Insight into the molecular basis of substrate recognition by the wall teichoic acid glycosyltransferase TagA

Received for publication, August 11, 2021, and in revised form, November 18, 2021 Published, Papers in Press, December 2, 2021,
https://doi.org/10.1016/j.jbc.2021.101464

Orlando E. Martinez[1,2], Brendan J. Mahoney[1,2], Andrew K. Goring[1,2], Sung-Wook Yi[1], Denise P. Tran[1], Duilio Cascio[2], Martin L. Phillips[1], Musleh M. Muthana[3], Xi Chen[3], Michael E. Jung[1,4], Joseph A. Loo[1,2,4], and Robert T. Clubb[1,2,4,*]

From the [1]Department of Chemistry and Biochemistry, and [2]UCLA-DOE Institute of Genomics and Proteomics, University of California, Los Angeles, Los Angeles, California, USA; [3]Department of Chemistry, University of California, Davis, California, USA; [4]Molecular Biology Institute, University of California, Los Angeles, Los Angeles, California, USA

Edited by Wolfgang Peti

Wall teichoic acid (WTA) polymers are covalently affixed to the Gram-positive bacterial cell wall and have important functions in cell elongation, cell morphology, biofilm formation, and β-lactam antibiotic resistance. The first committed step in WTA biosynthesis is catalyzed by the TagA glycosyltransferase (also called TarA), a peripheral membrane protein that produces the conserved linkage unit, which joins WTA to the cell wall peptidoglycan. TagA contains a conserved GT26 core domain followed by a C-terminal polypeptide tail that is important for catalysis and membrane binding. Here, we report the crystal structure of the *Thermoanaerobacter italicus* TagA enzyme bound to UDP-*N*-acetyl-D-mannosamine, revealing the molecular basis of substrate binding. Native MS experiments support the model that only monomeric TagA is enzymatically active and that it is stabilized by membrane binding. Molecular dynamics simulations and enzyme activity measurements indicate that the C-terminal polypeptide tail facilitates catalysis by encapsulating the UDP-*N*-acetyl-D-mannosamine substrate, presenting three highly conserved arginine residues to the active site that are important for catalysis (R214, R221, and R224). From these data, we present a mechanistic model of catalysis that ascribes functions for these residues. This work could facilitate the development of new antimicrobial compounds that disrupt WTA biosynthesis in pathogenic bacteria.

Staphylococcus aureus and other Gram-positive bacteria are surrounded by a thick murein sacculus that is densely functionalized with wall teichoic acid (WTA) polymers (1–5). These anionic glycopolymers are essential components of the bacterial cell and involved in cell elongation, morphogenesis, cation homeostasis, pathogenesis, and autolysin localization (6–13). They are polymers of alditol phosphate repeating units that are covalently joined to the surface peptidoglycan *via* a conserved linkage unit that is composed of 1 to 3 glycerol-3-phosphate (GroP) groups appended to an *N*-acetyl-D-mannosamine (ManNAc) (β1 → 4) *N*-acetyl-D-glucosamine (GlcNAc) disaccharide monophosphate (2, 14–16). The WTA biosynthetic pathway has drawn considerable interest as a drug target, as genetically eliminating WTA production in clinically important methicillin-resistant *S. aureus* resensitizes it to β-lactam antibiotics and attenuates its virulence (6, 7).

WTA polymers are synthesized on the cytoplasmic face of the cell membrane by enzymes that sequentially elaborate a membrane-embedded undecaprenyl phosphate carrier molecule (2, 3, 5). In *Bacillus subtilis* and *S. aureus*, the linkage unit is synthesized by the sequential action of the TagO, TagA, and TagB enzymes (originally designated as Tar enzymes in *S. aureus*). WTA synthesis is initiated by TagO, which catalyzes the reversible transfer of GlcNAc-1-P from UDP–GlcNAc to the undecaprenyl (Und) phosphate scaffold to produce lipid-α (GlcNAcα–PP-Und) (17). The TagA ManNAc transferase then appends ManNAc from a sugar nucleotide donor, UDP–ManNAc, producing a ManNAc (β1 → 4) GlcNAcα–PP-Und product, called lipid-β (18–20). The linkage unit synthesis is completed by TagB, which adds a GroP to lipid-β using a CDP–glycerol donor substrate (18, 21). Polymerase enzymes then extend the (GroP)–ManNAc (β1 → 4) GlcNAcα-1-P linkage unit from the terminal GroP at the nonreducing end to construct the body of the polymer with either ribitol- or glycerol-phosphate repeating units, which can vary substantially among species (22, 23). The mature WTA is modified with monosaccharides at free hydroxyl groups on the main chain polymer (24). WTAs are then transported across the membrane to the cell surface by the TagGH transporter (2, 25). Finally, in the cell wall, the GlcNAcα-1-P at the reducing end of the linkage unit is covalently attached to the 6′ hydroxyl of the *N*-acetylmuramic acid component in the peptidoglycan *via* a phosphodiester bond, and in some instances, further tailored with D-alanine modifications (2, 15, 26, 27).

The TagA enzyme is a promising drug target because it catalyzes the first committed step in WTA biosynthesis and methicillin-resistant *S. aureus* strains in which it is genetically deleted are attenuated in virulence and resensitized to β-lactam antibiotics (7, 28). Biochemical studies have shown that *B. subtilis* TagA is a metal-independent inverting

* For correspondence: Robert T. Clubb, rclubb@mbi.ucla.edu.
Present address for Musleh M. Muthana: Division of Immunotherapy, Institute of Human Virology and Department of Pharmacology, University of Maryland School of Medicine, Baltimore, Maryland 21201, USA.

glycosyltransferase (GT) that catalyzes the transfer of ManNAc from UDP–ManNAc to the 4′ hydroxyl of GlcNAc in lipid-α (19). Catalysis occurs *via* a Bi–Bi mechanism in which the UDP–ManNAc donor first binds to the enzyme, and the UDP product is released last (19). *S. aureus* TagA is part of the WecB/TagA/CpsF GT family (PFAM03808), whose >6000 members catalyze the synthesis of WTAs and other important surface glycopolymers such as capsular polysaccharides in group B *Streptococcus* and the enterobacterial common antigen in *Escherichia coli* (29–31). Members of this large family are classified as GT26 enzymes in the Carbohydrate-Active enZYmes database (www.cazy.org), but little is known about the molecular basis of their enzymatic activity (32–35). Recently, we reported the structure of the TagA enzyme from *Thermoanaerobacter italicus*, which represents the only known structure of a GT26 enzyme (36, 37). The GT26 domain adopts a unique fold that is distinct from other GTs (*e.g.*, GT-A, GT-B, GT-C, and GT-D folds) and is followed by a 49-amino acid C-terminal tail (CTT) that targets the enzyme to the cell membrane (36). Functional studies showed that the CTT is essential for activity and identified D65 and R221 as important determinants for catalysis as their alteration causes more than a 20-fold reduction in activity as compared with the wildtype enzyme (36). However, the molecular basis of substrate recognition and catalysis remains incompletely understood.

In this study, we used a combination of biochemical, structural, and computational approaches to determine how TagA recognizes its substrates and how flexibility in the CTT mediates substrate binding and catalysis. Using a solubility-enhanced enzyme variant, we determined the structure of TagA bound to its natural substrate, UDP–ManNAc. Computational modeling, molecular dynamics (MD) simulations, and biochemical experiments of the full-length enzyme provide insight into the role of the CTT in catalysis, revealing that it presents conserved amino-acid side chains to contact the UDP–ManNAc substrate. Paired with native MS of TagA in the presence of micelles, we provide additional support for a membrane-induced activation mechanism of the enzyme. Bilayer association *via* the CTT appears to stabilize the monomeric form of the enzyme so that it can glycosylate its membrane-embedded glycophospholipid substrate. These results could provide a foundation for the discovery of new antibiotics that work by inhibiting WTA biosynthesis.

Results

Solution-state studies of TagA

In previously reported crystal structures, *T. italicus* TagA$^{\Delta C}$ (TagA, residues M1-G195) adopted both dimeric and trimeric oligomeric states (36). To probe its oligomerization state in solution, we used NMR. TagA$^{\Delta C}$ contains the core domain but is missing 49 residues at its C terminus that have been shown to promote membrane interactions and facilitate catalysis (Fig. 1*A*) (36). A 0.6 mM sample of [U-^{2}H (70%), ^{13}C, ^{15}N] TagA$^{\Delta C}$ was overexpressed and purified, and the chemical shifts of its backbone atoms were assigned using triple resonance methods. The ^{1}H–^{15}N transverse relaxation optimized spectroscopy (TROSY)–heteronuclear single quantum coherence (HSQC) spectrum of TagA$^{\Delta C}$ is reasonably well resolved, enabling ∼61% of the amino acids in its primary sequence to be specifically assigned (Fig. 2*A*). An analysis of the secondary chemical shifts reveals that the positioning of the secondary structural elements in the protein match those visualized in the crystal structure. Based on ^{15}N NMR relaxation measurements, TagA$^{\Delta C}$ has a molecular correlation time (τ_c) of 27 ns, corresponding to a dimer. This is consistent with previously reported sedimentation equilibrium–analytical ultracentrifugation (SE–AUC) experiments, which showed that apo-TagA$^{\Delta C}$ exists in a monomer–dimer equilibrium in

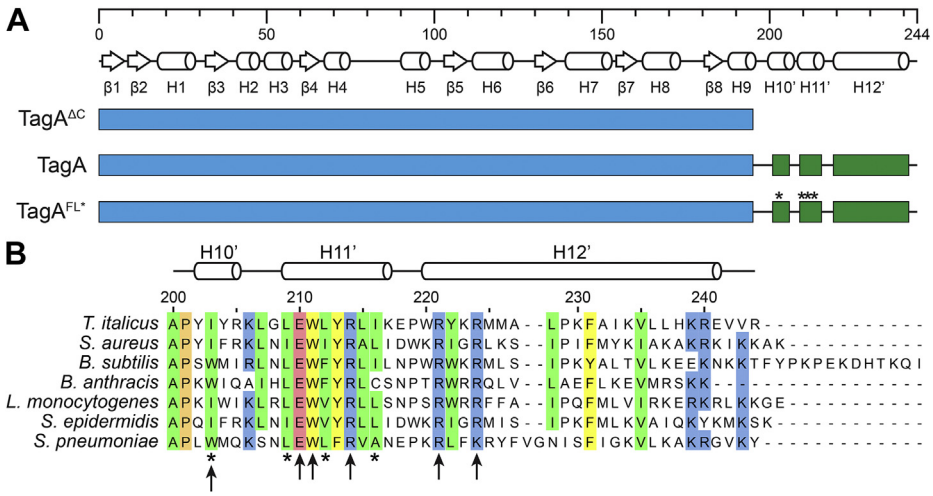

Figure 1. TagA protein constructs and sequence alignment. *A*, three *Thermoanaerobacter italicus* TagA proteins were used in this study: TagA$^{\Delta C}$, residues M1–G195; TagA, residues M1–R244; and TagA^{FL}*, residues M1–R244 with four amino-acid substitutions (I203E/L209Q/L212K/I216E). The folded core domain of TagA (*blue*) and computationally predicted helices (*green*) in the CTT are shown as *horizontal bars*. The secondary structure elements are shown with the predicted helices in the tail labeled H10′, H11′, and H12′. *Asterisks* indicate the location of the amino-acid substitutions in TagA^{FL}*. *B*, primary sequence alignment of select TagA proteins showing the CTT. Residues with high similarity across species are colored as follows: nonpolar (*green*), basic (*blue*), acidic (*red*), aromatic (*yellow*), and proline (*orange*). *Asterisks* indicate the location of the amino-acid substitutions in TagA^{FL}*, and *arrows* indicate conserved residues that were altered in this study. CTT, C-terminal tail.

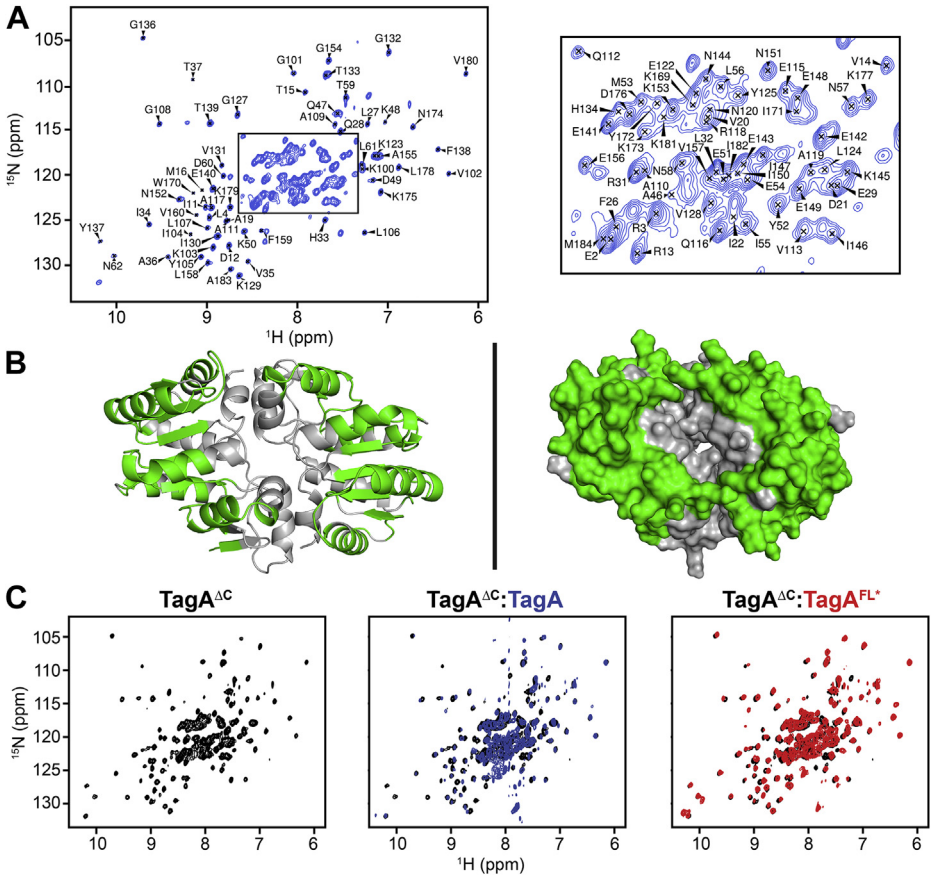

Figure 2. NMR studies of TagA. *A*, assigned ^{1}H–^{15}N TROSY–HSQC spectrum of TagA$^{\Delta C}$, with the center of the spectrum expanded and shown on the *right*. About 61% of amide residues were assigned for TagA$^{\Delta C}$. *B*, crystal structure of the TagA$^{\Delta C}$ dimer colored to show amino acids whose chemical shifts were assigned (*shaded green*). The TagA$^{\Delta C}$ dimer is shown in *cartoon* (*left*) and *surface* (*right*) representations (Protein Data Bank code: 5WB4). *C*, overlays of ^{1}H–^{15}N TROSY–HSQC spectra: TagA$^{\Delta C}$ alone (*left*), overlaid with the spectrum of TagA (*center*, *blue*) and with spectrum of TagA^{FL}* (*right*, *red*). HSQC, heteronuclear single quantum coherence; TROSY, transverse relaxation optimized spectroscopy.

solution (dissociation constant [K_D] = 7.4 ± 0.7 μM) (36). Interestingly, backbone amide signals for 81 of 195 (74 of 188 non-proline) residues in the primary sequence of TagA$^{\Delta C}$ are absent in the NMR data. Two structures of TagA$^{\Delta C}$ have been reported, a dimeric structure of the protein in its apo-state (Protein Data Bank [PDB]: 5WB4) and a trimeric structure in which it is bound to its UDP product (PDB: 5WFG) (37). Mapping the location of the missing amide signals onto the crystal structure of dimeric TagA$^{\Delta C}$ reveals that they are localized at the intersubunit interface, suggesting that signals for these residues are broadened beyond detection because of conformational exchange caused by dimer dissociation (Fig. 2*B*). In contrast, when the missing residues are mapped onto the trimeric structure of the TagA$^{\Delta C}$:UDP complex, many of the missing residues are not located at the intersubunit interface, suggesting that this form of the protein is not as prevalent. Thus, we conclude from the NMR data that in solution the dimeric form of TagA$^{\Delta C}$ predominates and that it resembles the previously reported crystal structure of apo-TagA$^{\Delta C}$.

We next used NMR and SE–AUC experiments to investigate full-length TagA (M1–R244), the functional form of the enzyme that associates with the membrane *via* residues within its CTT (Fig. 2*C*) (36). SE–AUC experiments indicate that TagA exists as a mixture of aggregated heterogeneous species, as its weight-averaged molecular weight decreases as the centrifugation speed increases (Fig. S1) (38). NMR studies of a 0.3 mM sample of [U-^{15}N] TagA are consistent with the SE–AUC data, as its ^{1}H–^{15}N TROSY–HSQC spectrum exhibits severe line broadening (Fig. 2*C*, *middle*). The positions of the cross peaks in the spectra of TagA$^{\Delta C}$ and TagA are similar, indicating that the structure of the core domain is retained. Notably, signals for the 49 residues in the CTT are absent in the spectra of TagA, and new signals in the indole region of the spectrum that presumably originate from the side chains of W211 and W220 within the CTT are significantly broadened. From these data, we conclude that the full-length protein adopts a conformation that is similar to that observed in the crystal structure of dimeric TagA$^{\Delta C}$, but residues in the CTT are likely disordered and mediate nonspecific interactions that promote aggregation in solution.

Crystal structure of solubility-enhanced TagA (TagA^{FL}) bound to its sugar–nucleotide substrate*

Only the structure of TagA$^{\Delta C}$ lacking the CTT has been determined because the full-length protein exhibits low solubility. To overcome this problem, we modeled the structure of

full-length TagA to identify potential locations on the protein's surface that might promote aggregation. We first generated a model of full-length *B. subtilis* TagA using the Generative Regularized Models of Proteins (GREMLIN) method, which predicts protein structures by integrating sequence conservation and coevolutionary patterns in a multiple sequence alignment (36, 39). The full-length *T. italicus* TagA computational model ($TagA^{CM}$) used in this study was then generated by homology modeling using the GREMLIN-derived structure as a template for input into the program I-TASSER (Iterative Threading Assembly Refinement) (40). Coevolution analyses are a robust approach to model the structure of full-length *T. italicus* TagA, as structures similar to $TagA^{CM}$ are generated using coevolution-derived residue–residue distance probabilities implemented in the RaptorX program (Fig. S2) (41, 42). $TagA^{CM}$ is monomeric with the CTT forming three helices (H10', H11', and H12') that pack against the oligomerization interface observed in the crystal structure and is positioned against the body of the protein by several pairs of coevolving residues (Fig. 3, *A* and *B* and Table S1). Inspection of $TagA^{CM}$ reveals that the CTT contains many nonpolar amino acid side chains that colocalize to form a hydrophobic patch that might promote protein aggregation in solution. Prior cell fractionation studies of *B. subtilis* revealed that several residues within the patch are essential for membrane association (36). To improve protein solubility in solution, we constructed $TagA^{FL*}$, a variant of TagA in which four nonpolar residues in the patch are exchanged with hydrophilic amino acids (I203E/L209Q/L212K/I216E) (Figs. 1*B* and 3*C*). Indeed, $TagA^{FL*}$ exhibits markedly improved solubility and no longer requires detergents to remain soluble. Moreover, unlike native full-length TagA, which nonspecifically aggregates (Fig. S1), $TagA^{FL*}$ is in equilibrium between monomeric and dimeric states; SE–AUC experiments indicate that the $TagA^{FL*}$ dimer has a K_D of 210 ± 50 μM (Fig. 3*D*). This supports previous size-exclusion chromatography experiments

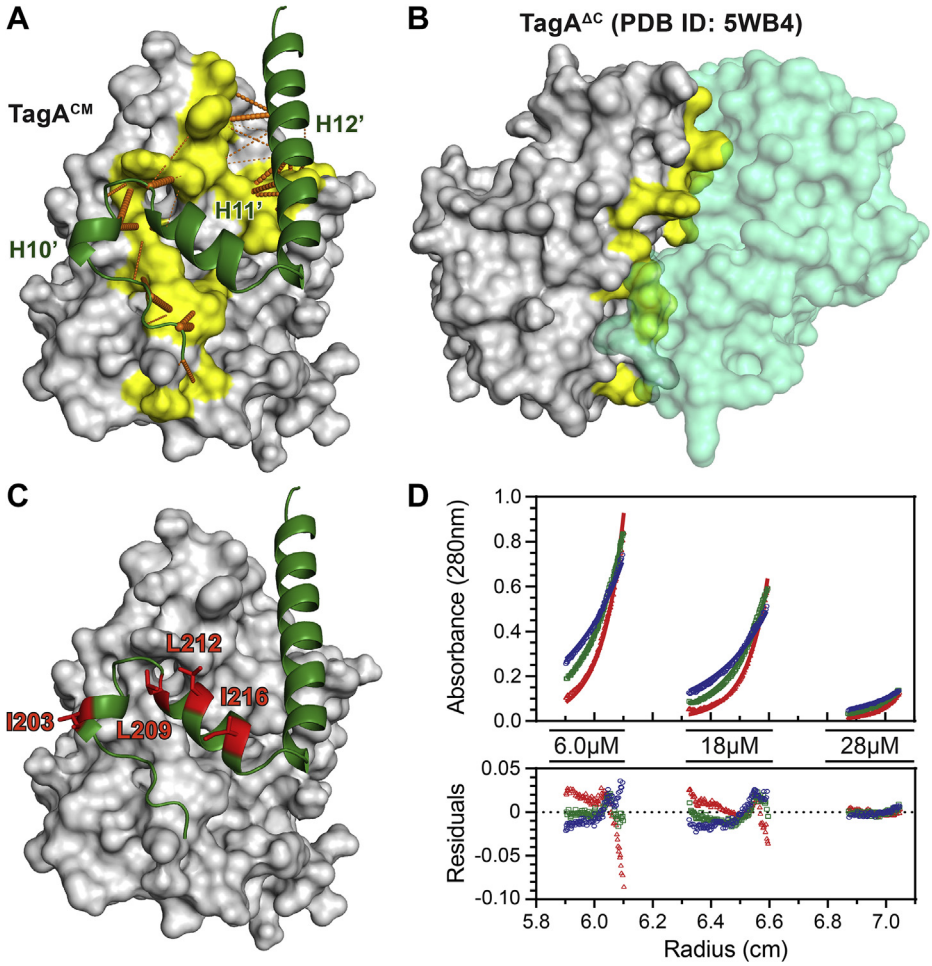

Figure 3. A model of the monomeric form of full-length TagA and its use in creating a solubility-enhanced TagA protein. *A*, $TagA^{CM}$, a computationally derived model of the TagA monomer. *Orange bars* mapped onto the structure connect coevolving amino acids between the core domain (*yellow surface*) and the CTT (*green cartoon*) that were identified using the program RaptorX (41). The thickness of the bars indicates the probability of residues being within 8 Å (50%—thin, 100%—thick). *B*, *surface* representation of the crystal structure of the $TagA^{\Delta C}$ dimer showing coevolving core domain residues in *yellow*. The predicted CTT binding surface on the monomeric form of the enzyme conflicts with the dimer interface. *C*, $TagA^{CM}$ model showing the location of the nonpolar amino acids (*red*) that were altered to create the solubility-enhanced protein ($TagA^{FL*}$, I203E/L209Q/L212K/I216E). *D*, SE–AUC experiments of $TagA^{FL*}$. Data were collected using three sample concentrations at three rotor speeds: 15,000 (*blue*), 19,000 (*green*), and 24,000 (*red*) rpm. Residuals after fitting the data to a monomer–dimer equilibrium are shown. CTT, C-terminal tail; SE-AUC, sedimentation equilibrium–analytical ultracentrifugation.

that asserted that TagA^{FL*} contains both monomeric and dimeric species in aqueous solution, and the monomeric species is more abundant compared with native TagA (36). TagA^{FL*} also has *in vitro* enzymatic activity that is superior to TagA$^{\Delta C}$, albeit lower than that of the native enzyme (discussed below).

To gain insight into the molecular basis of substrate binding, we determined the crystal structure of the solubility-enhanced TagA^{FL*} enzyme bound to its UDP–ManNAc substrate. The crystal structure of the TagA^{FL*}:UDP–ManNAc complex was resolved to 3.3 Å and solved using molecular replacement with the coordinates of TagA$^{\Delta C}$ (PDB: 5WB4) as a search model (data collection and refinement statistics in Table 1). The diffraction dataset contained moderate anisotropy, so we applied anisotropic scaling and isotropic *B*-factor sharpening prior to molecular replacement (43). The complex crystallized with three molecules in its asymmetric unit, with each intersubunit interface burying only ∼430 Å^2 of surface area (Fig. 4*A*). These small protein–protein interfaces are not deemed to be biologically relevant based on an analysis using the Evolutionary Protein–Protein Interface Classifier server (44). Only coordinates for the core domain could be modeled (residues M1–R199), as no interpretable electron density for the CTT was observable. The reason for its absence is unclear, but a large volume of nonuniform density is located adjacent to the C terminus of each protein molecule, suggesting that the CTT may be disordered in the crystal. It is also conceivable that during the crystallization process, which took approximately 1 month, the CTT was degraded by proteolysis. The positioning of the UDP–ManNAc substrate is best defined by the electron density data in chain A and is discussed further (Figs. 4*B* and S3*A*) (45). The uracil base of UDP–ManNAc is stabilized by hydrogen bonding with the terminal amino group of K166, the backbone carbonyl of G136, and the backbone amide of A109, along with π-stacking interactions with the Y137 tyrosyl ring (Fig. 4*C*). The ribose in the substrate is stabilized by two strong hydrogen bonds, one between the 3′ hydroxyl hydrogen donor and the terminal carboxylate acceptor of D191 and the other between the 2′ hydroxyl and the terminal hydroxyl of Y137 (Fig. 4*C*). Finally, the ManNAc sugar accepts two hydrogen bonds donated from the side-chain amino groups of N39 and Q167 to the substrate 3′ hydroxyl oxygen and the *N*-acetyl carbonyl oxygen of ManNAc, respectively (Fig. 4*C*). The enzyme surface that binds the substrate is near residue D65, which has been shown to be important for catalysis, consists of an electronegative portion that contacts ManNAc and an electropositive region that interacts with the uracil and ribose rings (Fig. 4*D*) (36).

Table 1
Crystal data collection and structure refinement statistics

Protein complex	TagA^{FL*}:UDP–GlcNAc	TagA^{FL*}:UDP–ManNAc
Data collection		
PDB code	7MPK	7N41
Space group	$P\,3_1\,2\,1$	$P\,3_1\,2\,1$
Cell dimensions		
a, *b*, *c* (Å)	113.24, 113.24, 114.45	113.24, 113.24, 118.55
α, *β*, *γ* (°)	90.00, 90.00, 120.00	90.00, 90.00, 120.00
Resolution (Å)	74.5–3.0 (3.1–3.0)[a]	98.4–3.3 (3.4–3.3)[a]
Wavelength (Å)	0.9797	0.9791
R_{merge} (%)	12.0 (88.1)	8.0 (128.8)
$I/\sigma I$	16.2 (2.9)	14.7 (1.6)
$CC_{1/2}$	99.8 (83.0)	99.8 (66.6)
Completeness (%)	99.7 (97.3)	99.9 (99.8)
Redundancy	9.8 (9.1)	7.4 (7.2)
Wilson *B*-factor (Å^2)	76.05	84.37
Refinement		
Resolution (Å)	74.5–3.0	98.4–3.3
No. of reflections	17,470	13,694
R_{work}/R_{free} (%)[b]	20.6/25.3	19.1/24.7
No. of atoms	4755	4612
Protein	4519	4456
Ligand/ion	234	156
Water	2	0
B-factors (Å^2) (all atoms)	71.2	91.8
Protein	71.4	91.6
Ligand/ion	66.4	97.5
Water	54.1	N/A
RMSD		
Bond lengths (Å)	0.013	0.010
Bond angles (°)	1.55	1.26
Ramachandran favored (%)	98.1	96.4
Ramachandran allowed (%)	1.9	3.6
Ramachandran outliers (%)	0.0	0.0

Abbreviation: N/A, not applicable.
[a] Values in parentheses are for the highest-resolution shell.
[b] R_{merge} calculated using 5% of collected experimental data.

To investigate how enzyme contacts to the substrate's *N*-acetyl group affect sugar donor binding, we also determined the structure of TagA^{FL*} bound to UDP–GlcNAc, a C2' epimer of UDP–ManNAc where the *N*-acetyl group is inverted. The TagA^{FL*}:UDP–GlcNAc complex crystallized with three molecules in the asymmetric similar to the TagA^{FL*}:UDP–ManNAc complex and was resolved to 3.0 Å. The structures of the core domain in each complex are nearly identical with a C_α RMSD of 0.28 Å. However, in the TagA^{FL*}:UDP–GlcNAc complex, the GlcNAc adopts two distinct conformations: conformer 1 in which it is displaced from the body of the protein (60% abundance) and conformer 2 that resembles the positioning of the substrate in the UDP–ManNAc complex (40% abundance) (Figs. 4*B* and S3*B*). However, as compared with the TagA^{FL*}:UDP–ManNAc structure, the most abundant GlcNAc conformer is positioned farther away from the catalytically important D65 residue, and both GlcNAc conformers form distinct contacts to the enzyme from their *N*-acetyl groups (Fig. S3, *C* and *D*). As TagA can only utilize UDP–ManNAc as a substrate, it seems likely that core domain contacts to the *N*-acetyl and 3′ hydroxyl groups in the substrate from the side chains of Q167 and N39 are important specificity determinants.

MD simulations provide insight into the function of the CTT

The crystal structure of TagA^{FL*} bound to UDP–ManNAc lacks density for the CTT that is required for efficient catalysis *in vitro* (36). A comparison of the TagA^{FL*}:UDP–ManNAc complex and TagA^{CM} structures suggests that the CTT may be involved in substrate binding. To gain insight into the function of the CTT, we performed MD simulations of full-length monomeric TagA in its apo and UDP–ManNAc bound states. The program MODELLER was used to

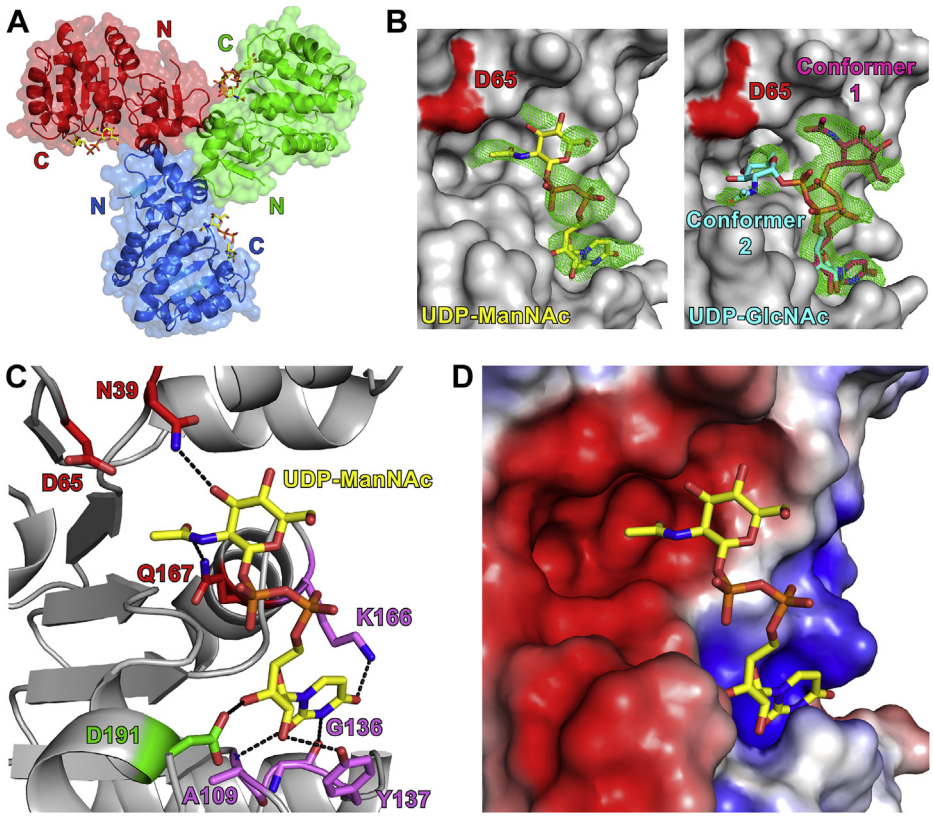

Figure 4. Structures of the TagA^{FL}*:UDP–ManNAc and TagA^{FL}*:UDP–GlcNAc complexes. *A*, the asymmetric unit of the TagA^{FL}*:UDP–ManNAc complex (Protein Data Bank code: 7N41). Protein subunits are shown in *red*, *blue*, and *green*. The protein in the TagA^{FL}*:UDP–GlcNAc complex (Protein Data Bank: 7MPK) adopts a very similar structure. *B*, iterative-build 2mF_o–DF_c composite omit maps showing the location of the UDP–ManNAc (*left*, *yellow*) and UDP–GlcNAc (*right*, *cyan* and *purple*) ligands in the TagA^{FL}* complexes (contoured at 1.0 σ) (45). UDP–GlcNAc adopts two conformations, 1 and 2, where the sugar is oriented away from or toward the catalytic pocket containing D65. Additional simulated annealing omit maps for the structures are presented in Fig. S3. *C*, image showing enzyme–substrate interactions in the TagA^{FL}*:UDP–ManNAc complex. A *cartoon* representation of the protein is shown in *gray*. UDP–ManNAc and amino acid side chains are shown in *stick* format. Amino acid side chains are colored based on the part of the substrate they contact: uracil base (*purple*), ribose (*green*), and ManNAc moiety (*red*). *D*, electrostatic surface of the substrate binding site in the TagA^{FL}*:UDP–ManNAc complex. Coloring is as follows: anionic (*red*), neutral (*white*), and cationic (*blue*). GlcNAc, *N*-acetyl-D-glucosamine; ManNAc, *N*-acetyl-D-mannosamine.

generate starting models for MD using the coordinates of the TagA^{FL}*:UDP–ManNAc complex and TagA^{CM} (described in the Experimental procedures section) (46). The simulations were equilibrated until the protein's backbone RMSD indicated good convergence, requiring 20 ns for the apo state and 200 ns for the complex. Production simulations were then performed for a duration of 1.0 μs for both systems. For simulations of the complex, the positioning of the substrate relative to the core domain was held fixed using distance restraints.

The simulation of the apoenzyme shows that the CTT remains packed against the body of the protein but transiently moves to expose the UDP–ManNAc binding site on the core domain. The position of the CTT against the core domain resembles the TagA^{CM} model derived solely from coevolutionary restraints, suggesting that it represents a low-energy state. In particular, H12' primarily rests in a groove formed by core domain helices H2 and H4, and helix H11' is positioned adjacent to the surface formed by residues in the β4–H4 and H4–H5 loops. Interestingly, the root mean square fluctuation of the apoprotein's backbone atoms during the trajectory reveals significant mobility in the H10' and H11' helices, as their backbone atom coordinates fluctuate by as much as 4 Å (Fig. 5*A*). Large coordinate variations are also observed in residues located at the C-terminal end of H12' (H238–R244), which are poorly conserved and transiently unwind. To gain deeper insight into the nature of the H10'–H11' fluctuations, we clustered conformers in the trajectory that adopted similar structures (47). More than 95% of conformers in the MD trajectory can be grouped into six clusters of related structures, in which members of each cluster exhibit backbone coordinate RMSDs that are less than 0.9 Å. Comparing the central structure from each cluster reveals that helices H10' and H11' in the CTT move as a unit that toggles between two extrema: an "in state" in which they are positioned nearer the β4–H4 and H4–H5 loop elements on the core domain and an "out state" in which they are further displaced from the body of the protein (Fig. 5, *B–D*). As expected, the two extrema are in fairly low abundance (2.2% and 1.2% of the trajectory for the "in state" and the "out state," respectively), and the vast majority of the simulation (>90% of the trajectory) falls in transient stages between the extremes. Upon closer inspection, multiple periods of transition occur between the "in state" and "out state" for the H10' and H11' helices throughout the simulation (Fig. S4*A*). Transient enzyme toggling between states exposes the surface that binds

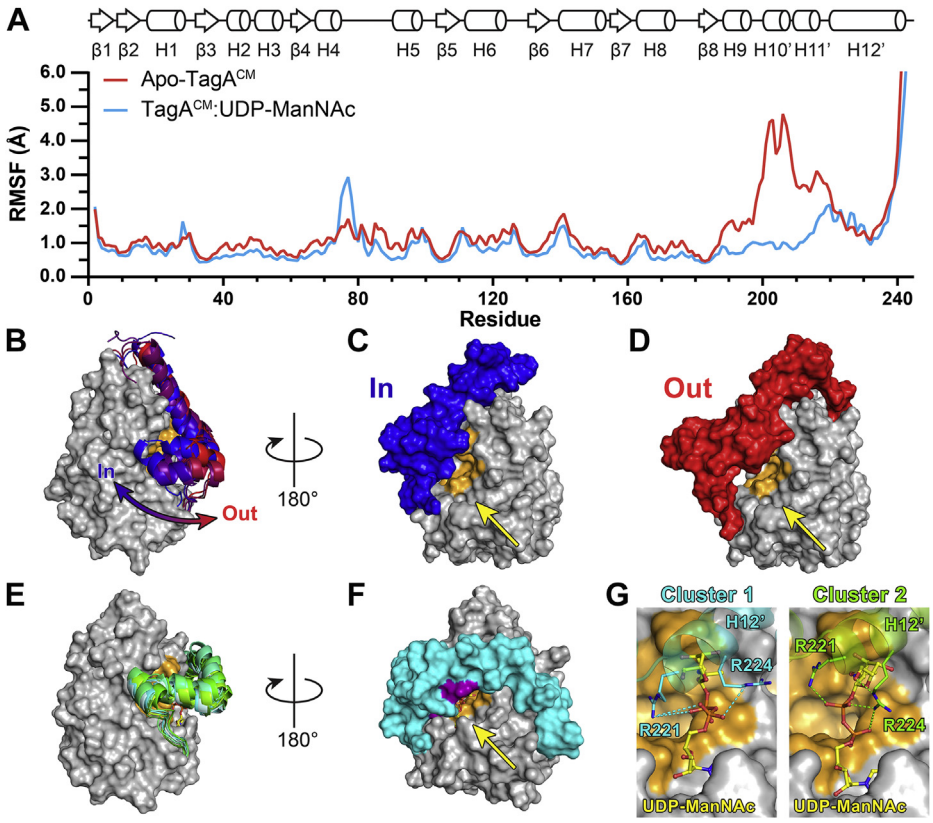

Figure 5. MD simulations of apo-TagA^{CM} and the TagA^{CM}:UDP–ManNAc complex. *A*, a plot showing the root mean square fluctuation (RMSF) differences of TagA backbone coordinates during apo and complex MD simulations. *B*, representative clusters of the apo-TagA^{CM} simulation showing the fluctuations of the C-terminal tail (CTT). The surface of the core domain is colored *gray*, residues within the catalytic pocket are colored *orange*, and helices H10′–H12′ in the CTT are colored based on their positioning ("in state" [*blue*, *C*] or "out state" [*red*, *D*]). The UDP–ManNAc binding surface is indicated by *yellow arrows*. *E*, representative clusters of the TagA^{CM}:UDP–ManNAc complex simulation. The H10′ and H11′ helices are stable and pack against the core in a conformation that resembles the "in state" observed in the simulation of apo-TagA^{CM}. The CTT for each cluster is represented in *cartoon* format and shaded from *light blue* to *green*. *F*, the surface representation of the primary cluster (*cyan*) from the complex simulation. Coloring as in panels (*B–D*). Conserved CTT arginine residues (R214, R221, and R224) are colored *purple*. *G*, enzyme–substrate contacts observed in the MD simulation of the TagA^{CM}:UDP–ManNAc complex. In the two most populated clusters in the trajectory (cluster 1, 51% of the trajectory, *cyan*; cluster 2, 25% of the trajectory, *green*), the diphosphate group in UDP–ManNAc interacts with two highly conserved arginine residues, R221 and R224. ManNAc, *N*-acetyl-D-mannosamine; MD, molecular dynamics.

UDP–ManNAc in the crystal structure of the complex, suggesting that these motions may have a functional role in promoting substrate access to the active site (Fig. 5, *C* and *D*).

Simulations of the TagA^{CM}:UDP–ManNAc complex reveal that the CTT plays a prominent role in binding UDP–ManNAc and that it becomes immobilized upon substrate binding. Substrate-dependent CTT immobilization is evident from a plot of the root mean square fluctuation values for the backbone atoms during the simulation of the complex, as they generally exhibit small magnitudes throughout the protein (Fig. 5*A*). Moreover, more than 95% of the conformers in the trajectory (represented by eight clusters) contain a CTT that adopts an "in state" similar to that observed in the apoenzyme simulation, and the coordinates of the H10′–H11′ unit are stable throughout the simulation (Figs. 5*E* and S4*B*). In addition, the C-terminal H12′ helix is displaced upon substrate binding, moving from the groove formed by core helices H2 and H4 in the apoform to a groove formed by helices H2 and H8 in the complex (Fig. 5*F*). As helix H12′ no longer packs against helix H4 of the core domain, the newly freed H4 and its following H4–H5 loop exhibit increased coordinate fluctuations between residues 74 and 82 (Fig. 5*A*). The most deviant coordinate fluctuations appear again for the C terminus of helix H12′ (H238–R244), which transiently unwind similar to the apo simulation. Conformational ordering of the CTT appears to be triggered by substrate contacts from two residues located at the N terminus of helix H12′, R221, and R224, which form ionic and hydrogen bond interactions with the diphosphate group in UDP–ManNAc. In the representative clusters for the complex simulation, R221 and R224 work in tandem to stabilize the anionic charges of the substrate (Fig. 5*G*). R221 most often makes contacts with the β-phosphate oxygen of UDP–ManNAc and is sometimes proximal to the α-phosphate oxygen atoms. In the eight clusters representing 95% of the trajectory, R224 appears to occupy two different side-chain conformations throughout the simulation. The most populated orientation of R224 (represented by cluster 1) participates in cation–π interactions with F231, and the ε-amine donates a hydrogen bond to an α-phosphate oxygen (Figs. 5*G* and S5*A*). The minor orientation (represented by cluster 2) makes ionic interactions near α- and β-phosphate oxygen atoms by a terminal amino group, whereas the ε-amine donates a new hydrogen bond to the 6′ hydroxyl oxygen of ManNAc (Figs. 5*G* and S5*B*). The ionic interactions of these arginine residues with the substrate may be important for catalysis, as they are highly conserved in the primary sequences

of TagA homologs (Fig. 1*B*). Substrate binding causes the CTT to adopt a more compact state as it engages the core domain; in more than 95% of the conformers in the trajectory, helix H10' packs against the H4–H5 loop, and helix H11' is placed near helix H4. Substrate-induced closure of the CTT over the nucleotide substrate positions several highly conserved amino acids in CTT near the active site whose significance in catalysis is investigated later (Figs. 1*B* and S5).

Highly conserved residues within the CTT are required for catalysis

TagA homologs contain five highly conserved amino acids in the CTT whose side chains project toward the bound substrate or the core domain in the MD simulation of the enzyme–substrate complex (E210, W211, R214, R221, and R224) (Figs. 1*B* and S5). To assess their importance in catalysis, we determined the *in vitro* GT activity of TagA variants containing amino-acid substitutions at these sites (Fig. 6*A*). A previously described enzyme-coupled end-point activity assay was employed to measure the rate of UDP product formation (19, 36). The TagA variants studied include five single amino-acid mutants that alter conserved amino acids in the CTT (E210A, W211A, R214E, R221E, and R224E), a variant that serves as a negative control because it alters a surface-exposed side chain in the CTT that is not expected to be involved in catalysis (I203E) and the $TagA^{\Delta C}$ and $TagA^{FL*}$ constructs. Severe reductions in activity relative to native TagA occur when the CTT is removed ($TagA^{\Delta C}$ is approximately eightfold less active), or key reactants are withheld from the assay (~15-fold activity reductions are observed when the lipid substrate, MnaA epimerase, or TagA enzyme are not present) (Fig. 6*B*). These reductions in activity also occur when the three conserved arginine residues in the CTT are altered, with R214E and R221E substitutions causing the largest decline (Fig. 6*B*). In the complex MD simulation, R221 and R224 directly interact with the UDP–ManNAc substrate, whereas the side chain of R214 may be involved in lipid-α binding (see Discussion section). Altering other conserved residues within the CTT (E210A and W211A) also reduces activity, albeit to a much lesser extent (Fig. 6*B*). An I203E variant serves as negative control as its side chain projects into the solvent in the complex MD simulation and, as expected, it exhibits wildtype-level activity. The $TagA^{FL*}$ variant that alters four nonpolar residues in the CTT to improve enzyme solubility has reduced activity; however, it is still nearly three times more active than $TagA^{\Delta C}$ in which the CTT is removed (Fig. 6*B*). The activities of the R221A and $TagA^{\Delta C}$ variants have been measured previously (36). The new data presented here substantiate and expand upon these findings by showing that two additional highly conserved arginine residues in the CTT are important for catalysis.

Native MS suggests that monomeric TagA interacts with membranes

Cell fractionation studies previously revealed that the *B. subtilis* TagA enzyme is a peripheral membrane protein and that nonpolar residues in a hydrophobic patch on CTT affect interactions with the lipid bilayer (36). We investigated how the closely related *T. italicus* TagA interacts with membranes using native MS, which enables the study of intact proteins and their complexes using electrospray ionization MS. A negative ion mode MS analysis of 10 μM TagA in aqueous solution resolves monomeric and dimeric forms of the protein, with a slight enrichment for the dimer (Fig. 7*A*). This is consistent with NMR and SE–AUC studies of TagA, which were performed at much higher concentrations that presumably cause the protein to nonspecifically aggregate. The behavior of native TagA in the presence of *n*-dodecyl-β-D-maltoside (DDM), an MS-compatible detergent that forms micelles at higher concentrations, was monitored by native MS (48–50). With DDM concentrations below its critical micelle concentration, relatively small effects on the TagA monomer–dimer equilibrium are observed by native MS. Interestingly, as the DDM concentration is increased above its critical micelle concentration (ca. 170 μM), there is a shift in TagA toward its monomeric state, suggesting this form of the enzyme is stabilized by the DDM micelles. Combined, the MS, MD, and activity data suggest that the structure of monomeric TagA resembles the $TagA^{CM}$ model, and that this represents the active form of the enzyme that can associate with lipid bilayers *via* its CTT.

Discussion

TagA ManNAc transferases catalyze the first committed step in WTA biosynthesis and are a promising antimicrobial drug target (6, 7). At present, only the structure of the TagA homolog from *T. italicus* has been reported, revealing that these enzymes contain a unique GT core domain that is followed by a conserved ~50 amino acid CTT of unknown structure (Fig. 1) (36). The CTT targets the enzyme to the cell membrane and is required for catalysis *in vitro* (36). However, its role in catalysis is not well understood because only a structure of a C-terminally truncated variant in which the CTT was removed has been determined ($TagA^{\Delta C}$) (36). $TagA^{\Delta C}$ forms dimeric and trimeric crystals when the protein is in its apo- and UDP-bound states, respectively (36). Our working hypothesis is that only the monomeric form of TagA is biologically active, since a computationally derived model of the monomeric structure of TagA (called $TagA^{CM}$) places several highly conserved and functionally important residues in the CTT near the presumed active site on the core domain (Fig. 3*B*). Here, we used a combination of crystallography, molecular modeling, MS, MD simulations, and functional assays to gain insight into how the core domain and CTT in TagA work together to catalyze glycosylation at the cell membrane.

NMR and AUC studies of the full-length protein reveal that the CTT promotes nonspecific aggregation (Figs. 2*C* and S1). Using the $TagA^{CM}$ model as a guide, we constructed $TagA^{FL*}$, which is significantly more soluble than the native protein because it eliminates four nonpolar residues in the CTT that are predicted to form a hydrophobic patch (Fig. 3*C*). Unlike the native protein, $TagA^{FL*}$ is in equilibrium between monomeric and dimeric states (Fig. 3*D*). A crystal structure of the

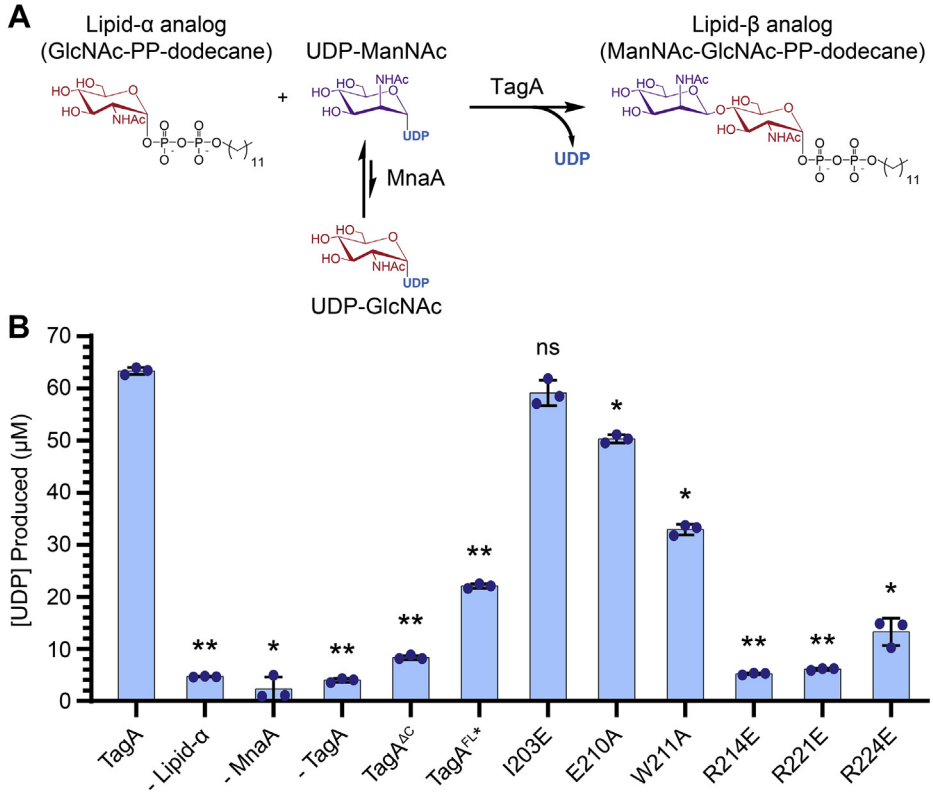

Figure 6. Enzyme activities of TagA variants. *A*, schematic of the *in vitro* enzyme-coupled TagA activity assay. Synthetic lipid-α analog is coincubated with UDP–GlcNAc and an MnaA epimerase before addition of the TagA enzyme. UDP product accumulation is quantified by absorbance at 262 nm. *B*, a chart of UDP product formation for a series of variants of the *Thermoanaerobacter italicus* TagA enzyme following an end-point activity assay. Bars labeled "-lipid-α," "-MnaA," and "-TagA" indicate assays in which these components were not present. All single amino-acid substitutions were introduced in the native full-length TagA enzyme. Each experiment was performed in technical triplicate, and error bars represent the standard deviation of measurements. The statistical significance (*p* value) between native TagA and variant activity datasets was determined using the Analysis ToolPak in Microsoft Excel. *Asterisks* indicate the statistical significance between the activity of native TagA and variant datasets (*$p < 0.005$, **$p < 0.0001$, and ns). GlcNAc, *N*-acetyl-D-glucosamine; ns, not significant.

TagA^{FL*}:UDP–ManNAc complex provides insight into the mechanism of substrate binding. This structure contains a core domain that adopts a Rossmann-like fold in which the UDP–ManNAc sugar donor is positioned at the edge of the sheet (coordinated by residues in loops β5–H6, β6–H7, β7–H8 and β8–H9, and helices H2 and H9) (Fig. 4*C*). The substrate is positioned near residue D65 in the core domain, an essential catalytic residue. UDP–ManNAc binds to TagA similar to its UDP product, which was visualized previously in the crystal structure of the TagA$^{\Delta C}$:UDP complex (36). However, in the product complex, the diphosphate group of UDP contacts the protein's core domain, whereas in the structure of the UDP–ManNAc complex, it does not. TagA cannot utilize UDP–GlcNAc as a substrate, a closely related sugar nucleotide that differs from UDP–ManNAc only in the stereochemistry of the C2' atom that is bonded to the *N*-acetyl group. The structure of the TagA^{FL*}:UDP–ManNAc complex reveals sugar contacts to the TagA core domain that contribute to stereoselectivity. In particular, the *N*-acetyl group of ManNAc is projected into a pocket where its carbonyl oxygen accepts a hydrogen bond from the amino group in Q167, and the 3′ hydroxyl oxygen accepts a hydrogen bond from the amino group in N39 (Fig. 4*C*). In contrast, the distinct stereochemistry of UDP–GlcNAc precludes these contacts, and instead the GlcNAc adopts two distinct conformers that are presumably enzymatically inactive (Fig. 4*B*).

Our results reveal that the CTT plays a major role in facilitating catalysis by encapsulating the UDP–ManNAc substrate and by contributing catalytically important side chains to the active site. We modeled the structure of the full-length TagA enzyme containing the CTT using coevolution restraints and the coordinates of the TagA^{FL*}:UDP–ManNAc complex. MD simulations of the complex reveal that the CTT covers the UDP–ManNAc molecule, forming three α-helices (H10'–H12') that project five highly conserved residues toward the active site: E210, W211, R214, R221, and R224 (Fig. S5). Prior work demonstrated that R221 in the CTT is crucial for catalysis, but the role of other conserved amino acids remained unknown. In this study, we demonstrate that all conserved arginine residues in the CTT are critical for catalysis (R214, R221, and R224). MD simulations shed light onto the functions of R221 and R224, as both residues form favorable and long-lived ionic interactions with the diphosphate group of UDP–ManNAc (Fig. 5*G*). The conserved R214 residue in the CTT is also extremely important for catalysis and may be involved in binding lipid-α (see later). In the MD simulation of the complex, the other conserved residues in the CTT (E210 and W211) pack against the core domain and may stabilize the

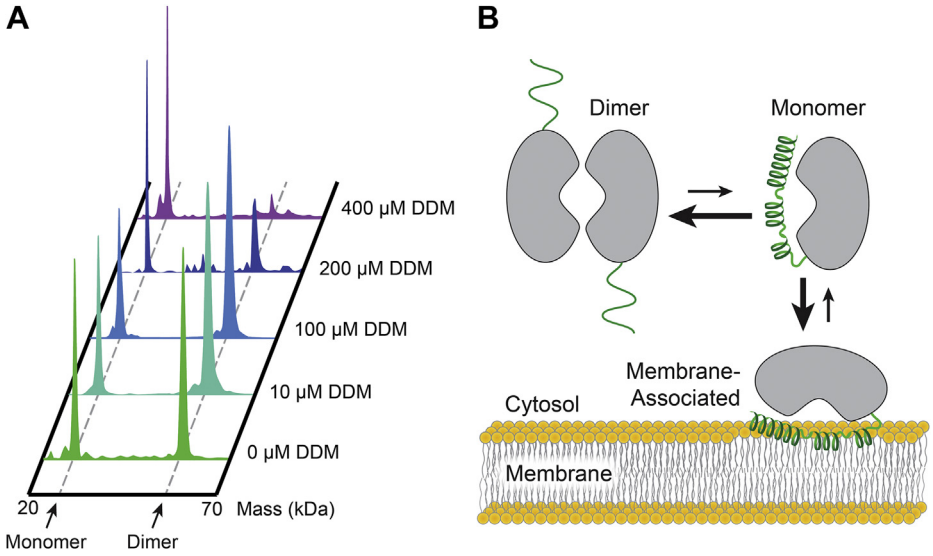

Figure 7. Native MS studies of micelle binding by TagA. *A*, negative ion mode MS of 10 μM TagA. Deconvoluted mass spectra of TagA with varying concentrations of DDM detergent (critical micelle concentration [CMC]: 170 μM) are shown with monomeric and dimeric peaks identified. *B*, proposed model of membrane association by TagA based on the results of experimental and computational studies. In solution, TagA exists in a monomer–dimer equilibrium. When a membrane is present, a hydrophobic patch on the CTT that exists only in the monomeric form of the enzyme favorably interacts with the membrane and is stabilizing. As a monomer on the membrane, TagA is poised to bind to its lipid-α substrate. CTT, C-terminal tail; DDM, *n*-dodecyl-β-D-maltoside.

positioning of the CTT over the UDP–ManNAc substrate. However, these interactions appear to be less important for catalysis as E210A and W211A variants retain ∼80% and ∼50% activity as compared with wildtype TagA, respectively (Fig. 6*B*).

The data presented here provide new insight into the molecular basis of catalysis. From this work and a previously published study, a total of four highly conserved residues in TagA have been shown to be important for catalysis, and when altered, they reduce enzyme activity to less than 25% of the wildtype enzyme (36). These residues include R214, R221, and R224 in the CTT and D65 located in the core domain (Fig. 6*B*) (36). TagA is a metal-independent inverting GT and therefore presumably catalyzes an S_N2-like direct displacement reaction at the anomeric carbon of the UDP–ManNAc sugar donor (19). In this reaction, the nucleophilic 4′ hydroxyl group of GlcNAc in lipid-α may be deprotonated for catalysis by the side chain of D65 located in a conserved pocket on the core domain. Indeed, the $TagA^{FL*}$:UDP–ManNAc structure shows that the D65 side chain is near the C1′ of ManNAc in a solvent-accessible pocket that the GlcNAc sugar of lipid-α presumably occupies when bound. Following activation by D65, the 4′ oxygen in lipid-α attacks the C1′ anomeric carbon in UDP–ManNAc, promoting the breakage of its phosphodiester bond and transfer of ManNAc to lipid-α. Based on our MD simulations of the complex, R221 and R224 in the CTT likely play a key role in this process by stabilizing the UDP-leaving group through ionic and hydrogen-bonding interactions with its diphosphate group (Fig. 5*G*). How the enzyme recognizes lipid-α remains unclear, as it may also involve enzyme–bilayer interactions that are mediated by the CTT. In the MD simulation of the complex, a small pore formed between the CTT and core domain could serve as the entry point for lipid-α. Intriguingly, this pore connects the nonpolar surface on the CTT that is predicted to contact the membrane to a cavity that contains the C1′ atom in UDP–ManNAc. Moreover, it is lined with the conserved E210 and R214 residues in the CTT as well as D65 located in the core domain that is conserved and has previously been shown to be important for catalysis (Figs. 6*B* and S5) (36). It is thus tempting to speculate that the pore serves as the entry point for lipid-α whose positioning would be stabilized by favorable electrostatic interactions between the diphosphate group of lipid-α and the guanidinium group of R214. A comparison of the MD trajectories of TagA in its free and UDP–ManNAc bound states reveals that UDP–ManNAc binding quenches motions in the H10′ and H11′ helices of the CTT (Fig. 5*A*). When bound, the substrate locks the CTT in an "in state" conformation, in which it presents highly conserved residues that either interact with the core domain (E210 and W211) or the UDP–ManNAc substrate (R221 and R224). This substrate-induced structural transition may explain the results of steady-state kinetics measurements that concluded that the *B. subtilis* TagA enzyme works *via* an ordered Bi–Bi mechanism in which UDP–ManNAc binds to the enzyme first, followed by lipid-α (19). However, additional studies are needed to investigate the molecular basis of this second binding event.

TagA and other members of the GT26 family are structurally unique GTs that share some commonalities with the recently characterized GT-D type enzymes. Structural analyses have identified four distinct GT superfamilies based on their folds, GT-A through GT-D (33, 51, 52). The majority of GT enzymes form water-soluble structures that fall into the GT-A or GT-B superfamilies or are members of the membrane-integrated GT-C superfamily (51). The GT-D superfamily contains a single characterized enzyme, DUF1792, which

transfers glucose from UDP-glucose to post-translationally modify a serine-rich repeat glycoprotein Fap1 in *Streptococci* and *Staphylococci* (52, 53). As previously described, TagA has a similar β-sheet topology and limited structural similarity to the GT-D domain in DUF1792 (DALI *Z*-score: 8.7) (Fig. S6) (36, 54, 55). However, the proteins differ markedly at their C termini. DUF1792 contains an extended loop region between the H10 and H11 helices in its core domain that encloses the substrate binding site and coordinates the catalytically essential Mn^{2+} cation with an acidic glutamate residue (Fig. S6) (52). In contrast, TagA enzymes eschew this extended loop in favor of a CTT appendage that encapsulates the donor sugar and contributes conserved arginine residues that promote metal-independent GT activity. Based on primary sequence alignments, the CTT is conserved among GT26 enzymes, with ∼81% of these enzymes retaining this structural element and its essential conserved arginine residues. The surface-exposed hydrophobic patch is also conserved in these enzymes, suggesting that like TagA, they glycosylate membrane-embedded lipid substrates. Thus, we conclude that TagA and other GT26 enzymes adopt a novel GT-E type structural fold that is well suited for recognizing membrane-associated sugar acceptors.

The results presented here further support the hypothesis that TagA functions as a monomer on the cytoplasmic membrane. Native MS experiments of TagA at protein concentrations likely to occur within the cell (∼10 μM) demonstrate that it is in equilibrium between monomeric and dimeric states (Fig. 7*A*). However, in the presence of micelles, the equilibrium is shifted to the monomeric form, presumably because bilayer interactions with the hydrophobic surface of CTTs in the monomer stabilize its structure and association with the core domain. As the CTT and potential intersubunit surfaces on the core domain are coincident, stabilizing CTT binding can be expected to limit oligomerization. This notion is compatible with SE-AUC studies of solubilized variants of TagA ($TagA^{\Delta C}$ and $TagA^{FL*}$), which show that removing the CTT and its ability to compete for binding to the intersubunit interface promotes dimerization (the K_D for dimer association is 210 ± 50 μM for $TagA^{FL*}$ and 7.4 ± 0.7 μM for $TagA^{\Delta C}$). Finally, MD simulations and biochemical studies further validate the structure of the $TagA^{CM}$ model, which is shown to represent an energetically stable form of the protein that explains the importance of several conserved CTT (R214, R221, and R224) and core domain (D65) residues that are required for catalysis.

In conclusion, our data provide new insight into how TagA enzymes synthesize the linkage unit that joins WTA to the bacterial cell wall. When removed from the membrane and at low protein concentrations, TagA is in an equilibrium between monomeric and dimeric states (36). Upon encountering the cytoplasmic membrane that houses its lipid-α substrate, its monomeric form is stabilized *via* interactions between a hydrophobic patch on the CTT and the lipid bilayer (Fig. 7*B*). This form of the protein is enzymatically active as it enables the CTT to encapsulate the UDP–ManNAc substrate that is bound to the surface of the core domain. Both the CTT and core domain are instrumental in constructing the active site that contains UDP–ManNAc, with each presenting conserved residues that are important for catalysis (core domain: D65; CTT: R214, R221, and R224). Steady-state kinetics suggest that UDP–ManNAc binds first, which in MD simulations is shown to stabilize the closure of the CTT and the formation of a potential lipid-α binding surface. Based on modeling studies, we ascribe tentative functions for the conserved residues, with the arginine residues acting to stabilize the phosphate groups in UDP–ManNAc (R221 and R224) and lipid-α (R214) substrates. Meanwhile, D65 in the core domain may act as a general base, which deprotonates the terminal GlcNAc sugar in lipid-α. As other members of the GT26 family also contain a conserved CTT segment, they may glycosylate membrane lipid substrates using a similar mechanism. The results of this study could facilitate the development of new antimicrobial compounds that work by disrupting TagA function and, in turn, WTA biosynthesis.

Experimental procedures

Protein expression and purification

All TagA constructs were cloned and expressed using standard methods as described previously (36). Briefly, each construct contained an N-terminal 6×His-tag and tobacco etch virus protease recognition sequence (ENLYFQS) in the pMAPLe4 expression vector and was expressed in *E. coli* BL21(DE3) cells. Cultures were grown in the presence of 50 μg/ml kanamycin at 37 °C in a shaking incubator to an absorbance of 0.6 to 0.8 at 600 nm, before induction with 1 mM IPTG. Expression proceeded at 17 °C for 16 h. For isotopically labeled samples, the cell pellets were exchanged into M9 media supplemented with $^{15}NH_4Cl$, ^{13}C-glucose, and 70% or 99% deuterium oxide (Cambridge Isotope Laboratories) before induction (56). The cells were pelleted by centrifugation, and cells were resuspended in 40 mM CHAPS, 50 mM Tris–HCl, pH 7.5, 500 mM NaCl supplemented with 400 μl of protease inhibitor cocktail (Sigma), and 2 mM phenylmethanesulfonylfluoride, 1 mg egg white lysozyme, and 0.5 mg *Serratia marcescens* nuclease per liter of culture (57). Resuspended cells were lysed using an Emulsiflex high pressure homogenizer (Avestin), and soluble TagA lysate was clarified by centrifugation.

TagA protein was purified by passing over Co^{2+}–NTA HisPur resin (Fisher) and washed with 8 mM CHAPS, 50 mM Tris–HCl, pH 7.5, and 500 mM NaCl buffer. Protein was eluted from the column using the same buffer supplemented with 200 mM imidazole. Eluted protein was concentrated using Amicon centrifugal filters (Fisher). The 6×His-tag was proteolytically removed using 0.5 mg tobacco etch virus protease and buffer exchanged into a 10% v/v glycerol, 50 mM Tris–HCl, pH 7.5, and 500 mM NaCl buffer. The protein was passed over Co^{2+}–NTA again and washed off the column with 10% v/v glycerol, 50 mM Tris–HCl, pH 7.5, and 500 mM NaCl buffer. Wash fractions containing protein were concentrated, and the protein was subjected to size-exclusion chromatography using a Superdex 75 preparation grade column (GE). Protein was concentrated prior to storage or use. TagA

variants were tested for their *in vitro* activity using a previously described HPLC assay (19, 36).

NMR assignments and relaxation experiments

Isotopically labeled samples were dissolved in NMR buffer at pH 6.8 (50 mM sodium phosphate, pH 6.8, and 100 mM NaCl). TagA$^{\Delta C}$ was uniformly isotopically labeled with ^{15}N or ^{15}N/^{13}C/^{2}H (70%), whereas TagA^{FL*} was uniformly isotopically labeled with ^{15}N or ^{15}N/^{13}C/^{2}H. Spectra were acquired at 303 K on Bruker Avance III HD 600 MHz (14.1 T) and Bruker Avance NEO 800 MHz (18.8 T) spectrometers equipped with triple resonance cryogenic probes. Backbone assignments of TagA$^{\Delta C}$ were determined by carrying out TROSY-enhanced variants of the following experiments: ^{15}N-HSQC, HNCO, HN(CA)CO, HNCA, HN(CO)CA, HNCACB, and HN(CO)CACB (58, 59). Data were processed using NMRPipe, and CARA was used to perform sequential assignment (60, 61). Secondary structures were predicted from secondary ^{13}C chemical shifts using TALOS-N (62).

The rotational correlation time (τ_c) of TagA$^{\Delta C}$ was estimated from a series of ^{15}N-TRACT experiments (63). Spectra were acquired using a 0.48 mM sample of ^{15}N-labeled TagA$^{\Delta C}$, with 2048 complex points, 128 transients, and 100 experiments for each spin state, and the relevant delay incremented by 1.5 ms. The decay of integrated amide proton signals was fitted to an exponential decay resulting in a transverse cross-correlated relaxation rate (η_{xy}) of 28 Hz. This allowed calculation of τ_c ~27 ns *via* algebraic solutions (64).

SE–AUC

SE–AUC experiments were performed on an Optima XL-A analytical ultracentrifuge (Beckman Coulter). Three concentrations for each sample (TagA: 4.2, 13, and 19 μM; TagA^{FL*}: 6.0, 18, and 28 μM) were subjected to three ultracentrifuge speeds (TagA: 12,000, 15,000, and 19,000 rpm; TagA^{FL*}: 15,000, 19,000, and 24,000 rpm) and allowed to reach SE at 4 °C. Data regression analysis was performed using the Beckman-Coulter Optima Analytical Ultracentrifuge Origin Data Analysis Package. The data were fit to multiexponential and single-exponential models. TagA^{FL*} was best represented by a monomer–dimer multiexponential model that was calculated using the predicted monomeric molecular weight of 27,800 Da by the ExPASy ProtParam tool (65). The dissociation constant (K_D) was determined to be the inverse of $K_{A(conc)}$ using the following equation (66, 67):

$$K_{A(\mathrm{conc})} = K_{A(\mathrm{abs})}\left((\varepsilon l)^{n-1} / n\right)$$

where the molar extinction coefficient (ε) was determined to be 21,720 cm^{-1} M^{-1} using the ExPASy ProtParam tool, l is the path length of 1.2 cm, n is the order of oligomerization, and $K_{A(abs)}$ is the absorbance association constant.

X-ray crystallography

The TagA^{FL*}:UDP–GlcNAc and TagA^{FL*}:UDP–ManNAc complexes were cocrystallized from solutions containing 15 mg/ml or 11 mg/ml TagA^{FL*}, respectively, dissolved in 50 mM Tris–HCl, pH 7.5, 500 mM NaCl buffer, and supplemented with either UDP–GlcNAc or UDP–ManNAc at five-fold molar excess. The UDP–ManNAc used to form the complex was produced using a chemoenzymatic method as described previously (68). Screening was done using the hanging-drop vapor diffusion method at 4 °C. The TagA^{FL*}:UDP–GlcNAc complex crystallized in 100 mM HEPES, pH 7.5, 10% w/v PEG 8000, and 8% v/v ethylene glycol. The TagA^{FL*}:UDP–ManNAc complex crystallized in 40 mM potassium phosphate monobasic, pH 7.5, 10% w/v PEG 8000, and 20% v/v ethylene glycol. Crystals were cryoprotected when harvested in reservoir solution containing 45% glycerol. Both datasets were obtained at the Advanced Photon Source beamline 24-ID-C. Data were acquired at 100 K with 0.25° oscillations and detector distances of 300 and 450 mm for the TagA^{FL*}:UDP–ManNAc and TagA^{FL*}:UDP–GlcNAc complexes, respectively.

XDS/XSCALE was used to perform indexing, integration, and scaling of each dataset in the P3$_1$21 space group (69). Dataset anisotropy was corrected using the UCLA Diffraction Anisotropy Server and a −50 Å^2 isotropic *B*-factor sharpening was applied to improve the apparent *B*-factors of the TagA^{FL*}:UDP–ManNAc crystal structure (43). For both complexes, the asymmetric unit contained three molecules of TagA^{FL*}. Phases were determined by molecular replacement using Phaser and the previously published coordinates of TagA$^{\Delta C}$ (PDB: 5WB4) (70, 71). Structures were iteratively improved by rounds of manual model building in Coot and automated refinement using BUSTER (Global Phasing Ltd) and PHENIX (71–73). The GRADE server was used to parameterize UDP–ManNAc for subsequent refinement (73). To assess the accuracy of ligand placement, composite iterative-build omit maps were generated by refining the structures of TagA^{FL*}:UDP–GlcNAc and TagA^{FL*}:UDP–ManNAc using the AutoBuild program such that the resulting electron density in the omitted ligand region is unbiased by the atomic model (45). Additional simulated-annealing composite omit maps were generated using PHENIX (71). Statistics for data collection and structure refinement are given in Table 1.

MD simulations

A model of the full-length *T. italicus* protein for MD simulations was constructed as follows. First, the GREMLIN method was used to model the structure of the full-length *B. subtilis* TagA protein using coevolutionary patterns in a multiple sequence alignment (39). Second, the structure of the full-length *T. italicus* TagA protein was constructed by homology modeling using the program I-TASSER and the GREMLIN-derived structure of the *B. subtilis* enzyme as a template (40). Finally, because the coordinates of the core domain in *T. italicus* TagA are known, the final structure used in the MD simulations contains the experimentally derived coordinates of the core domain in the TagA^{FL*}:UDP–GlcNAc crystal structure and the coordinates of the CTT derived from TagA^{CM}. The program MODELLER was used to

combine the coordinates (residues 2–186 and 140–244 were combined using the coordinates of the $TagA^{FL*}$:UDP–GlcNAc crystal structure and $TagA^{CM}$, respectively) (46). The final merged coordinates were used in the MD simulations of the apoenzyme. The structure of the complex used in the simulations was obtained by positioning the ligand in a manner similar to that observed in the $TagA^{FL*}$:UDP–ManNAc crystal structure.

Explicit-solvent MD simulations of the generated model were run using GROMACS 2021 using the CHARMM36m force field (74, 75). The system was solvated with TIP3P, energy minimized, and equilibrated in two stages (NVT followed by NPT). For the complex model, the equilibration was further divided into two steps with both macromolecules held *via* position restraints and then restraints on the protein alone. 1.0 μs production MD simulations were then performed on both the apo and complexed systems. For the complex, pairwise distances between the protein and UDP–ManNAc as observed in the crystal structure were restrained using three pull groups (N39 terminal amine nitrogen to UDP–ManNAc sugar 3′ oxygen, E41 C_α to UDP–ManNAc *N*-acetyl oxygen and Y137 C_ε to UDP–ManNAc uracil C5' carbon). Global backbone RMSD calculated over the course of the simulations revealed that they reached convergence after 20 ns (apo simulation) or 200 ns (complex simulation) of NPT simulation. Clustering was performed using the GROMACS cluster module with the GROMOS method and a 0.9 Å cutoff threshold where the least squares fit was calculated for backbone atoms of the core domain only (residues 1–195) (47).

Enzyme-coupled activity assay

The TagA enzyme-coupled assay was described previously and is briefly summarized here (19, 36). The reaction was analyzed using a paired ion chromatography HPLC assay and under the following conditions: 0.2 μM TagA; 100 μM lipid-α analog, 100 μM UDP–GlcNAc, 3 μM UDP-*N*-acetyl-D-glucosamine 2-epimerase (MnaA), 50 mM Tris–HCl, pH 7.8, and 250 mM NaCl (19, 76). Reactions were preincubated without TagA enzyme for 15 min at 30 °C. Then the TagA enzyme was added and incubated for 40 min at 30 °C before quenching with 4 M urea. The end-point reactions were separated with an Agilent Eclipse Plus C18 column using a buffer gradient of 65% buffer A (15 mM sodium phosphate, 10 mM tetrabutylammonium hydrogen sulfate, and pH 7.0) and 35% buffer B (35 mM sodium phosphate, 10 mM tetrabutylammonium hydrogen sulfate, 30% acetonitrile, and pH 7.0) to 100% buffer B over 4 min. UDP–ManNAc and UDP elution peaks were monitored at 262 nm and integrated to determine turnover rate.

Native mass spectrometry

TagA was exchanged into 500 mM ammonium acetate buffer, pH 7.4, and incubated at 4 °C overnight, then diluted to a concentration of 10 μM with varying concentrations of DDM (0–400 μM). Nanoelectrospray ionization native MS analysis was performed on a G2-Si Synapt (Waters Corporation) in negative ion mode with the following parameters: 2000 to 8000 *m/z*, capillary: 1 kV; source temperature: 50 °C; desolvation temperature: 150 °C; sampling cone: 20 V; trap collision energy: 6 to 50 V; and transfer collision energy: 2 V. The collision energies in the "trap" region of the mass spectrometer used to remove DDM adducts from the protein were: 6, 20, 20, 30, and 50 V for 0 μM, 10 μM, 100 μM, 200 μM, and 400 μM DDM, respectively. The *m/z* raw data were centered (15 channels, 80% centroid top) in MassLynx (version 4.1; Waters Corporation) prior to input and deconvolution in UniDec (77). The deconvolution parameters are as follows: peak full-width half-maximum: 50 Th; peak shape function: gaussian; beta: 1000; charge smooth width: 1.0; point smooth width: 100; and maximum number of iterations: 100.

Data availability

UniProt code: *T. italicus* TagA, D3T4E0; *B. subtilis* MnaA, P39131.

PDB code: 7MPK; 7N41.

Biological Magnetic Resonance Data Bank code: 51022.

Supporting information—This article contains supporting information.

Acknowledgments—We thank Michael Collazo for assistance in crystallography with crystallization, data processing, and refinement. This research was supported by the resources at the University of California, Los Angeles—Department of Energy (UCLA-DOE) Institute's X-ray and EM Structure Determination Core and the Macromolecular NMR Core, which is supported by the US DOE Office of Science, Office of Biological and Environmental Research program, under award number DE-FC02-02ER63421. We also acknowledge NIH shared instrumentation grants S10OD025073, S10OD016336, and S10OD018504. The X-ray diffraction experiments were performed at the Northeastern Collaborative Access Team beamlines, which are funded by the National Institute of General Medical Sciences from the NIH (P30 GM124165). This research used resources of the Advanced Photon Source, a US DOE Office of Science User Facility operated for the DOE Office of Science by Argonne National Laboratory under contract no. DE-AC02-06CH11357.

Author contributions—O. E. M., B. J. M., J. A. L., and R. T. C. conceptualization; O. E. M., B. J. M., A. K. G., D. P. T., and R. T. C. methodology; O. E. M., B. J. M., D. C., and M. L. P. software; O. E. M., B. J. M., D. C., M. L. P., J. A. L., and R. T. C. validation; O. E. M., B. J. M., and M. L. P. formal analysis; O. E. M., B. J. M., A. K. G., D. P. T., and D. C. investigation; S.-W. Y., M. M. M., X. C., and M. E. J. resources; O. E. M., B. J. M., and R. T. C. writing–original draft; O. E. M., B. J. M., A. K. G., X. C., J. A. L., and R. T. C. writing–review & editing; O. E. M., B. J. M., and D. P. T. visualization; R. T. C. project administration; D. C., J. A. L., and R. T. C. funding acquisition.

Funding and additional information—This research was supported by grants from the National Institutes of Health (NIH) (AI52217 [to R. T. C.]; GM103479 [to J. A. L.]). O. E. M. was supported by the NIH Chemistry-Biology Training Fellowship (T32GM008496). The content is solely the responsibility of the authors and does not

necessarily represent the official views of the National Institutes of Health.

Conflict of interest—The authors declare that they have no conflicts of interest with the contents of this article.

Abbreviations—The abbreviations used are: CTT, C-terminal tail; DDM, *n*-dodecyl-β-D-maltoside; DOE, Department of Energy; GlcNAc, *N*-acetyl-D-glucosamine; GREMLIN, Generative Regularized Models of Proteins; GroP, glycerol-3-phosphate; GT, glycosyltransferase; HSQC, heteronuclear single quantum coherence; I-TASSER, Iterative Threading Assembly Refinement; ManNAc, *N*-acetyl-D-mannosamine; MD, molecular dynamics; NIH, National Institutes of Health; PDB, Protein Data Bank; SE–AUC, sedimentation equilibrium–analytical ultracentrifugation; TROSY, transverse relaxation optimized spectroscopy; Und, undecaprenyl; WTA, wall teichoic acid.

References

1. Formstone, A., Carballido-Lopez, R., Noirot, P., Errington, J., and Scheffers, D. J. (2008) Localization and interactions of teichoic acid synthetic enzymes in Bacillus subtilis. *J. Bacteriol.* **190**, 1812–1821
2. Brown, S., Santa Maria, J. P., and Walker, S. (2013) Wall teichoic acids of gram-positive bacteria. *Annu. Rev. Microbiol.* **67**, 313–336
3. Sewell, E. W. C., and Brown, E. D. (2014) Taking aim at wall teichoic acid synthesis: New biology and new leads for antibiotics. *J. Antibiot.* **67**, 43–51
4. Schade, J., and Weidenmaier, C. (2016) Cell wall glycopolymers of Firmicutes and their role as nonprotein adhesins. *FEBS Lett.* **590**, 3758–3771
5. Swoboda, J. G., Campbell, J., Meredith, T. C., and Walker, S. (2010) Wall teichoic acid function, biosynthesis, and inhibition. *Chembiochem* **11**, 35–45
6. Brown, S., Xia, G., Luhachack, L. G., Campbell, J., Meredith, T. C., Chen, C., Winstel, V., Gekeler, C., Irazoqui, J. E., Peschel, A., and Walker, S. (2012) Methicillin resistance in Staphylococcus aureus requires glycosylated wall teichoic acids. *Proc. Natl. Acad. Sci. U. S. A.* **109**, 18909–18914
7. Farha, M. A., Leung, A., Sewell, E. W., D'Elia, M. A., Allison, S. E., Ejim, L., Pereira, P. M., Pinho, M. G., Wright, G. D., and Brown, E. D. (2013) Inhibition of WTA synthesis blocks the cooperative action of PBPs and sensitizes MRSA to β-lactams. *ACS Chem. Biol.* **8**, 226–233
8. Heckels, J. E., Lambert, P. A., and Baddiley, J. (1977) Binding of magnesium ions to cell walls of Bacillus subtilis W23 containing teichoic acid or teichuronic acid. *Biochem. J.* **162**, 359–365
9. Holland, L. M., Conlon, B., and O'Gara, J. P. (2011) Mutation of tagO reveals an essential role for wall teichoic acids in Staphylococcus epidermidis biofilm development. *Microbiology* **157**, 408–418
10. Misawa, Y., Kelley, K. A., Wang, X., Wang, L., Park, W. B., Birtel, J., Saslowsky, D., and Lee, J. C. (2015) Staphylococcus aureus colonization of the mouse gastrointestinal tract is modulated by wall teichoic acid, capsule, and surface proteins. *PLoS Pathog.* **11**, e1005061
11. Peschel, A., Vuong, C., Otto, M., and GöTz, F. (2000) The d-alanine residues of Staphylococcus aureus teichoic acids alter the susceptibility to vancomycin and the activity of autolytic enzymes. *Antimicrob. Agents Chemother.* **44**, 2845–2847
12. Schirner, K., Marles-Wright, J., Lewis, R. J., and Errington, J. (2009) Distinct and essential morphogenic functions for wall- and lipo-teichoic acids in Bacillus subtilis. *EMBO J.* **28**, 830–842
13. Wanner, S., Schade, J., Keinhörster, D., Weller, N., George, S. E., Kull, L., Bauer, J., Grau, T., Winstel, V., Stoy, H., Kretschmer, D., Kolata, J., Wolz, C., Bröker, B. M., and Weidenmaier, C. (2017) Wall teichoic acids mediate increased virulence in Staphylococcus aureus. *Nat. Microbiol.* **2**, 16257
14. Kojima, N., Araki, Y., and Ito, E. (1985) Structure of the linkage units between ribitol teichoic acids and peptidoglycan. *J. Bacteriol.* **161**, 299–306
15. Neuhaus, F. C., and Baddiley, J. (2003) A continuum of anionic charge: Structures and functions of D-alanyl-teichoic acids in gram-positive bacteria. *Microbiol. Mol. Biol. Rev.* **67**, 686–723
16. Yokoyama, K., Miyashita, T., Araki, Y., and Ito, E. (1986) Structure and functions of linkage unit intermediates in the biosynthesis of ribitol teichoic acids in Staphylococcus aureus H and Bacillus subtilis W23. *Eur. J. Biochem.* **161**, 479–489
17. Soldo, B., Karamata, D., and Lazarevic, V. (2002) tagO is involved in the synthesis of all anionic cell-wall polymers in Bacillus subtilis 168. *Microbiology* **148**, 2079–2087
18. Ginsberg, C., Zhang, Y.-H., Yuan, Y., and Walker, S. (2006) *In vitro* reconstitution of two essential steps in wall teichoic acid biosynthesis. *ACS Chem. Biol.* **1**, 25–28
19. Zhang, Y. H., Ginsberg, C., Yuan, Y., and Walker, S. (2006) Acceptor substrate selectivity and kinetic mechanism of Bacillus subtilis TagA. *Biochemistry* **45**, 10895–10904
20. D'Elia, M. A., Henderson, J. A., Beveridge, T. J., Heinrichs, D. E., and Brown, E. D. (2009) The N-acetylmannosamine transferase catalyzes the first committed step of teichoic acid assembly in Bacillus subtilis and Staphylococcus aureus. *J. Bacteriol.* **191**, 4030–4034
21. Bhavsar, A. P., Truant, R., and Brown, E. D. (2005) The TagB protein in Bacillus subtilis 168 is an intracellular peripheral membrane protein that can incorporate glycerol phosphate onto a membrane-bound acceptor *in vitro*. *J. Biol. Chem.* **280**, 36691–36700
22. Endl, J., Seidl, H. P., Fiedler, F., and Schleider, K. H. (1983) Chemical composition and structure of cell wall teichoic acids of staphylococci. *Arch. Microbiol.* **135**, 215–223
23. Naumova, I. B., Shashkov, A. S., Tul'Skaya, E. M., Streshinskaya, G. M., Kozlova, Y. I., Potekhina, N. V., Evtushenko, L. I., and Stackebrandt, E. (2001) Cell wall teichoic acids: Structural diversity, species specificity in the genus Nocardiopsis, and chemotaxonomic perspective. *FEMS Microbiol. Rev.* **25**, 269–283
24. Xia, G., Maier, L., Sanchez-Carballo, P., Li, M., Otto, M., Holst, O., and Peschel, A. (2010) Glycosylation of wall teichoic acid in Staphylococcus aureus by TarM. *J. Biol. Chem.* **285**, 13405–13415
25. Chen, L., Hou, W.-T., Fan, T., Liu, B., Pan, T., Li, Y.-H., Jiang, Y.-L., Wen, W., Chen, Z.-P., Sun, L., Zhou, C.-Z., and Chen, Y. (2020) Cryo-electron microscopy structure and transport mechanism of a wall teichoic acid ABC transporter. *mBio* **11**, e02749-19
26. Schirner, K., Stone, L. K., and Walker, S. (2011) ABC transporters required for export of wall teichoic acids do not discriminate between different main chain polymers. *ACS Chem. Biol.* **6**, 407–412
27. Gale, R. T., Li, F. K. K., Sun, T., Strynadka, N. C. J., and Brown, E. D. (2017) B. subtilis LytR-CpsA-Psr enzymes transfer wall teichoic acids from authentic lipid-linked substrates to mature peptidoglycan in vitro. *Cell Chem. Biol.* **24**, 1537–1546.e4
28. Campbell, J., Singh, A. K., Santa Maria, J. P., Kim, Y., Brown, S., Swoboda, J. G., Mylonakis, E., Wilkinson, B. J., and Walker, S. (2011) Synthetic lethal compound combinations reveal a fundamental connection between wall teichoic acid and peptidoglycan biosyntheses in Staphylococcus aureus. *ACS Chem. Biol.* **6**, 106–116
29. Finn, R. D., Coggill, P., Eberhardt, R. Y., Eddy, S. R., Mistry, J., Mitchell, A. L., Potter, S. C., Punta, M., Qureshi, M., Sangrador-Vegas, A., Salazar, G. A., Tate, J., and Bateman, A. (2016) The Pfam protein families database: Towards a more sustainable future. *Nucleic Acids Res.* **44**, D279–D285
30. Haft, R. F., Wessels, M. R., Mebane, M. F., Conaty, N., and Rubens, C. E. (1996) Characterization of cpsF and its product CMP-N-acetylneuraminic acid synthetase, a group B streptococcal enzyme that can function in K1 capsular polysaccharide biosynthesis in Escherichia coli. *Mol. Microbiol.* **19**, 555–563
31. Rahman, A., Barr, K., and Rick, P. D. (2001) Identification of the structural gene for the TDP-Fuc4NAc:Lipid II Fuc4NAc transferase involved in synthesis of enterobacterial common antigen in Escherichia coli K-12. *J. Bacteriol.* **183**, 6509–6516
32. Cantarel, B. L., Coutinho, P. M., Rancurel, C., Bernard, T., Lombard, V., and Henrissat, B. (2009) The carbohydrate-active EnZymes database (CAZy): An expert resource for glycogenomics. *Nucleic Acids Res.* **37**, D233–D238

33. Liu, J., and Mushegian, A. (2003) Three monophyletic superfamilies account for the majority of the known glycosyltransferases. *Protein Sci.* **12**, 1418–1431
34. Campbell, J. A., Davies, G. J., Bulone, V., and Henrissat, B. (1997) A classification of nucleotide-diphospho-sugar glycosyltransferases based on amino acid sequence similarities. *Biochem. J.* **326**, 929–939
35. Coutinho, P. M., Deleury, E., Davies, G. J., and Henrissat, B. (2003) An evolving hierarchical family classification for glycosyltransferases. *J. Mol. Biol.* **328**, 307–317
36. Kattke, M. D., Gosschalk, J. E., Martinez, O. E., Kumar, G., Gale, R. T., Cascio, D., Sawaya, M. R., Philips, M., Brown, E. D., and Clubb, R. T. (2019) Structure and mechanism of TagA, a novel membrane-associated glycosyltransferase that produces wall teichoic acids in pathogenic bacteria. *PLoS Pathog.* **15**, e1007723
37. [dataset] Kattke, M. D., Gosschalk, J. E., Martinez, O. E., Kumar, G., Gale, R. T., Cascio, D., Sawaya, M. R., Philips, M., Brown, E. D., and Clubb, R. T. (2019) Structure and mechanism of TagA, a novel membrane-associated glycosyltransferase that produces wall teichoic acids in pathogenic bacteria. *Protein Data Bank*, 5WB4 and 5WFG
38. McRorie, D. K., and Voelker, P. J. (2004) *Self-Associating Systems in the Analytical Ultracentrifuge*. Beckman Instruments, Inc, Brea, California
39. Ovchinnikov, S., Kamisetty, H., and Baker, D. (2014) Robust and accurate prediction of residue–residue interactions across protein interfaces using evolutionary information. *Elife* **3**, e02030
40. Roy, A., Kucukural, A., and Zhang, Y. (2010) I-TASSER: A unified platform for automated protein structure and function prediction. *Nat. Protoc.* **5**, 725–738
41. Wang, S., Sun, S., Li, Z., Zhang, R., and Xu, J. (2017) Accurate de novo prediction of protein contact map by ultra-deep learning model. *PLoS Comput. Biol.* **13**, e1005324
42. Xu, J. (2019) Distance-based protein folding powered by deep learning. *Proc. Natl. Acad. Sci. U. S. A.* **116**, 16856–16865
43. Strong, M., Sawaya, M. R., Wang, S., Phillips, M., Cascio, D., and Eisenberg, D. (2006) Toward the structural genomics of complexes: Crystal structure of a PE/PPE protein complex from Mycobacterium tuberculosis. *Proc. Natl. Acad. Sci. U. S. A.* **103**, 8060–8065
44. Duarte, J. M., Srebniak, A., Schärer, M. A., and Capitani, G. (2012) Protein interface classification by evolutionary analysis. *BMC Bioinformatics* **13**, 334
45. Terwilliger, T. C., Grosse-Kunstleve, R. W., Afonine, P. V., Moriarty, N. W., Adams, P. D., Read, R. J., Zwart, P. H., and Hung, L.-W. (2008) Iterative-build OMIT maps: Map improvement by iterative model building and refinement without model bias. *Acta Crystallogr. D Biol. Crystallogr.* **64**, 515–524
46. Webb, B., and Sali, A. (2016) Comparative protein structure modeling using MODELLER. *Curr. Protoc. Bioinformatics* **54**, 5.6.1–5.6.37
47. Daura, X., Gademann, K., Jaun, B., Seebach, D., Gunsteren, W. F.v., and Mark, A. E. (2004) Peptide folding: When simulation meets experiment. *Angew. Chem. Int. Ed. Engl.* **38**, 236–240
48. Laganowsky, A., Reading, E., Hopper, J. T. S., and Robinson, C. V. (2013) Mass spectrometry of intact membrane protein complexes. *Nat. Protoc.* **8**, 639–651
49. Le Maire, M., Champeil, P., and Møller, J. V. (2000) Interaction of membrane proteins and lipids with solubilizing detergents. *Biochim. Biophys. Acta Biomembr.* **1508**, 86–111
50. Privé, G. G. (2009) Lipopeptide detergents for membrane protein studies. *Curr. Opin. Struct. Biol.* **19**, 379–385
51. Lairson, L. L., Henrissat, B., Davies, G. J., and Withers, S. G. (2008) Glycosyltransferases: Structures, functions, and mechanisms. *Annu. Rev. Biochem.* **77**, 521–555
52. Zhang, H., Zhu, F., Yang, T., Ding, L., Zhou, M., Li, J., Haslam, S. M., Dell, A., Erlandsen, H., and Wu, H. (2014) The highly conserved domain of unknown function 1792 has a distinct glycosyltransferase fold. *Nat. Commun.* **5**, 4339
53. Zhang, H., Zhou, M., Yang, T., Haslam, S. M., Dell, A., and Wu, H. (2016) New helical binding domain mediates a glycosyltransferase activity of a bifunctional protein. *J. Biol. Chem.* **291**, 22106–22117
54. Holm, L. (2020) DALI and the persistence of protein shape. *Protein Sci.* **29**, 128–140
55. [dataset] Zhang, H., Zhu, F., Yang, T., Ding, L., Zhou, M., Li, J., Haslam, S. M., Dell, A., Erlandsen, H., and Wu, H. (2014) The highly conserved domain of unknown function 1792 has a distinct glycosyltransferase fold. *Protein Data Bank*, 4PHR
56. Marley, J., Lu, M., and Bracken, C. (2001) A method for efficient isotopic labeling of recombinant proteins. *J. Biomol. NMR* **20**, 71–75
57. Friedhoff, P., Gimadutdinow, O., Ruter, T., Wende, W., Urbanke, C., Thole, H., and Pingoud, A. (1994) A procedure for renaturation and purification of the extracellular Serratia marcescens nuclease from genetically engineered Escherichia coli. *Protein Expr. Purif.* **5**, 37–43
58. Salzmann, M., Wider, G., Pervushin, K., Senn, H., and Wüthrich, K. (1999) TROSY-type triple-resonance experiments for sequential NMR assignments of large proteins. *J. Am. Chem. Soc.* **121**, 844–848
59. Eletsky, A., Kienhöfer, A., and Pervushin, K. (2001) TROSY NMR with partially deuterated proteins. *J. Biomol. NMR* **20**, 177–180
60. Delaglio, F., Grzesiek, S., Vuister, G., Zhu, G., Pfeifer, J., and Bax, A. (1995) NMRPipe: A multidimensional spectral processing system based on UNIX pipes. *J. Biomol. NMR* **6**, 277–293
61. Keller, R. L. J. (2004) *The Computer Aided Resonance Assignment Tutorial*, 1st ed, CANTINA Verlag, Goldau, Switzerland
62. Shen, Y., and Bax, A. (2013) Protein backbone and sidechain torsion angles predicted from NMR chemical shifts using artificial neural networks. *J. Biomol. NMR* **56**, 227–241
63. Lee, D., Hilty, C., Wider, G., and Wüthrich, K. (2006) Effective rotational correlation times of proteins from NMR relaxation interference. *J. Magn. Reson.* **178**, 72–76
64. Robson, S. A., Dağ, Ç., Wu, H., and Ziarek, J. J. (2021) TRACT revisited: An algebraic solution for determining overall rotational correlation times from cross-correlated relaxation rates. *J. Biomol. NMR* **75**, 293–302
65. Hoogland, C., Gattiker, A., Duvaud, S. E., Wilkins, M. R., Appel, R. D., and Bairoch, A. (2005) *Protein Identification and Analysis Tools on the ExPASy Server*. Humana Press, Inc, Totowa, NJ: 571–607
66. Becerra, S. P., Kumar, A., Lewis, M. S., Widen, S. G., Abbotts, J., Karawya, E. M., Hughes, S. H., Shiloach, J., and Wilson, S. H. (1991) Protein-protein interactions of HIV-1 reverse transcriptase: Implication of central and C-terminal regions in subunit binding. *Biochemistry* **30**, 11707–11719
67. Ross, P. D., Howard, F. B., and Lewis, M. S. (1991) Thermodynamics of antiparallel hairpin-double helix equilibria in DNA oligonucleotides from equilibrium ultracentrifugation. *Biochemistry* **30**, 6269–6275
68. Muthana, M. M., Qu, J., Li, Y., Zhang, L., Yu, H., Ding, L., Malekan, H., and Chen, X. (2012) Efficient one-pot multienzyme synthesis of UDP-sugars using a promiscuous UDP-sugar pyrophosphorylase from Bifidobacterium longum (BLUSP). *Chem. Commun.* **48**, 2728
69. Kabsch, W. (2010) XDS. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 125–132
70. McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C., and Read, R. J. (2007) Phaser crystallographic software. *J. Appl. Crystallogr.* **40**, 658–674
71. Liebschner, D., Afonine, P. V., Baker, M. L., Bunkóczi, G., Chen, V. B., Croll, T. I., Hintze, B., Hung, L.-W., Jain, S., McCoy, A. J., Moriarty, N. W., Oeffner, R. D., Poon, B. K., Prisant, M. G., Read, R. J., *et al.* (2019) Macromolecular structure determination using X-rays, neutrons and electrons: Recent developments in Phenix. *Acta Crystallogr. D Struct. Biol.* **75**, 861–877
72. Emsley, P., and Cowtan, K. (2004) Coot: Model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.* **60**, 2126–2132
73. Smart, O. S., Womack, T. O., Flensburg, C., Keller, P., Paciorek, W., Sharff, A., Vonrhein, C., and Bricogne, G. (2012) Exploiting structure similarity in refinement: Automated NCS and target-structure restraints in BUSTER. *Acta Crystallogr. D Biol. Crystallogr.* **68**, 368–380

74. Abraham, M. J., Murtola, T., Schulz, R., Páll, S., Smith, J. C., Hess, B., and Lindahl, E. (2015) GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **1-2**, 19–25
75. Huang, J., Rauscher, S., Nawrocki, G., Ran, T., Feig, M., De Groot, B. L., Grubmüller, H., and Mackerell, A. D. (2017) CHARMM36m: An improved force field for folded and intrinsically disordered proteins. *Nat. Methods* **14**, 71–73
76. Soldo, B., Lazarevic, V., Pooley, H. M., and Karamata, D. (2002) Characterization of a Bacillus subtilis thermosensitive teichoic acid-deficient mutant: Gene mnaA (yvyH) encodes the UDP-N-acetylglucosamine 2-epimerase. *J. Bacteriol.* **184**, 4316–4320
77. Marty, M. T., Baldwin, A. J., Marklund, E. G., Hochberg, G. K. A., Benesch, J. L. P., and Robinson, C. V. (2015) Bayesian deconvolution of mass and ion mobility spectra: From binary interactions to polydisperse ensembles. *Anal. Chem.* **87**, 4370–4376

Orlando E. Martinez is a PhD student in the Department of Chemistry & Biochemistry at the University of California, Los Angeles. His research is focused on understanding molecular mechanisms used by pathogenic bacteria to synthesize cell wall-associated glycopolymers. The results of his research could facilitate the discovery of new therapeutics to treat infections caused by multi-drug resistant bacterial pathogens.