Manuscript version: Author's Accepted Manuscript
The version presented in WRAP is the author's accepted manuscript and may differ from the published version or Version of Record.

Persistent WRAP URL:
http://wrap.warwick.ac.uk/163073

How to cite:
Please refer to published version for the most recent bibliographic citation information. If a published version is known of, the repository item page linked to above, will contain details on accessing it.

Copyright and reuse:
The Warwick Research Archive Portal (WRAP) makes this work by researchers of the University of Warwick available open access under the following conditions.

Copyright © and all moral rights to the version of the paper presented here belong to the individual author(s) and/or other copyright owners. To the extent reasonable and practicable the material made available in WRAP has been checked for eligibility before being made available.

Copies of full items can be used for personal research or study, educational, or not-for-profit purposes without prior permission or charge. Provided that the authors, title and full bibliographic details are credited, a hyperlink and/or URL is given for the original metadata page and the content is not changed in any way.

Publisher's statement:
Please refer to the repository item page, publisher's statement section, for further information.

For more information, please contact the WRAP Team at: wrap@warwick.ac.uk.

Structural basis of Lipopolysaccharide Maturation by the O-Antigen Ligase

Khuram U. Ashraf[1], Rie Nygaard[1], Owen N. Vickery[2,3,#], Satchal K. Erramilli[4,#], Carmen M. Herrera[5,#], Thomas H. McConville[6,#], Vasileios I. Petrou[7,8], Sabrina I. Giacometti[1], Meagan Belcher Dufrisne[1], Kamil Nosol[4], Allen P. Zinkle[1], Chris L.B. Graham[2], Michael Loukeris[9], Brian Kloss[9], Karolina Skorupinska-Tudek[10], Ewa Swiezewska[10], David I. Roper[2,1], Oliver B. Clarke[1,11], Anne-Catrin Uhlemann[6], Anthony A. Kossiakoff[4], M. Stephen Trent[5*], Phillip J. Stansfeld[2,3*] and Filippo Mancia[1*]

[1]Department of Physiology and Cellular Biophysics, Columbia University Irving Medical Center, New York, NY 10032, USA.

[2]School of Life Sciences, University of Warwick, Coventry, CV4 7AL, UK.

[3]Department of Chemistry, University of Warwick, Coventry, CV4 7AL, UK.

[4]Department of Biochemistry and Molecular Biology, University of Chicago, Chicago, IL, 60637, USA.

[5]Department of Infectious Diseases, College of Veterinary Medicine, University of Georgia, Athens, GA, 30602, USA.

[6]Department of Medicine, Division of Infectious Diseases, Columbia University Medical Center, New York, New York, 10032, USA.

[7]Department of Microbiology, Biochemistry, and Molecular Genetics, New Jersey Medical School, Rutgers Biomedical Health Sciences, Newark, NJ, 07103, USA.

[8]Center for Immunity and Inflammation, New Jersey Medical School, Rutgers Biomedical Health Sciences, Newark, NJ, 07103, USA.

[9]New York Consortium on Membrane Protein Structure, New York Structural Biology Center, 89 Convent Avenue, New York, NY, 10027, USA.

[10]Institute of Biochemistry and Biophysics, Polish Academy of Sciences, 02-106 Warsaw, Poland.

[11]Department of Anesthesiology, Columbia University Irving Medical Center, New York, NY, 10032, USA.

[#]These authors contributed equally.

*Correspondence to be addressed to: strent@uga.edu (M.S.T.), phillip.stansfeld@warwick.ac.uk (P.J.S.), fm123@cumc.columbia.edu (F.M.).

Keywords:

Lipopolysaccharide, single-particle cryo-electron microscopy, molecular dynamics simulations, structural biology, undecaprenyl pyrophosphate, WaaL ligase, lipid A, O-antigen, membrane proteins, transglycosylation, glycosyltransferase

Abstract

The outer membrane of Gram-negative bacteria has an external leaflet largely composed of lipopolysaccharides (LPS), that provides a selective permeation barrier, particularly against antimicrobials[1]. The final and crucial step in LPS biosynthesis is the addition of a species-dependent O-antigen to the lipid A core oligosaccharide, catalyzed by the O-antigen ligase WaaL[2]. Here we present structures of WaaL from *Cupriavidus metallidurans,* both apo and complexed with its lipid-carrier undecaprenyl pyrophosphate, determined by single-particle cryo-electron microscopy to 3.5 Å and 3.2 Å resolution, respectively. The structures reveal that WaaL is comprised of 12 transmembrane helices and a predominantly α-helical periplasmic region, which we show contains many of the conserved residues required for catalysis. We observe a conserved fold within the GT-C family of glycosyltransferases and hypothesize a common mechanism for shuttling the undecaprenyl-based carrier to and from the active site. The structures, combined with genetic, biochemical, bioinformatics and molecular dynamics simulation experiments, provide molecular details on how the ligands come in apposition, and allows us to propose a mechanistic model for catalysis. Our work presents a structural basis for LPS maturation by a novel member of the GT-C superfamily of glycosyltransferases.

Main

The external leaflet of the outer membrane of Gram-negative bacteria is comprised mainly of lipopolysaccharides (LPS), making it a robust permeability barrier. LPS is critical for bacterial fitness during infection, can impact pathogenicity, and serves as one of the conserved microorganism-associated molecular patterns recognized by the mammalian innate immune system[1,3,4]. LPS is comprised of a lipid anchor (termed lipid A), an oligosaccharide (OS) core, and an O-antigen, which is a highly variable polysaccharide composed of repeating sequences of three to six sugar moieties[3,5,6]. The O-antigen domain of LPS has been shown to contribute to bacterial evasion of complement-mediated killing, impact host autoimmunity through molecular mimicry, and alter bacterial adherence to host tissue[7-9]. Within *E. coli* alone, more than 187 different O-antigens have been identified[10], and many of the corresponding O-antigen structures have been determined[11].

The polysaccharide repeats that constitute the O-antigen are synthesised in the bacterial cytoplasm by a series of glycosyltransferases, and then attached to the lipid carrier undecaprenyl pyrophosphate (Und-PP), before being flipped to the periplasmic leaflet of the inner membrane[3,12]. This occurs via the Wzy polymerase, ABC transporter or the less well-characterized Synthase pathway, according to the composition of the O-antigen and the nature of transport and elongation (**Extended Data Fig. 1a**)[13]. In all cases, the final step is the addition of O-antigen to the lipid A core OS, carried out by the O-antigen ligase WaaL, via a metal-independent glycosyltransferase reaction (**Fig 1a** and **Extended Data Fig. 1b**)[2,14]. Mutations within the gene encoding for WaaL prevent this ligation step, are devoid of O-antigen and accumulate Und-PP-linked precursors in the periplasm[15]. WaaL is predicted to have 12 transmembrane (TM) helices and a significant periplasmic region between TM helix 9 and TM helix 10[5,16,17], shown to be essential for the formation of mature LPS[1,7].

To investigate the molecular details of the final assembly step in LPS biosynthesis, we determined structures of WaaL by single-particle cryo-electron microscopy (cryo-EM), in its ligand-bound and apo state. Combining structural information with genetics, biochemical assays, and molecular dynamics (MD) simulations, we provide a rationale for substrate binding as well as a hypothesis for the reaction mechanism.

Structure determination of WaaL

We screened WaaL orthologues from ~200 different species for expression and stability in detergent to select suitable candidates for structural studies. Out of nineteen shortlisted, WaaL from *Cupriavidus metallidurans* (*Cm*WaaL; 44kDa) was identified as the most promising. We confirmed that *Cm*WaaL is an O-antigen ligase by deleting the gene encoding *Cm*WaaL and showing loss of O-antigen ligase activity that could be restored when *Cm*WaaL was expressed *in trans* (**Extended Data Fig. 1c**). *Cm*WaaL was purified in detergent and reconstituted into lipid-filled nanodiscs for structure determination by cryo-EM (**Extended Data Fig. 1d, e**). In order to overcome current size limitations for this technique and to provide fiducials for particle alignment[18], we screened a synthetic phage display library to select for recombinant antigen-binding fragments (Fabs) against *Cm*WaaL[19]. Seven high-affinity Fab candidates were screened for complex formation with *Cm*WaaL, and WaB10 was chosen due to its high binding affinity (**Extended Data Fig. 1f-h**). Data processing of particles picked from 2378 micrographs allowed us to separate them into two distinct classes resulting in density maps to 3.5 and 3.2 Å resolution (**Extended Data Table 1** and **Extended Data Fig. 2**). We built first into the higher resolution map (**Fig. 1b** and **Extended Data Fig. 3**) residues 5-407 of the 413 amino acids of *Cm*WaaL, and then adopted the model to fit the density of the second map.

The structure of Und-PP-bound *Cm*WaaL

The structure of *Cm*WaaL derived from the 3.2 Å map shows a monomer with twelve TM helices and one short cytoplasmic helix (CH) connecting TM helices 2 and 3 (**Fig. 1c, d**). Both the N- and C-termini are on the cytoplasmic side of the membrane. There is also a ~90 amino acid region between TM helices 9 and 10 on the periplasmic side which is mainly α-helical, divided into four short α-helical segments (PH 1-4), two of which (PH1 and PH2) are amphipathic and parallel to the membrane (**Fig. 1c, d**). The Fab required for structure determination binds to this periplasmic region (**Fig. 1b** and **Extended Data Fig. 4a, b**).

The overall TM core of the protein is tightly packed and comprises TM helices 1-8 and 10-12. In contrast, TM helix 9 protrudes from the TM core (**Fig. 1c, d**), and forms a conduit to a central cavity (cavity 1) that is lined with conserved charged and polar residues at the periplasmic surface, and hydrophobic residues within the membrane core (**Fig. 2a, Extended Data Fig. 4c** and **Extended Data Fig. 5a-c**). A cluster of four conserved arginines – R139, R191, R242, and R265 – provide a positive net electrostatic charge within this region (**Fig. 2a** and **Extended Data Fig. 5a**). Here, we observed clear non-protein density, which we modeled as two isoprenyl groups and a pyrophosphate, and interpreted it as being a partially ordered Und-PP (**Fig. 2b**). The presence and identity of Und-PP, which co-purified with the protein, was confirmed by thin-layer chromatography (TLC) (**Extended Data Fig. 5d**). R191 and R265 coordinate the di-phosphate of Und-PP (**Fig. 2b**) and the polyprenyl tail appears to interact with several conserved hydrophobic residues, including I373, L376, T377 (TM helix 11), and V195 (TM helix 8) (**Extended Data Fig. 5e**).

The complete Und-PP was modelled by superposition onto the two isoprenyl groups and the pyrophosphate coordinates described above. This WaaL-Und-PP complex was subjected to MD simulations within a lipid membrane. The simulations are consistent with the density map, illustrating that the portion of Und-PP that is visible in the structure is stabilized in the cavity, while the remainder of the undecaprenyl tail is highly mobile within the membrane (**Extended Data Fig. 5f**).

To study the effect on enzymatic activity of the four conserved arginines and other potentially functionally important residues in the Und-PP binding cavity, we shifted our attention to *E. coli* (*Ec*) K12, as this organism is more amenable to genetic manipulation and functional characterization. *Ec*WaaL could not be cross-complemented by *Cm*WaaL, likely reflecting differences in chemical composition of the lipid A core OS, which aids in determining WaaL specificity[5,15]. We therefore generated a homology model of *Ec*WaaL based on the *Cm*WaaL structure, guided by co-evolutionary analysis and the coordinates of a structure modelled by AlphaFold[20] (**Extended Data Fig. 6**), and utilized this to select mutants to test functionally in a ligation assay. We observed that the ligase activity is abolished for *Ec*WaaL when R161, R215, R265 and R288 – corresponding to R139, R191, R242 and R265 in *Cm*WaaL – are mutated to alanine (**Extended Data Fig. 7a-c**).

We next evaluated the role of these four arginines in *Cm*WaaL ligase activity directly in *Cm*. In this organism, when the WaaL KO was complemented with R191A and R265A *in trans*, ligase activity was abolished and almost abolished in R242A, whereas when the ΔwaaL was complemented with R139A, activity appeared to be retained (**Fig. 2c** and **Extended Data Fig. 7d**).

The structure of apo *Cm*WaaL

Analysis of the lower (3.5 Å) resolution map showed no evidence of density corresponding to Und-PP (**Extended Data Fig. 8a**), likely representing the apo state of the enzyme. The structure of *Cm*WaaL shows a similar overall fold in both states, with the exception of TM9 (**Extended Data Fig. 8b, c**), which is poorly resolved in the apo state likely due to its flexibility in the absence of Und-PP. R191 and R242 also appear to change conformation (**Fig. 2d)**. The density of R265 together with a nearby cluster of conserved histidines (H311, H313 and H383) is weaker, probably because of the more dynamic nature of the apo state (**Extended Data Fig. 8d**).

The structures in apo and Und-PP-bound states of *Cm*WaaL, coupled to functional experiments, suggest that R191 and R242 (*Ec*R215 and *Ec*R265) might be involved in shuttling the Und-PP in and/or out of its binding pocket. Once the substrate is in place, R191 (*Ec*R215), together with R265 (*Ec*R288) appear to play a dominant role in coordinating it.

*Cm*R139 and the equivalent *Ec*R161 could instead have different roles in the two orthologues. *Cm*R139 is in a flexible linker between TM helix 5 and 6, and is the most distant of the four arginine residues from the bound Und-PP. In the model, *Ec*R161 is directed towards *Ec*D389, previously shown to be important for function[2]. *Ec*D389 likely corresponds to *Cm*Q378, which when mutated to alanine does not affect activity (**Extended Data Fig. 7e**)[3], suggesting inter-species differences which may correlate with diversity in chemical composition of LPS.

Putative lipid A-binding site

A second cavity (cavity 2) is located between the first three TM helices and extends over the periplasmic surface of the enzyme to cavity 1 (**Fig. 2a**). We hypothesize that this may be the binding site for lipid A. By performing a focused classification around cavity 2, we identified density consistent with lipid within the proposed lipid A binding site (**Fig. 3a**). In this region of the density map, we also observed a membrane deformation in the periplasmic leaflet adjacent to TM helices 1-3 (**Fig. 3a**). Membrane deformation was also observed in MD simulations, with and without bound lipid A. This appears to be induced by the presence of two short TM helices (1 and 2) in conjunction with the buried basic residues, R28 and R92. In these simulations, the phospholipid head groups deform ~8.5 Å towards both arginines to accommodate the hydrophobic mismatch induced by the two short TM helices (**Fig. 3b-d**). The thinning of the membrane could favour the capture of the shorter, C10 to C14, acyl tails of lipid A, in comparison to the longer, inner membrane phospholipid species. This is comparable to the membrane thinning – from 38 to 33 Å thickness – observed around outer membrane proteins when embedded and simulated in a model phospholipid membrane[21]. In addition, the distance between the two arginine CZ atoms, at ~8 Å, is ideally positioned to dynamically coordinate the phosphate groups of lipid A, which are ~10 Å apart. Furthermore, both residues are also located in the same plane of the membrane.

The two short TM helices and the basic residues in *Cm*WaaL are also conserved in *Ec*WaaL (K45 and K111). However, alanine mutations of these residues in *Ec*WaaL did not abolish O-antigen ligation (**Extended Data Fig. 7a-c**), although the *Cm*WaaL R92A mutant does show a slight decrease in activity (**Extended Data Fig. 7d, e**), as does the double mutant K45A and K111A in *Ec*WaaL (**Extended Data Fig. 7a-c**). Tryptophan mutations on residues at the interface between cavity 1 and cavity 2 were also carried out to monitor their effect on ligase activity. T168W and A394W did not abolish O-antigen ligation, whereas T170W did (**Extended Data Fig. 7a-c and f**). This could be due to the T170W mutant having a steric impact on Und-PP binding (**Extended Data Fig. 7g**). We also expect that the activity of WaaL could partly be retained due to the large size and complexity of lipid A and the nature of the multifaceted binding interface between the lipid A core OS and WaaL. Alternatively, lipid A could bind in cavity 1, near Und-PP (**Extended Data Fig. 7h**), a hypothesis we cannot exclude *a priori*. However, this could result in steric clashes between both substrates. Furthermore, the geometry of the two ligands would likely be somewhat constrained, with the groups to be ligated positioned at an acute angle, when compared with the more linear arrangement observed if the substrates are modelled bound to two distinct sites.

Proposed mechanism of action

The reaction catalyzed by WaaL is known to be independent of ATP and metal ions[2], and the energy required to drive it has been suggested to result from cleavage of the bond between the distal phosphate group of Und-PP and its O-antigen cargo[22]. This is analogous to classical glycosyltransferases in which a nucleotide diphosphate sugar is the substrate[23]. In inverting glycosyltransferases, nucleophilic attack by the acceptor hydroxy group leads to an inversion of stereochemistry at the anomeric center of the donor substrate. These enzymes typically use a catalytic base and a bound metal ion that stabilizes the leaving phosphate groups. However, metal ion-independent inverting glycosyltransferases also exist and use basic amino acids such as arginine, lysine, or histidine to perform a similar function to that of the metal ion[2,16].

We propose a mechanism for *Ec*WaaL in which the absolutely conserved *Ec*H338 abstracts a proton from the leading hydroxyl of the terminal outer core of LPS. This would enable the oxygen to perform a nucleophilic attack at the C1 carbon of the Und-PP-linked O-antigen sugar ring and induce cleavage of the sugar-phosphate bond **(Fig. 4a, b**). *Ec*H338 appears to be stabilised by a network of hydrogen bonds involving nearby *Ec*H336 and *Ec*E343. The enzyme would then reset as the proton moves from *Ec*H338 to the pyrophosphate of Und-PP. The Und-PP moiety is held in position by *Ec*R215 and *Ec*R288, as detailed above (**Fig. 4b**). The mutation *Ec*H338A abolishes enzymatic activity (**Fig. 4c**), as does mutagenesis of the corresponding residue in *Cm*WaaL, *Cm*H313A (**Extended Data Fig. 7d, e)**. The residues *Ec*H336 and *Ec*E343, which are adjacent to *Ec*H338 in the structure, also significantly reduce overall activity when mutated to alanine. We speculate that they might stabilise the position of *Ec*H338, in order for the mechanism to proceed (**Fig. 4c, d**). Inspection of the structure shows that highly conserved residues *Ec*N339 and *Ec*E340 are also in the vicinity of the active site; however, alanine mutants of these residues showed no change in activity compared to the wild-type enzyme and thus are not included in our proposed mechanism (**Fig. 4c, d**).

In addition to MD simulations with bound Und-PP, we also performed MD simulations of the apo enzyme (**Extended Data Fig. 9a**). Based on these experiments, we observed TM helix 9 in two discrete states, which we defined as the closed and the open state (**Extended Data Fig. 9b**). In the absence of Und-PP, an outward movement of TM helix 9 is observed, thereby transitioning from a closed to an open state. This appears to be driven by an electrostatic repulsion of *Cm*R242 (*Ec*R265) from the adjacent basic groups of the active site, in particular *Cm*R191 (*Ec*R215) (**Extended Data Fig. 9b**).

Features of GT-C enzymes that use Und-PP The structure of WaaL reveals that it shares an overall similar fold as *T. thermophilus (Tt)* RodA[24], with an Cα RMSD of 4.73 Å over 248 residues (**Extended Data Fig. 10a**). RodA is a transglycosylase of the shape, elongation, division and sporulation (SEDS) family of proteins[25], critically required for peptidoglycan cell wall formation. Mechanistically, WaaL is a transferase whereas RodA is a processive polymerase. Both utilise Und-PP substrates and appear to share a similar helical structural arrangement with conserved arginines lining the binding pocket. Indeed, R242 in *Cm*WaaL and R203 in *Tt*RodA are both highly conserved and could adopt the same mechanism of shuttling the Und-PP into or out of the active site (**Extended Data Fig. 10b**). The major structural distinction between WaaL and RodA is that the latter lacks the first two TM helices, which we propose are integral to binding the lipid A core OS (**Extended Data Fig. 10a**).

We also noticed structural and topological similarities between WaaL, PglB and ArnT glycosyltransferases (**Extended Data Fig. 10b**)[26-28]. ArnT catalyzes the transfer of a sugar moiety via a undecaprenyl-phosphate (Und-P) carrier to lipid A[23], while the metal-dependent PglB is a protein glycosyltransferase[28]. For all three enzymes, residues in the periplasmic loops have been shown to be critical for substrate recognition and catalysis[2,23,26], suggesting commonalities in structure, function and mechanism.

Discussion

We have determined the structures of Und-PP bound and apo *Cm*WaaL by cryo-EM, to 3.2 Å and 3.5 Å resolution, respectively. The structure reveals two major cavities on the periplasmic side, which we observe and suggest, respectively, are the binding sites for Und-PP and lipid A. There are conserved residues bridging these two cavities, that have previously been suggested to be involved in the ligation of the O-antigen to the lipid A core OS[2,5,7,17].

We have combined structural evidence, MD simulations and biochemical experiments to propose a reaction mechanism in which the O-antigen linked Und-PP is shuttled into and coordinated within the binding pocket by a set of conserved arginines. Subsequently, the O-antigen is transferred from the Und-PP onto the lipid A core OS via a reaction carried out by a highly conserved histidine, in turn coordinated by another histidine and a glutamate. WaaL shares features with metal ion-independent inverting glycosyltransferases to bring about catalysis, but utilizes a membrane embedded Und-PP substrate as opposed to the soluble nucleotide diphosphate moiety of classical nucleoside glycosyltransferases[2,29]. In both cases – Und-PP-linked, or nucleotide diphosphate-linked sugars – the leaving group upon glycosyl transfer is a diphosphate molecule and the proximal sugar to the distal phosphate is O-linked to the terminal sugar of the acceptor[9,23].

We propose that the lipid A binding site is on the opposite side of the structure from the location of the bound Und-PP. This site is coordinated by two basic residues and is shaped by two unique short TM helices (1 and 2). These helices induce a membrane deformation that appears to draw the phospholipids towards two basic residues, *Cm*R28 and *Cm*R92, that are appropriately spaced to coordinate the phosphate groups of the lipid A moiety of LPS. The distance between this binding site and that of the Und-PP is ~25 Å, which is comparable to the dimensions of the core OS of LPS. This suggests that carbohydrates of the immature LPS might rest on the periplasmic surface of WaaL to position the terminal outer core sugar for the attachment of O-antigen.

The structure of *Cm*WaaL in its Und-PP bound state, along with the previously published structures of RodA-PBP2, PglB and ArnT, has allowed us to identify structural similarities[24,26,28]. In all these cases, we observe a TM helix protruding from the helical bundle, and a short amphipathic helix parallel to the membrane connecting the protruding helix back to the helical bundle, creating a binding cavity for the ligand (**Extended Data Fig. 10**)[24]. This suggests an evolutionary relationship between these enzymes which are all members of the GT-C family of glycosyltransferases[23].

The structure of *Cm*WaaL provides a framework for understanding how O-antigen ligases perform the final, critical, step in LPS maturation of Gram-negative bacteria and offers a structural basis for designing inhibitors of this process in bacterial cell envelope biosynthesis.

Main Text References

1 Valvano, M. A. Export of O-specific lipopolysaccharide. *Front. Biosci* **8**, s452-s471 (2003).

2 Ruan, X., Loyola, D. E., Marolda, C. L., Perez-Donoso, J. M. & Valvano, M. A. The WaaL O-antigen lipopolysaccharide ligase has features in common with metal ion-independent inverting glycosyltransferases. *Glycobiology* **22**, 288-299 (2012).

3 Whitfield, C. & Trent, M. S. Biosynthesis and export of bacterial lipopolysaccharides. *Annual review of biochemistry* **83**, 99-128 (2014).

4 Woodward, L. & Naismith, J. H. Bacterial polysaccharide synthesis and export. *Current opinion in structural biology* **40**, 81-88 (2016).

5 Kaniuk, N. A., Vinogradov, E. & Whitfield, C. Investigation of the Structural Requirements in the Lipopolysaccharide Core Acceptor for Ligation of O Antigens in the Genus Salmonella WAAL “LIGASE” IS NOT THE SOLE DETERMINANT OF ACCEPTOR SPECIFICITY. *Journal of Biological Chemistry* **279**, 36470-36480 (2004).

6 Raetz, C. R., Reynolds, C. M., Trent, M. S. & Bishop, R. E. Lipid A modification systems in gram-negative bacteria. *Annu. Rev. Biochem.* **76**, 295-329 (2007).

7 Hong, Y. & Reeves, P. R. Model for the controlled synthesis of O-antigen repeat units involving the WaaL ligase. *MSphere* **1** (2016).

8 Lundstedt, E., Kahne, D. & Ruiz, N. Assembly and Maintenance of Lipids at the Bacterial Outer Membrane. *Chemical Reviews* (2020).

9 Whitfield, C., Williams, D. M. & Kelly, S. D. Lipopolysaccharide O-antigens-bacterial glycans made to measure. *Journal of Biological Chemistry*, jbc. (2020).

10 Liu, B. *et al.* Structure and genetics of Escherichia coli O antigens. *FEMS microbiology reviews* **44**, 655-683 (2020).

11 Feldman, M. F. *et al.* The activity of a putative polyisoprenol-linked sugar translocase (Wzx) involved in Escherichia coli O antigen assembly is independent of the chemical structure of the O repeat. *Journal of Biological Chemistry* **274**, 35129-35138 (1999).

12 Bertani, B. & Ruiz, N. Function and biogenesis of lipopolysaccharides. *EcoSal Plus* **8** (2018).

13 Schmid, J., Sieber, V. & Rehm, B. Bacterial exopolysaccharides: biosynthesis pathways and engineering strategies. *Frontiers in microbiology* **6**, 496 (2015).

14 Ruan, X. & Valvano, M. A. In vitro O-antigen ligase assay. In *Glycosyltransferases* 185-197 (Springer, 2013).

15 Abeyrathne, P. D., Daniels, C., Poon, K. K., Matewish, M. J. & Lam, J. S. Functional characterization of WaaL, a ligase associated with linking O-antigen polysaccharide to the core of Pseudomonas aeruginosa lipopolysaccharide. *Journal of bacteriology* **187**, 3002-3012 (2005).

16 Pérez, J. M., McGarry, M. A., Marolda, C. L. & Valvano, M. A. Functional analysis of the large periplasmic loop of the Escherichia coli K- 12 WaaL O- antigen ligase. *Molecular microbiology* **70**, 1424-1440 (2008).

17 Islam, S. T., Taylor, V. L., Qi, M. & Lam, J. S. Membrane topology mapping of the O-antigen flippase (Wzx), polymerase (Wzy), and ligase (WaaL) from Pseudomonas aeruginosa PAO1 reveals novel domain architectures. *MBio* **1**, e00189-00110 (2010).

18 Nygaard, R., Kim, J. & Mancia, F. Cryo-electron microscopy analysis of small membrane proteins. *Current Opinion in Structural Biology* **64**, 26-33 (2020).

19 Dominik, P. K. & Kossiakoff, A. A. in *Methods in enzymology* Vol. 557 219-245 (Elsevier, 2015).

20 Jumper, J. *et al.* Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583-589 (2021).

21 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 research* **47**, D390-D397 (2019).

22 Lazarus, M. B., Nam, Y., Jiang, J., Sliz, P. & Walker, S. Structure of human O-GlcNAc transferase and its complex with a peptide substrate. *Nature* **469**, 564-567 (2011).

23 Valvano, M. A. Glycan translational reactions in the periplasmic space. in *Recent Trends in Carbohydrate Chemistry* 37-49 (Elsevier, 2020).

24 Sjodt, M. *et al.* Structural coordination of polymerization and crosslinking by a SEDS–bPBP peptidoglycan synthase complex. *Nature Microbiology* **5**, 813-820 (2020).

25 Meeske, A. J. *et al.* SEDS proteins are a widespread family of bacterial cell wall polymerases. *Nature* **537**, 634-638 (2016).

26 Petrou, V. I. *et al.* Structures of aminoarabinose transferase ArnT suggest a molecular basis for lipid A glycosylation. *Science* **351**, 608-612 (2016).

27 Tavares-Carreón, F., Fathy Mohamed, Y., Andrade, A. & Valvano, M. A. ArnT proteins that catalyze the glycosylation of lipopolysaccharide share common features with bacterial N-oligosaccharyltransferases. *Glycobiology* **26**, 286-300 (2016).

28 Napiórkowska, M. *et al.* Molecular basis of lipid-linked oligosaccharide recognition and processing by bacterial oligosaccharyltransferase. *Nature structural & molecular biology* **24**, 1100-1106 (2017).

29 Ruan, X., Monjarás Feria, J., Hamad, M. & Valvano, M. A. Escherichia coli and Pseudomonas aeruginosa lipopolysaccharide O- antigen ligases share similar

membrane topology and biochemical properties. *Molecular microbiology* **110**, 95-113 (2018).

30 Voss, N. R. & Gerstein, M. 3V: cavity, channel and cleft volume calculator and extractor. *Nucleic acids research* **38**, W555-W562 (2010).

Main Text Figure Legends

Figure 1 | The central role of WaaL in lipopolysaccharide biosynthesis in bacteria and the structure of *Cm*WaaL. a, Atomic representation of a *O1* O-antigen ligation to the *R1* lipid A core OS. The O-antigen is transferred to the lipid A core OS by WaaL, forming the mature LPS. After ligation, Und-PP is recycled. **b,** Cryo-EM density map of the Und-PP bound *Cm*WaaL-Fab complex. Density corresponding to the variable region of the Fab (vFAB) is shown in grey and *Cm*WaaL in rainbow color. Representation is in rainbow color from the N-terminus (blue) to the C-terminus (red). **c,** Schematic diagram showing the topology of WaaL, consisting of 12 TM helices and a large periplasmic loop between TM helix 9 and 10. **d,** The 3.2 Å cryo-EM structure of Und-PP bound WaaL shown in two different orientations, with the 12 TM helices colored as in **e**. The N- and C- termini are labeled. The right-hand side shows a 90° rotation with helices numbered, as viewed from the top. Und-PP is shown as sticks (in gold). Approximate membrane boundaries are represented with dotted lines.

Figure 2 | Significant structural features of *Cm*WaaL. a, Cavities within the *Cm*WaaL structure. *Cm*WaaL is shown in ribbon representation in grey, and the two cavities as a semi-transparent surface, with cavity 1 in green and cavity 2 in blue. Volumes were calculated using the Voss Volume Voxelator (3V) server[30], using probes with 10 and 2 Å radii, corresponding to the outer and inner probe, respectively. The approximate membrane boundaries are represented with dotted lines. Und-PP (in gold) and the key arginine (R191, R242, R265) residues for *Cm*WaaL are shown as sticks. **b,** Close up view of R191 and R265 coordinating the Und-PP (in gold) within the binding pocket. The cryo-EM density assigned to bound Und-PP is displayed as a semi-transparent grey mesh. **c,** Functional analysis of *Cm*WaaL ligase activity in whole cells. LPS gel showing O-antigen extension in *C. metallidurans* expressing *Cm*WaaL variants (mutated *Cm*WaaL residues denoted in blue). Ø indicates empty plasmid. **d,** Close-up view of the Und-PP binding pocket in the apo structure (pink) and the Und-PP bound (grey), superimposed, showing R139, R191, R242 and R265 in sticks with CPK coloring.

Figure 3 | A putative binding site for lipid A. a, Cryo-EM density of WaaL (grey) incorporated into a nanodisc (salmon), showing a slice through the nanodisc where membrane deformation can be observed and outlined with a dashed blue box. Also, in salmon a lipid tail density is seen in the putative lipid A binding site. The density for the Fab is not included. **b,** The mean phosphate atom positions, highlighting a strong thinning of the membrane of ~8.5 Å around the two short helices (TM helix 1 and 2) in the apo *Cm*WaaL MD simulations. The mean phosphate position is coloured according to the deviation from the mean (blue = outward, red = inward). **c,** Contact of lipid A phosphates to *Cm*WaaL in MD simulation. Throughout the MD simulations, the phosphates of lipid A remain in contact with R28 and to a lesser extent with R92 and Q384. The residue contacts with threshold above 0.5 are shown as sticks. **d,** Left, top view of *Cm*WaaL, in ribbon, highlighting conserved residues, shown in sticks. Distance between R28 and R92 and between the phosphates of lipid A core and the Und-PP are shown as black lines. Right, lipid A chemical structure, with distances marked as black lines.

Figure 4 | Mechanism of catalysis for WaaL. a, Schematic representation of the ligation of the O-antigen linked Und-PP to the lipid A core OS in *Ec*WaaL, viewed from the periplasmic side of the membrane. In all cases, key arginine (R161, R215, R265, R288) and histidines (H338, H336) are shown as grey sticks. Left panel shows a surface representation for both lipid A (peach) with its inner (cyan) and outer (purple) core OS and the O-antigen (green) linked Und-PP (yellow) approaching their binding sites on the apo-state of WaaL. Middle panel shows the coordination of both O-antigen linked Und-PP and lipid A core OS in both sites. Right panel shows the mature LPS, with the Und-PP (yellow) product still bound, as shown in the cryo-EM structure. **b,** Residues within the active site of *Ec*WaaL are shown around the Und-PP-linked *O1* O-antigen N-acetylglucosamine and terminal *R1* outer core glucose. H338 (highlighted in green) is coordinated by a H-bond network between H336 and E343, which permits the abstraction of a proton from the terminal hydroxyl of the *R1* outer core glucose to protonate H338. The deprotonated oxygen may then perform a nucleophilic attack on the *C1* of the *O1* N-acetylglucosamine. This allows cleavage of the GlcNAc-phosphate bond. To reset the enzyme, H338 will deprotonate, with the proton possibly transferring to the phosphate of Und-PP, which then leaves the active site. **c,** Functional analysis of *Ec*WaaL ligase variants in whole cells by LPS gel analysis. WaaL proteins were expressed from plasmid pWSK29 in the W3110 Δ*waaL* strain. Ø indicates empty plasmid. **d,** Key residues involved in the putative ligation mechanism and ligand coordination within *Ec*WaaL. The *Ec*WaaL homology model is based on the *Cm*WaaL structure and shown as cartoon. Key residues R215, R265, R288, E343, H336, H338, and D389 are shown as sticks.

Methods

Target identification and cloning. Nineteen putative WaaL genes corresponding to a wide genomic background were identified based on a bioinformatics approach, as previously described[31]. Ligation independent cloning (LIC) was used to clone these nineteen orthologues into five different LIC-adapted expression vectors (pNYCOMPS-Nterm, pNYCOMPS-Cterm, pMCSG7, pNYCOMPS-N23 and pNYCOMPS-C23), all bearing protease-cleavable decahistidine tags for metal-affinity chromatography-based purification. Small scale expression and purification tests were performed as previously described[32], to identify eleven unique targets as “expression positive” from the small-scale expression tests. These eleven candidates were carried forward for midscale expression, purification and detergent screens as previously described[32], for the selection of *Cupriavidus metallidurans* (*Cm*) as the best candidate (protein accession code: WP_011517284.1). All cloning and initial protein characterization were performed at the protein production core facility of the former New York Consortium on Membrane Protein Structure (NYCOMPS)[33], now the Center on Membrane Protein Production and Analysis (COMPPÅ).

Protein expression and purification in detergent. WaaL from *Cupriavidus metallidurans* (*Cm*WaaL), cloned in the pNYCOMPS-Cterm vector, was used to transform BL21(DE3)pLysS *E. coli* competent cells, and grown overnight in 2xYT medium supplemented with 100 µg/mL ampicillin and 35 µg/mL chloramphenicol at 37°C with shaking (240 r.p.m.). The next day, 800 mL (large scale for protein expression) 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 OD600 reached 0.8-1.2. 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 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 of *Cm*WaaL, 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 with an Emulsiflex C3 homogenizer (Avestin) and the lysate was solubilized for 2 hours with n-dodecyl-β-D-maltopyranoside (DDM; Anatrace) added to a final concentration of 1% (w/v), in a volume of approximately 40 mL per cell pellet from 800 mL culture. 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). The supernatant after addition of 40mM 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 (GE Healthcare).

Nanodisc incorporation after detergent purification. The imidazole-containing buffer of the protein eluted from metal-affinity chromatography was desalted using a PD-10 column into the final protein buffer. *Cm*WaaL 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)[34,35]. 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 (Bio-Rad) at 4°C overnight with constant rotation. Bio-beads were removed and the nanodisc reconstitution mixture was bound again to Ni^{2+}-NTA resin at 4°C for 2 h to remove empty nanodiscs. The resin was washed with 10 column volumes of wash buffer (20 mM HEPES pH 7.5, 150 mM NaCl, and 60 mM imidazole) followed by four column volumes of elution buffer (20 mM HEPES pH 7.0, 150 mM NaCl, and 300 mM imidazole). Subsequently, the protein was incubated with TEV protease[36] to cleave the decahistidine tag (~0.5 mg TEV protease added per pellet equivalent from 800 mL of initial bacterial culture) overnight at 4^{o}C, while exchanging into an imidazole free buffer using a 3.5K cut-off Slide-A-Lyzer Dialysis Cassettes (Thermofisher). The sample was then passed through a column containing Ni^{2+}-NTA beads to separate the cleaved decahistidine tag, the TEV protease and uncleaved protein. The cleaved protein was further purified by loading onto a Superdex 200 Increase 10/300 GL size-exclusion column (GE Healthcare Life Sciences) 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 (**Extended Data Fig. 1d**).

Phage display to identify *Cm*WaaL-specific Fab fragments (WaB10). *Cm*WaaL was reconstituted into nanodiscs formed using chemically biotinylated MSP1E3D1, which were prepared as previously described[19]. Biotinylation efficiency was evaluated by a pull-down assay using streptavidin-coated paramagnetic particles (Promega). Phage displayed synthetic antigen binder (sAB) Library E[37,38] (kindly prepared by S. Mukherjee) was used as the naïve starting pool. Library E and the target were both diluted into selection buffer (20 mM Hepes, pH 7.4, 200 mM NaCl, 1% BSA) and selections were performed using a protocol adapted from prior publications[39,40]. In round one, biopanning was performed manually using 300 nM of *Cm*WaaL immobilized onto magnetic beads. Following a one-hour incubation, beads were washed three times with selection buffer and then used to subsequently infect log-phase *E. coli* XL-1 Blue cells. Infected cells were used to inoculate 30 mL cultures containing ampicillin (100 μg/mL) and M13-K07 helper phage (10^9 pfu/mL), and phage were amplified overnight. Four additional rounds of biopanning were performed by stepwise reduction of the target concentration to increase the stringency of selection. Rounds 2-5 were performed semi-automatically using a Kingfisher Beads Handling Robot (Thermo), and phage from each preceding round were amplified and used as the input pools. In addition, the amplified phage pool was pre-cleared using 200 μL of streptavidin beads, and in all rounds 1.5 μM of empty MSP1E3D1 nanodiscs were used as competitors in solution. Finally, in the last four rounds, phage were eluted from magnetic beads using a 15 min incubation with 1% Fos-choline 12 (Anatrace).

Single-point phage ELISA. *E. coli* XL-1 Blue cells were infected with phage from the 4th and 5th round pools and plated on LB-agar supplemented with 100 μg/mL ampicillin. The following day, individual colonies harboring phagemids were used to inoculate 400 μL of 2xYT media supplemented with ampicillin (100 μg/mL) and M13-K07 helper phage (10^9 pfu/mL). Phage were amplified overnight with shaking (280 r.p.m.). Single-point phage

ELISA was subsequently performed as previously described (**Supplementary Table 1**) [39,40]. All amplified phage were tested against *Cm*WaaL in MSP1E3D1 nanodiscs (30 nM), empty 1E3D1 nanodiscs (50 nM), or buffer alone to determine specific target binding. Bound phage particles were detected by a colorimetric assay using an HRP-conjugated anti-M13 monoclonal antibody (GE Healthcare). Binders with high target and low non-specific signal were chosen for subsequent experiments.

sAB cloning, expression, and purification. All specific binders were sequenced and then cloned into the sAB expression vector RH2.2 (kindly gifted by Sachdev Sidhu) as previously described[40]. All phage and expression vector subclones were sequence-verified at the University of Chicago Comprehensive Cancer Center DNA sequencing facility. sABs were expressed and purified as previously described[40] and subsequently dialyzed overnight in buffer with 20 mM Hepes, pH 7.5, 150 mM NaCl.

Multi-point sAB ELISA. To estimate apparent binding affinity of each sAB, multi-point ELISA was performed in triplicate as previously described[39,40] for each unique sAB. All sABs were again verified for specific binding by testing against *Cm*WaaL in MSP1E3D1 nanodiscs (30 nM), empty MSP1E3D1 nanodiscs (50 nM), or buffer alone. The amount of bound sAB was measured by a colorimetric assay using an HRP-conjugated anti-Fab monoclonal antibody (Jackson ImmunoResearch). Measured absorbance (A_{450}) values were plotted against the log sAB concentration, and estimated binding affinities (EC_{50}) were determined using a variable slope model with sigmoidal dose response in Prism (GraphPad software) (**Supplementary Table 1**).

***Cm*WaaL complex formation with the Fab WaB10.** TEV cleaved *Cm*WaaL incorporated into nanodiscs was incubated with the WaB10 Fab on ice for 1 h in a 1:2 molar ratio of protein to Fab. The *Cm*WaaL-Fab complex was concentrated and filtered, and then loaded on a Superdex 200 Increase 10/300 GL size-exclusion column in gel filtration buffer (20 mM HEPES pH 7.0 and 150 mM NaCl).

Single-particle cryo-EM vitrification and data acquisition. Purified *Cm*WaaL-Fab complex was concentrated to 1.2 mg/ml using a 50-kDa concentrator (Amicon). 3 µl of sample was added to a plasma-cleaned (Gatan Solarus) 1.2/1.3 µm holey gold grid (Quantifoil UltrAuFoil) and blotted using filter paper on one side for 3.5 s using the Vitrobot (Thermofisher) with a blot force of 3 and a wait time of 30 sec, before plunging immediately into liquid ethane for vitrification. The plunger was operating at 4°C with >90% humidity to minimize evaporation and sample degradation. Images were recorded using a Titan Krios electron microscope (FEI), at the Simons Electron Microscopy Center, equipped with an energy filter and a K2 direct electron detector (Gatan K3-BioQuantum) using a 1.061Å pixel size. An energy filter slit width of 20 eV was used during the collection and was aligned automatically every hour using Leginon[41]. Data collection was performed using a dose of ~70.15 $e^-/Å^2$ across 50 frames (200 ms per frame) at a dose rate of ~7.0 e^-/pix/s, using a set defocus range of -1.3 µm to -2.8 µm. A 100 µm objective aperture was used. In total, 2,378 micrographs were recorded over a single two-day collection using an image beam shift data collection strategy. Ice thickness was monitored after every 4th exposure using the Leginon zero-loss peak (ZLP) algorithm and was determined to be 23.1 ± 9.1 nm (SD).

Data processing. Contrast transfer function (CTF) estimation was performed using Patch CTF as implemented in cryoSPARC v.2.8. Blob picker in cryoSPARC v.2.8 were used to pick particles and inspect picks was used to curate the picks. This resulted in 844,438 particles which were then subjected to 2D classification in cryoSPARC v.2.8. 132,664 particles were selected for further processing from 2D classes with well-defined high-resolution features. One round of *ab initio* reconstruction was performed in cryoSPARC v.2.8 using four classes, with a maximum resolution set at 7 Å and an initial resolution at 9 Å, the best class was selected resulting in a stack of 46,362 particles. Hetrogeneous refinement was carried out on this particle stack, after which this stack was then extracted with a 320 pixel size box. This particle stack was subjected to non-uniform refinement in cryoSPARC v.2.8 resulting in a 3.9 Å reconstruction. Using a mask covering *Cm*WaaL and the variable region of the Fab, local refinement using non-uniform regularization was performed in cryoSPARC v.3.2, resulting in a 3.2 Å density map (**Extended data Fig. 2a**).

On initial observation it was hypothesized that the particles resulting in the 46,362 particle stack had a mixed population of bound and unbound WaaL particles contributing to this stack, which was mainly due to the density that we attributed to the geranyl diphosphate part of the Und-PP and the density for TM9. We observed that depending on the particles that were selected from the initial 132,664 particles the density of the Und-PP and TM9 had either weak or strong density, indicating that a mixed population of Und-PP bound and unbound may be present within the 132,664 particle stack. To address this, 3D classification was then performed using the initial 132,664 particles, which were extracted with a 400 pixel box size for further processing. These extracted particles were then imported into Relion v3.1.1 for 3D classification using a mask that only included TM9-11 as well as PH1-4, helices that surrounded the Und-PP. In total six classes were obtained, with the initial low pass set at 20 Å and the T value set at 40, the two best classes where then selected resulting in a stack of 39,844 and 30,514 particles. Both particle stacks were imported back into cryoSPARC v.3.2 and using a mask covering *Cm*WaaL and the variable region of the Fab for the 39,844 particle stack, and a mask covering *Cm*WaaL for the 30,514 particle stack, local refinement using non-uniform regularization was performed, resulting in a 3.23 Å density map for the

39,844 particle stack and 3.5 Å density map for the 30,514 particle stack (**Extended data Fig. 2a**).

A 3D classification followed by a focused refinement was performed to obtain the lipid density in our proposed lipid A binding site (**Fig. 3a**). The initial 132,664 particles, which were extracted with a 400 pixel size box were imported into Relion v3.1.1 for 3D classification using a mask surrounding the proposed lipid A site, including residues 29-79 and 315-352. In total six classes were obtained, with the initial low pass set at 20 Å and the T value set at 40, the best class containing 50,857 particles was selected. The particle stack was imported into cryoSPARC v.3.2 and using a mask covering *Cm*WaaL and the variable region of the Fab a local refinement using non-uniform regularization was performed, resulting in a density map at 3.22 Å resolution that has a clear lipid tail density within the binding cleft of cavity 2 (**Fig. 3a**).

Structural model building and refinement. To build the *Cm*WaaL model, Coot was used for manual model building. We observed extra density within the map of the 39,844 particle stack and were able to fit two isoprenyl groups and a pyrophosphate for the Und-PP (Pubchem ID 5280604) into the density. The map obtained from the 39,844 stack was the local refinement with the *Cm*WaaL and the variable region of the Fab (vFAB) mask created that gave the 3.23 Å map that was used to build the backbone and majority of the sidechains. The map from the 30,514 particle stack was a local refinement map created from the CmWaaL mask that yielded the 3.5 Å map. *Cm*WaaL was modeled *de novo* in Coot using secondary structure predictions from the XtalPred server[42] as a guide. Subsequent model refinement and adjustment was performed in Coot[43-45] and Phenix[46,47] iteratively.

Model analysis. A cavity search using the Solvent Extractor from Voss Volume Voxelator server[30] was performed using an outer-probe radius of 10 Angstrom and inner-probe radius of 2 Angstrom. Chimera[48] and ChimeraX[49] were used to visualize the structures in the figures.

Mutagenesis. Mutations of the *E. coli* K12 *Ec*WaaL (protein accession code: WP_001395405.1),in pWSK29 vector, were generated with an in-house method using KOD polymerase and custom primers.

Expression test of *E. coli* K12 *Ec*WaaL mutants. WaaL (WT and all mutants) from *E. coli* K12 (*Ec*WaaL), cloned in the pWSK29 vector, were used to transform BL21(DE3)pLysS *E. coli* competent cells, and grown overnight in 2xYT medium supplemented with 100 µg/mL ampicillin and 35 µg/mL chloramphenicol at 37°C with shaking (240 r.p.m.). The next day, 50 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 r.p.m.), until OD600 reached 0.8-1.2. 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 phosphate buffered saline (PBS) and centrifuged again to produce a solid pellet. Cell pellets were resuspended in 2 ml 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 by sonication and the lysate was solubilized for two hours with DDM, added to a final concentration of 1% (w/v). The solubilized material was clarified by centrifugation for 30 min (13,000 r.p.m.) at 4 °C, and the supernatant was mixed with Ni-NTA agarose beads (Qiagen) and binding allowed to proceed overnight in the presence of 40mM imidazole. The beads were then loaded on a column and washed with five 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 two column volumes of 20 mM HEPES pH 7.0, 150 mM NaCl, 300 mM imidazole, and 0.03% (w/v) DDM (100 μL). 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 by concentrating the protein (Amicon Ultra 0.5 mL; 50 Kda cut-off) and washing the sample through with 5 column volumes of buffer (500 μL), and repeating this 5 times, until the final step where the samples were concentrated to 20 μL. The eluates were then separated on a 14% SDS-PAGE gel, to confirm expression (**Extended Data Fig. 7c**).

Preparation of Und-PP ($C_{55}H_{101}N_3O_7P_2$). The procedure was modified from Danilov *et al.* (1989)[50]. Prenol composed of 11 isoprenoid units ($C_{55}H_{90}O$) and trichloroacetonitrile was dissolved in dichloromethane and added to tetra-n-butylammonium dihydrogen phosphate and the mixture was stirred for 90 min at room temperature. The solvent was evaporated, and the residue was dissolved in tetrahydrofuran. An ammonia solution (25%) was added to the mixture to precipitate the inorganic phosphate, and left overnight at 4^0 C. The following day, the supernatant was removed, and the pellet was washed twice with tetrahydrofuran. The tetrahydrofuran extracts were pooled and evaporated. The residue was dissolved in a chloroform:methanol mixture (2:1 v/v) and applied onto a DEAE-Sephadex A-25 (acetal form) column. Any unreacted undecaprenol was eluted with a chloroform:methanol mixture (2:1 v/v). The final product was eluted with increasing concentrations of ammonium acetal (0-55 mM). Fractions containing Und-PP were collected and analysed by thin-layer chromatography TLC. The solvent was then evaporated, and the residues were dissolved in a mixture of chloroform:methanol (2:1 v/v) with the addition of a 3% ammonia solution. The Und-PP was stored at -20^0 C until further use.

Thin-layer chromatography analysis of Und-PP. TLC analysis was utilized to identify the ligand, as mass spectrometry was not successful. Lipids were extracted from detergent purified *Cm*WaaL by adding chloroform:methanol (2:1) in a 1:1 ratio. The solvent for the mobile phase of the TLC on silica gel plates (Millipore; TLC Silica gel 60 F_{254}) was a solution of chloroform, methanol and water (65:25:4 per volume ratio). TLC plates were dried for five minutes and placed in an enclosed container along with iodine crystals for staining.

CRISPR-Cas9 gene editing in *Cupriavidius metallidurans*. To allow for efficient genetic manipulation in *Cupriavidius metallidurans* we adapted our recently optimised single plasmid CRISPR-Cas9 / lambda red recombineering system [51-57]. We utilised a pBBR1MCS2 origin of replication, *Streptococcus* cas9 controlled by an araBAD promoter; a guide RNA specific to the *Cm*WaaL gene controlled by a pTAC promoter for constitutive expression, a 764 base pair (bp) area of homology to *WaaL*, the lambda red recombineering genes to improve the efficiency of homologous recombination, and an ampicillin resistance cassette[54,58,59]. For the knock-out, we first analysed the WT *Cm*WaaL sequence via the CRISPR direct website to identify an appropriate N20 sequence, which was incorporated into *Cm*WaaL sgRNA. The homology was engineered to contain a 183 bp deletion surrounding the cas9 cut site. The *Cm*WaaL specific sgRNA and homology cassettes were cloned into the pBBR1_CRISPR vector. The sequence confirmed plasmid was inserted into the CH34 *Cupriavidius metallidurans* isolate via electroporation and appropriate transformants were identified through colony PCR. Transformants were grown at 30°C under ampicillin selection in tryptic soy broth (TSB) and induced with 0.2% arabinose after two hours. Following 12-24 hours of induction, the cultures were diluted 1:100 and plated on tryptic soy agar (TSA) + ampicillin at 100 mg/mL and 0.2% arabinose. Appropriate mutants were identified by colony

PCR and Sanger sequencing (Genewiz). Mutants were cured of the CRISPR plasmid with serial passage on non-selective media TSA.

Expression test of *Cm*WaaL WT and Mutants. WT and mutant *Cm*WaaL, which was cured of the CRISPR plasmid and transformed with pBBR1_WaaL complementation plasmids – all bearing a C-terminal FLAG tag fused to *Cm*WaaL WT and mutant constructs for detection purposes – were grown in 5 mL of TSB for 3-4 days and supplemented with 100 μg/mL tetracycline at 30^{o}C while shaking (80 r.p.m.). The next day, 50 mL of the same medium were inoculated with the starter culture at 1:100 ratio, and left to grow at 30^{o}C with shaking (80 r.p.m.), for 3-4 days. Cells were harvested by centrifugation (3,700 r.p.m. for 15 min) at 4^{o}C to produce a solid pellet. Cell pellets were resuspended in 2 ml 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 Complete Mini EDTA-free protease inhibitor cocktail (Roche) according to the manufacturer's instructions. Cells were lysed using a sonicator and the lysate was solubilized for two hours with DDM added to a final concentration of 1% (w/v). The solubilized material was then spun on a tabletop centrifuge for 30 min (13,000 r.p.m.) and a portion of the supernatant was loaded on to a 14% SDS-PAGE gel. Protein expression was confirmed by western blot analysis (**Extended Data Fig. 7d**). Proteins from SDS-PAGE gels were transferred electrophoretically onto nitrocellulose membrane overnight at 200 mA. The membrane was blocked for one hour with 5% BSA and immobilized proteins were probed with an anti-Flag mouse antibody (Sigma's Anti-FLAG® M2 Antibody: #8146)) used at 1:5000 dilution. Immobilized protein:FLAG antibody complexes were detected using a secondary IR-labelled, goat anti-mouse green antibody (IRDye 800CW secondary antibodies: #926-32210) diluted 1:10,000. The immunoblot was developed and quantified using an Odyssey system (LI-COR Biosciences, Lincoln, NE) (**Extended Data Fig. 7d**).

Functional analysis of WaaL variants by LPS gel. The *Ec* W3110 Δ*waaL* mutant was generated by P1 *vir* phage transduction[60] using the *waaL*::km mutant from the Keio collection[61] and the strain confirmed by PCR. *Ec* cells were grown in lysogeny broth (LB) broth at 37°C and *Cm* in tryptic soy (TB) broth at 30°C supplemented with antibiotics. Cells were harvested at OD_{600} of 1.0 and resuspended in 100 µl 1X LDS sample buffer (Novex), containing 4% 2-mercaptoethanol (Sigma). Whole cells were treated with proteinase K as previously described[62]. LPS samples were separated by 4-12% NUPAGE Bis-Tris gel and visualized by silver staining (for *Ec* samples)[62] or using ProQ Emerald 300 from Molecular Probes (for *Cm*). For *E. coli* work only, all strains also contained plasmid pMF19 for expression of WbbL, a rhamonsyltransferease required for O-antigen synthesis, as the *wbbL* gene in K12 strains contain an IS*5* insertion mutation.

Statistics and Reproducibility. For reproducibility all O-antigen ligase assay gels where run in triplicates (**Fig. 2c, 4c, Extended Data Fig. 1c, 7a,** and **7e**), all SDS page gels and westerns were run in duplicates (**Extended data Fig. 1e, 7c** and **7d**). The initial screen single point ELISA to identify potential binders was carried out once (**Extended data Fig. 1g**), and the multi-point sAB ELISA on the initial hits was run in triplicates (**Extended data Fig. 1h**).

Bioinformatics sequence analysis and homology modelling *Ec*WaaL. WaaL homologues were identified using Hhblits[63], from the UniClust UniRef30 sequence database[64]. The resulting sequence alignment was illustrated using Weblogo3[65]. Co-evolution analysis for WaaL was performed using MapPred[66] and visualised with PyMOL. The Hhblits sequence alignment, in combination with the model of *Ec*WaaL from the AlphaFold database[20], was used to create and refine a pairwise sequence alignment between *Cm*WaaL and *Ec*WaaL. The sequence alignment was visualised using ESPript [67]. Modeller 9.24[68] was used to generate the

*Ec*WaaL homology model, with the co-evolutionary data used to evaluate the pairwise residue distances in the resulting structure.

Molecular dynamics simulations setup. All simulations were run using GROMACS 2020[69]. The Martini 2.2 force field[69] was used to run a Coarse-Grained (CG) MD simulation to permit the assembly and equilibration of a palmitoyl-oleolyl-phosphatidylglycerol (POPG) and palmitoyl-oleoyl-phosphatidylethanolamine (POPE) (1:4 mole ratio) bilayer around *Cm*WaaL[70] in the apo, Und-PP bound, Und-PP and lipid A core OS bound, or WaB10 Fab-bound states. The initial glycerophospholipid bilayer was created using the *insane* python script[71,72] after which the *Cm*WaaL binding partners were added manually. An elastic network of 1000 kJ mol^{-1} nm^{-2} was applied between all backbone beads between 0.5 and 1 nm. Electrostatics were described using the reaction field method, with a cut-off of 1.1 nm using the potential shift modifier and the Van der Waals interactions were shifted between 0.9-1.1 nm. The systems were first energy minimised by steepest descent algorithm to 1000 kJ mol^{-1} nm^{-2} and then simulated for a total of 200 ns. Temperature and pressure were kept constant throughout the simulation at 310 K and 1 bar respectively, with protein, lipids and water/ions coupled individually to a temperature bath by the V-rescale method[73] and a semi-isotropic Parrinello-Rahman barostat[74]. The final snapshots from the CG simulations were then converted back to an atomistic description using CG2AT2 using the protein-aligned method[75].

Atomistic molecular dynamics simulations. All ionisable groups were simulated with default protonation states, unless otherwise mentioned. The CHARMM36m forcefield[76] was employed, with the use of a 2 fs time step during the simulations. Electrostatics were described using PME, with a cut-off of 1.2 nm and the van der Waals interactions were shifted between 1-1.2 nm. The TIP3P water model was used and water bond angles and distances were constrained by SETTLE[77]. All other bonds were constrained using the LINCS algorithm[78]. The systems were then equilibrated for an additional 1 ns using a 2 fs timestep, with positional restraints of 1000 kJ mol^{-1} nm^{-2} on the protein heavy atoms, in an NPT ensemble, with temperature V-rescale coupling at 310 K with protein, lipids and water/ions coupled individually[73] and semi-isotropic Parrinello-Rahman barostat at 1 bar[74]. Where present during the equilibration, the non-hydrogen atoms of the first isoprenyl unit and pyrophosphate were additionally restrained with positional restraints of 1000 kJ mol^{-1} nm^{-2} applied to the coordinates taken from the cryo-EM structure. The parameters for Und-PP are part of the CHARMM36m force field[76,79]. The simulations for the four states were performed without position restraints for a total of 500 ns and run in triplicate.

Molecular dynamics simulation analysis. RMSD and RMSF calculations were performed on the backbone of *Cm*WaaL using the gmx rms and gmx rmsf tools, respectively. The dynamics of Und-PP over a total of 1.5 μs were calculated on a per atom basis from the three replicates via the gmx rmsf tool[69]. The TM helix 9 movement was calculated via the distance between the geometric centers of residues 184-295 and 229-245 using the PLUMED v2.5 software package[80]. Membrane deformation was calculated by mapping the xy coordinates for each phosphate over a total of 1.5 μs of simulation onto a 1 Å resolution grid. The grid represented by beads is coloured according to the deviation from the average membrane phosphate z coordinate for each leaflet.

Methods and Supplementary Material References

31 Punta, M. *et al.* Structural genomics target selection for the New York consortium on membrane protein structure. *Journal of structural and functional genomics* **10**, 255 (2009).

32 Mancia, F. & Love, J. High-throughput expression and purification of membrane proteins. *Journal of structural biology* **172**, 85-93 (2010).

33 Mancia, F. & Love, J. High throughput platforms for structural genomics of integral membrane proteins. *Current opinion in structural biology* **21**, 517-522 (2011).

34 Bayburt, T. H., Grinkova, Y. V. & Sligar, S. G. Self-assembly of discoidal phospholipid bilayer nanoparticles with membrane scaffold proteins. *Nano letters* **2**, 853-856 (2002).

35 Denisov, I. G., Grinkova, Y. V., Lazarides, A. A. & Sligar, S. G. Directed self-assembly of monodisperse phospholipid bilayer Nanodiscs with controlled size. *Journal of the American Chemical Society* **126**, 3477-3487 (2004).

36 Kapust, R. B., Tözsér, J., Copeland, T. D. & Waugh, D. S. The P1′ specificity of tobacco etch virus protease. *Biochemical and biophysical research communications* **294**, 949-955 (2002).

37 Rizk, S. S. *et al.* Allosteric control of ligand-binding affinity using engineered conformation-specific effector proteins. *Nature structural & molecular biology* **18**, 437 (2011).

38 Miller, K. R. *et al.* T cell receptor-like recognition of tumor in vivo by synthetic antibody fragment. *PloS one* **7**, e43746 (2012).

39 Dominik, P. K. *et al.* Conformational chaperones for structural studies of membrane proteins using antibody phage display with nanodiscs. *Structure* **24**, 300-309 (2016).

40 Kim, J. *et al.* Structure and drug resistance of the Plasmodium falciparum transporter PfCRT. *Nature* **576**, 315-320 (2019).

41 Suloway, C. *et al.* Automated molecular microscopy: the new Leginon system. *Journal of structural biology* **151**, 41-60 (2005).

42 Slabinski, L. *et al.* XtalPred: a web server for prediction of protein crystallizability. *Bioinformatics* **23**, 3403-3405 (2007).

43 Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta crystallographica section D: biological crystallography* **60**, 2126-2132 (2004).

44 Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallographica Section D: Biological Crystallography* **66**, 486-501 (2010).

45 Casañal, A., Lohkamp, B. & Emsley, P. Current developments in Coot for macromolecular model building of Electron Cryo- microscopy and Crystallographic Data. *Protein Science* **29**, 1055-1064 (2020).

46 Afonine, P. V. *et al.* Towards automated crystallographic structure refinement with phenix. refine. *Acta Crystallographica Section D: Biological Crystallography* **68**, 352-367 (2012).

47 Afonine, P. V. *et al.* Real-space refinement in PHENIX for cryo-EM and crystallography. *Acta Crystallographica Section D: Structural Biology* **74**, 531-544 (2018).

48 Pettersen, E. F. *et al.* UCSF Chimera—a visualization system for exploratory research and analysis. *Journal of computational chemistry* **25**, 1605-1612 (2004).

49 Goddard, T. D. *et al.* UCSF ChimeraX: Meeting modern challenges in visualization and analysis. *Protein Science* **27**, 14-25 (2018).

50 Danilov, L., Druzhinina, T., Kalinchuk, N., Maltsev, S. & Shibaev, V. Polyprenyl phosphates: synthesis and structure-activity relationship for a biosynthetic system of Salmonella anatum O-specific polysaccharide. *Chemistry and physics of lipids* **51**, 191-203 (1989).

51 Jiang, W., Bikard, D., Cox, D., Zhang, F. & Marraffini, L. A. RNA-guided editing of bacterial genomes using CRISPR-Cas systems. *Nature Biotechnology* **31**, 233-239, doi:10.1038/nbt.2508 (2013).

52 Jiang, Y. *et al.* Multigene Editing in the Escherichia coli Genome via the CRISPR-Cas9 System. *Applied and Environmental Microbiology* **81**, 2506-2514, (2015).

53 Selle, K. & Barrangou, R. Harnessing CRISPR–Cas systems for bacterial genome editing. *Trends in Microbiology* **23**, 225-232, (2015).

54 Jiang, X. *et al.* Vector promoters used in Klebsiella pneumoniae. *Biotechnology and Applied Biochemistry* **63**, 734-739, (2016).

55 Zhao, D. *et al.* Development of a fast and easy method for Escherichia coli genome editing with CRISPR/Cas9. *Microbial Cell Factories* **15**, 205, (2016).

56 Wang, Y. *et al.* CRISPR-Cas9 and CRISPR-Assisted Cytidine Deaminase Enable Precise and Efficient Genome Editing in Klebsiella pneumoniae. *Applied and Environmental Microbiology* **84**, e01834-01818, (2018).

57 McConville, T. H. *et al.* CrrB Positively Regulates High-Level Polymyxin Resistance and Virulence in Klebsiella pneumoniae. *Cell Reports* **33**, 108313, (2020).

58 Mijnendonckx, K. *et al.* Characterization of the Survival Ability of Cupriavidus metallidurans and Ralstonia pickettii from Space-Related Environments. *Microbial Ecology* **65**, 347-360, (2013).

59 Schmidt, C., Schwarzenberger, C., Große, C. & Nies, D. H. FurC Regulates Expression of for the Central Zinc Importer ZupT of Cupriavidus metallidurans. *Journal of Bacteriology* **196**, 3461-3471, (2014).

60 Sambrook, J. & Rusell, D. Molecular Cloning: A Laboratory Manual 3rd Ed Cold Spring Harbor Lab. *Press, Plainview, New York* (2001).

61 Baba, T. *et al.* Construction of Escherichia coli K- 12 in- frame, single- gene knockout mutants: the Keio collection. *Molecular systems biology* **2**, 2006.0008 (2006).

62 Hitchcock, P. J. & Brown, T. M. Morphological heterogeneity among Salmonella lipopolysaccharide chemotypes in silver-stained polyacrylamide gels. *Journal of bacteriology* **154**, 269-277 (1983).

63 Remmert, M., Biegert, A., Hauser, A. & Söding, J. HHblits: lightning-fast iterative protein sequence searching by HMM-HMM alignment. *Nature methods* **9**, 173-175 (2012).

64 Mirdita, M. *et al.* Uniclust databases of clustered and deeply annotated protein sequences and alignments. *Nucleic acids research* **45**, D170-D176 (2017).

65 Crooks, G. E., Hon, G., Chandonia, J.-M. & Brenner, S. E. WebLogo: a sequence logo generator. *Genome research* **14**, 1188-1190 (2004).

66 Wu, Q. *et al.* Protein contact prediction using metagenome sequence data and residual neural networks. *Bioinformatics* **36**, 41-48 (2020).

67 Robert, X. & Gouet, P. Deciphering key features in protein structures with the new ENDscript server. *Nucleic acids research* **42**, W320-W324 (2014).

68 Šali, A. & Blundell, T. L. Comparative protein modelling by satisfaction of spatial restraints. *Journal of molecular biology* **234**, 779-815 (1993).

69 Abraham, M. J. *et al.* GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **1**, 19-25 (2015).

70 Vogeley, L. *et al.* Structural basis of lipoprotein signal peptidase II action and inhibition by the antibiotic globomycin. *Science* **351**, 876-880 (2016).

71 Wassenaar, T. A., Ingólfsson, H. I., Böckmann, R. A., Tieleman, D. P. & Marrink, S. J. Computational lipidomics with insane: a versatile tool for generating custom membranes for molecular simulations. *Journal of chemical theory and computation* **11**, 2144-2155 (2015).

72 *https://github.com/Tsjerk/Insane*.

73 Bussi, G., Donadio, D. & Parrinello, M. Canonical sampling through velocity rescaling. *The Journal of chemical physics* **126**, 014101 (2007).

74 Parrinello, M. & Rahman, A. Polymorphic transitions in single crystals: A new molecular dynamics method. *Journal of Applied physics* **52**, 7182-7190 (1981).

75 Stansfeld, P. J. & Sansom, M. S. From coarse grained to atomistic: a serial multiscale approach to membrane protein simulations. *Journal of chemical theory and computation* **7**, 1157-1166 (2011).

76 Huang, J. *et al.* CHARMM36m: an improved force field for folded and intrinsically disordered proteins. *Nature methods* **14**, 71-73 (2017).

77 Miyamoto, S. & Kollman, P. A. Settle: An analytical version of the SHAKE and RATTLE algorithm for rigid water models. *Journal of Computational Chemistry* **13**, 952-962, (1992).

78 Hess, B., Bekker, H., Berendsen, H. J. C. & Fraaije, J. G. E. M. LINCS: A linear constraint solver for molecular simulations. *Journal of Computational Chemistry* **18**, 1463-1472, (1997).

79 Jo, S., Kim, T., Iyer, V. G. & Im, W. CHARMM-GUI: A web-based graphical user interface for CHARMM. *Journal of Computational Chemistry* **29**, 1859-1865, (2008).

80 Bonomi, M. *et al.* Promoting transparency and reproducibility in enhanced molecular simulations. *Nature methods* **16**, 670-673 (2019).

81 Whitney, J. & Howell, P. Synthase-dependent exopolysaccharide secretion in Gram-negative bacteria. *Trends in microbiology* **21**, 63-72 (2013).

82 Whitfield, C. Biosynthesis and assembly of capsular polysaccharides in Escherichia coli. *Annu. Rev. Biochem.* **75**, 39-68 (2006).

83 Cuthbertson, L., Kos, V. & Whitfield, C. ABC transporters involved in export of cell surface glycoconjugates. *Microbiology and Molecular Biology Reviews* **74**, 341-362 (2010).

84 Pérez-Burgos, M. *et al.* Characterization of the exopolysaccharide biosynthesis pathway in Myxococcus xanthus. *Journal of bacteriology* **202** (2020).

85 Mi, W. *et al.* Structural basis of MsbA-mediated lipopolysaccharide transport. *Nature* **549**, 233-237 (2017).

86 Fellouse, F. A. *et al.* High-throughput generation of synthetic antibodies from highly functional minimalist phage-displayed libraries. *Journal of molecular biology* **373**, 924-940 (2007).

87 Rubinstein, J. L. & Brubaker, M. A. Alignment of cryo-EM movies of individual particles by optimization of image translations. *Journal of structural biology* **192**, 188-195 (2015).

Acknowledgements

We gratefully acknowledge the assistance of members of the Mancia lab, and of the Columbia University cryo-EM facility. We would like to thank Professor Gideon Davies, University of York, for his precious input on the glycosyl transferase mechanism. This work was funded by NIH grants GM132120 (to F.M.), AI150098, AI138576, and AI129940 (to M.S.T.), GM117372 (to A.A.K.) and GM116799 (to Wayne A. Hendrickson), U54 DK104309 (ACU), T32 AI100852, and K08 AI146284 (THM). Research in P.J.S.'s lab was funded by Wellcome (208361/Z/17/Z), the MRC (MR/S009213/1) and BBSRC (BB/P01948X/1, BB/R002517/1 and BB/S003339/1). This project made use of time on ARCHER and JADE granted via the UK High-End Computing Consortium for Biomolecular Simulation, HECBioSim (http://hecbiosim.ac.uk), supported by EPSRC (grant no. EP/R029407/1). P.J.S. acknowledges Athena and Sulis at HPC Midlands+, which were funded by the EPSRC on grants EP/P020232/1 and EP/T022108/1, and the University of Warwick Scientific Computing Research Technology Platform for computational access. C.L.B.G. is funded by BBSRC studentship grant BB/M01116X/1 and D.I.R. is funded by a Schaefer Research Scholars Program Awards to Columbia University and MRC grant MR/N002679/1. Some of the work was performed at the Center for Membrane Protein Production and Analysis (COMPPÅ) and at the National Resource for automated Molecular Microscopy at the National Resource for Molecular Microscopy at the Simons Electron Microscopy Center, both located at the New York Structural Biology Center.

Author contributions

K.U.A. with help from B.K., M.B.D. and A.P.Z., performed the genomic expansion screen, protein expression and purification. S.K.E., K.N. and A.A.K. identified and purified the Fabs. K.U.A. produced and analyzed the cryo-EM data, and built the model with help from O.B.C.. Mutational analysis were performed by K.U.A., V.I.P. and S.I.G.. Gene editing for *Cupriavidius metallidurans* were performed by T.H.M. and A.C.U.. Assessment of WaaL function was carried out by C.M.H and M.S.T.. All molecular dynamics simulations were performed by P.J.S. and O.N.V.. The ligands for TLC analysis were synthesized by K.S.T. under the guidance of E.S., and K.U.A. performed the TLC analysis. K.U.A., F.M., P.J.S., O.N.V., C.M.H., and M.S.T. designed experiments and wrote the paper with R.N., C.L.B.G., and D.I.R.. Oversight for the entire project was provided by F.M.

Competing interests

The authors declare no competing interests.

Correspondence and requests for materials

Should be addressed to strent@uga.edu (M.S.T.), phillip.stansfeld@warwick.ac.uk (P.J.S.) or fm123@cumc.columbia.edu (F.M.).

Data availability

All raw movie frames have been deposited into EMPIAR, with accession code EMPIAR-10938. The density maps have been deposited into EMDB, with accession code EMD-26054 for the Und-PP bound *Cm*WaaL and EMD-26057 for the apo *Cm*WaaL. Both models have been deposited in the PDB, with accession code 7TPG for the Und-PP bound, and 7TPJ for the apo *Cm*WaaL model. All raw gels are available in the supplementary information.

Extended Data Figure Legends

Extended Data Table 1 | Cryo-EM Data. Cryo-EM data collection, refinement and validation statistics.

Extended Data Figure 1 | Functional validation of *Cm*WaaL, identification of WaaL-specific Fabs and preparation of nanodisc-reconstituted WaaL-Fab complex for structural analysis. a, Schematic representation of O-antigen synthesis and transfer to the periplasmic leaflet of the inner membrane by the three different pathways, the arrows represent the direction of the Und-PP linked O-antigen takes in each pathway. Individual lipid-linked O-antigen repeat units are ligated to the lipid carrier Und-PP by glycosyltransferase enzymes. In the Wzy-dependent pathway, the O units are transported into the periplasm by the flippase Wzx. Wzy then catalyses the polymerization of O-antigen repeats, while Wzz controls the preferred modal length of the final O-antigen polymer[9]. The synthase dependent pathway is the least well characterized pathway[81], the O-antigen is assembled at the cytoplasmic face of the inner membrane by a synthase that is also involved in its transportation across the membrane. In the ABC-dependent pathway, the polymerized Und-PP-O-antigen molecule is flipped to the periplasmic face of the inner membrane by an ABC transporter, Wzm-Wzt flippase[81-84]. It is important to note that the chemical composition of the *C. metallidurans* O-antigen is unknown. **b,** Schematic of WaaL function. On the right, the lipid A core OS is synthesized in the cytoplasm and flipped to the periplasm via MsbA[85]. On the left, the lipid A core OS and the O-antigen, irrespective of the pathway of origin, are ligated by WaaL. **c,** Functional analysis of *Cm*WaaL ligase activity in whole cells. LPS gel showing that O-antigen ligase activity is abolished when *Cm waaL* is deleted, and activity is restored by plasmid complementation. **d,** Size exclusion chromatography elution profiles of purified *Cm*WaaL in detergent (blue), *Cm*WaaL incorporated into a nanodisc (red), and *Cm*WaaL incorporated into a nanodisc with Fab (WaB10) bound (black). **e,** SDS-PAGE gel of *Cm*WaaL purification. First lane is *Cm*WaaL purified in DDM, second lane is *Cm*WaaL reconstituted into nanodiscs (MSP1E3D1 and POPG), and third lane is *Cm*WaaL reconstituted into nanodiscs (MSP1E3D1 and POPG) with Fab (WaB10) bound. **f,** Complementarity-determining region (CDR) sequences of unique synthetic antigen binders (sABs) from biopanning against *Cm*WaaL in MSPE3D1 nanodiscs. sABs were selected following multiple rounds of phage display starting from Fab Library E[37,38]. Enriched YSGW residues are highlighted by coloured boxes (yellow, red, green, and blue, respectively). YSGW residues have been previously shown to play dominant roles in highly specific and high affinity antigen recognition[86]. **g,** Single-point ELISA measuring the binding of phage-displayed sABs to *Cm*WaaL in MSP1E3D1 nanodiscs (red), empty nanodiscs (light grey), or buffer (empty wells, dark grey). ELISA signal measured at 450 nm absorbance, see Supplementary Table 1. **h,** Multi-point sAB ELISA: EC_{50} estimation for purified sAB binding to *Cm*WaaL incorporated into MSP1E3D1 nanodiscs, showing high affinity binding of WaE8 (green, 6.6±0.045 nM), WaB10 (red, 1.87±0.07 nM), WaC9 (orange, 6.26±0.18 nM), WaG11 (cyan, 3.31±0.06 nM), and WaC10 (magenta, 3.90±0.09 nM), and modest affinity binding of WaF10 (blue, 279.5±0.68 nM) and WaB12 (brown, 154±0.11 nM), see Supplementary Table 1. EC_{50} values represent the mean of three independent experiments +/- standard error (n=3).

Extended Data Figure 2 | Cryo-EM analysis of the *Cm*WaaL-Fab complex. a, Flow chart outlining cryo-EM data processing and refinement performed to obtain a structure of a nanodisc-reconstituted *Cm*WaaL with the Fab WaB10, both for the apo and the Und-PP bound structures. **b,** On the left, representative micrograph (2.44 µm defocus). On right, representative two-dimensional class averages from CryoSPARC two-dimensional classification[86,87]. **c,** Fourier shell correlation (FSC) curves for the Und-PP bound *Cm*WaaL-Fab complex. **d,** Local resolution display of unsharpened reconstructions of Und-PP bound *Cm*WaaL in complex with the WaB10 Fab, in orthogonal views. **e,** Euler angle distribution of all Und-PP bound particles used in the final map reconstruction. Final map shown in green. Each orientation is represented by a cylinder, with each cylinder's height and color (from blue to red) proportional to the number of particles for that specific direction.

Extended Data Figure 3 | Fit of cryo-EM density with model. Cryo-EM densities (mesh) are superimposed on TM and PH helices of the *Cm*WaaL model. The model is rendered as sidechain, coloured in rainbow, as in **Fig. 1d**. Und-PP (gold) is shown as sticks.

Extended Data Figure 4 | Interaction of *Cm*WaaL and Fab in the complex, and Sequence Alignment WebLogo. a, *Cm*WaaL-WaB10 complex structure shown in ribbon with WaaL in grey and WaB10 in pink. Only the variable domain of WaB10 was modelled into the map. **b,** Interface between *Cm*WaaL (grey) and WaB10 (pink). Residues shown in sticks (I137, D245, S261 for *Cm*WaaL, and N31, Y93, F104 for Fab). **c,** A Weblogo for orthologues of *Cm*WaaL annotated with TM and soluble helices. The numbering for *Cm*WaaL is shown.

Extended Data Figure 5 | Structural features and analysis of *Cm*WaaL. a, *Cm*WaaL rendered in surface representation coloured by electrostatic potential on a range of ±5 kBT/e. **b,** *Cm*WaaL surface coloured by Wimley-White hydrophobicity, on a cyan (very hydrophilic) to gold (very hydrophobic) scale. **c,** *Cm*WaaL surface coloured by residue conservation on a green (no conservation) to purple (absolute conservation) scale. **d,** TLC analysis of detergent purified *Cm*WaaL. *Cm*WaaL was purified in detergent and run on the

TLC plate after organic-phase extraction of lipids. POPG, Und-P, Und-PP and an unrelated control protein expressed in *E. coli* were run in separate lanes as standards. MCR-1 was chosen as it utilizes a lipid donor (phosphatidylethanolamine) to modify the lipid A domain of LPS, but does not utilize Und-PP. **e,** Und-PP binding site. Residues coordinating the first two isoprenyls in the Und-PP tail, shown as sticks (blue). Und-PP shown as sticks coloured golden. **f,** (Top) MD simulations showing the Und-PP binding to *Cm*WaaL from two views, and the flexibility of the Und-PP tail beyond the first two isoprenyl groups. The Und-PP shows increasing mobility away from the pyrophosphate. Here, the Root Mean Squared Fluctuation (RMSF) of Und-PP is shown for 3 repeats of 500 ns simulation. The non-hydrogen atoms of Und-PP are coloured by RMSF from blue to red. (Bottom) The RMSF of all atoms of the Und-PP are shown. The portion of the Und-PP resolved in the cryo-EM density map is highlighted in red.

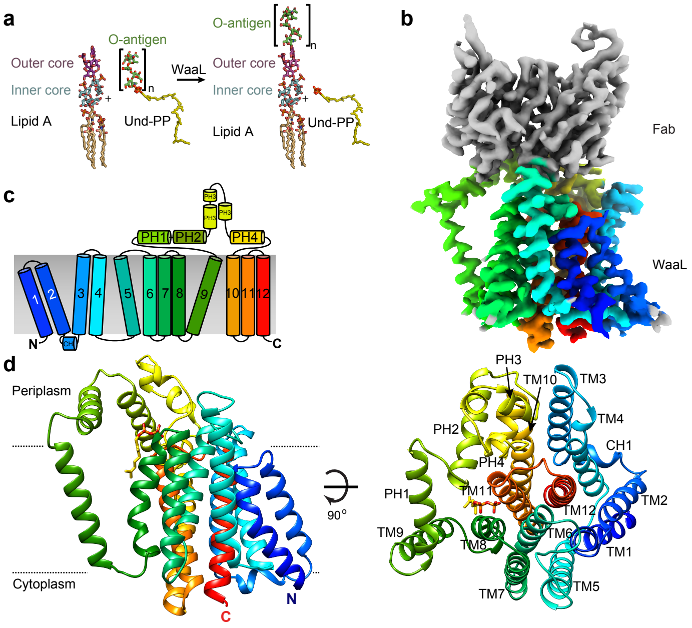
Extended Data Figure 6 | *Ec*WaaL homology model. a, Co-evolutionary analysis for *Cm*WaaL calculated using MapPred and mapped onto the cryo-EM structure of *Cm*WaaL, using a threshold of 0.2[66]. Predicted contacts between Cα atoms of given residues are shown as dashes for distances less than 10 Å (green), between 10 and 12 Å (cyan), between 12 and 20 Å (yellow), and above 20 Å (red). The *Cm*WaaL structure is shown as a grey ribbon representation. **b,** Comparison of the *Cm*WaaL structure with the homology model of *Ec*WaaL, both coloured in rainbow from N- to C-termini. Und-PP is shown bound to both structures and coloured in gold, as well as the key arginine residues that surround Und-PP, show as sticks. **c,** Sequence alignment and secondary structure of *Cm*WaaL and *Ec*WaaL. Conserved residues are highlighted in black.

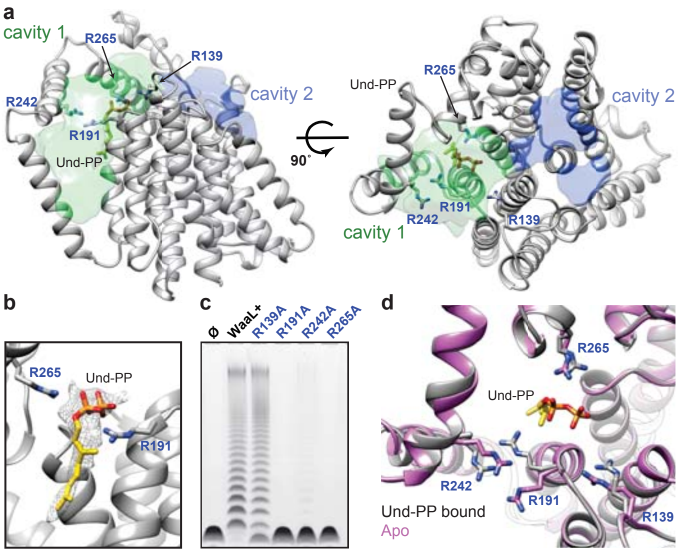
Extended Data Figure 7 | Analysis of *Cm*WaaL and *Ec*WaaL ligase activity. **a,** Functional analysis of *Ec*WaaL ligase activity in whole cells by LPS gel analysis. *Ec* LPS profile. W3110 Δ*waaL* containing either empty vector pWSK29 (Ø), pWSK29::*Ec*WaaL (WT) or pWSK19::*Ec*WaaL-variants[101] was evaluated for O-antigen extension. W3110 *Ec*WaaL point mutations that cause loss of ligase activity. **b,** Table showing key residues in *Cm*WaaL and their corresponding residues in *Ec*WaaL. **c,** SDS-PAGE gel of all *Ec*WaaL mutants that were purified to verify expression. **d,** Western blot analysis, using a mouse monoclonal anti-Flag antibody, of Flag purified WT *Cm*WaaL and mutants, grown in *C. metallidurans*. **e,** Functional analysis of *Cm*WaaL ligase activity in whole cells by LPS gel analysis *C. metallidurans* Δ*waaL* containing either empty vector pBBR1(Ø), pBBR1:*Cm*WaaL (WT) or pBBR1:*Cm*WaaL-variants was evaluated for O-antigen extension. **f**, Top view of *Ec*WaaL showing the residues mutated in the two right panels in panel **a**. **g,** Top view of the *Ec*WaaL model, highlighting T170 (left panel) and when mutated to Trp (right panel). **h,** Representative views of lipid A bound to *Ec*WaaL within the interface of cavity 2 (left panel) and an alternative binding site within the Und-PP pocket of cavity 1 (right panel). Lipid A-core shown as sticks, Und-PP (gold) and H338 shown as spheres.

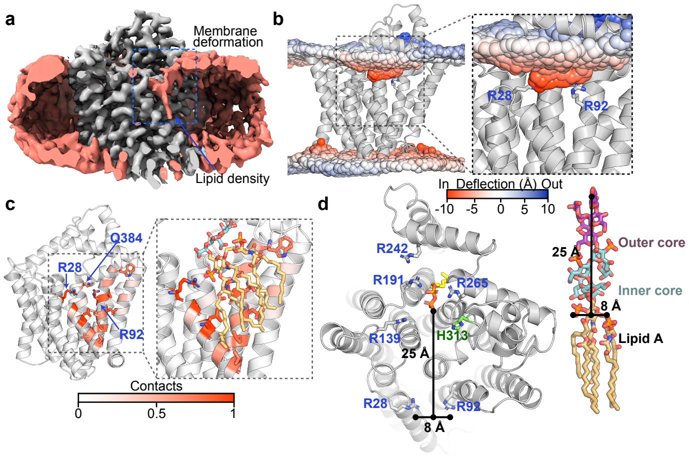
Extended Data Figure 8 | Comparison of the Und-PP bound and the apo WaaL *Cm* structures. **a**, Top Views of the Und-PP bound (grey) and the apo (pink) *Cm*WaaL structures, showing the density (mesh) of the Und-PP (yellow) in the ligand bound structure in comparison to the apo, structure. **b,** Cryo-EM density maps for the Und-PP bound (grey) and the apo (pink) *Cm*WaaL. Density maps were prepared in chimeraX[49], by deleting any density within a 4 Å radius of the Fab in the final model. **c,** Side views of the Und-PP bound (grey) and the apo (pink) *Cm*WaaL shown as ribbon. The Und-PP (yellow) is shown as sticks in the bound structure. **d,** Top views of the Und-PP bound (grey) and the apo (pink) *Cm*WaaL showing the key residues that we hypothesize play a role in either binding/shuttling or ligation of the substrates. The density for the selected residues is shown as grey mesh. On the right an overlay of bound and apo states are shown with highlighted residues shown in stick representation.

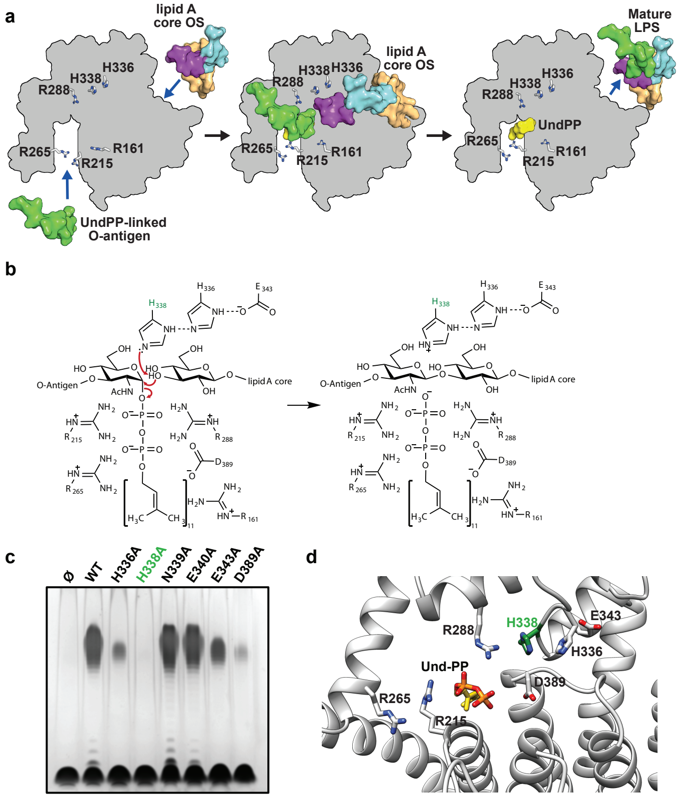
Extended Data Figure 9 | *Cm*WaaL Molecular Dynamics Simulations. a, The RMSF of the backbone of *Cm*WaaL. The RMSF measurements were averaged across 3 repeats of 500 ns simulation. The grey shading refers to the standard deviation across the repeats. Two highly mobile domains are highlighted in yellow (TM helix 9 and PH1) and purple (PH3). These domains have been highlighted on the ribbon structure for reference. The ribbon structures shown are coloured by their respective RMSF from blue to red. **b,** Representative frames for the closed (red) and open (blue) states are shown, derived from the simulations. The mobility of TM helix 9 is demonstrated by the histograms of the distance between the geometric centers of residues 229-244 and 184-195. The simulations of *Cm*WaaL with Und-PP bound show a stabilisation of the closed state, while TM helix 9 separates from the core of *Cm*WaaL in the apo state, opening an access channel to the active site. The initial distance from the cryo-EM structure is highlighted by a black line.

Extended Data Figure 10 | A putative common shuttling mechanism between *Cm*WaaL, RodA, ArnT and PglB. a, Structural comparison between *Cm*WaaL and *Tt*RodA (PDB ID 6PL5), coloured on a blue to red rainbow from N- to C-terminus. The two additional helices (TMs 1 and 2) of *Cm*WaaL are in grey. **b,** A comparison of the putative access pathway for polyprenyl-linked-phosphate containing ligands for *Cm*WaaL, *Tt*RodA, *Cm*ArnT (PDB ID 5F15) and *Cl*PglB (PDB ID 5OGL). The equivalent TM helix is shown in red and periplasmic helix in blue. Conserved residues, that may be involved in the shuttling and/or coordination of polyprenyl-phosphate containing ligands in *Cm*WaaL, *Tt*RodA, *Cm*ArnT and *Cl*PglB are shown in sticks and highlighted in blue.

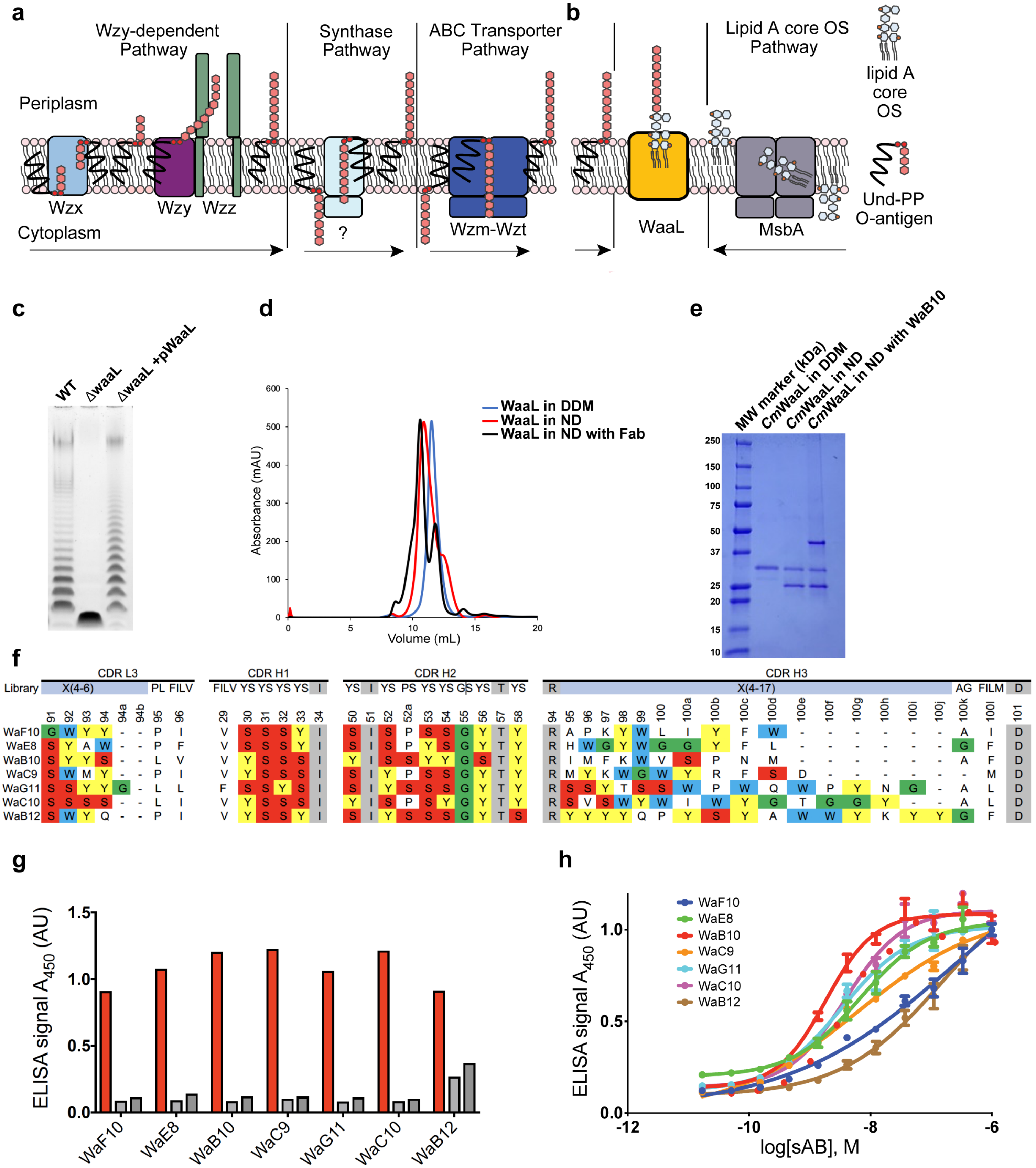

a
O-antigen
Outer core
Inner core
Lipid A
+
n
Und-PP
WaaL
O-antigen
n
Outer core
Inner core
+
Lipid A
Und-PP
b
Fab
WaaL
c
PH3
PH3
PH3
PH1
PH2
PH4
1
2
3
4
5
6
7
8
9
10
11
12
N
C
d
Periplasm
Cytoplasm
C
N
90°
PH3
TM10
TM3
TM4
PH2
PH4
CH1
PH1
TM11
TM2
TM12
TM9
TM8
TM6
TM1
TM7
TM5


a
cavity 1
R265
R139
cavity 2
R242
R191
Und-PP
90˚
Und-PP
R265
cavity 2
R191
R242
R139
cavity 1
b
R265
Und-PP
R191
c
ϕ
WaaL+
R139A
R191A
R242A
R265A
d
R265
Und-PP
R242
R191
R139
Und-PP bound
Apo


a
Membrane deformation
Lipid density
b
R28
R92
In Deflection (Å) Out
-10 -5 0 5 10
c
Q384
R28
R92
Contacts
0 0.5 1
d
R242
R191
R265
H313
R139
25 Å
R28
R92
8 Å
25 Å
Outer core
Inner core
8 Å
Lipid A


a
lipid A
core OS
H336
H338
R288
R265
R215
R161
UndPP-linked
O-antigen
lipid A
core OS
H338
H336
R288
R265
R215
R161
Mature
LPS
H338
H336
R288
UndPP
R265
R215
R161
b
H338
H336
E343
O-Antigen
AcHN
lipid A core
R215
R288
D389
R265
R161
c
Ø
WT
H336A
H338A
N339A
E340A
E343A
D389A
d
R288
H338
E343
H336
Und-PP
D389
R265
R215

	WaaL-WaB10 (Und-PP bound) (EMD – 26054) (PDB – 7TPG)	**WaaL-WaB10 (apo)** (EMD – 26057) (PDB – 7TPJ)
Data collection and processing		
Magnification	130,000	13,000
Voltage (kV)	300	300
Electron exposure (e-/Å^2)	70	70
Exposure time (s)	2.5	2.5
Dose rate (e-/pixel/s)	16	16
Nominal defocus range (μm)	1-2	1-2
Defocus range (μm)	0.8-1.2	0.8-1.2
Pixel size (Å)	1.061	1.061
Symmetry imposed	C1	C1
Number of micrographs	2,378	2,378
Initial particle images (no.)	844,438	844,438
Final particle images (no.)	39,844	30,514
Map resolution (Å)	3.23	3.46
FSC threshold	0.143	0.143
Sphericity of 3DFSC	0.944	0.944
Refinement		
Map sharpening *B* factor (Å^2)	-24.07	-12.22
Residue range	5-407 (*Cm*WaaL) 2-105 and 5-123 (Fab)	
Model composition		
Non-hydrogen atoms	4778	4759
Protein residues	626	626
Ligands	1	0
B factors (Å^2)		
Protein	43.04	35.95
Ligand	61.05	----
R.m.s. deviations		
Bond lengths (Å)	0.002	0.004
Bond angles (°)	0.621	0.700
Validation		
MolProbity score	1.63	1.83
Clashscore	7.19	9.09
Poor rotamers (%)	0.20	0.00
EM-Ringer Score	2.45	1.60
Ramachandran plot		
Favored (%)	96.45	95.00
Allowed (%)	3.55	5.00
Disallowed (%)	0.00	0.00



CDR L3

	91	92	93	94	94a	94b	95	96
Library	X(4-6)						PL	FILV
WaF10	G	W	Y	Y	-	-	P	I
WaE8	S	Y	A	W	-	-	P	F
WaB10	S	Y	Y	S	-	-	L	V
WaC9	S	W	M	Y	-	-	P	I
WaG11	S	S	Y	Y	G	-	L	L
WaC10	S	S	S	S	-	-	L	I
WaB12	S	W	Y	Q	-	-	P	I

CDR H1

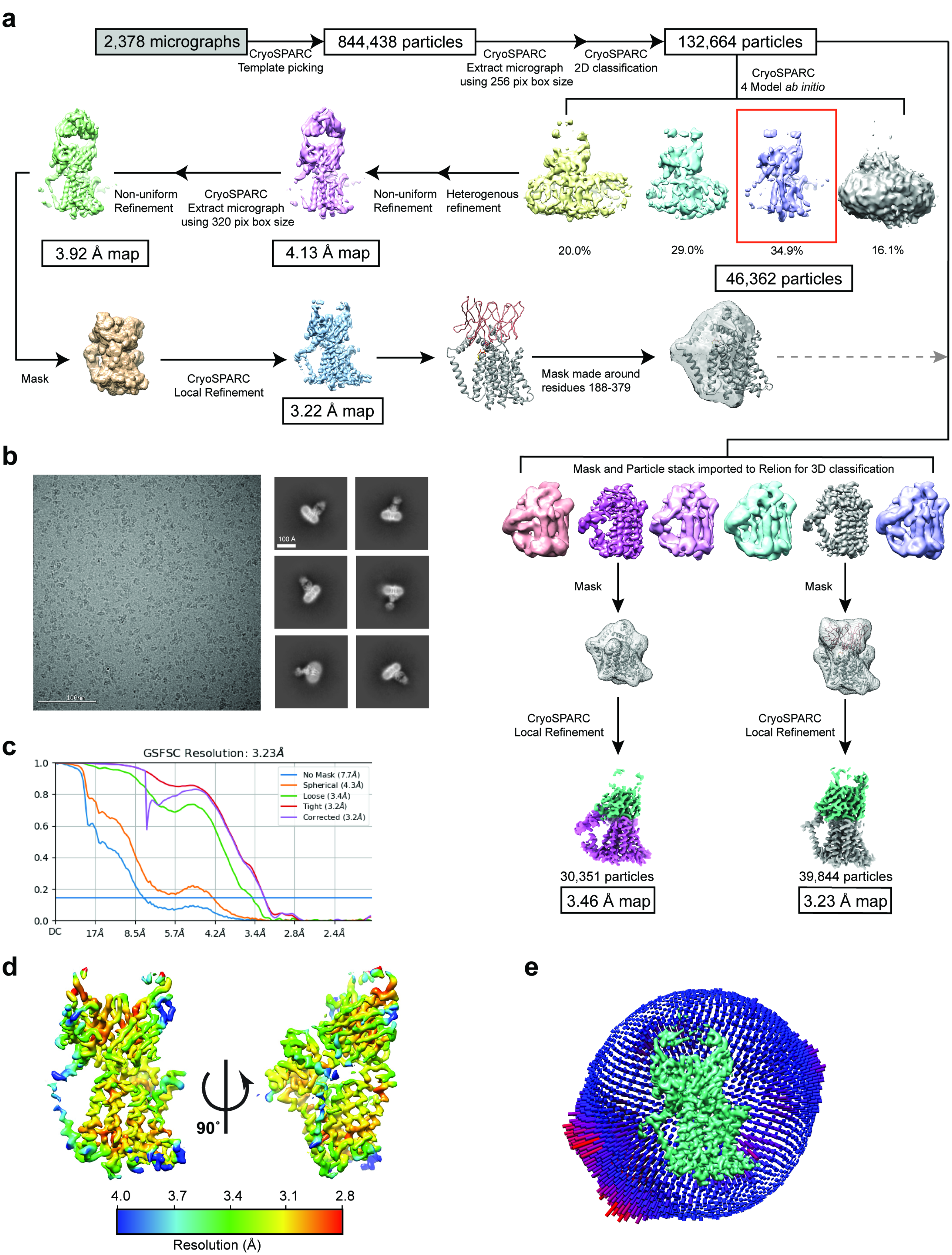
	29	30	31	32	33	34
Library	FILV	YS	YS	YS	YS	I
WaF10	V	S	S	S	Y	I
WaE8	V	S	S	S	Y	I
WaB10	V	Y	S	S	S	I
WaC9	V	Y	S	S	S	I
WaG11	F	S	S	Y	S	I
WaC10	V	Y	S	S	S	I
WaB12	V	Y	S	S	Y	I

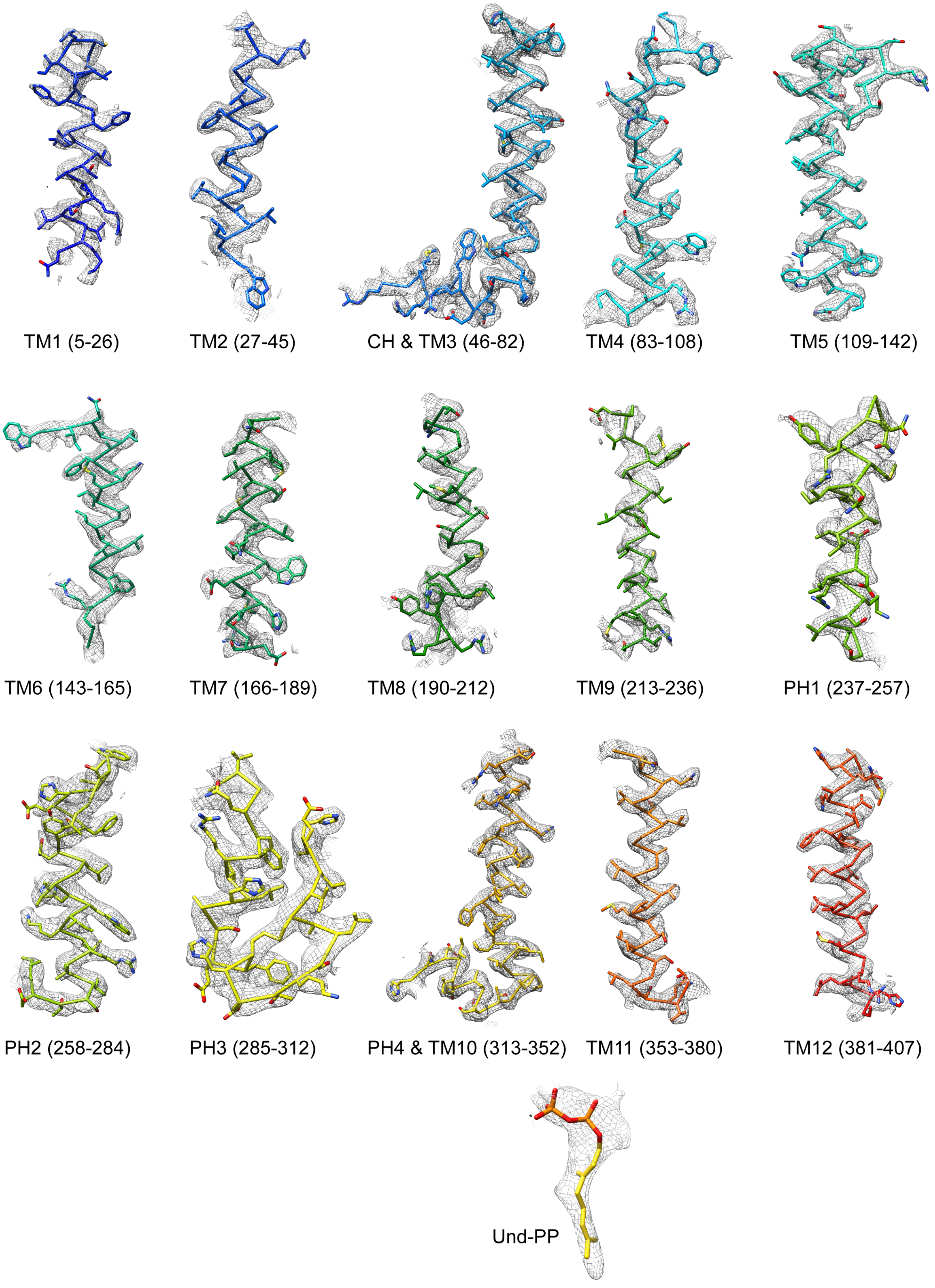
CDR H2

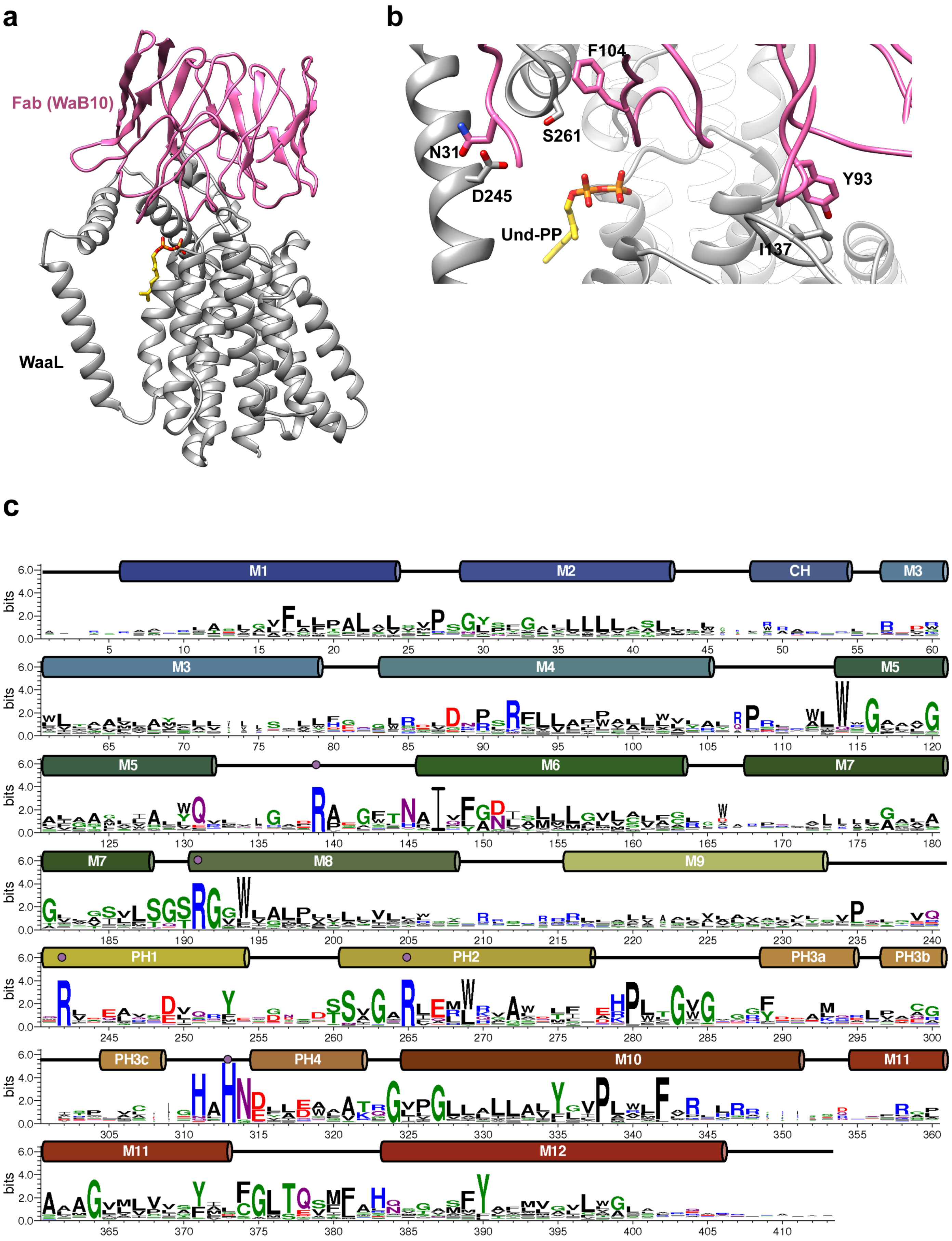
	50	51	52	52a	53	54	55	56	57	58
Library	YS	I	YS	PS	YS	YS	GS	YS	T	YS
WaF10	S	I	S	P	S	S	G	Y	T	Y
WaE8	S	I	S	P	Y	S	G	Y	T	Y
WaB10	Y	I	S	S	Y	Y	G	S	T	Y
WaC9	S	I	Y	P	S	S	G	Y	T	Y
WaG11	S	I	Y	S	S	S	G	Y	T	Y
WaC10	Y	I	S	P	S	Y	G	Y	T	Y
WaB12	S	I	Y	S	S	S	G	Y	T	S

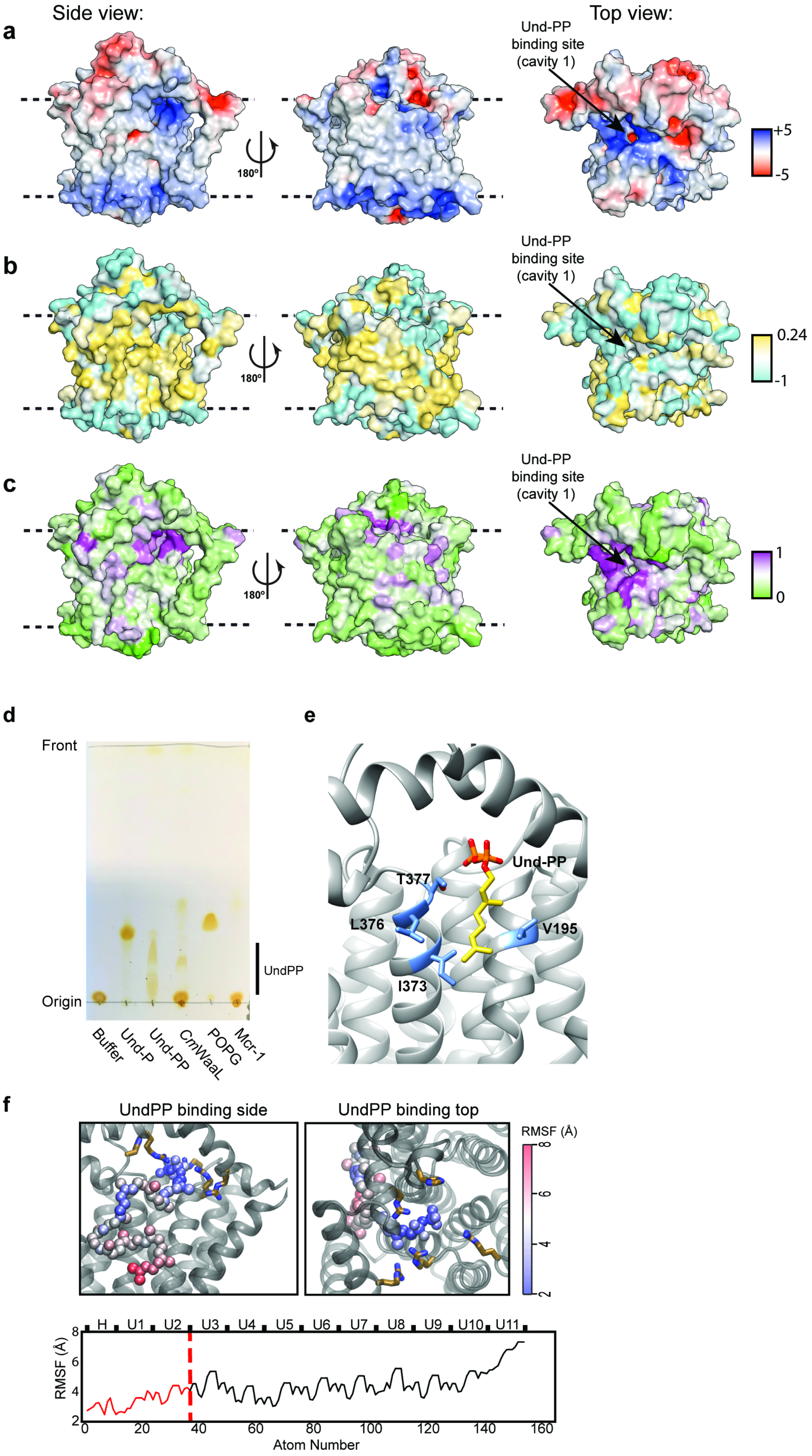
CDR H3

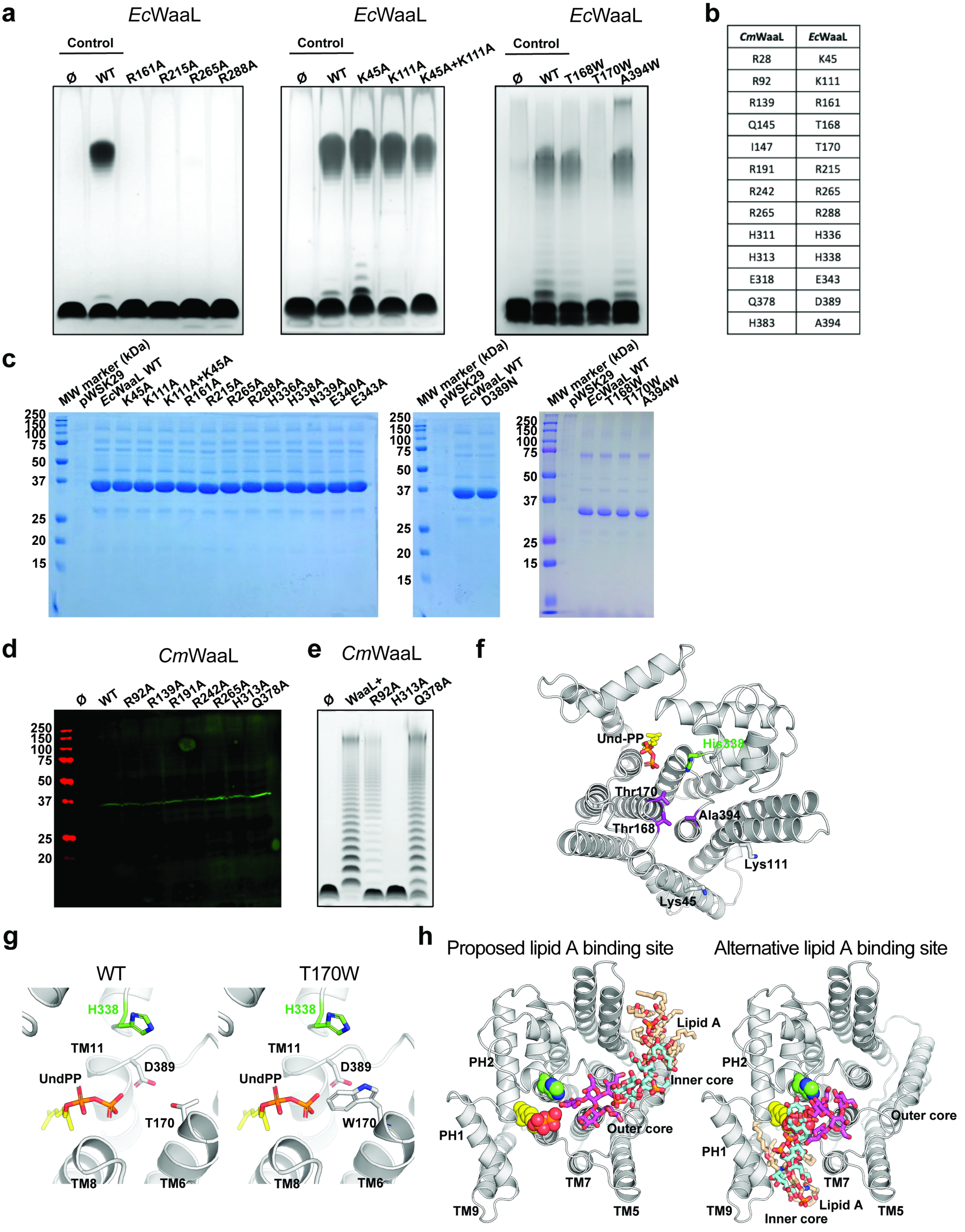
	94	95	96	97	98	99	100	100a	100b	100c	100d	100e	100f	100g	100h	100i	100j	100k	100l	101
Library	R	X(4-17)																AG	FILM	D
WaF10	R	A	P	K	Y	W	L	I	Y	F	W	-	-	-	-	-	-	A	I	D
WaE8	R	H	W	G	Y	W	G	G	Y	F	L	-	-	-	-	-	-	G	F	D
WaB10	R	I	M	F	K	W	V	S	P	N	M	-	-	-	-	-	-	A	F	D
WaC9	R	M	Y	K	W	G	W	Y	R	F	S	D	-	-	-	-	-	-	M	D
WaG11	R	S	S	Y	T	S	S	W	P	W	Q	W	P	Y	N	G	-	A	L	D
WaC10	R	S	V	S	W	Y	W	I	W	Y	G	T	G	G	Y	-	-	A	L	D
WaB12	R	Y	Y	Y	Y	Q	P	Y	S	Y	A	W	W	Y	K	Y	Y	G	F	D


a
2,378 micrographs
CryoSPARC
Template picking
844,438 particles
CryoSPARC
Extract micrograph
using 256 pix box size
CryoSPARC
2D classification
132,664 particles
CryoSPARC
4 Model *ab initio*
20.0%
29.0%
34.9%
16.1%
46,362 particles
Heterogenous
refinement
Non-uniform
Refinement
4.13 Å map
CryoSPARC
Extract micrograph
using 320 pix box size
Non-uniform
Refinement
3.92 Å map
Mask
CryoSPARC
Local Refinement
3.22 Å map
Mask made around
residues 188-379
Mask and Particle stack imported to Relion for 3D classification
Mask
Mask
CryoSPARC
Local Refinement
CryoSPARC
Local Refinement
30,351 particles
3.46 Å map
39,844 particles
3.23 Å map
b
100nm
100 Å
c
GSFSC Resolution: 3.23Å
1.0
0.8
0.6
0.4
0.2
0.0
DC
17Å
8.5Å
5.7Å
4.2Å
3.4Å
2.8Å
2.4Å
No Mask (7.7Å)
Spherical (4.3Å)
Loose (3.4Å)
Tight (3.2Å)
Corrected (3.2Å)
d
90˚
4.0
3.7
3.4
3.1
2.8
Resolution (Å)
e

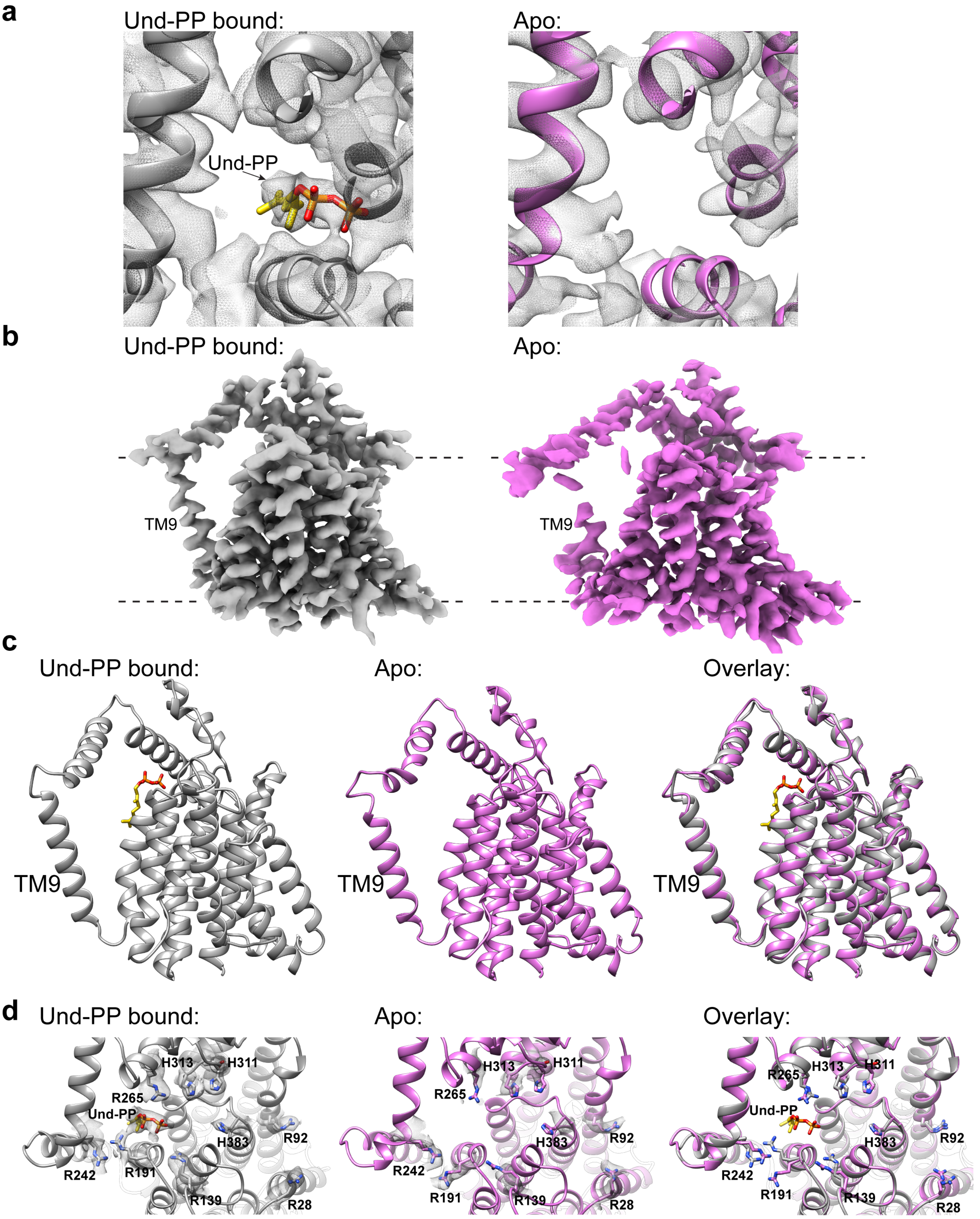

TM1 (5-26)
TM2 (27-45)
CH & TM3 (46-82)
TM4 (83-108)
TM5 (109-142)
TM6 (143-165)
TM7 (166-189)
TM8 (190-212)
TM9 (213-236)
PH1 (237-257)
PH2 (258-284)
PH3 (285-312)
PH4 & TM10 (313-352)
TM11 (353-380)
TM12 (381-407)
Und-PP

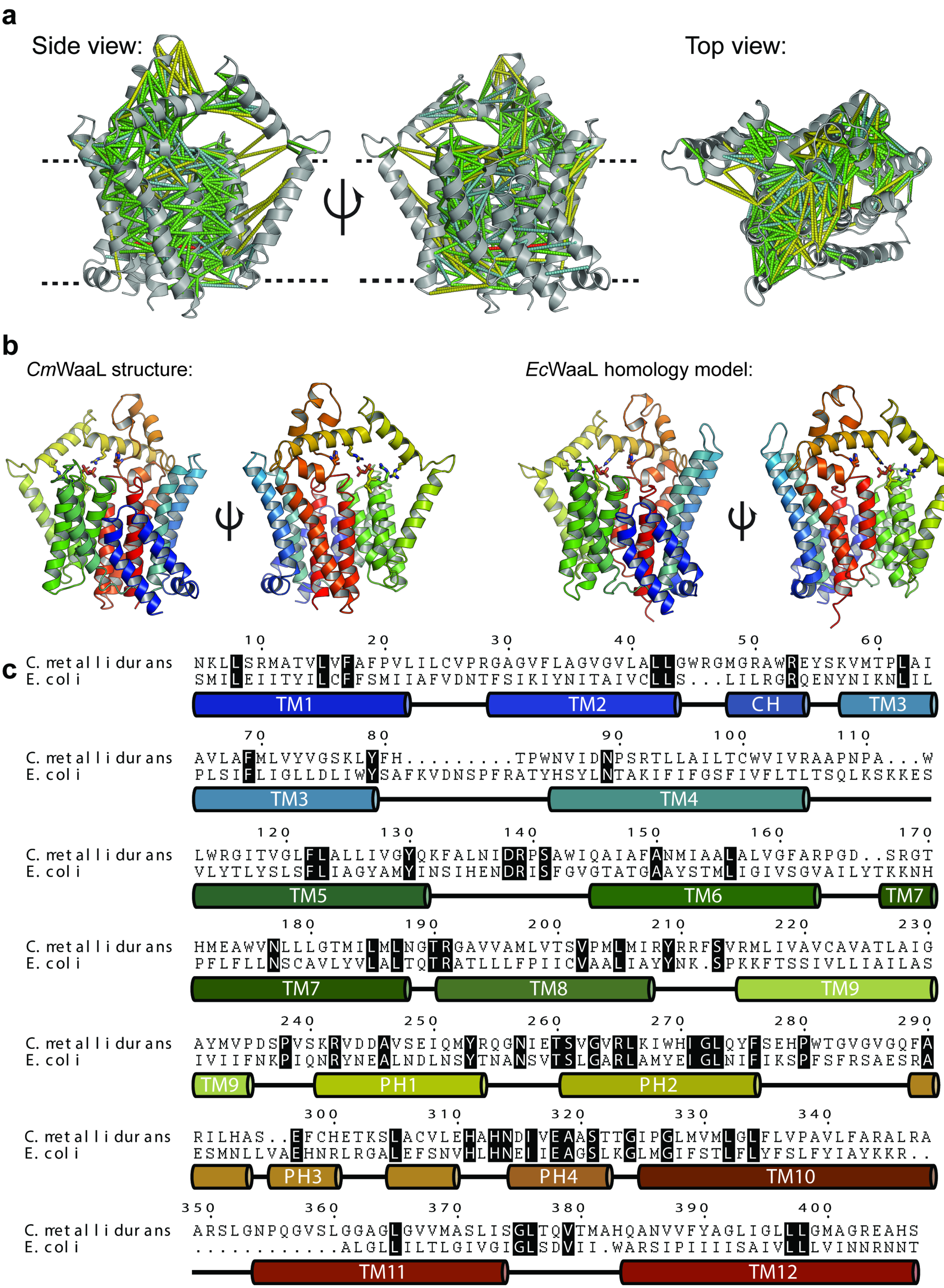

a
Fab (WaB10)
WaaL
b
F104
N31
S261
D245
Und-PP
Y93
I137
c
bits
M1
M2
CH
M3
M4
M5
M6
M7
M8
M9
PH1
PH2
PH3a
PH3b
PH3c
PH4
M10
M11
M12

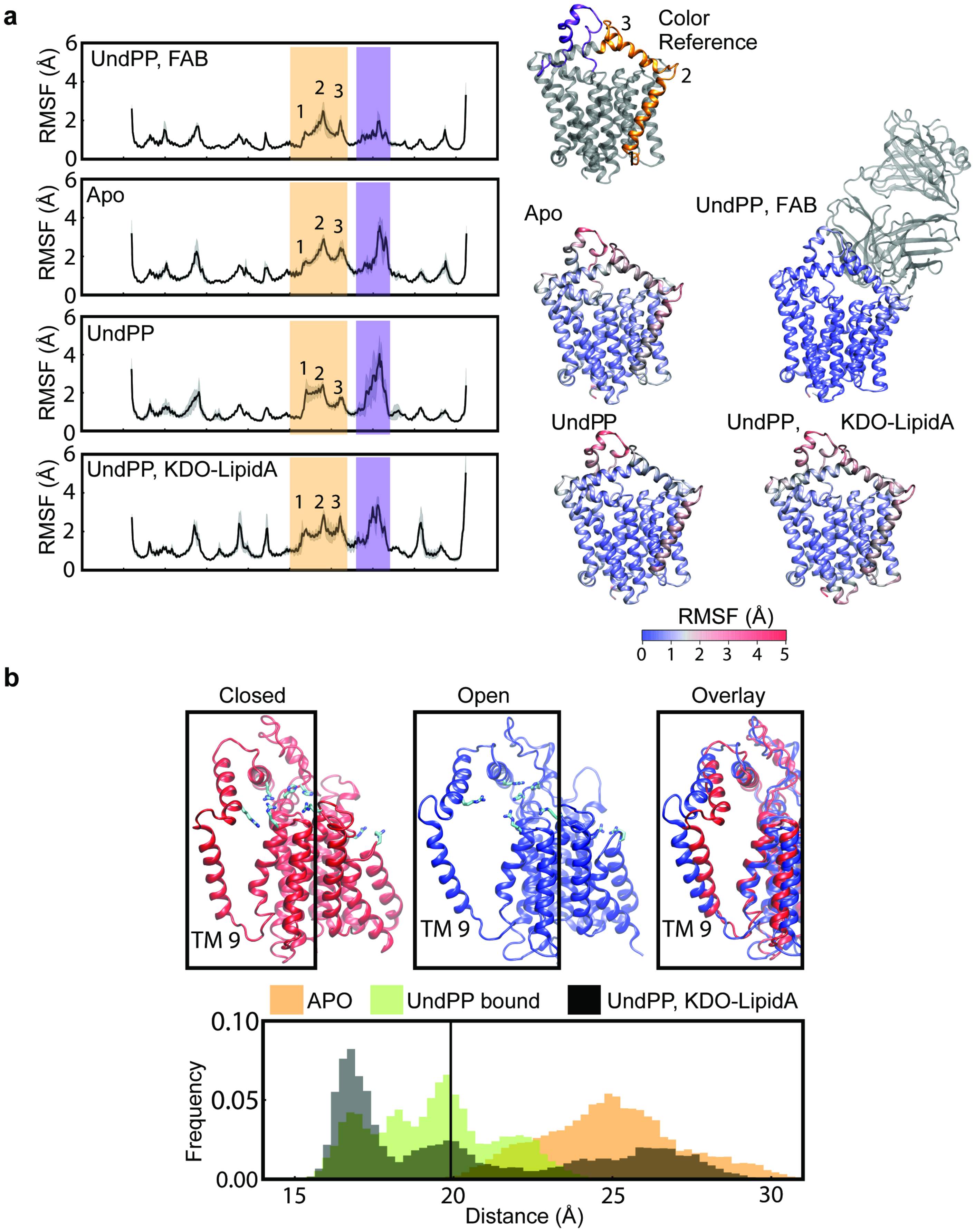

Side view:
Top view:
a
Und-PP binding site (cavity 1)
180°
+5
-5
b
Und-PP binding site (cavity 1)
180°
0.24
-1
c
Und-PP binding site (cavity 1)
180°
1
0
d
Front
Origin
UndPP
Buffer
Und-P
Und-PP
CmWaaL
POPG
Mcr-1
e
Und-PP
T377
L376
V195
I373
f
UndPP binding side
UndPP binding top
RMSF (Å)
8
6
4
2
H
U1
U2
U3
U4
U5
U6
U7
U8
U9
U10
U11
RMSF (Å)
8
6
4
2
0
20
40
60
80
100
120
140
160
Atom Number

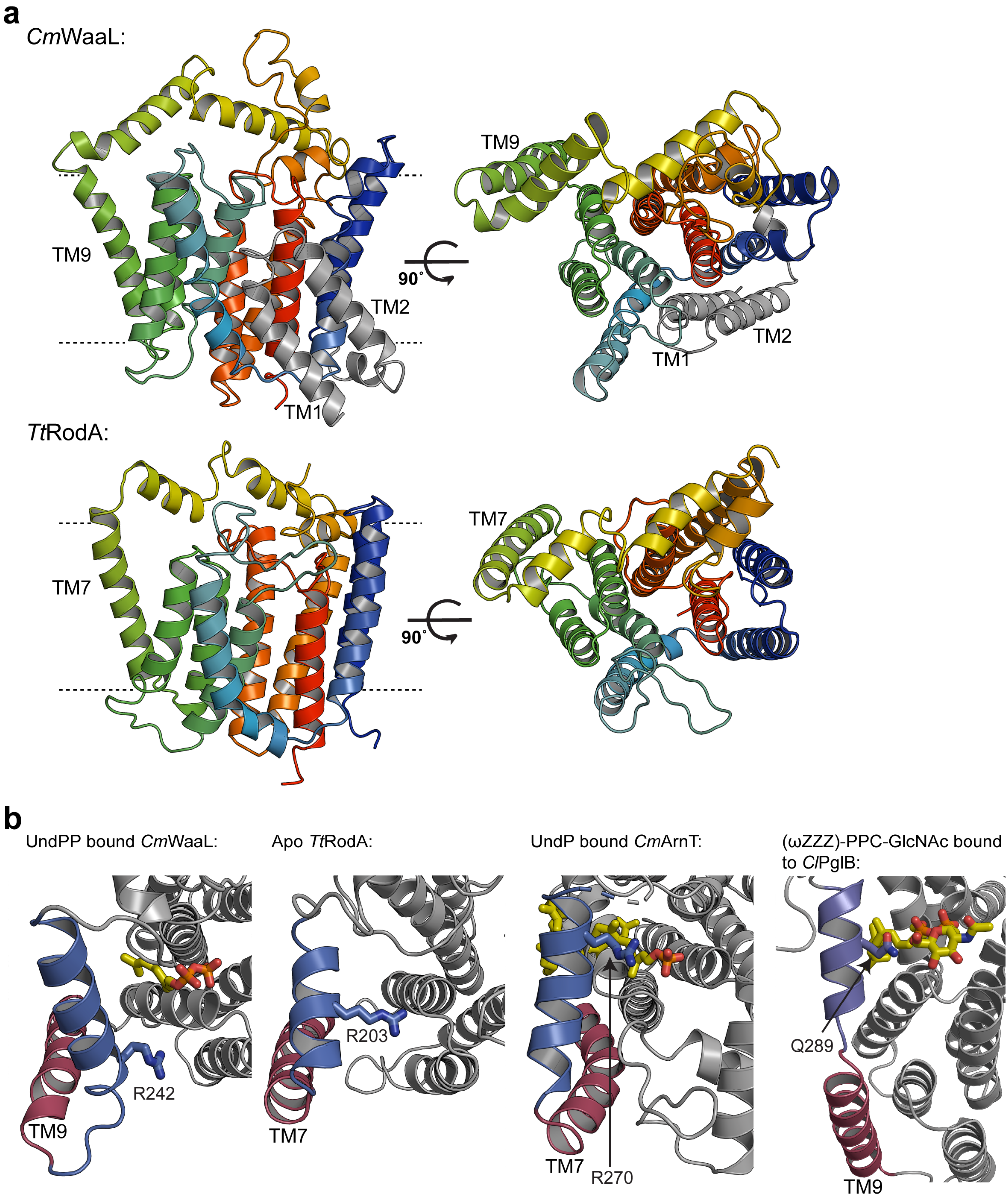


*Cm*WaaL	*Ec*WaaL
R28	K45
R92	K111
R139	R161
Q145	T168
I147	T170
R191	R215
R242	R265
R265	R288
H311	H336
H313	H338
E318	E343
Q378	D389
H383	A394


a
Und-PP bound:
Und-PP
Apo:
b
Und-PP bound:
TM9
Apo:
TM9
c
Und-PP bound:
TM9
Apo:
TM9
Overlay:
TM9
d
Und-PP bound:
H313
H311
R265
Und-PP
H383
R92
R242
R191
R139
R28
Apo:
H313
H311
R265
H383
R92
R242
R191
R139
R28
Overlay:
R265
H313
H311
Und-PP
H383
R92
R242
R191
R139
R28


a
Side view:
Top view:
b
*Cm*WaaL structure:
*Ec*WaaL homology model:
c
10 20 30 40 50 60
C. metallidurans NKLLSRMATVLVFAFPVLILCVPRGAGVFLAGVGVLALLGWRGMGRAWREYSKVMTPLAI
E. coli SMILEIITYILCFFSMIIAFVDNTFSIKIYNITAIVCLLS...LILRGRQENYNIKNLIL
TM1 TM2 CH TM3
70 80 90 100 110
C. metallidurans AVLAFMLVYYVGSKLYFH.........TPWNVIDNPSRTLLAILTCWVIVRAAPNPA...W
E. coli PLSIFLIGLLDLIWYSAFKVDNSPFRATYHSYLNTAKIFIFGSFIVFLTLTSQLKSKKES
TM3 TM4
120 130 140 150 160 170
C. metallidurans LWRGITVGLFLALLIVGYQKFALNIDRPSAWIQAIAFANMIAALALVGFARPGD..SRGT
E. coli VLYTLYSLSFLIAGYAMYINSIHENDRISFGVGTATGAAYSTMLIGIVSGVAILYTKKNH
TM5 TM6 TM7
180 190 200 210 220 230
C. metallidurans HMEAWVNLLLGTMILMLNGTRGAVVAMLVTSVPMLMIRYRRFSVRMLIVAVCAVATLAIG
E. coli PFLFLLNSCAVLYVLALTQTRATLLLFPIICVAALIAYYNK.SPKKFTSSIVLLIAILAS
TM7 TM8 TM9
240 250 260 270 280 290
C. metallidurans AYMVPDSPVSKRVDDAVSEIQMYRQGNIETSVGVRLKIWHIGLQYFSEHPWTGVGVGQFA
E. coli IVIIFNKPIQNRYNEALNDLNSYTNANSVTSLGARLAMYEIGLNIFIKSPFSFRSAESRA
TM9 PH1 PH2
300 310 320 330 340
C. metallidurans RILHAS..EFCHETKSLACVLEHAHNDIVEAASTTGIPGLMVMLGLFLVPAVLFARALRA
E. coli ESMNLLVAEHNRLRGALEFSNVHLHNEIIEAGSLKGLMGIFSTLFLYFSLFYIAYKKR..
PH3 PH4 TM10
350 360 370 380 390 400
C. metallidurans ARSLGNPQGVSLGGAGLGVVMASLISGLTQVTMAHQANVVFYAGLIGLLLGMAGREAHS
E. coli ............ALGLLILTLGIVGIGLSDVII.WARSIPIIIISAIVLLLVINNRNNT
TM11 TM12


a
UndPP, FAB
Apo
UndPP
UndPP, KDO-LipidA
RMSF (Å)
0
2
4
6
1
2
3
Color Reference
Apo
UndPP, FAB
UndPP
UndPP, KDO-LipidA
RMSF (Å)
0
1
2
3
4
5
b
Closed
Open
Overlay
TM 9
APO
UndPP bound
UndPP, KDO-LipidA
Frequency
0.00
0.05
0.10
15
20
25
30
Distance (Å)


a
*Cm*WaaL:
TM9
TM2
TM1
90˚
TM9
TM2
TM1
*Tt*RodA:
TM7
90˚
TM7
b
UndPP bound *Cm*WaaL:
R242
TM9
Apo *Tt*RodA:
R203
TM7
UndP bound *Cm*ArnT:
TM7
R270
(ωZZZ)-PPC-GlcNAc bound to *Cl*PglB:
Q289
TM9

SI Guide

Structural basis of Lipopolysaccharide Maturation by O-Antigen Ligase WaaL

Khuram U. Ashraf[1], Rie Nygaard[1], Owen N. Vickery[2,3,#], Satchal K. Erramilli[4,#], Carmen M. Herrera[5,#], Thomas H. McConville[6,#], Vasileios I. Petrou[7,8], Sabrina I. Giacometti[1], Meagan Belcher Dufrisne[1], Kamil Nosol[4], Allen P. Zinkle[1], Chris L.B. Graham[2], Michael Loukeris[9], Brian Kloss[9], Karolina Skorupinska-Tudek[10], Ewa Swiezewska[10], David I. Roper[2,1], Oliver B. Clarke[1,11], Anne-Catrin Uhlemann[6], Anthony A. Kossiakoff[4], M. Stephen Trent[5*], Phillip J. Stansfeld[2,3*] and Filippo Mancia[1*]

[1]Department of Physiology and Cellular Biophysics, Columbia University Irving Medical Center, New York, NY 10032, USA.

[2]School of Life Sciences, University of Warwick, Coventry, CV4 7AL, UK.

[3]Department of Chemistry, University of Warwick, Coventry, CV4 7AL, UK.

[4]Department of Biochemistry and Molecular Biology, University of Chicago, Chicago, IL, 60637, USA.

[5]Department of Infectious Diseases, College of Veterinary Medicine, University of Georgia, Athens, GA, 30602, USA.

[6]Department of Medicine, Division of Infectious Diseases, Columbia University Medical Center, New York, New York, 10032, USA.

[7]Department of Microbiology, Biochemistry, and Molecular Genetics, New Jersey Medical School, Rutgers Biomedical Health Sciences, Newark, NJ, 07103, USA.

[8]Center for Immunity and Inflammation, New Jersey Medical School, Rutgers Biomedical Health Sciences, Newark, NJ, 07103, USA.

[9]New York Consortium on Membrane Protein Structure, New York Structural Biology Center, 89 Convent Avenue, New York, NY, 10027, USA.

[10]Institute of Biochemistry and Biophysics, Polish Academy of Sciences, 02-106 Warsaw, Poland.

[11]Department of Anesthesiology, Columbia University Irving Medical Center, New York, NY, 10032, USA.

[#]These authors contributed equally.

*Correspondence to be addressed to: strent@uga.edu (M.S.T.), Phillip.Stansfeld@warwick.ac.uk (P.J.S.), fm123@cumc.columbia.edu (F.M.)

Table of Content:

***Red, dashed, boxes indicate how gels where cropped.**

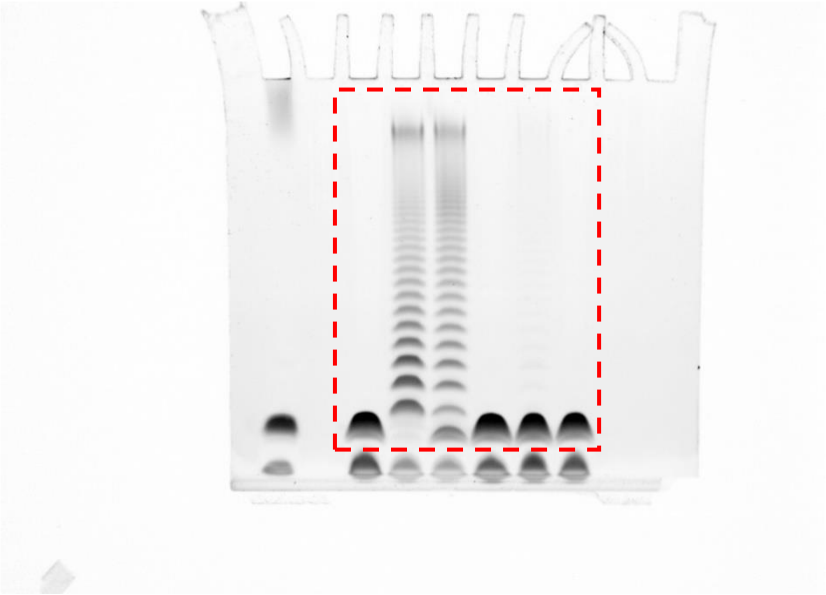
SI Figure 1 | Uncropped LPS assay gels from Main Figures. a, LPS assay gel from Fig. 2c showing analysis of *Cm*WaaL ligase activity in whole cells. LPS gel showing O-antigen extension in *C. metallidurans* expressing *Cm*WaaL variants. With and empty plasmid as a control. **b,** Functional analysis gel from Fig. 4c of *Ec*WaaL ligase variants in whole cells by LPS gel analysis. WaaL proteins were expressed from plasmid pWSK29 in the W3110 ΔwaaL strain. With and empty plasmid as a control.

SI Figure 2 | Uncropped LPS assay gels from Extended Data Figures. a, LPS assay gel from Extended Data Fig. 1c showing functional analysis of CmWaaL ligase activity in whole cells, showing O-antigen ligase activity is abolished when the gene encoding WaaL is deleted and activity is restored by plasmid complementation. **b,** Functional analysis of *Ec*WaaL ligase activity from Extended Data Fig. 7a carried out in whole cells by LPS gel analysis. W3110 ΔwaaL containing either empty vector pWSK29 (Ø), pWSK29::*Ec*WaaL (WT) as controls and pWSK19::*Ec*WaaL- mutant variants to test for O-antigen ligase activity. **c,** Functional analysis gel from Extended Data Fig. 7e showing *Cm*WaaL ligase activity in whole cells. *C. metallidurans ΔwaaL* containing either empty vector pBBR1(Ø), pBBR1:*Cm*WaaL (WT) as controls for evaluating *Cm*WaaL-variants for O-antigen extension.

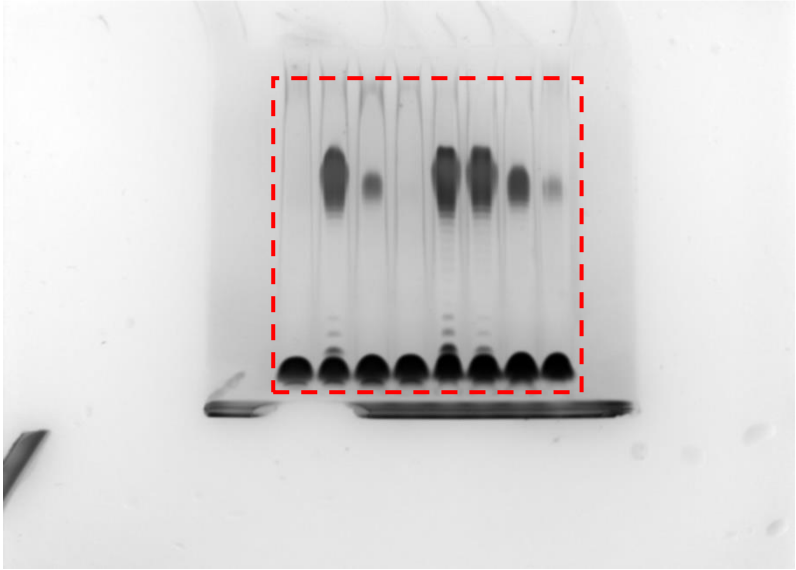
SI Figure 3 | Uncropped SDS-page gels from Extended Data Figures. a, SDS-PAGE gel from Extended Data Fig. 1e depicting *Cm*WaaL purified in detergent, *Cm*WaaL reconstituted into nanodiscs (MSP1E3D1 and POPG), and CmWaaL reconstituted into nanodiscs (MSP1E3D1 and POPG) with Fab (WaB10) bound. Molecular weight markers can also be observed and are labelled. **b,** SDS-PAGE gels from Extended Data Fig. 7c of all *Ec*WaaL mutants used for LPS ligase analysis, that were purified to verify expression.

SI Figure 4 | Uncropped TLC and Western Blot analysis from Extended Data Figures. a, TLC analysis from Extended Data Fig. 5d of detergent purified CmWaaL along with suitable controls: POPG, Und-P, Und-PP and an unrelated control protein expressed in *E. coli* were run in separate lanes. **b,** Western blot analysis, from Extended Data Fig. 7d, using a mouse monoclonal anti-Flag antibody, of Flag purified WT *Cm*WaaL and mutants, grown in *C. metallidurans,* for testing in the O-antigen ligase assay. Red markers indicate molecular weight markers.

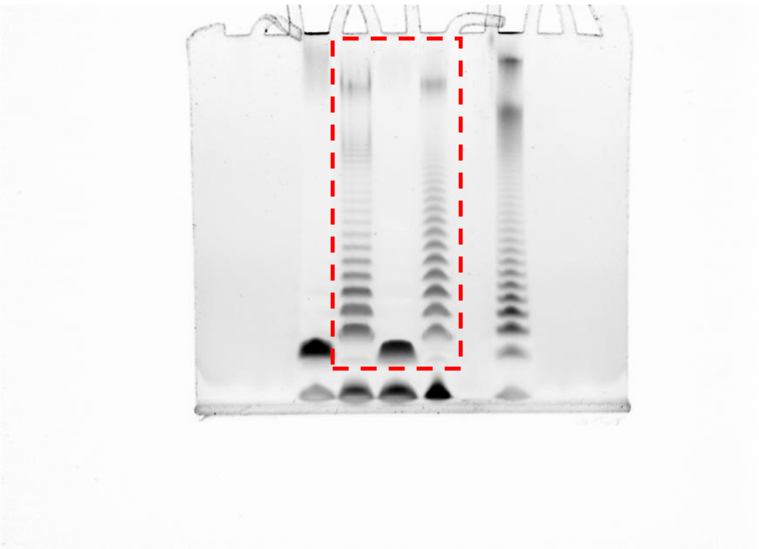
Supplementary Table 1 | Raw Single-point ELISA and Multi-point sAB ELISA data



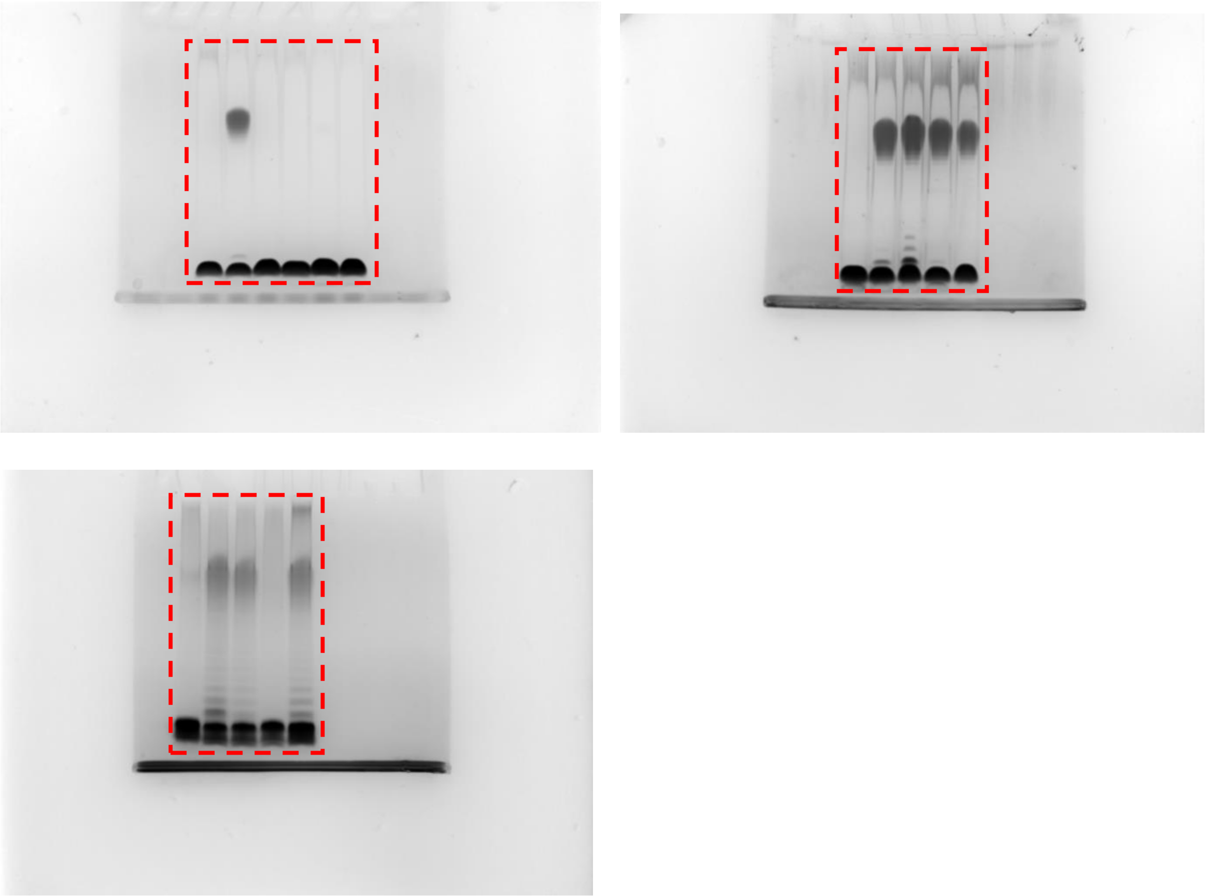
Supplementary Figure 1a | Uncropped gel from Figure 2c.



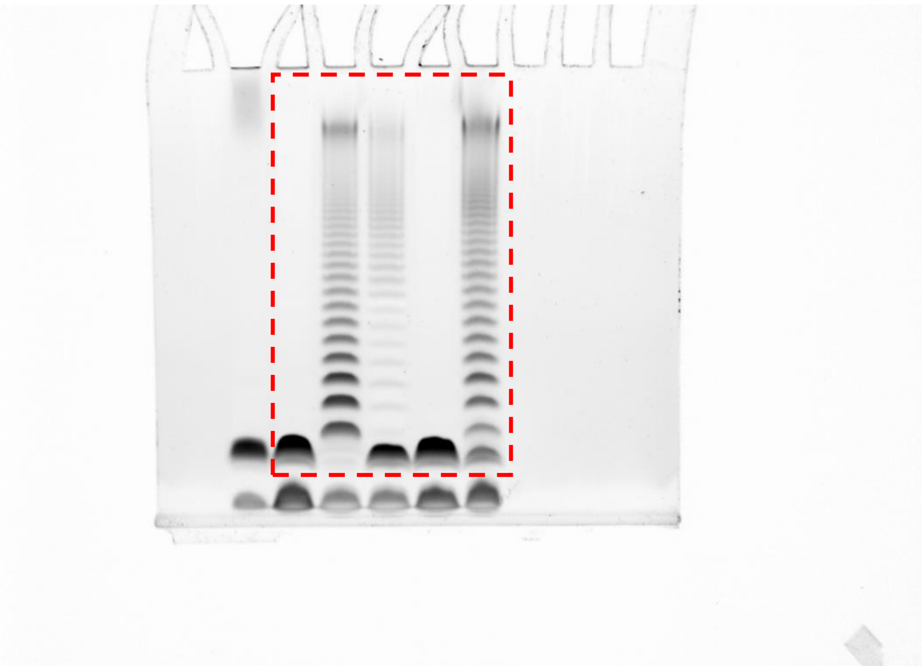
Supplementary Figure 1b | Uncropped gel from Figure 4c.



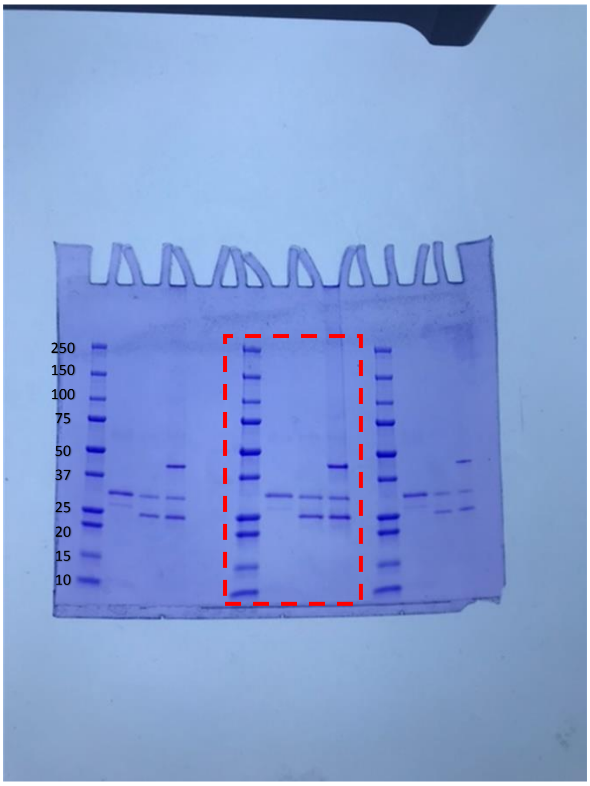
Supplementary Figure 2a | Uncropped gel from Extended Data Fig. 1c



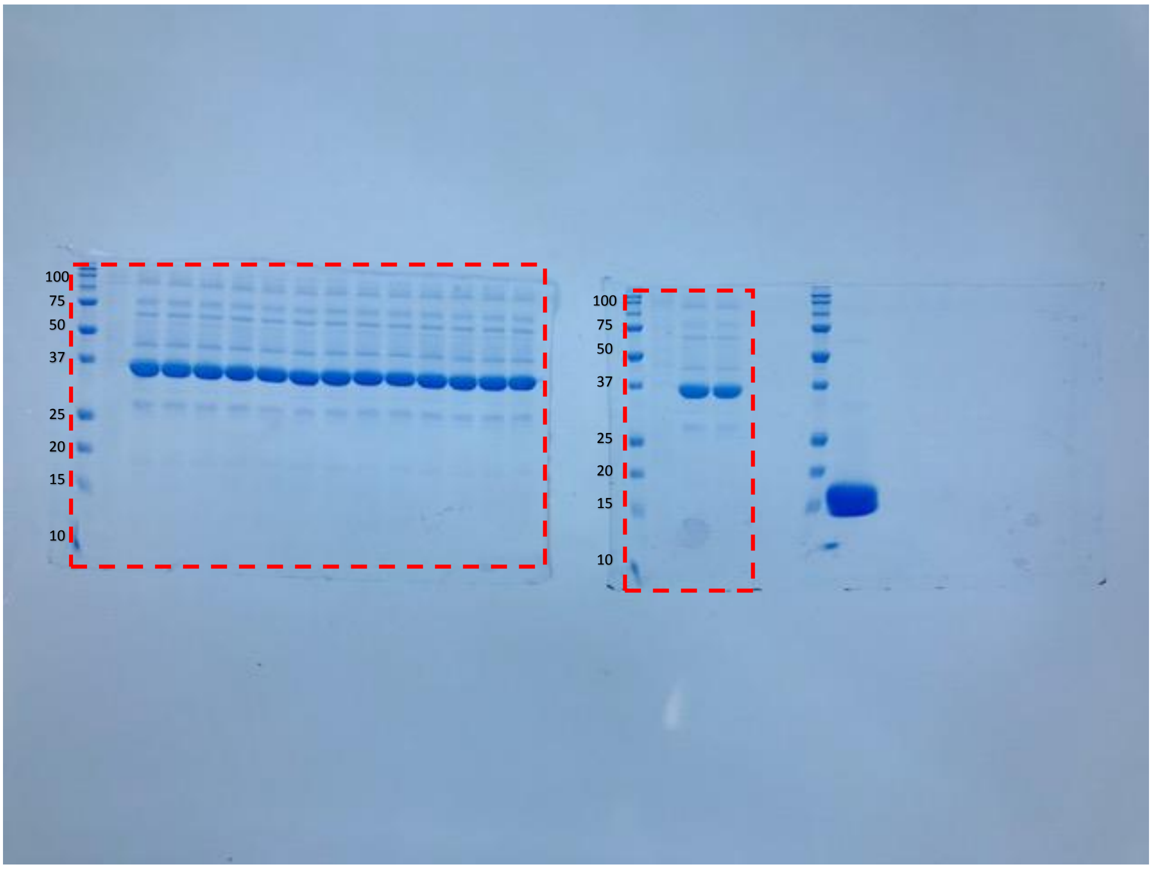
Supplementary Figure 2b | Uncropped gels from Extended Data Fig. 7a

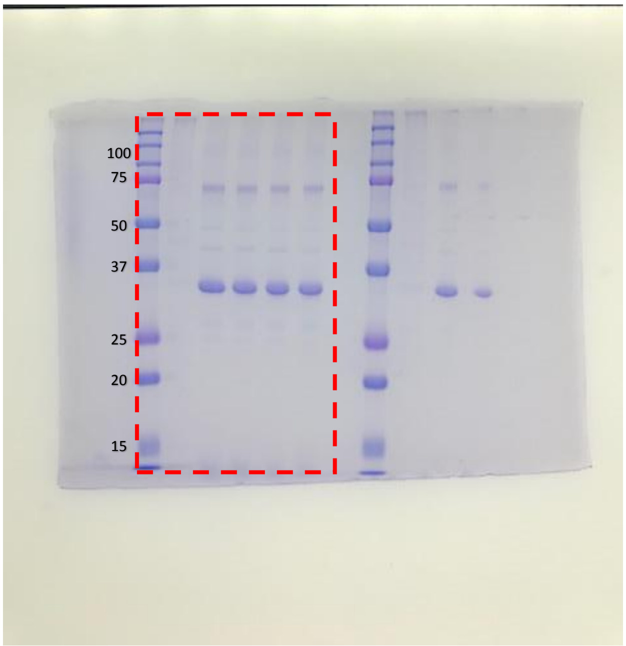


Supplementary Figure 2c | Uncropped gels from Extended Data Fig. 7e

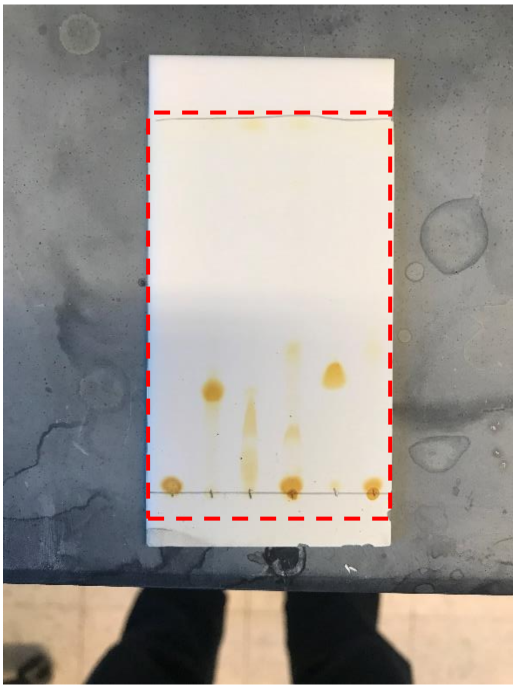


Supplementary Figure 3a | Uncropped gel from Extended Data Fig. 1e

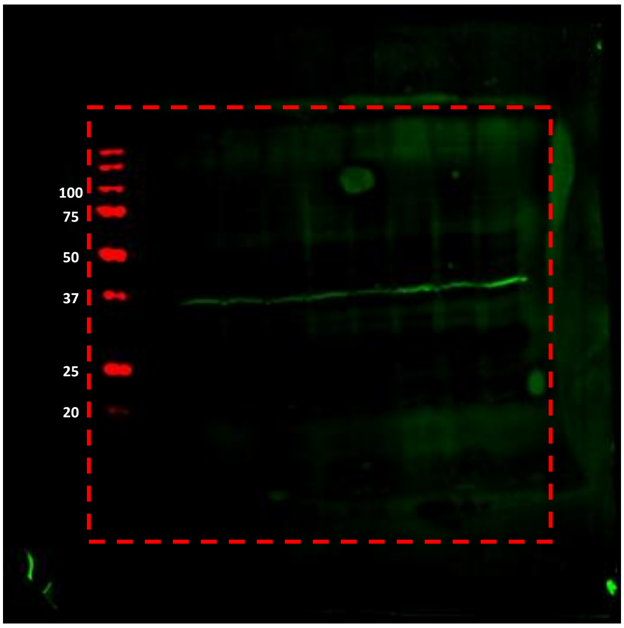
101 

102 

103 **Supplementary Figure 3b | Uncropped gel from Extended Data Fig. 7c**



Supplementary Figure 4a | Uncropped gel from Extended Data Fig. 5d



Supplementary Figure 4b | Uncropped gel from Extended Data Fig. 7d