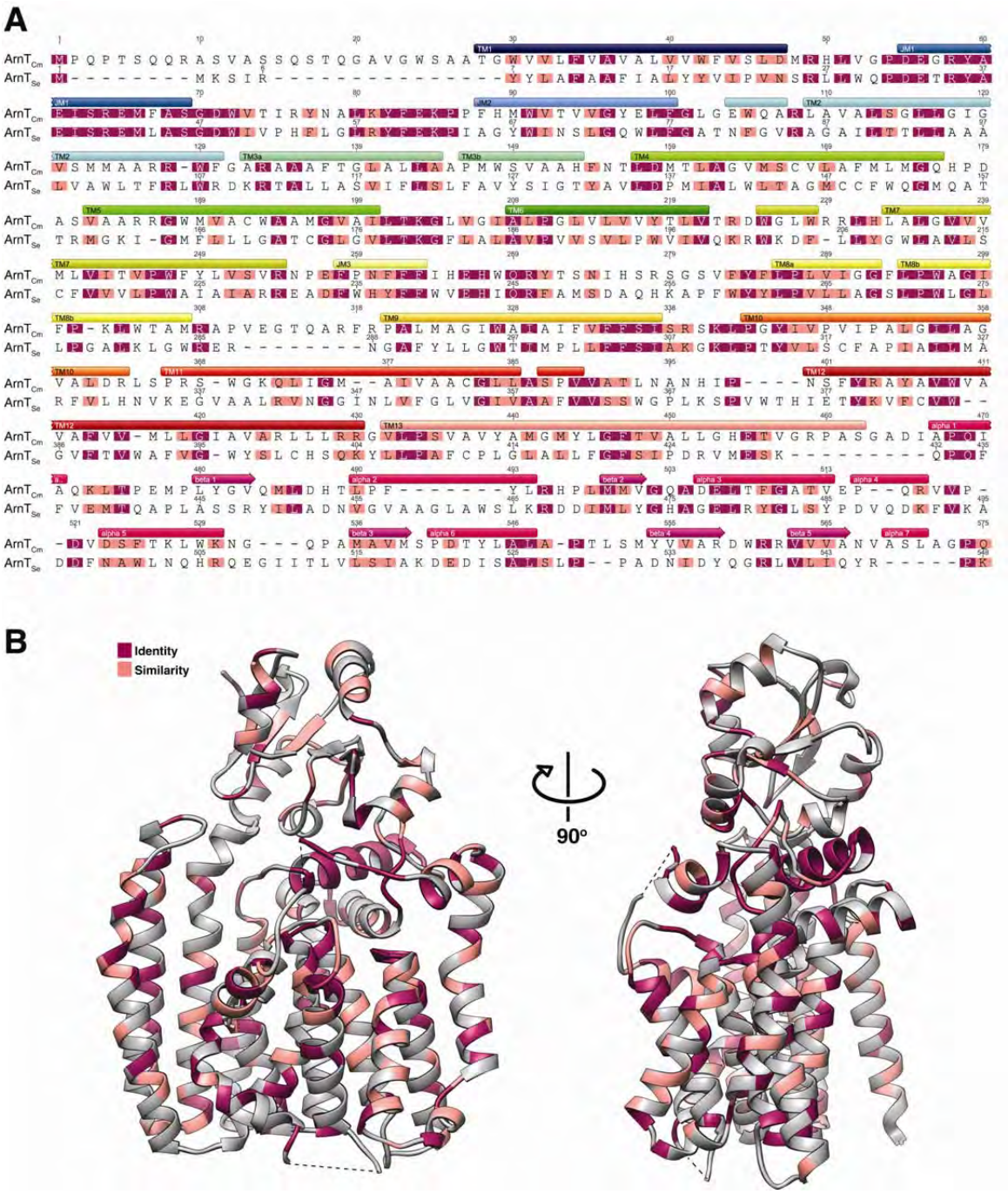

Fig. S3. Homology between ArnT$_{Cm}$ and ArnT$_{Se}$. (**A**) Sequence alignment between ArnT$_{Cm}$ and ArnT$_{Se}$ was produced using the MAFFT server (G-INS-i) and manually curated, also using ArnT sequences from *E. coli* and *B. cenocepacia* as evolutionary intermediates (not shown). The secondary structure is annotated according to the color scheme used in Fig. 1C for ArnT$_{Cm}$. Identical residues are shown in maroon and similar residues (according to the BLOSUM62

matrix) in pink. (**B**) Identity and similarity between ArnT$_{Cm}$ and ArnT$_{Se}$ color-mapped on the ArnT$_{Cm}$ structure.

Fig. S4

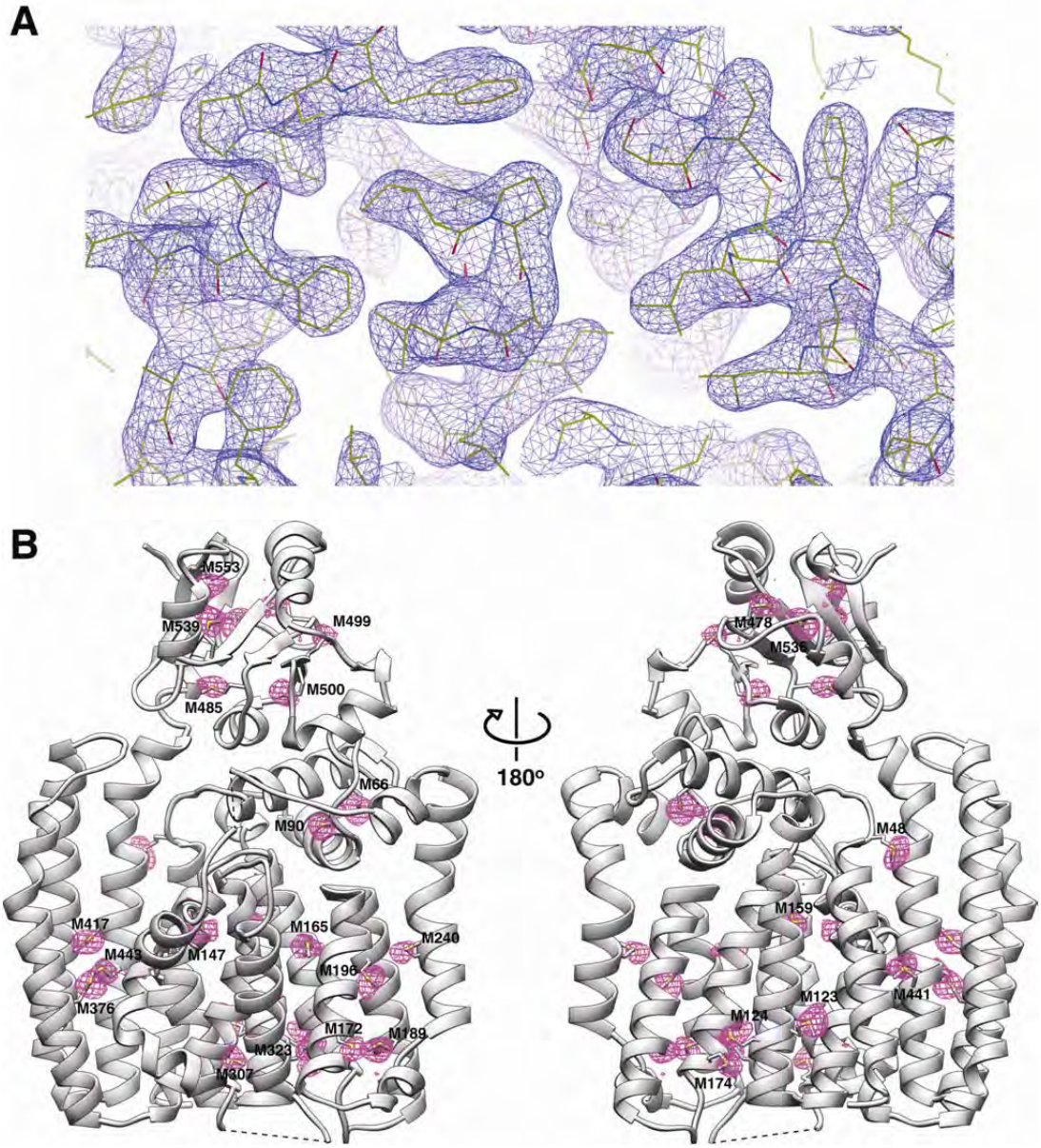

Fig. S4. Representative electron density and selenomethionine markers used for sequence assignment. (**A**) Sample electron density from a $2F_o$-F_c electron density map contoured at 1.5 RMSD. (**B**) An anomalous difference map calculated from the original selenomethionine dataset used for phasing, contoured at 3 RMSD within 5 Å from the model, is displayed as a magenta mesh. 26 anomalous peaks can be seen, corresponding to the 26 methionines present in the model, which are labeled.

Fig. S5

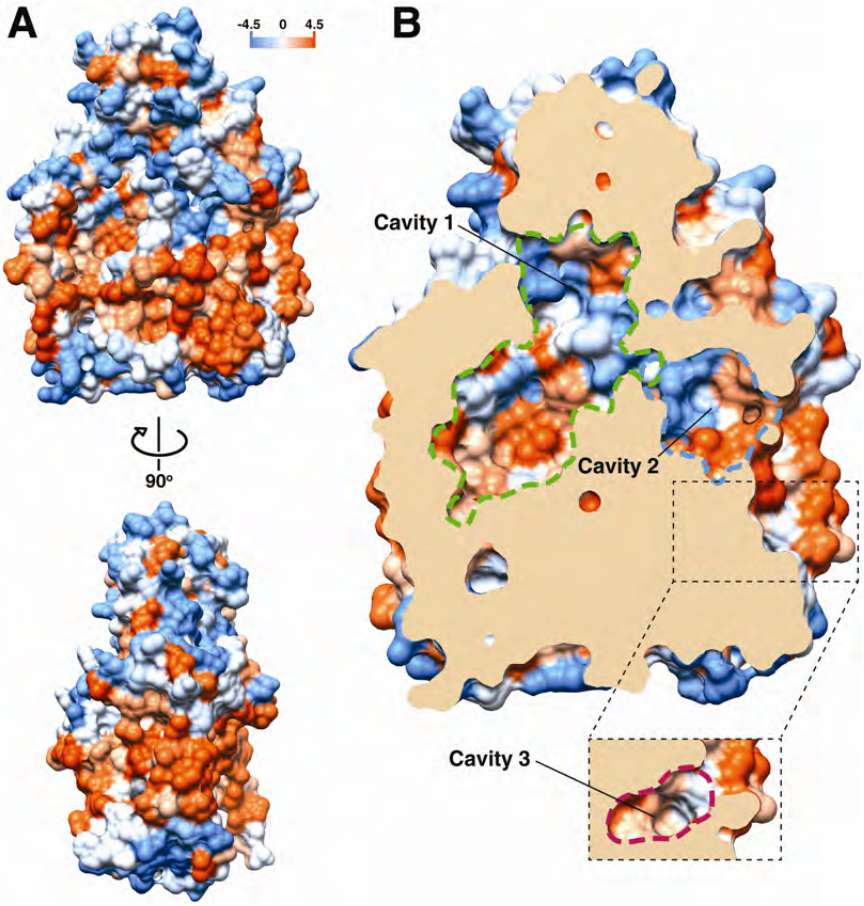

Fig. S5. Hydrophobicity profile of ArnT$_{Cm}$ surface and cavities. (**A**) Surface representation colored for hydrophobicity in the Kyte-Doolittle scale, ranging from blue (-4.5, polar) to white (0.0) to orange red (4.5, hydrophobic). (**B**) Slice of the ArnT$_{Cm}$ surface presenting the hydrophobicity of the inner cavities. The cavities are delineated by dashed lines, color-coded according to the cavity volume color scheme in Fig. 2A. Cavity 3 is located at a different depth in respect to the other two cavities and is presented in an inset for this reason.

Fig. S6

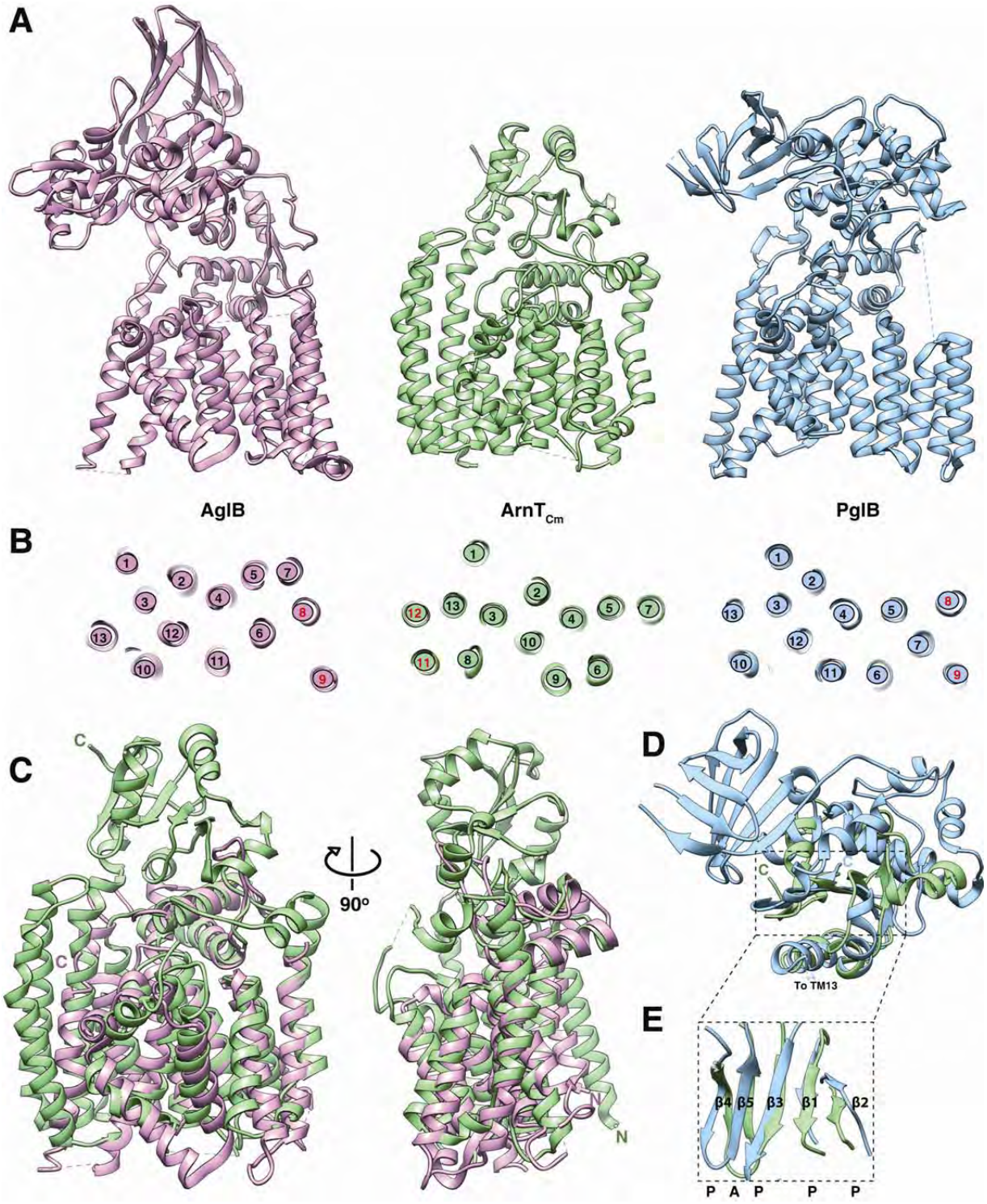

Fig. S6. Structural fold comparison between ArnT$_{Cm}$ and protein glycosyltransferases PglB and AglB. (**A**) Side-by-side comparison of *A. fulgidus* AglB (PDB: 3WAJ), ArnT$_{Cm}$ and *C. lari* PglB (PDB: 3RCE). All three proteins exhibit a two-domain architecture, with a TM domain containing 13 TMs and a periplasmic (extracellular for AglB) soluble-like domain. The flexible loop that covers the active site is located between TMs 7 and 8 in ArnT$_{Cm}$ and TMs 9 and 10 in AglB and PglB, but is maintained as a structural feature in all three enzymes. (**B**) Comparison of TMD organization for the three proteins through slices perpendicular to the transmembrane helices. The color scheme is maintained from (A). The sequence of the helices is given numerically, with the helices that are positioned uniquely for each protein highlighted in red. Overall, ArnT$_{Cm}$ has two unique helices on the left side (in the orientation presented here) (TM 11 and 13), whereas AglB and PglB have two unique TMs on the right side (TM 8 and 9) with slightly different arrangements. (**C**) Structural superposition of ArnT$_{Cm}$, in green and part of the AglB TMD, in purple (Overall RMSD: 2.24 Å for 257 residues). The TMD structure between TMs 2 and 7 (including the two juxtamembrane helices) superimposes well between the two molecules, and TM8 from ArnT$_{Cm}$ superimposes with TM10 from AglB. (**D**) Structural superposition of ArnT$_{Cm}$ and PglB periplasmic domains (Overall RMSD: 2.11 Å for 70 residues). The non-superimposable part of the PglB PD consists of two expansions of the loops between strands β2/β3 and β3/β4. (**E**) Detailed view showing superposition of the beta strands within the core of the PD. The arrangement of the beta strands is conserved in both proteins. Parallel and antiparallel strands are labeled as P and A respectively. All structural superpositions were prepared using the Chimera Matchmaker tool. A recent report also suggests a similarity between the folds of ArnT and PglB based on a bioinformatics approach (*41*).

Fig. S7

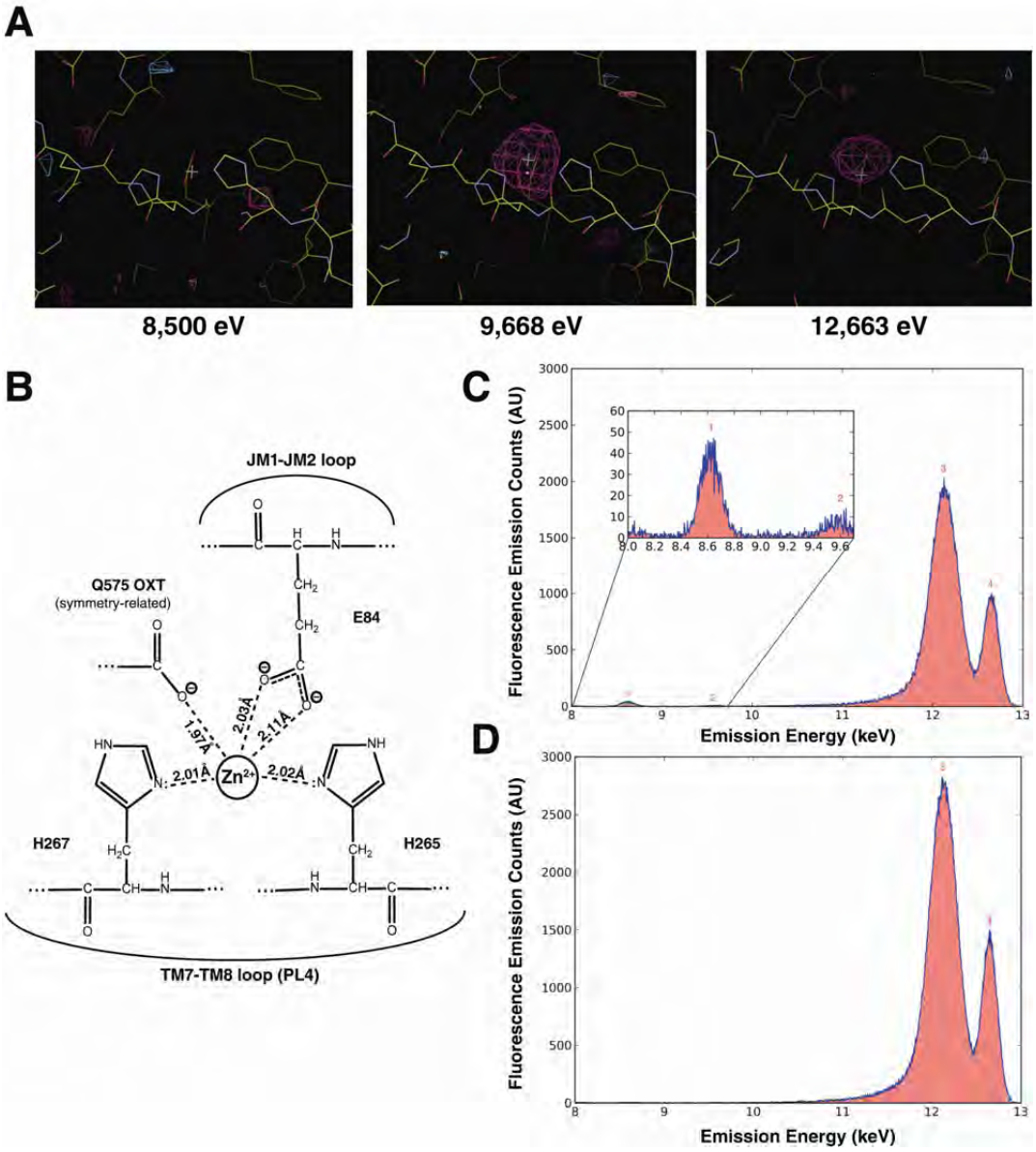

Fig. S7. Anomalous diffraction, coordination diagram and independent verification for the Zn^{+2} bound to $ArnT_{Cm}$. (**A**) Anomalous difference maps, calculated from datasets collected from isomorphous crystals at the three energies indicated, and contoured at 3 RMSD, are shown in proximity to the divalent coordination site. The second dataset collected at 9,668 eV, just above the K edge of Zn (9,658 eV), gives the strongest signal. (**B**) Schematic of metal coordination observed in the $ArnT_{Cm}$ structure, compatible with Zn^{+2} (*42*). (**C**) Energy-dispersive X-ray spectroscopy (EDS) was used to independently verify the identity of the metal bound to $ArnT_{Cm}$. An EDS spectrum collected on a cryo-loop containing crystals of $ArnT_{Cm}$ (in the apo state) is shown. Significant fluorescent emission lines are labeled 1 to 4. Following are the peak identities. Peak 1: 8.62keV, Zn K-alpha1; peak 2: 9.57keV, Zn K-alpha2; peak 3: 12.11keV,

Compton scattered X-rays; peak 4: 12.66 keV, incident X-rays. The inset shows a close-up of the region where Zn fluorescent emission lines are observed. (**D**) EDS spectrum collected on a cryo-loop containing monoolein lipid and crystallization solution only (negative control). Fluorescent emission lines were labeled as in (**C**). Only peaks 3 and 4 are observed, while Zn peaks are absent. Both anomalous diffraction and EDS identify the bound metal as zinc.

Fig. S8

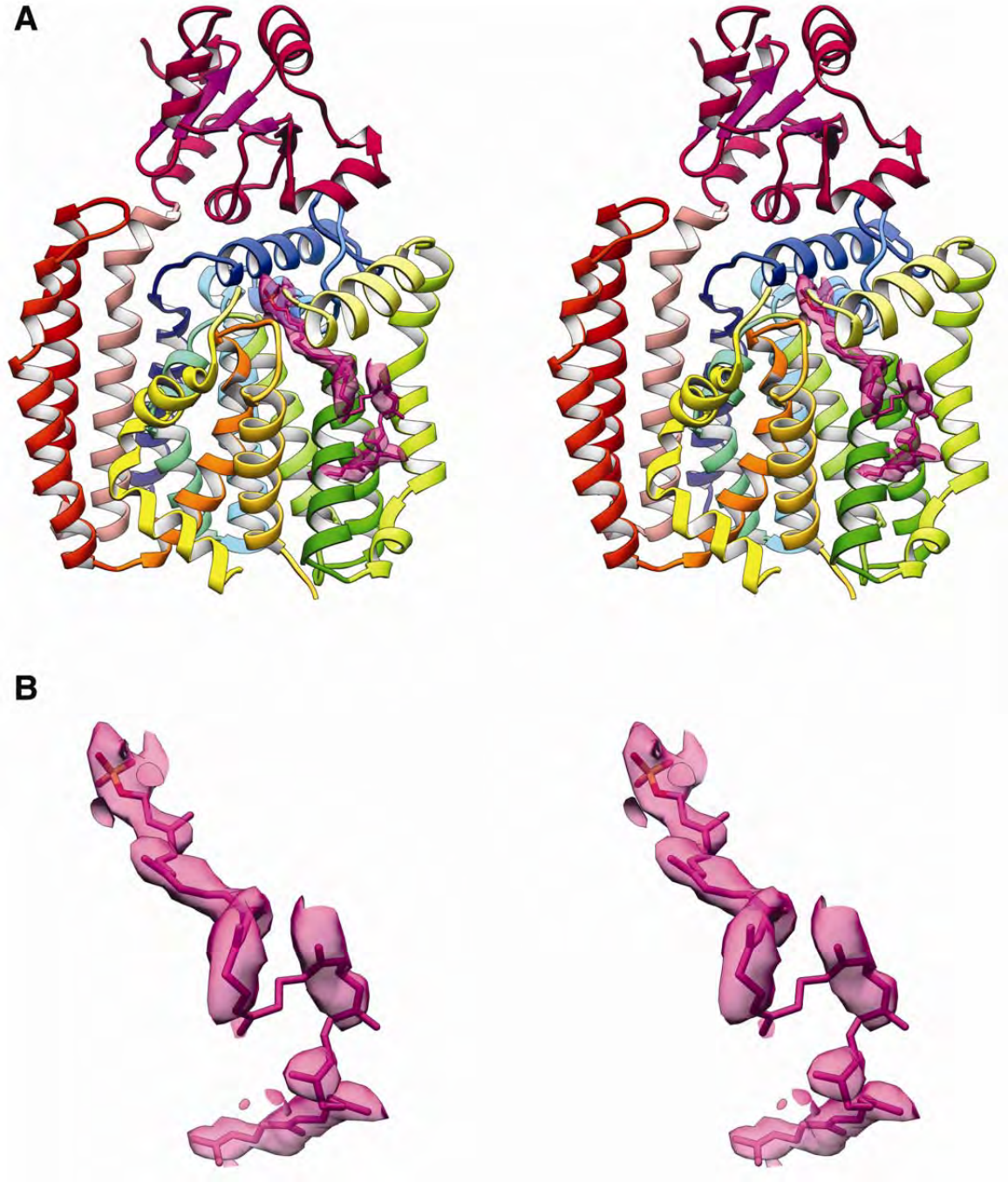

Fig. S8. Stereo representation of UndP bound to ArnT$_{Cm}$. (**A**) Wall-eyed stereo representation of the UndP-bound structure. The protein color scheme is as shown in Fig. 1A, whereas the UndP is shown in magenta. The lipid densities shown correspond to a 2Fo-Fc omit map, contoured at 1 RMSD, and drawn with a 2.3 Å zone around the UndP model. They are

represented as semi-transparent surfaces, also in magenta. (**B**) Wall-eyed stereo representation of UndP and corresponding densities, contoured and presented as above.

Fig. S9

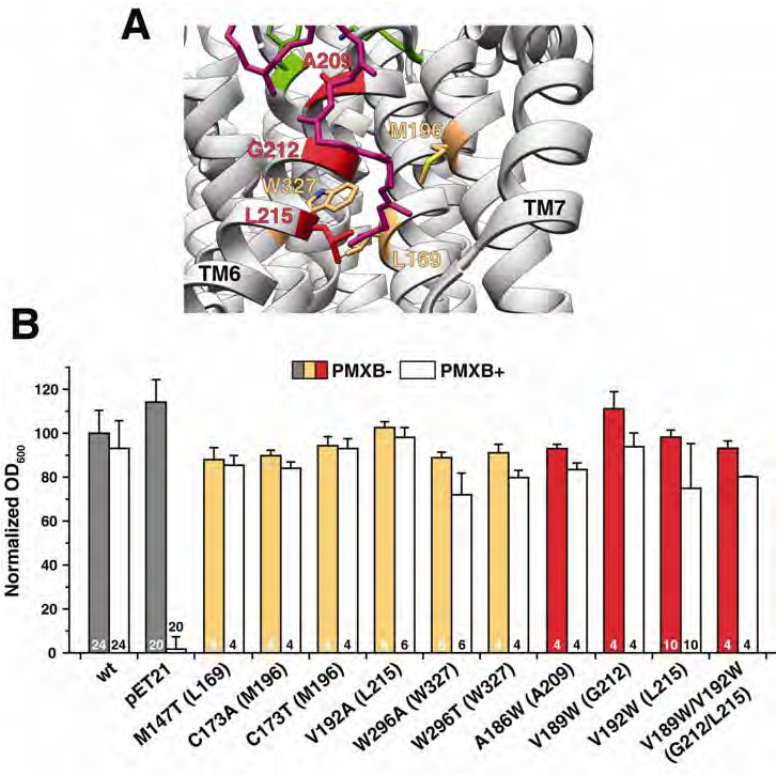

Fig. S9. Probing the functional importance of cavity 3. (**A**) Four residues lining cavity 3, shown in yellow, were mutated to either Ala or Thr. Three residues shown in red, located on the face of TM6 and oriented to face cavity 3, were mutated to Trp in ArnT$_{Se}$, in an attempt to block entry of UndP into the cavity. (**B**) Bar graph showing the effect of the mutations presented in (A) on ArnT$_{Se}$, tested using the PMXB growth assay**.** Only minor changes in PMXB growth were observed. Data presented are mean + SD. *N* is shown for each data column.

Fig. S10

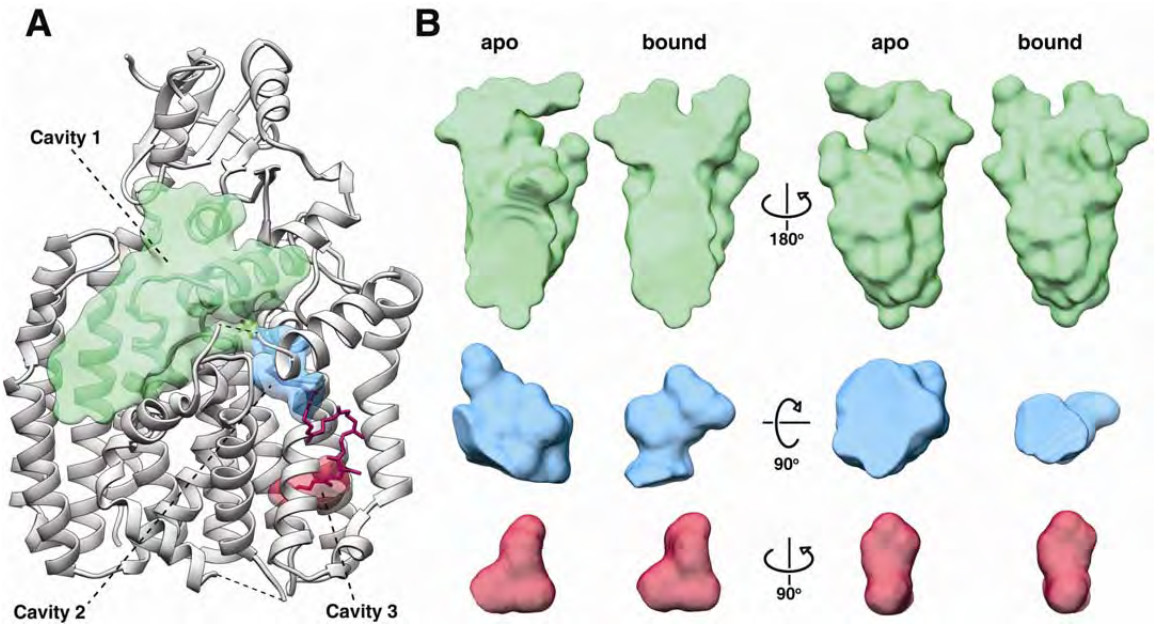

Fig. S10. Cavities of the ArnT$_{Cm}$ UndP-bound structure. (**A**) Ribbon representation of UndP-bound ArnT$_{Cm}$ showing the volumes of cavities 1, 2 and 3 in the structure. Volumes were calculated as in Fig. 2A, using the Voss Volume Voxelator server (*25*), using probes with 15 and 1.75 Å radii, corresponding to the outer and inner probe respectively. UndP is shown in stick representation, colored in magenta. (**B**) Side by side comparison of cavity 1, 2 and 3 volumes between apo and UndP-bound structures, each shown in two different orientations. The color scheme is the same as in (A).

Fig. S11

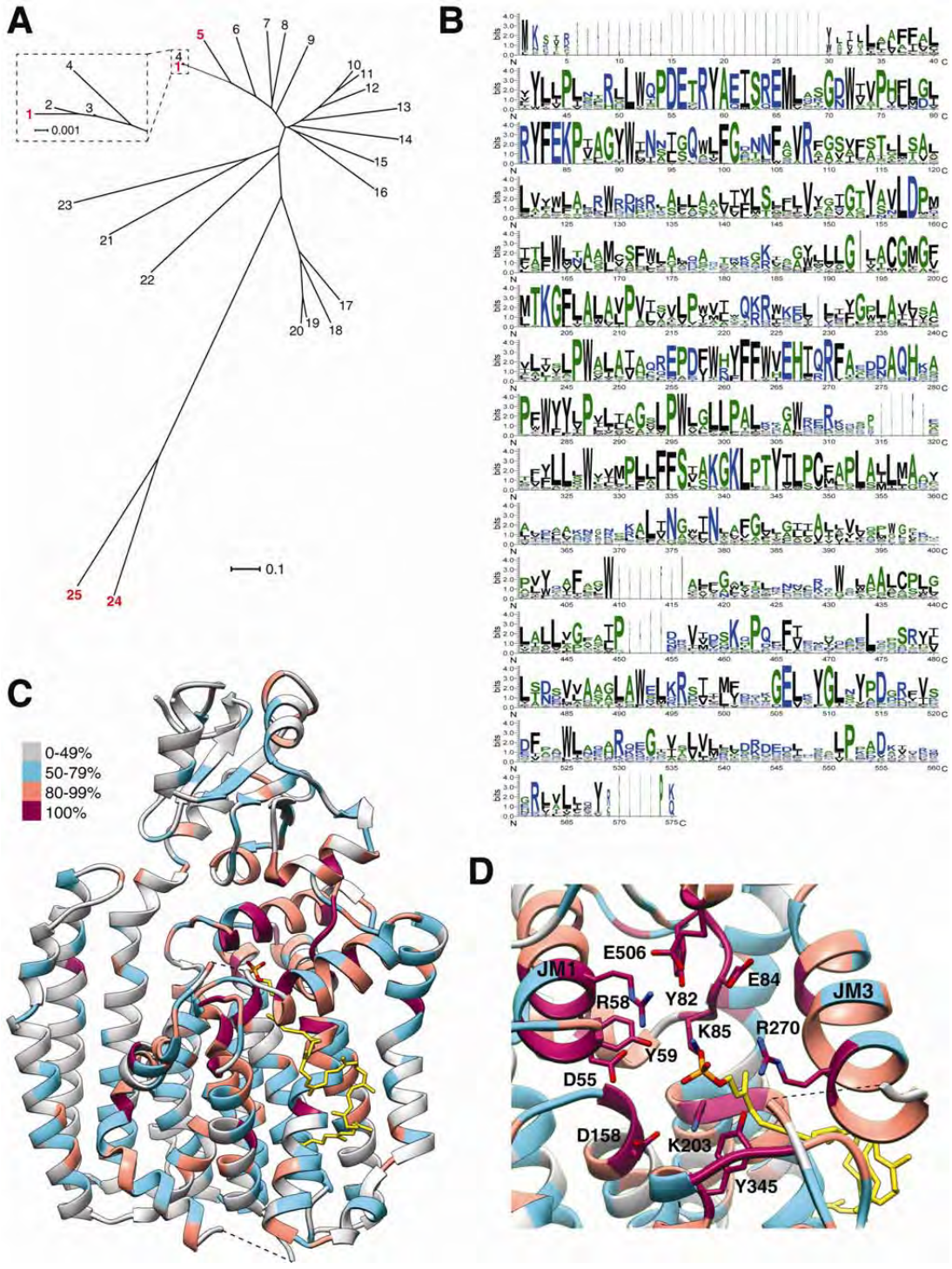

Fig. S11. Conservation within the ArnT family. (**A**) Phylogenetic tree showing genetic relationship between 25 ArnTs. Starting from 4 ArnTs from *E. coli*, *S. enterica* serovar

typhimurium, B. cenocepacia and *C. metallidurans CH34* (all shown in red), 100 other ArnTs were selected using MAFFT homolog (*43*). Closely related sequences were removed using CD-HIT, and the resulting list was manually curated to ensure appropriate genetic coverage. The final 25 genes shown are the following: 1. P76473; *E. coli*. 2. B2TW40; *S. boydii*. 3. Q3YZU9; *S. sonnei*. 4. Q32DT1; *S. dysenteriae*. 5. O52327; *S. enterica serovar typhimurium*. 6. A6TF96; *K. pneumoniae*. 7. A4WAM1; *Enterobacter sp.* 8. B2VBI7; *E. tasmaniensis*. 9. C5BDQ8; *E. ictaluri*. 10. Q8ZDX9; *Y. pestis*. 11. A1JPM4; *Y. enterocolitica*. 12. A8GDR9; *S. proteamaculans*. 13. Q2NWA9; *S. glossinidius*. 14. Q7N3Q9; *P. luminescens*. 15. C6DAW3; *P. carotovorum*. 16. A4SQX1; *A. salmonicida*. 17. Q9HY61; *P. aeruginosa*. 18. Q4KC80; *P. fluorescens*. 19. Q48HY9; *P. savastanoi*. 20. Q4ZSZ0; *P. syringae*. 21. B4EUL1; *P. mirabilis*. 22. A8FRR0; *S. sediminis*. 23. Q8D339; *W. glossinidia brevipalpis*. 24. A2VXS5; *B. cenocepacia*. 25. Q1LDT6; *C. metallidurans CH34*. The 25 sequences were aligned using the G-INS-i algorithm in MAFFT. ClustalW2 phylogeny (*44*) was used to generate the tree, which was depicted in an radial representation using Dendroscope (*45*). (**B**) Conservation logo. Polar, neutral and aliphatic residues are shown in blue, green and black, respectively. The alignment from (A) was cropped to the length of $ArnT_{Cm}$ (eliminating loop extensions). The logo was produced using the weblogo3 server (*46*). The last ~1/3 of the sequence logo corresponding to the last 2 TMs and the PD may be less reliable due to weak alignment between $ArnT_{Cm}$ and the other ArnTs used. (**C**) Ribbon representation of UndP-bound $ArnT_{Cm}$ structure colored by percent conserved residues in the 25 sequences described above. (**D**) Detail from (C) showing the active site region, colored as in (C). Experimentally determined UndP is shown in yellow. Residues discussed in the text are labeled.

Fig. S12

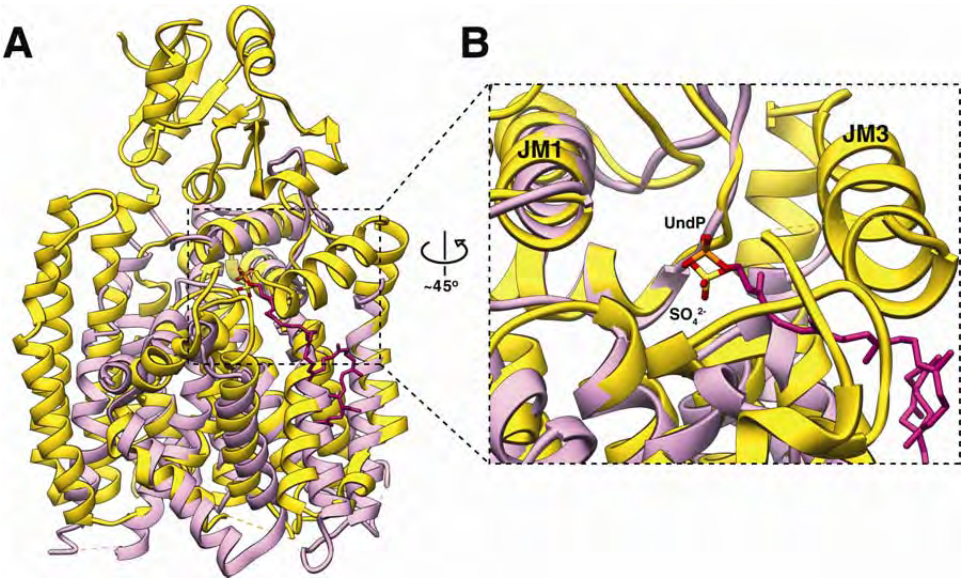

Fig. S12. AglB sulfate superposition on the ArnT$_{Cm}$ structure. (**A**) AglB was crystallized with a sulfate molecule, thought to correspond to the phosphate moiety of UndP (*20*). To compare its position with that of the phosphate moiety of UndP in ArnT$_{Cm}$, we superimposed the UndP-ArnT$_{Cm}$ structure with the inner core of AglB (similarly to Fig. S6C), and shown here with ArnT$_{Cm}$ in yellow AglB in light purple and UndP in dark purple (Overall RMSD: 2.27 Å for 259 residues). (**B**) Detail of SO_4^{2-} position compared to the phosphate moiety from UndP observed in the UndP-ArnT$_{Cm}$ structure.

Fig. S13

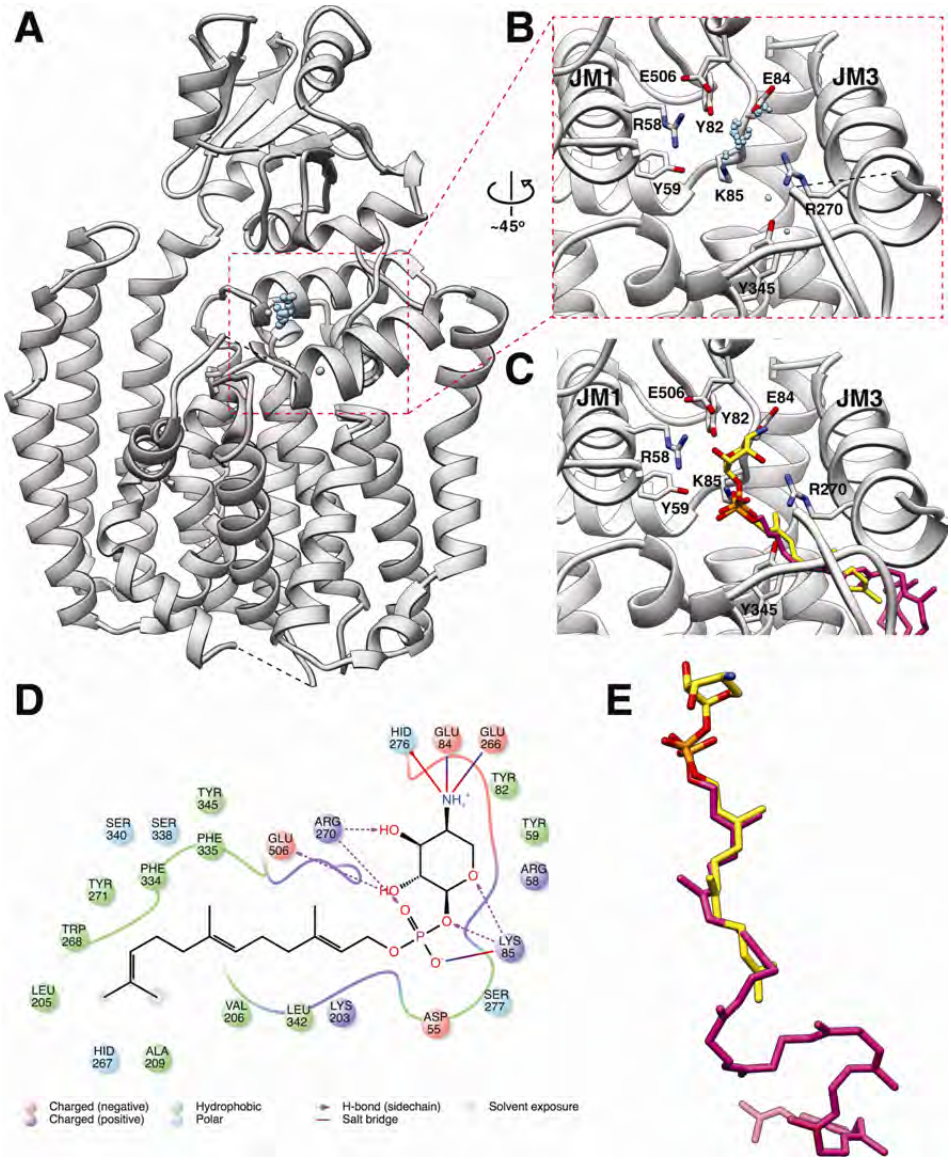

Fig. S13. *In silico* docking of L-Ara4N sugar on ArnT$_{Cm}$. (**A**) 38 poses generated by *in silico* docking of L-Ara4N-tri-prenyl phosphate on the ArnT$_{Cm}$ structure are shown with a blue dot each, marking the center of mass of the sugar ring (C1, C2, C3, C4, C5 and O1 atoms). (**B**) The inset shows a close-up view, with potential binding residues depicted in stick representation, labeled and colored by heteroatom. (**C**) Selected pose for L-Ara4N-tri-prenyl phosphate binding, based on energy, position of the phosphate moiety and of the 3 prenyl groups. Potential sugar binding residues as well as the phosphate binding residues determined experimentally are shown as in (**B**). (**D**) Schematic representation of the predicted L-Ara4N-triprenyl binding mode on ArnT$_{Cm}$ showing significant residue interactions. A key for different types of interactions is shown at the bottom of the panel. (**E**) Detail of the experimentally determined UndP (magenta)

and docked **L**-Ara4N-tri-prenyl phosphate (yellow), showing almost complete overlap for the phosphate and 3 prenyl groups.

Fig. S14

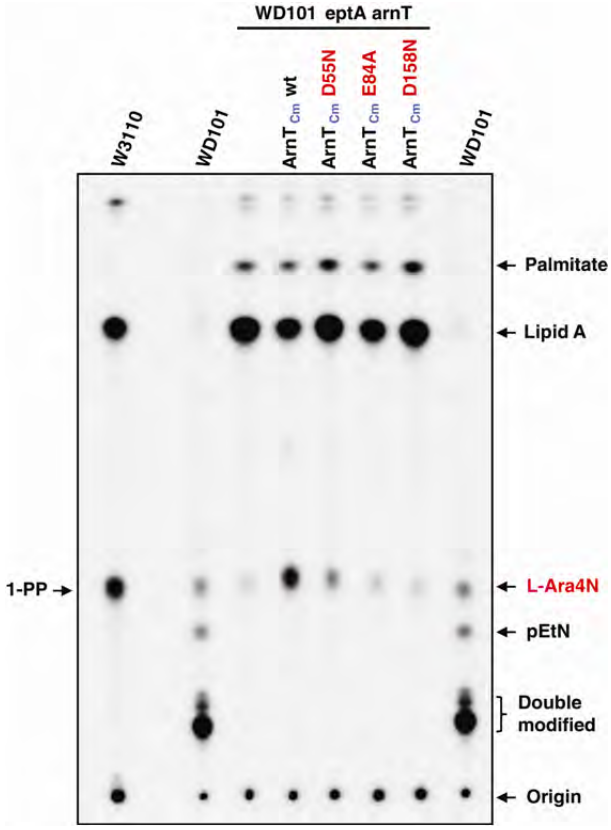

Fig. S14. Mutations of ArnT$_{Cm}$ abrogate lipid A modification in vitro. Analysis of ^{32}P-labeled lipid A by thin-layer chromatography (TLC). The different lipid A species and background strains are the same as used and described for the experiments shown in Fig. S2D. Mutants of the two putative catalytic residues (D55N and D158N) and of the metal-coordinating glutamate (E84A) were introduced in *arnT* and *eptA E. coli* knockout (WD101 *ΔarnTΔeptA*). E84A and D158N were unable to rescue lipid A modification by L-Ara4N, whereas D55N shows only minimal rescue of lipid A modification.

REFERENCES

1. A. P. Zavascki, L. Z. Goldani, J. Li, R. L. Nation, Polymyxin B for the treatment of multidrug-resistant pathogens: A critical review. *J. Antimicrob. Chemother.* **60**, 1206–1215 (2007). Medline doi:10.1093/jac/dkm357

2. D. Landman, C. Georgescu, D. A. Martin, J. Quale, Polymyxins revisited. *Clin. Microbiol. Rev.* **21**, 449–465 (2008). Medline doi:10.1128/CMR.00006-08

3. J. Mares, S. Kumaran, M. Gobbo, O. Zerbe, Interactions of lipopolysaccharide and polymyxin studied by NMR spectroscopy. *J. Biol. Chem.* **284**, 11498–11506 (2009). Medline doi:10.1074/jbc.M806587200

4. C. R. H. Raetz, C. M. Reynolds, M. S. Trent, R. E. Bishop, Lipid A modification systems in gram-negative bacteria. *Annu. Rev. Biochem.* **76**, 295–329 (2007). Medline doi:10.1146/annurev.biochem.76.010307.145803

5. K. A. Brogden, Antimicrobial peptides: Pore formers or metabolic inhibitors in bacteria? *Nat. Rev. Microbiol.* **3**, 238–250 (2005). Medline doi:10.1038/nrmicro1098

6. B. D. Needham, M. S. Trent, Fortifying the barrier: The impact of lipid A remodelling on bacterial pathogenesis. *Nat. Rev. Microbiol.* **11**, 467–481 (2013). Medline doi:10.1038/nrmicro3047

7. M. S. Trent, A. A. Ribeiro, S. Lin, R. J. Cotter, C. R. H. Raetz, An inner membrane enzyme in *Salmonella* and *Escherichia coli* that transfers 4-amino-4-deoxy-L-arabinose to lipid A: Induction on polymyxin-resistant mutants and role of a novel lipid-linked donor. *J. Biol. Chem.* **276**, 43122–43131 (2001). Medline doi:10.1074/jbc.M106961200

8. M. S. Trent, A. A. Ribeiro, W. T. Doerrler, S. Lin, R. J. Cotter, C. R. Raetz, 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). Medline doi:10.1074/jbc.M106962200

9. C. M. Herrera, J. V. Hankins, M. S. Trent, Activation of PmrA inhibits LpxT-dependent phosphorylation of lipid A promoting resistance to antimicrobial peptides. *Mol. Microbiol.* **76**, 1444–1460 (2010). Medline doi:10.1111/j.1365-2958.2010.07150.x

10. H. Lee, F.-F. Hsu, J. Turk, E. A. Groisman, The PmrA-regulated pmrC gene mediates phosphoethanolamine modification of lipid A and polymyxin resistance in *Salmonella enterica*. *J. Bacteriol.* **186**, 4124–4133 (2004). Medline doi:10.1128/JB.186.13.4124-4133.2004

11. F. Mancia, J. Love, High-throughput expression and purification of membrane proteins. *J. Struct. Biol.* **172**, 85–93 (2010). Medline doi:10.1016/j.jsb.2010.03.021

12. M. Caffrey, Crystallizing membrane proteins for structure determination: Use of lipidic mesophases. *Annu. Rev. Biophys.* **38**, 29–51 (2009). Medline doi:10.1146/annurev.biophys.050708.133655

13. N. A. Impellitteri, J. A. Merten, L. E. Bretscher, C. S. Klug, Identification of a functionally important loop in *Salmonella typhimurium* ArnT. *Biochemistry* **49**, 29–35 (2010). Medline doi:10.1021/bi901572f

14. L. E. Bretscher, M. T. Morrell, A. L. Funk, C. S. Klug, Purification and characterization of the L-Ara4N transferase protein ArnT from *Salmonella typhimurium*. *Protein Expr. Purif.* **46**, 33–39 (2006). Medline doi:10.1016/j.pep.2005.08.028

15. T. W. Cullen, D. K. Giles, L. N. Wolf, C. Ecobichon, I. G. Boneca, M. S. Trent, *Helicobacter pylori* versus the host: Remodeling of the bacterial outer membrane is required for survival in the gastric mucosa. *PLOS Pathog.* **7**, e1002454 (2011). Medline doi:10.1371/journal.ppat.1002454

16. X. Wang, S. C. McGrath, R. J. Cotter, C. R. H. Raetz, Expression cloning and periplasmic orientation of the *Francisella novicida* lipid A 4′-phosphatase LpxF. *J. Biol. Chem.* **281**, 9321–9330 (2006). Medline doi:10.1074/jbc.M600435200

17. F. Tavares-Carreón, K. B. Patel, M. A. Valvano, *Burkholderia cenocepacia* and *Salmonella enterica* ArnT proteins that transfer 4-amino-4-deoxy-L-arabinose to lipopolysaccharide share membrane topology and functional amino acids. *Sci. Rep.* **5**, 10773 (2015). 10.1038/srep10773 Medline doi:10.1038/srep10773

18. L. L. Lairson, B. Henrissat, G. J. Davies, S. G. Withers, Glycosyltransferases: Structures, functions, and mechanisms. *Annu. Rev. Biochem.* **77**, 521–555 (2008). Medline doi:10.1146/annurev.biochem.76.061005.092322

19. C. Lizak, S. Gerber, S. Numao, M. Aebi, K. P. Locher, X-ray structure of a bacterial oligosaccharyltransferase. *Nature* **474**, 350–355 (2011). Medline doi:10.1038/nature10151

20. S. Matsumoto, A. Shimada, J. Nyirenda, M. Igura, Y. Kawano, D. Kohda, Crystal structures of an archaeal oligosaccharyltransferase provide insights into the catalytic cycle of N-linked protein glycosylation. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 17868–17873 (2013). Medline doi:10.1073/pnas.1309777110

21. M. A. Lomize, I. D. Pogozheva, H. Joo, H. I. Mosberg, A. L. Lomize, OPM database and PPM web server: Resources for positioning of proteins in membranes. *Nucleic Acids Res.* **40** (Database issue), D370–D376 (2012). Medline doi:10.1093/nar/gkr703

22. N. R. Voss, M. Gerstein, 3V: Cavity, channel and cleft volume calculator and extractor. *Nucleic Acids Res.* **38** (Web Server), W555–W562 (2010). Medline doi:10.1093/nar/gkq395

23. Z. Zhou, A. A. Ribeiro, C. R. H. Raetz, High-resolution NMR spectroscopy of lipid A molecules containing 4-amino-4-deoxy-L-arabinose and phosphoethanolamine substituents: Different attachment sites on lipid A molecules from NH_4VO_3-treated *Escherichia coli* versus *kdsA* mutants of *Salmonella typhimurium*. *J. Biol. Chem.* **275**, 13542–13551 (2000). Medline doi:10.1074/jbc.275.18.13542

24. J. B. Thoden, H. M. Holden, Active site geometry of glucose-1-phosphate uridylyltransferase. *Protein Sci.* **16**, 1379–1388 (2007). Medline doi:10.1110/ps.072864707

25. C. Lizak, S. Gerber, G. Michaud, M. Schubert, Y. Y. Fan, M. Bucher, T. Darbre, M. Aebi, J. L. Reymond, K. P. Locher, Unexpected reactivity and mechanism of carboxamide activation in bacterial N-linked protein glycosylation. *Nat. Commun.* **4**, 2627 (2013). Medline doi:10.1038/ncomms3627

26. M. Punta, J. Love, S. Handelman, J. F. Hunt, L. Shapiro, W. A. Hendrickson, B. Rost, Structural genomics target selection for the New York Consortium on Membrane Protein Structure. *J. Struct. Funct. Genomics* **10**, 255–268 (2009). Medline doi:10.1007/s10969-009-9071-1

27. J. Love, F. Mancia, L. Shapiro, M. Punta, B. Rost, M. Girvin, D. N. Wang, M. Zhou, J. F. Hunt, T. Szyperski, E. Gouaux, R. MacKinnon, A. McDermott, B. Honig, M. Inouye, G. Montelione, W. A. Hendrickson, The New York Consortium on Membrane Protein Structure (NYCOMPS): A high-throughput platform for structural genomics of integral membrane proteins. *J. Struct. Funct. Genomics* **11**, 191–199 (2010). Medline

28. R. B. Kapust, J. Tözsér, T. D. Copeland, D. S. Waugh, The P1′ specificity of tobacco etch virus protease. *Biochem. Biophys. Res. Commun.* **294**, 949–955 (2002). Medline doi:10.1016/S0006-291X(02)00574-0

29. W. Kabsch, Integration, scaling, space-group assignment and post-refinement. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 133–144 (2010). Medline doi:10.1107/S0907444909047374

30. G. M. Sheldrick, A short history of SHELX. *Acta Crystallogr. A* **64**, 112–122 (2008). Medline doi:10.1107/S0108767307043930

31. T. Pape, T. R. Schneider, HKL2MAP : A graphical user interface for macromolecular phasing with SHELX programs. *J. Appl. Cryst.* **37**, 843–844 (2004). doi:10.1107/S0021889804018047

32. A. J. McCoy, R. W. Grosse-Kunstleve, P. D. Adams, M. D. Winn, L. C. Storoni, R. J. Read, Phaser crystallographic software. *J. Appl. Cryst.* **40**, 658–674 (2007). Medline doi:10.1107/S0021889807021206

33. A. Vagin, A. Teplyakov, Molecular replacement with MOLREP. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 22–25 (2010). Medline doi:10.1107/S0907444909042589

34. T. C. Terwilliger, Maximum-likelihood density modification. *Acta Crystallogr. D Biol. Crystallogr.* **56**, 965–972 (2000). Medline doi:10.1107/S0907444900005072

35. P. R. Evans, G. N. Murshudov, How good are my data and what is the resolution? *Acta Crystallogr. D Biol. Crystallogr.* **69**, 1204–1214 (2013). Medline doi:10.1107/S0907444913000061

36. P. D. Adams, P. V. Afonine, G. Bunkóczi, V. B. Chen, I. W. Davis, N. Echols, J. J. Headd, L. W. Hung, G. J. Kapral, R. W. Grosse-Kunstleve, A. J. McCoy, N. W. Moriarty, R. Oeffner, R. J. Read, D. C. Richardson, J. S. Richardson, T. C. Terwilliger, P. H. Zwart, PHENIX: A comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 213–221 (2010). Medline doi:10.1107/S0907444909052925

37. P. Emsley, B. Lohkamp, W. G. Scott, K. Cowtan, Features and development of Coot. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 486–501 (2010). Medline doi:10.1107/S0907444910007493

38. E. F. Pettersen, T. D. Goddard, C. C. Huang, G. S. Couch, D. M. Greenblatt, E. C. Meng, T. E. Ferrin, UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004). Medline doi:10.1002/jcc.20084

39. P. V. Afonine, N. W. Moriarty, M. Mustyakimov, O. V. Sobolev, T. C. Terwilliger, D. Turk, A. Urzhumtsev, P. D. Adams, FEM: Feature-enhanced map. *Acta Crystallogr. D Biol. Crystallogr.* **71**, 646–666 (2015). Medline doi:10.1107/S1399004714028132

40. Z. Zhou, S. Lin, R. J. Cotter, C. R. Raetz, Lipid A modifications characteristic of *Salmonella typhimurium* are induced by NH_4VO_3 in *Escherichia coli* K12. Detection of 4-amino-4-deoxy-L-arabinose, phosphoethanolamine and palmitate. *J. Biol. Chem.* **274**, 18503–18514 (1999). Medline doi:10.1074/jbc.274.26.18503

41. F. Tavares-Carreón, Y. Fathy Mohamed, A. Andrade, M. A. Valvano, ArnT proteins that catalyze the glycosylation of lipopolysaccharide share common features with bacterial N-oligosaccharyltransferases. *Glycobiology* cwv095 (2015). Medline doi:10.1093/glycob/cwv095

42. M. Laitaoja, J. Valjakka, J. Jänis, Zinc coordination spheres in protein structures. *Inorg. Chem.* **52**, 10983–10991 (2013). Medline doi:10.1021/ic401072d

43. K. Katoh, H. Toh, Recent developments in the MAFFT multiple sequence alignment program. *Brief. Bioinform.* **9**, 286–298 (2008). Medline doi:10.1093/bib/bbn013

44. M. A. Larkin, G. Blackshields, N. P. Brown, R. Chenna, P. A. McGettigan, H. McWilliam, F. Valentin, I. M. Wallace, A. Wilm, R. Lopez, J. D. Thompson, T. J. Gibson, D. G. Higgins, Clustal W and Clustal X version 2.0. *Bioinformatics* **23**, 2947–2948 (2007). Medline doi:10.1093/bioinformatics/btm404

45. D. H. Huson, C. Scornavacca, Dendroscope 3: An interactive tool for rooted phylogenetic trees and networks. *Syst. Biol.* **61**, 1061–1067 (2012). Medline doi:10.1093/sysbio/sys062

46. G. E. Crooks, G. Hon, J.-M. Chandonia, S. E. Brenner, WebLogo: A sequence logo generator. *Genome Res.* **14**, 1188–1190 (2004). Medline doi:10.1101/gr.849004