Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

HHS Public Access

Author manuscript

Structure. Author manuscript; available in PMC 2023 October 23.

Published in final edited form as:

Structure. 2023 October 05; 31(10): 1166–1173.e6. doi:10.1016/j.str.2023.07.010.

Insights into substrate coordination and glycosyl transfer of poplar cellulose synthase-8

Preeti Verma[1], **Albert L. Kwansa**[2], **Ruoya Ho**[1,3], **Yaroslava G. Yingling**[2], **Jochen Zimmer**[1,3,4,*]

[1]Department of Molecular Physiology and Biological Physics, University of Virginia, Charlottesville, VA 22903, USA

[2]Department of Materials Science and Engineering, North Carolina State University, Raleigh, NC 27695, USA

[3]Howard Hughes Medical Institute, University of Virginia, Charlottesville, VA 22903, USA

[4]Lead contact

SUMMARY

Cellulose is an abundant cell wall component of land plants. It is synthesized from UDP-activated glucose molecules by cellulose synthase, a membrane-integrated processive glycosyltransferase. Cellulose synthase couples the elongation of the cellulose polymer with its translocation across the plasma membrane. Here, we present substrate- and product-bound cryogenic electron microscopy structures of the homotrimeric cellulose synthase isoform-8 (CesA8) from hybrid aspen (poplar). UDP-glucose binds to a conserved catalytic pocket adjacent to the entrance to a transmembrane channel. The substrate's glucosyl unit is coordinated by conserved residues of the glycosyltransferase domain and amphipathic interface helices. Site-directed mutagenesis of a conserved gating loop capping the active site reveals its critical function for catalytic activity. Molecular dynamics simulations reveal prolonged interactions of the gating loop with the substrate molecule, particularly across its central conserved region. These transient interactions likely facilitate the proper positioning of the substrate molecule for glycosyl transfer and cellulose translocation.

Graphical Abstract

This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

*Correspondence: jz3x@virginia.edu.

AUTHOR CONTRIBUTIONS

P.V. performed all mutagenesis experiments. R.H. prepared cryo grid samples and collected all EM data. J.Z. processed the data and built the models. A.L.K. and Y.G.Y. performed the MD analysis. All authors evaluated the data. J.Z. wrote the initial manuscript and all authors edited the manuscript.

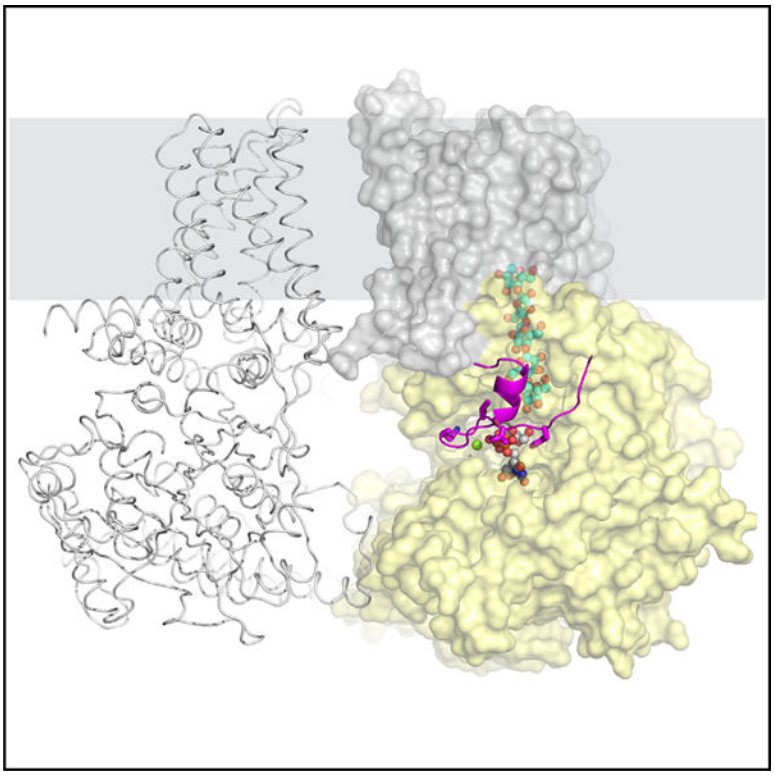
SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.str.2023.07.010.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript



In brief

Verma et al. determined the cryo-electron microscopy structure of UDP-glucose-bound poplar cellulose synthase-8 that provides insights into substrate selectivity and glycosyl transfer. Site-directed mutagenesis of a conserved but structurally flexible gating loop reveals residues critical for cellulose biosynthesis. Molecular dynamics simulations of substrate-bound cellulose synthase-8 reveal direct gating loop-substrate interactions.

INTRODUCTION

Cellulose is an abundant biopolymer that is produced primarily by land plants as a structural cell wall component. Because plants produce cellulose from photosynthetically synthesized glucose molecules, the polysaccharide is a major atmospheric carbon dioxide sink as well as a significant renewable energy resource.[1]

Cellulose's glucosyl units are connected via β-(1,4)-glycosidic linkages that enable an approximately 180° rotation of neighboring sugar units within the polymer.[2] The resulting amphipathic polysaccharide can be organized into cable-like fibrillar structures, so-called cellulose micro- and macrofibrils, that are spun around the cell as a load-bearing wall component.[3] Cellulose is synthesized from UDP-activated glucose (UDP-Glc) by cellulose synthase (CesA), a membrane-integrated processive family-2 glycosyltransferase (GT).[4,5]

CesA catalyzes glucosyl transfer from UDP-Glc (the donor sugar) to the C4 hydroxyl group at the non-reducing end of the nascent cellulose polymer (the acceptor). Following

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

chain elongation, CesA also facilitates cellulose translocation across the plasma membrane through a pore formed by its own transmembrane (TM) segment. To couple cellulose synthesis with secretion, CesA's catalytic GT domain packs against a channel-forming TM region via three conserved amphipathic interface helices (IF1–3).[6,7]

Cellulose biosynthesis is evolutionarily conserved, with homologous pathways found in prokaryotes, oomycetes, and some animals. Previous work on bacterial cellulose biosynthetic systems from *Gluconacetobacter xylinus* (formerly *Acetobacter xylinum*),[8,9] *Rhodobacter sphaeroides*,[10] and *Escherichia coli* (*E. coli*)[11] provided detailed insights into the reaction mechanism and enzyme regulation,[12–15] as well as cellulose secretion,[7] assembly,[16,17] and modification.[18] Further, recent cryogenic electron microscopy (cryo-EM) studies on trimeric plant CesA complexes confirmed an evolutionarily conserved enzyme architecture, in support of an equally conserved catalytic reaction mechanism.[19,20]

To delineate principles underlying substrate selectivity and catalysis, we determined cryo-EM structures of the full-length poplar CesA isoform-8 (CesA8) bound to either UDP-Glc or the product and competitive inhibitor UDP. The obtained complexes demonstrate substrate binding to a conserved pocket at the interface between the enzyme's cytosolic catalytic domain and TM region. However, in contrast to the bacterial homolog BcsA where a conserved ''gating loop'' stabilizes UDP-Glc at the active site,[7] this loop is disordered in the substrate-bound CesA8 complex. Extensive mutagenesis analyses of the loop's conserved FxVTxK motif in *Rhodobacter sphaeroides* BcsA as well as poplar CesA8 underscore its importance for catalytic activity. All-atom molecular dynamics simulations indeed confirm prolonged interactions of the loop with the substrate molecule and suggest its role in proper positioning of the substrate for glycosyl transfer.

RESULTS

Cryo-EM analyses of nucleotide-bound poplar CesA8

Poplar CesA8 was expressed in Sf9 insect cells and purified in the detergents lauryl maltose neopentyl glycol/cholesteryl hemisuccinate and glyco-diosgenin as previously described[19] and summarized in the STAR methods. Under these conditions, the enzyme is catalytically active, synthesizing cellulose *in vitro* in the presence of UDP-Glc and magnesium ions. Alongside cellulose, CesA generates UDP as a second reaction product of the glycosyl transfer reaction. Previous studies on bacterial and plant cellulose synthases as well as other related family-2 GTs demonstrated that UDP competitively inhibits the enzymes, due to its interactions with the catalytic pocket.[10,21,22] This observation has been exploited to obtain UDP-inhibited structures of hyaluronan and cellulose synthases.[14,23]

We determined UDP and UDP-Glc-bound poplar CesA8 structures by incubating the purified enzyme with 5 mM of nucleotide and 20 mM $MgCl_2$ prior to cryo grid preparation (see STAR methods, Figure S1). The ligand-bound CesA complexes were imaged and processed as described before[19] (Figures S1 and S2; Table S1). Overall, the trimeric organization of CesA8 is preserved in UDP and UDP-Glc-bound states, suggesting that the complex indeed represents a biologically functional unit (Figure 1A). Within the resolution limits of our cryo-EM maps (approximately 3.5 Å), the UDP moiety adopts

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

the same binding pose in the UDP-only and UDP-Glc-bound states (Figure S3A); hence, the following discussion focuses on the substrate-bound conformation. The UDP-Glc-bound structure is very similar to the previously solved apo CesA8 structure,[19] with a root-mean-square deviation between Cα atoms of 0.8 Å (Figure 1B). Each CesA8 protomer also contains a nascent cellulose polymer within the TM channel. The polymer's first five glucosyl units, starting at the non-reducing end near the catalytic pocket, are sufficiently well ordered to allow modeling (Figure S2). As described previously,[19] the terminal acceptor glucosyl unit rests next to Trp718 of the conserved QxxRW motif, right above the substrate binding pocket (Figure 1C).

CesA8 positions the donor sugar beneath the acceptor glucosyl unit

CesA8's GT domain forms a classical GT-A fold with a central mixed β-sheet surrounded by α-helices.[24] The bound substrate molecule is coordinated by conserved residues distributed throughout the GT domain[6] (Figure 1C). First, the substrate's uridine group is sandwiched between Glu265 and Lys436 and fits into a groove created by Ser258 and Val260 of the conserved STVDP motif belonging to the first β-strand of the GT-A fold. Second, Asp294 of the invariant DDG motif terminating β-strand #2 is in hydrogen bond distance to the Nε ring nitrogen of the uracil moiety. Third, the conserved DxD motif (Asp460 and Asp462), following β-strand #5, contributes to the coordination of a magnesium cation, which is also in contact with the substrate's β-phosphate. Additionally, the nucleotide's diphosphate group interacts with Arg717 of the QxxRW motif originating from IF-2 (Figure 1C).

The donor sugar fits into a polar pocket directly underneath the acceptor glucosyl unit of the nascent cellulose chain. This pocket is proximal to the water-membrane interface formed from IF-2, the finger helix that is N-terminally capped with the invariant VTED motif (residues 673 to 676), as well as the backbone of Val529, Gly530, and Thr531 belonging to the conserved YVGTG motif (Figure 1C). Potential hydrogen bond donors and acceptors from protein side chain and back-bone regions surround the donor glucosyl unit. However, all observed distances to the donor's hydroxyl groups exceed 3.5 Å, suggesting that the substrate molecule is not fully inserted into the catalytic pocket. Accordingly, the distance between the acceptor's C4 hydroxyl and the donor's C1 carbon exceeds 5.5 Å (Figure 1C). The observed substrate coordination is consistent with interactions delineated for bacterial BcsA bound to a non-hydrolysable UDP-Glc phosphonate analog[7] (Figure 1D), as well as UDP-N-acetylglucosamine bound to chitin and hyaluronan synthases[23,25,26] (Figures S3B–S3D).

The flexible gating loop is required for catalytic activity

CesA8's IF3 is connected to TM helix 5 via a ~20 residue long cytosolic gating loop that runs roughly across the opening of the catalytic pocket. The gating loop contains a conserved FxVTxK motif but is not resolved in all cryo-EM maps of CesAs, most likely due to conformational flexibility (Figure 2A).

Crystallographic analyses of *Rhodobacter sphaeroides* BcsA revealed different conformations of the gating loop.[6,14] The loop retracts from the catalytic pocket in a

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

nucleotide-free state and inserts into it in the presence of either UDP or a substrate analog. In the inserted state, the conserved FxVTxK motif contacts the UDP moiety, thereby likely stabilizing it at the active site. In addition to substrate stabilization, gating loop insertion into the catalytic pocket also facilitates translocation of the nascent polysaccharide between elongation steps.[7] Unlike for BcsA, nucleotide binding to CesA8 does not stabilize the gating loop in a similar manner (Figure 2A).

To probe the functional significance of the gating loop's FxVTxK motif, we performed site-directed mutagenesis and *in vitro* functional analyses of the bacterial BcsA and poplar CesA8 enzymes. For BcsA, replacing Phe503 of the FxVTxK motif with Ala or the bulky hydrophobic residue Ile abolishes catalytic activity (Figures 2B and S4). Substituting the conserved Val505 residue with Ala or Leu dramatically reduces catalytic activity to about 20% and 10%, respectively, relative to the wild-type enzyme. A drastic reduction is observed when the following Thr506 residue is replaced with Ala, yet its substitution with Ser retains about 60% relative catalytic activity. The Lys508 residue at the C terminus of the FxVTxK motif is also critical for function; neither an Ala nor an Arg residue at this position supports enzymatic activity (Figure 2B).

Upon insertion into the catalytic pocket, Phe503 of BcsA's gating loop forms cation-π interactions with Arg382 of the conserved QxxRW motif located in IF2 at the cytosolic water-lipid interface.[14] In this position, Arg382 forms a salt bridge with the substrate's diphosphate group, as also observed in CesA8 (Figures 1D and 2A). This residue is critical for catalytic activity as its substitution with Ala renders BcsA inactive, while its substitution with Phe retains about 16% activity (Figure 2B).

A similar analysis was performed of CesA8's FxVTxK motif. Here, we replaced Phe851 with Ile, V853 with Leu, Thr854 with Ser or Ala, and Lys856 with Arg. Due to increased background readings in membrane vesicles, the mutant enzymes were purified and analyzed for catalytic activity in a micelle-solubilized state, as previously described.[19] Of the generated mutants, only the V853L and T854S substitutions retain about 30% catalytic activity, relative to the wild-type enzyme. Further, as observed for BcsA, replacing Arg717 of CesA8's QxxRW motif with Ala renders the enzyme catalytically inactive (Figure 2C).

Molecular dynamics simulations reveal persistent gating loop-substrate interactions

To probe the interactions between CesA8's gating loop and a nucleotide at the active site, we examined the dynamics of the gating loop in the presence of UDP and UDP-Glc. The initial conformation of the loop was modeled based on the loop's position in the substrate-bound BcsA crystal structure (PDB: 5EIY) as well as weak discontinuous gating loop density observed at low contour levels in the CesA8 cryo-EM map (Figures 3 and S5). All-atom molecular dynamics simulations of UDP- and UDP-Glc-bound CesA8 monomers with a phospholipid bilayer, water, and NaCl were performed for 500 ns, in triplicate, as detailed in the STAR methods. Relative to its starting conformation, the gating loop remains in close contact with the nucleotide during the simulations, with greatest fluctuations observed for regions N and C terminal to the conserved FxVTxK motif (Figure 3A). For the UDP-bound case, notable contacts are observed for all residues of the FxVTxK motif, except for Thr852. For the UDP-Glc-bound case, a similar set of residues of the FxVTxK motif are found to

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

have influential interactions. In both cases, residues Phe851, Val853, Thr854, and Lys856 are predicted to interact the most with UDP or UDP-Glc, associated with relatively high mean contact scores. Phe851 and Val853 are positioned to surround the substrate's uracil group, while Thr854 interacts with the ligand's alpha phosphate (Figure 3B). Furthermore, the C-terminal Lys856 of the FxVTxK motif is predicted to interact primarily with the ligand's beta phosphate. The mean contact score metric employed in PyContact provides a measure of contact persistence over time, contact proximity, and contact area. Thus, mean contact scores associated with the gating loop residues and UDP/UDP-Glc/Mg^{2+} reveal relatively high contact persistence, proximity, and/or area, especially for the four conserved residues of the FxVTxK motif (Figure 3C; Tables S2 and S3).

DISCUSSION

CesA catalyzes multiple reactions: First, the formation of a linear β-(1,4)-glucan; second, the secretion of the polysaccharide into the extracellular milieu; and third, due to the self-assembly of CesAs into supramolecular complexes, the coalescence of cellulose polymers into fibrillar structures.[4,5]

Polymer secretion is achieved by closely associating the catalytic cytosolic GT domain with a channel-forming TM segment, with residues from both the regions contributing to donor and acceptor coordination.

The observed binding pose of the substrate UDP-Glc at CesA8's active site is consistent with other structures of processive and non-processive GT-A enzymes, including hyaluronan and chitin synthases[23,25,26] (Figures 4A and 4B). CesA coordinates UDP with invariant sequence motifs, primarily localized at the edge of its central GT-A β-sheet. The donor sugar, attached to UDP, is positioned in a hydrophilic pocket near the water-lipid interface. This pocket is created by the GT domain as well as amphipathic interface helices that establish the transition from the cytosolic to the TM region. The observed substrate coordination contrasts a recently reported UDP-Glc-bound crystal structure of the GT core of *Arabidopsis* CesA3,[27] which exhibits an unphysiological catalytic pocket.

Compared to UDP or substrate-bound states of bacterial BcsA, CesA8's conserved gating loop remains disordered in the substrate-bound cryo-EM structure. Weak map density ''above'' the entrance to the catalytic pocket (Figure S5A) indicates flexibility of the gating loop, as also observed for BcsA in nucleotide-free states.[14] Site-directed mutagenesis of the gating loop, however, demonstrates its profound importance for cellulose biosynthesis. We hypothesize that the gating loop inserts transiently into the catalytic pocket to position the substrate for glycosyl transfer and perhaps decreasing the distance between the donor and acceptor glucosyl units (Figure 4C). Substrate repositioning may also (a) favor protein interactions with the donor's hydroxyl groups to foster substrate specificity, and (b) facilitate conformational changes of the donor's pyranose ring to increase the reactivity of its C1 carbon.

Hyaluronan and chitin synthases also contain putative gating loops. The corresponding sequence is WGTR/KG instead of FxVTxK (Figure 4B). This motif is located in a loop

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

following IF3 near the active site. It is disordered in hyaluronan synthase structures or bridges a dimerization motif in chitin synthase (Figure 4A). Substrate-bound structures of these enzymes also suggest incomplete insertions of the substrate molecules into the catalytic pockets, based on the assumed acceptor binding site formed by the Trp of the conserved QxxRW motif. As for CesA, transient gating loop insertion could enforce a substrate conformation favorable for glycosyl transfer (Figure 4C).

Cellulose microfibrils are synthesized from CesA complexes (CSC) resembling 6-fold symmetric particles.[28] The repeat unit is likely represented by the trimeric cryo-EM structures obtained for poplar and cotton CesAs.[19,29] Within a CSC, multiple CesAs synthesize and secrete cellulose polymers to facilitate their alignment into a microfibril. It is currently unknown whether the individual CesA activities are coordinated. Our substrate-bound cryo-EM structures do not suggest allosteric regulation of the catalytic activities within a CesA trimer. However, we cannot exclude interprotomer crosstalk in the context of a fully assembled CSC, perhaps mediated by currently unresolved regions.

STAR★METHODS

RESOURCE AVAILABILITY

Lead contact—Further information and request for resources and reagents should be directed to and will be fulfilled by the lead contact, Jochen Zimmer (jz3x@virginia.edu).

Materials availability—This study did not generate new unique reagents. Plasmids generated in this study will be available upon request to the lead contact.

Data and code availability

- The coordinates of UDP-bound CesA8 and UDP-Glucose bound CesA8 have been deposited to the Protein Data Bank under accession numbers 8G27 and 8G2J. CryoEM maps of the UDP-bound CesA8 and UDP-Glucose bound CesA8 have been deposited in the Electron Microscopy Data Bank under the accession numbers 29678 and 29679 respectively. All the deposited data are publicly available as of the date of publication and accession numbers are also listed in the key resources table.
- This paper analyzes existing, publicly available data. These accession numbers for the datasets are listed in the key resources table.
- This paper does not report original code data.
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

Bacterial strains—*E. coli* Rosetta 2 (DE3) cells (Novagen) were used in this study for recombinant protein production. Cells were cultured in ZYP-5052 auto-induction media[42] supplemented with necessary antibiotics.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Cell lines—Hybrid Aspen (poplar) CesA8 was expressed in *Spodoptera frugiperda* (Sf9) cells infected with recombinant baculovirus (pACEBac1). SF9 cells were grown in ESF921 medium (Expression systems) at 27°C and cells were harvested 72 h after the infection when cell viability was dropped down to 70%.

METHOD DETAILS

Mutagenesis—The BcsA gating loop (F503A/I, V505A/L, T506A/S, K508A/R) and R382A/F mutants were generated from the wild-type (WT) construct as described earlier[7] via QuikChange mutagenesis. The WT type construct contains both the *Rhodobacter sphaeroides* cellulose synthase (bcs)A and bcsB genes expressed in pETDuet-1 vector, wherein BcsA was expressed with a C-terminal dodeca-histidine tag. Similarly, the hybrid aspen CesA8 mutants were generated from an existing WT pACEBac-CesA8 plasmid[19] by the same method. Not all mutations generated for BcsA were also introduced into CesA8. The CesA8 mutants were designed based on activity results obtained for the BcsA mutants.

All the oligonucleotides used in generating the BcsA and CesA8 mutants are provided in Table S5.

Expression and purification of CesA8 and its variants in SF9 insect cells—Expression of CesA8 and its mutants was performed in SF9 insect cells and the purification was carried out as described previously,[19] with some modifications. Briefly, the cell pellet from 1–1.5 L of culture was resuspended in the modified buffer [Buffer A: 20 mM Tris pH 7.5, 100 mM NaCl, 5 mM sodium phosphate, 5 mM sodium citrate, 1 mM TCEP] containing the detergents 1% lauryl maltose neopentyl glycol (LMNG, Anatrace) and 0.2% cholesteryl hemisuccinate (CHS, Anatrace), supplemented with protease inhibitor cocktail (0.8 μM Aprotinin, 5 μM E–64, 10 μM Leupeptin, 15 μM Bestatin-HCl, 100 μM AEBSF-HCl, 2 mM Benzamidine-HCl and 2.9 mM Pepstatin A). The entire mixture was lysed with 20 strokes of a tight fitting 100 mL Dounce homogenizer. After solubilization at 4°C for 1 h, the insoluble material was removed by centrifugation at 42,000 rpm for 45 min in a Ti45 rotor. All subsequent steps including the exchange of detergent from LMNG-CHS to 0.02% glycol-diosgenin (GDN, Anatrace) during the wash buffers were carried out as described before,[19] except for the omission of size exclusion chromatography. Instead, a dialysis step was performed wherein the Ni-NTA eluent was concentrated to approximately 5 mL using a 100-kDa spin concentrator (Millipore) and then dialyzed overnight against Buffer A containing 0.02% GDN. The following day, the dialyzed protein was concentrated to roughly 1.6–1.9 mg/mL and flash-frozen in small aliquots in liquid nitrogen and then stored at −80°C until further use.

Expression and preparation of inverted membrane vesicles of *Rhodobacter sphaeroides* BcsA-B complex and its variants—Expression and inverted membrane vesicle (IMV) preparation was carried out as described previously[6,10] for the wild-type BcsA-B complex. Briefly, the BcsA-B complex was expressed in *E. coli* Rosetta 2 cells in auto-induction medium. The 2 L cell pellet obtained for each protein construct was resuspended in Resuspension Buffer (RB) containing 20 mM Tris pH 7.2, 100 mM NaCl, and 10% Glycerol, supplemented with 1 mM phenylmethylsulfonyl fluoride (PMSF). Cells

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

were lysed in a microfluidizer followed by centrifugation at 12,500 rpm in a Beckman JA-20 rotor for 20 min. The supernatant was carefully recovered and roughly 25 mL was layered over a 1.8 M sucrose cushion made in RB buffer, followed by centrifugation at 42,000 rpm for 120 min in a Ti45 rotor. The dark brown ring formed at the sucrose cushion was carefully withdrawn, diluted 5-fold in the RB buffer, and finally, the membrane vesicles were sedimented via centrifugation at 42,000 rpm for 90 min in a Ti45 rotor. The pellet fraction was rinsed with RB buffer, resuspended in 1 mL RB, and homogenized using a no. 6 paintbrush followed by douncing in a 2 mL grinder. The vesicles were aliquoted in small quantities and flash frozen in liquid N_2 until further use. All the steps were performed at 4°C.

Normalization of protein levels for activity assays—To normalize the concentration of BcsA amongst the WT and mutant IMVs for performing the activity assays, freeze-thawed IMVs were treated with 2% SDS to aid in solubilization and proper migration during SDS-PAGE. Gel loading samples were made from these solubilized IMVs and western blotting against the BcsA-His tag was performed. The band intensities after Western blotting were used to calibrate the amounts of IMVs used for activity assays.

CesA8 protein concentrations for WT and its mutants were normalized based on quantitative SDS-PAGE analysis of the purified proteins. The gel was imaged after Coomassie-staining and further analyzed using LI-COR Odyssey Imaging system to compare the band intensities.

In vitro cellulose synthesis

For *Rhodobacter* BcsA-B, WT and mutant IMVs were used for assaying cellulose biosynthetic activity as described.[10] In the beginning, a time course was conducted using the wild-type BcsA-B IMVs to find an incubation time in the linear phase of product accumulation. Synthesis reaction was performed by incubating IMVs with 5 mM UDP-Glucose and 0.25 μCi of UDP-[^{3}H]-glucose, 30 μM cyclic-di-GMP (c-di-GMP) and 20 mM $MgCl_2$ in RB buffer lacking glycerol. Here, cellulose biosynthesis reaction was monitored for different time periods starting from 0 to 180 min. Aliquots were spotted onto Whatman-2MM chromatography paper and developed by descending paper chromatography using 60% ethanol. The polymer retained at the origin was quantified by scintillating counting. Based on this time course, all standard reactions were carried out at 37°C for 30 min. All reactions were performed in triplicate and error bars represent deviations from the means.

For CesA8, WT and mutants, the purified, micelle solubilized protein was used for synthesis reactions. The assays were performed as described previously.[19] Briefly, 5 μM of purified protein was incubated with 5 mM UDP-Glucose and 0.25 μCi of UDP-[^{3}H]-glucose in 20 mM $MgCl_2$ in buffer A containing 0.02% GDN. Then, the reactions were spotted onto Whatman-2MM paper, which was developed in an aqueous solution of 60% ethanol. The product retained at the origins was quantified by scintillation counting. As for assays with the bacterial enzyme, we determined a suitable assay time point during the linear product accumulation using the WT and mutants (V853L and T854S). Based on this time course,

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

all subsequent reactions were performed for 30 min at 30°C. To confirm the formation of authentic cellulose, enzymatic degradations of the *in vitro* synthesized glucan were performed using a commercial cellulase (endo-(1,4)-β-glucanase E-CELTR; Megazyme), wherein 5U of the enzyme was added at the beginning of synthesis reaction. All reactions were performed in triplicate and error bars represent deviations from the means.

Cryo-EM data collection

Ligand bound CesA8 complexes were generated by adding 20 mM $MgCl_2$ and 5 mM UDP or UDP-Glc to the purified protein followed by an incubation for 30 min on ice prior to cryo grid preparation. Cryo-EM analyses were performed as described before.[19] In short, 2.5 μL aliquot was applied to a glow-discharged (in the presence of amylamine) C-flat 400 mesh 1.2/1.3 holey carbon grid (Electron Microscopy Sciences), blotted with Vitrobot Mark IV (FEI, Thermo Fisher Scientific) with force 7 for 12–14 s at 4°C, 100% humidity, and flash frozen in liquid ethane. Grids were screened in-house for optimal ice thickness and particle distribution. High quality datasets were collected at the Brookhaven National Laboratory for BioMolecular Structure (LBMS) on a Titan Krios G3i equipped with an X-FEG electron source, Gatan K3 direct electron detector, and BioQuantum energy filter. Movies were collected in super-resolution mode with a pixel size of 0.4125 and 0.88 Å for the UDP-Glc and UDP-bound complexes, respectively. All movies were collected in counting mode at a magnification of 105,000k and 81K, respectively, and defocus range from −2.3 to −0.8 μm, with a total dose of 51e^{-}/$Å^2$.

Data processing

Cryo-EM data processing followed a similar workflow in cryoSPARC[31] as previously described.[19] Movies were full-frame motion corrected followed by CTF estimation. Exposures were manually curated based on estimated resolution, defocus, and drift as well as ice contamination.

Initial templates for particle picking were generated using ‘blob picking’ with inner and outer particle diameters of 200 and 350 Å. Particles were extracted with a box size of 600 pixels and Fourier cropped to a box size of 150 pixels. Following 2D classification, selected class averages were used for template-based particle picking. The new particle stack was inspected, extracted with 4-fold Fourier cropping, and classified in 2 and 3-dimensions. The best particles of the UDP-Glc-bound dataset were re-extracted using a 640 pixels box and Fourier cropped to a 320 pixels box. For the UDP complex, final particle stack was extracted at a box size of 400 pixels without cropping. Refinements followed standard non-uniform refinement with C3 symmetry.

Model building—PDB entry 6WLB was used as an initial model. The previous CesA8 trimer structure was rigid body docked into the EM map in Chimera[35] and manually adjusted in Coot.[32] UDP and UDP-alpha-D-Glc were placed and manually refined in Coot. The model was refined in phenix:refine[33] without imposing non-crystallographic symmtery (NCS). Coordinates and EM maps have been deposited at the Protein DataBank under accession codes 8G27 and 8G2J.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Molecular dynamics system construction—The MD workflow started from the cryo-EM structure of poplar CesA8 with a bound UDP, coordinated Mg^{2+}, and cellopentaose. MD simulations were performed with a monomeric CesA8 construct. Rather than a TM7 (932–958) of the same CesA8 monomer, this monomeric construct contained the TM7 of the adjacent CesA8 subunit to complete its TM channel architecture. Then, the unresolved gating loop was modeled by analogy to BcsA (PDB: 5EIY), representing an inserted state of the gating loop. Subsequently, the SWISS-MODEL web server was used to generate initial coordinates for the remaining residues of the gating loop (847–848 and 856–865); SWISS-MODEL employs the ProMod3 homology modeling engine, which uses a database of structural fragments derived from the Protein Data Bank or falls back to a Monte Carlo approach to generate coordinates.[43–45] A UDP-Glc-bound case was also prepared for MD simulation, based on the experimentally determined UDP-alpha-D-Glc-bound CesA8 structure described herein. The AMBER 2019 software package[36] was used for all subsequent system construction, force field implementation, and simulation tasks. The simulation environment was assembled around each initial model (CesA8 monomer with UDP/UDP-Glc, Mg^{2+}, and cellopentaose). Specifically, AMBER's packmol-memgen, with Packmol 18.169, was used to add a homogeneous dioleoylphosphatidylcholine (DOPC) phospholipid bilayer, water, and 0.15 M NaCl; unless indicated, default settings such as the padding distances for the lipid bilayer and water around the solute were used.[40,46] For each of the two cases (UDP- or UDP-Glc-bound), three independent systems were prepared by using a pseudo-random seed during the packmol-memgen system assembly stage, leading to a total of six modeled systems.

MD force fields—AMBER's tleap was then used to apply selected force fields (set of potential energy expressions, parameters, and structural libraries) to the constructed systems. The following force fields were used: ff14SB (CesA protein),[47] Lipid17 (DOPC),[36,48] GLYCAM06j (cellopentaose),[49] TIP3P model (water),[50] and the 12–6-4 Lennard-Jones (LJ) set of the Li-Merz monovalent and divalent ion parameters for TIP3P (Na^+, Cl^-, Mg^{2+}).[51–53] Base parameters for the uridine and phosphate groups of UDP were provided through AMBER's ''parm10.dat'', which includes the OL3 parameters for RNA.[54] Parameters for the glucosyl group of UDP-Glc were provided through GLYCAM06. The parameters describing the connection between the beta phosphate and the glucosyl group were assigned by analogy to GLYCAM06; specifically, one bond parameter, four angle parameters, and seven dihedral parameters were added based on methyl sulfate, ethyl sulfate, methoxy alkanes, and ether alcohols (Table S4). UDP parameters based on GLYCAM93 have been developed previously[55]; while this previous work was used as a general reference point, we sought to remain consistent with the newer GLYCAM06 and thus adopted comparable parameters directly from GLYCAM06. Furthermore, a modification was applied to the 12–6 LJ parameters of the hydroxyl hydrogen atom type ''HO'' as used in AMBER's ''all_modrna08.frcmod''[56] (Table S4); this was deemed necessary to avoid a known potential issue when using hydroxyl hydrogen atom types with Rmin and epsilon parameters of zero.

MD partial charges—For both UDP and UDP-Glc, the net charge was considered as −2.0 e, that is, a single protonation on the beta phosphate of UDP and no protonation of the

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

UDP-Glc phosphates. This selected protonation state of UDP is based on log(K_a) values that have been reported for adenosine diphosphate, ADP (7.02, 4.19, and 0.9 for single, double, and triple protonation of the phosphates) via ^{1}H NMR chemical shift data and non-linear regression.[57] The selected protonation state of UDP-Glc is based on pK_a values that have been reported for UDP-GlcNAc – an N-acetyl derivative of UDP-Glc (6.6 and 6.3 for the alpha and beta phosphates, respectively) via ^{31}P NMR chemical shift data[58]; thus, above these pKa values, one might expect a larger population of the phosphate-deprotonated UDP-Glc species. Partial charges for UDP and UDP-Glc (UDP moiety only) were obtained from project F-90 of the RESP ESP Charge Database (R.E.DD.B.)[59]; fragments 1 (‘‘POP’’, alpha and beta phosphates for UDP-Glc), 2 (‘‘P1’’, alpha phosphate for UDP), 3 (‘‘P1M’’, beta phosphate for UDP), and 47 (‘‘U5’’, uridine) of this project were used. Partial charges for the glucosyl group of UDP-Glc were obtained from a terminal alpha-D-glucose unit of GLYCAM06 (entry ‘‘0GA’’ in its library). However, for the carbon at the reducing end of this glucosyl group linked to the beta phosphate, a charge adjustment of +0.0102 was applied to this glucosyl ‘‘C1’’ atom to provide an integer net charge of −2.0 for the UDP-Glc molecule; within the modular or fragment-based framework of GLYCAM06 and related force fields, such charge adjustments at linking atoms have been used for carbohydrate derivatives, e.g., O-acetyl, O-methyl, O-sulfate, and N-sulfate modifications, and are applicable to other derivatives.

MD simulations—The MD simulation protocol employed is based on those used previously to simulate protein and lipid systems.[28,60–62] Specifically, a 10-stage protocol was used including up to 10,000 steps of energy minimization (5,000 steepest descent steps followed by conjugate gradient steps), gradual NVT heating to 300 K over 100 ps, NVT equilibration at 300 K for 200 ps, NPT equilibration at 300 K and 1 atm for 800 ps, and 500 ns of production MD at 300 K and 1 atm.[60] A cutoff of 1.0 nm, periodic boundary conditions (PBCs) in all directions, particle-mesh Ewald (PME),[63] and long-range Lennard-Jones correction[36] were used for all stages. The timestep was initially 1.0 fs but was increased to 2.0 fs during the last 700 ps of the NPT equilibration stages and during the NPT production stage. The SHAKE algorithm[64] was used to constrain bonds involving hydrogen atoms during all stages after the energy minimization stage. Step-wise-reduced harmonic positional restraints were used for all protein atoms (10, 5.0, 2.5, 1.0, 0.5, 0.1 kcal/mol/Å^2) and for the lipid head group phosphorus atoms (2.5, 2.5, 1.0, 0.5, 0.1, 0.0 kcal/mol/Å^2)[62] up to the end of the first 400 ps of the NPT equilibration, after which there were no restraints. The Langevin thermostat[65] (collision frequency = 1.0 ps^{-1}) and Berendsen barostat[66] (coupling constant = 1.0 ps, compressibility = 44.6*10^{-6} bar^{-1}, anisotropic scaling) were used, where applicable. Above, ‘‘NVT’’ and ‘‘NPT’’ refer to thermodynamic ensembles described by a constant number of particles-fixed volume-regulated temperature and constant number of particles-regulated pressure-regulated temperature, respectively. AMBER’s CPU-only PMEMD and single precision-fixed point/SPFP GPU-accelerated PMEMD programs were used for the energy minimization and subsequent simulation stages, respectively, where CPU = central processing unit and GPU = graphics processing unit.[67,68] These simulations were carried out with Exxact Corporation servers equipped with NVIDIA GeForce GTX GPUs. For each of the two cases (UDP- or UDP-Glc-bound), three independent MD simulations were conducted; further independence

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

from the system assembly stage was afforded by also using a pseudo-random seed for the Langevin thermostat during the simulations.

MD analysis—Contacts between residues of the CesA8 gating loop (847–869) and the ligand groups (uridine, alpha phosphate, beta phosphate, Mg^{2+}, and glucosyl group, as applicable) were analyzed using PyContact 1.0.4[37] with MDAnalysis 0.20.1.[38,39] Contact criteria included considering heavy atoms only and a contact distance cutoff of 0.5 nm.[37] Contact metrics were calculated and accumulated over atom-atom contacts to obtain residue-residue contact data, where ‘‘residue’’ here can refer to an amino acid residue or a ligand group as listed above. Contact metrics of interest included the mean contact score (based on a distance-weighted sigmoidal function), mean contact lifetime (ns), and total contact time (ns and %). The contact analysis was conducted using 1,000 evenly sampled frames from each simulation coordinate trajectory representing 500 ns of time. The contact metrics were then averaged over the three independent simulations for each case.

QUANTIFICATION AND STATISTICAL ANALYSIS

Average and standard deviation values were determined using AVERAGE and STDEV functions in Excel, and data plotting was performed using Microsoft Excel and GraphPad Prism 6.0.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGMENTS

The authors thank Balasubramanyam Chittoor for initial cloning of CesA8 mutants and purifications. We are grateful to the staff at the EM core facilities at the University of Virginia (MEMC, Kelly Dryden and Michael Purdy) as well as the Laboratory for Biomolecular Structure (LBMS) at Brookhaven National Laboratory (Liguo Wang). LBMS is supported by the DOE Office of Biological and Environmental Research (KP1607011). MD simulations by A.L.K. and Y.G.Y. and work on poplar CesA8 by R.H. and J.Z. were supported by the Center for Lignocellulose Structure and Formation, an Energy Frontier Research Center funded by the U.S. Department of Energy, Office of Science, Basic Energy Sciences (award DESC0001090). Work on bacterial BcsA performed by P.V. and J.Z. was supported by NIH grant R35GM144130 awarded to J.Z. J.Z. is an investigator of the Howard Hughes Medical Institute. This article is subject to HHMI’s Open Access to Publications policy. HHMI lab heads have previously granted a nonexclusive CC BY 4.0 license to the public and a sublicensable license to HHMI in their research articles. Pursuant to those licenses, the author-accepted manuscript of this article can be made freely available under a CC BY 4.0 license immediately upon publication.

REFERENCES

1. Carroll A, and Somerville C (2009). Cellulosic biofuels. Annu. Rev. Plant Biol 60, 165–182. 10.1146/annurev.arplant.043008.092125. [PubMed: 19014348]
2. Nishiyama Y, Sugiyama J, Chanzy H, and Langan P (2003). Crystal structure and hydrogen bonding system in cellulose I(alpha) from synchrotron X-ray and neutron fiber diffraction. J. Am. Chem. Soc 125, 14300–14306. 10.1021/ja037055w. [PubMed: 14624578]
3. Yang H, and Kubicki JD (2020). A density functional theory study on the shape of the primary cellulose microfibril in plants: effects of C6 exocyclic group conformation and H-bonding. Cellulose 27, 2389–2402.
4. Turner S, and Kumar M (2018). Cellulose synthase complex organization and cellulose microfibril structure. Philos. Trans. A Math. Phys. Eng. Sci 376, 20170048. 10.1098/rsta.2017.0048. [PubMed: 29277745]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

5. McNamara JT, Morgan JLW, and Zimmer J (2015). A molecular description of cellulose biosynthesis. Annu. Rev. Biochem 84, 895–921. [PubMed: 26034894]
6. Morgan JLW, Strumillo J, and Zimmer J (2013). Crystallographic snapshot of cellulose synthesis and membrane translocation. Nature 493, 181–186. [PubMed: 23222542]
7. Morgan JLW, McNamara JT, Fischer M, Rich J, Chen HM, Withers SG, and Zimmer J (2016). Observing cellulose biosynthesis and membrane translocation in crystallo. Nature 531, 329–334. 10.1038/nature16966. [PubMed: 26958837]
8. Brown RM, Willison JH, and Richardson CL (1976). Cellulose biosynthesis in Acetobacter xylinum: visualization of the site of synthesis and direct measurement of the in vivo process. Proc. Natl. Acad. Sci. USA 73, 4565–4569. [PubMed: 1070005]
9. Du J, Vepachedu V, Cho SH, Kumar M, and Nixon BT (2016). Structure of the Cellulose Synthase Complex of Gluconacetobacter hansenii at 23.4 A Resolution. PLoS One 11, e0155886. [PubMed: 27214134]
10. Omadjela O, Narahari A, Strumillo J, Mé lida, H., Mazur, O., Bulone, V., and Zimmer, J. (2013). BcsA and BcsB form the catalytically active core of bacterial cellulose synthase sufficient for in vitro cellulose synthesis. Proc. Natl. Acad. Sci. USA 110, 17856–17861. 10.1073/pnas.1314063110. [PubMed: 24127606]
11. Bokranz W, Wang X, Tscha€pe, H., and Rö mling, U. (2005). Expression of cellulose and curli fimbriae by Escherichia coli isolated from the gastrointestinal tract. J. Med. Microbiol 54, 1171–1182. [PubMed: 16278431]
12. Ross P, Weinhouse H, Aloni Y, Michaeli D, Weinberger-Ohana P, Mayer R, Braun S, de Vroom E, van der Marel GA, van Boom JH, and Benziman M (1987). Regulation of cellulose synthesis in Acetobacter xylinum by cyclic diguanylic acid. Nature 325, 279–281. [PubMed: 18990795]
13. Fang X, Ahmad I, Blanka A, Schottkowski M, Cimdins A, Galperin MY, Rö mling, U., and Gomelsky, M. (2014). GIL, a new c-di-GMP-binding protein domain involved in regulation of cellulose synthesis in enterobacteria. Mol. Microbiol 93, 439–452. 10.1111/mmi.12672. [PubMed: 24942809]
14. Morgan JLW, McNamara JT, and Zimmer J (2014). Mechanism of activation of bacterial cellulose synthase by cyclic di-GMP. Nat. Struct. Mol. Biol 21, 489–496. 10.1038/nsmb.2803. [PubMed: 24704788]
15. Richter AM, Possling A, Malysheva N, Yousef KP, Herbst S, von Kleist M, and Hengge R (2020). Local c-di-GMP Signaling in the Control of Synthesis of the E. coli Biofilm Exopolysaccharide pEtN-Cellulose. J. Mol. Biol 432, 4576–4595. 10.1016/j.jmb.2020.06.006. [PubMed: 32534064]
16. Nicolas WJ, Ghosal D, Tocheva EI, Meyerowitz EM, and Jensen GJ (2021). Structure of the bacterial cellulose ribbon and its assembly-guiding cytoskeleton by electron cryotomography. J. Bacteriol 203, e00371–20. 10.1128/JB.00371-20.
17. Abidi W, Decossas M, Torres-Sánchez L, Puygrenier L, Létoffé S, Ghigo JM, and Krasteva PV (2022). Bacterial crystalline cellulose secretion via a supramolecular BcsHD scaffold. Sci. Adv 8, eadd1170. 10.1126/sciadv.add1170.
18. Thongsomboon W, Serra DO, Possling A, Hadjineophytou C, Hengge R, and Cegelski L (2018). Phosphoethanolamine cellulose: A naturally produced chemically modified cellulose. Science 359, 334–338. 10.1126/science.aao4096. [PubMed: 29348238]
19. Purushotham P, Ho R, and Zimmer J (2020). Architecture of a catalytically active homotrimeric plant cellulose synthase complex. Science 369, 1089–1094. [PubMed: 32646917]
20. Yang Y, Liu J, Clarke BR, Seidel L, Bolla JR, Ward PN, Zhang P, Robinson CV, Whitfield C, and Naismith JH (2021). The molecular basis of regulation of bacterial capsule assembly by Wzc. Nat. Commun 12, 4349. 10.1038/s41467-021-24652-1. [PubMed: 34272394]
21. Purushotham P, Cho SH, Moreno SMKuma M, Nixon BT, Bulone V, and Zimmer J (2016). A single heterologously expressed plant cellulose synthase isoform is sufficient for cellulose microfibril formation in vitro. Proc. Natl. Acad. Sci. USA 113, 11360–11365. 10.1073/pnas.1606210113. [PubMed: 27647898]
22. Kumari K, and Weigel PH (1997). Molecular cloning, expression, and characterization of the authentic hyaluronan synthase from group C Streptococcus equisimilis. J. Biol. Chem 272, 32539–32546. [PubMed: 9405467]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

23. Maloney FP, Kuklewicz J, Corey RA, Bi Y, Ho R, Mateusiak L, Pardon E, Steyaert J, Stansfeld PJ, and Zimmer J (2022). Structure, substrate recognition and initiation of hyaluronan synthase. Nature 604, 195–201. 10.1038/s41586-022-04534-2. [PubMed: 35355017]
24. Lairson LL, Henrissat B, Davies GJ, and Withers SG (2008). Glycosyltransferases: structures, functions, and mechanisms. Annu. Rev. Biochem 77, 521–555. 10.1146/annurev.biochem.76.061005.092322. [PubMed: 18518825]
25. Ren Z, Chhetri A, Guan Z, Suo Y, Yokoyama K, and Lee SY (2022). Structural basis for inhibition and regulation of a chitin synthase from Candida albicans. Nat. Struct. Mol. Biol 29, 653–664. 10.1038/s41594-022-00791-x. [PubMed: 35788183]
26. Chen W, Cao P, Liu Y, Yu A, Wang D, Chen L, Sundarraj R, Yuchi Z, Gong Y, Merzendorfer H, and Yang Q (2022). Structural basis for directional chitin biosynthesis. Nature 610, 402–408. 10.1038/s41586-022-05244-5. [PubMed: 36131020]
27. Qiao Z, Lampugnani ER, Yan XF, Khan GA, Saw WG, Hannah P, Qian F, Calabria J, Miao Y, Grüber G, eet al. (2021). Structure of Arabidopsis CESA3 catalytic domain with its substrate UDP-glucose provides insight into the mechanism of cellulose synthesis. Proc. Natl. Acad. Sci. USA 118, e2024015118. 10.1073/pnas.2024015118. [PubMed: 33729990]
28. Nixon BT, Mansouri K, Singh A, Du J, Davis JK, Lee JG, Slabaugh E, Vandavasi VG, O'Neill H, Roberts EM, et al. (2016). Comparative structural and computational analysis supports eighteen cellulose synthases in the plant cellulose synthesis complex. Sci. Rep 6, 28696. 10.1038/srep28696. [PubMed: 27345599]
29. Zhang X, Xue Y, Guan Z, Zhou C, Nie Y, Men S, Wang Q, Shen C, Zhang D, Jin S, et al. (2021). Structural insights into homotrimeric assembly of cellulose synthase CesA7 from Gossypium hirsutum. Plant Biotechnol. J 19, 1579–1587. 10.1111/pbi.13571. [PubMed: 33638282]
30. Origin 2021b. OriginLab Corporation https://www.originlab.com/.
31. Punjani A, Rubinstein JL, Fleet DJ, and Brubaker MA (2017). cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. Nat. Methods 14, 290–296. 10.1038/nmeth.4169. [PubMed: 28165473]
32. Emsley P, and Cowtan K (2004). Coot: model-building tools for molecular graphics. Acta Crystallogr. D Biol. Crystallogr 60, 2126–2132. 10.1107/S0907444904019158. [PubMed: 15572765]
33. Adams PD, Afonine PV, Bunkóczi G, Chen VB, Davis IW, Echols N, Headd JJ, Hung LW, Kapral GJ, Grosse-Kunstleve RW, et al. (2010). PHENIX: a comprehensive Python-based system for macromolecular structure solution. Acta Crystallogr. D Biol. Crystallogr 66, 213–221. 10.1107/S0907444909052925. [PubMed: 20124702]
34. PyMol. The PYMOL Molecular Graphics System. PyMol v2.5.4. Schrö dinger, https://pymol.org/2/.
35. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, and Ferrin TE (2004). UCSF Chimera–a visualization system for exploratory research and analysis. J. Comput. Chem 25, 1605–1612. 10.1002/jcc.20084. [PubMed: 15264254]
36. Case DA, Ben-Shalom IY, Brozell SR, Cerutti DS, Cheatham TE III., Cruzeiro VWD, Darden TA, Duke RE, Ghoreishi D, Giambasu G, et al. (2019). AMBER 2019 (University of California)
37. Scheurer M, Rodenkirch P, Siggel M, Bernardi RC, Schulten K, Tajkhorshid E, and Rudack T (2018). PyContact: Rapid, customizable, and visual analysis of noncovalent interactions in MD simulations. Biophys. J 114, 577–583. 10.1016/j.bpj.2017.12.003. [PubMed: 29414703]
38. Michaud-Agrawal N, Denning EJ, Woolf TB, and Beckstein O (2011). MDAnalysis: a toolkit for the analysis of molecular dynamics simulations. J. Comp. Chem 32, 2319–2327. [PubMed: 21500218]
39. Gowers RJ, Linke M, Barnoud J, Reddy TJ, Melo MN, Seyler SL, and Beckstein O (2016). MDAnalysis: a Python package for the rapid analysis of molecular dynamics simulations. Proc. 15th Python Sci. Conf 98, 105.
40. Mart´ınez L, Andrade R, Birgin EG, and Martıńez JM (2009). PACKMOL: a package for building initial configurations for molecular dynamics simulations. J. Comput. Chem 30, 2157–2164. 10.1002/jcc.21224. [PubMed: 19229944]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

41. Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, Lopez R, McWilliam H, Remmert M, Soding J, et al. (2011). Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Mol Syst Biol 7, 539. 10.1038/msb.2011.75. [PubMed: 21988835]
42. Studier FW (2005). Protein production by auto-induction in high density shaking cultures. Protein Expr. Purif 41, 207–234. [PubMed: 15915565]
43. Waterhouse A, Bertoni M, Bienert S, Studer G, Tauriello G, Gumienny R, Heer FT, de Beer TAP, Rempfer C, Bordoli L, et al. (2018). SWISS-MODEL: homology modelling of protein structures and complexes. Nucleic Acids Res 46, W296–W303. 10.1093/nar/gky427. [PubMed: 29788355]
44. Biasini M, Schmidt T, Bienert S, Mariani V, Studer G, Haas J, Johner N, Schenk AD, Philippsen A, and Schwede T (2013). OpenStructure: an integrated software framework for computational structural biology. Acta Crystallogr. D Biol. Crystallogr 69, 701–709. 10.1107/S0907444913007051. [PubMed: 23633579]
45. Studer G, Tauriello G, Bienert S, Biasini M, Johner N, and Schwede T (2021). ProMod3-A versatile homology modelling toolbox. PLoS Comput. Biol 17, e1008667. 10.1371/journal.pcbi.1008667. [PubMed: 33507980]
46. Schott-Verdugo S, and Gohlke H (2019). PACKMOL-Memgen: A Simple-To-Use, Generalized Workflow for Membrane-Protein-Lipid-Bilayer System Building. J. Chem. Inf. Model 59, 2522–2528. 10.1021/acs.jcim.9b00269. [PubMed: 31120747]
47. Maier JA, Martinez C, Kasavajhala K, Wickstrom L, Hauser KE, and Simmerling C (2015). ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. J. Chem. Theory Comput 11, 3696–3713. 10.1021/acs.jctc.5b00255. [PubMed: 26574453]
48. Dickson CJ, Madej BD, Skjevik AA, Betz RM, Teigen K, Gould IR, and Walker RC (2014). Lipid14: The Amber Lipid Force Field. J. Chem. Theory Comput 10, 865–879. 10.1021/ct4010307. [PubMed: 24803855]
49. Kirschner KN, Yongye AB, Tschampel SM, González-Outeiriño J, Daniels CR, Foley BL, and Woods RJ (2008). GLYCAM06: a generalizable biomolecular force field. J. Comput. Chem 29, 622–655. 10.1002/jcc.20820. [PubMed: 17849372]
50. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, and Klein ML (1983). Comparison of simple potential functions for simulating liquid water. J. Chem. Phys 79, 926–935.
51. Li P, and Merz KM (2014). Taking into account the ion-induced dipole interaction in the nonbonded model of ions. J. Chem. Theory Comput 10, 289–297. 10.1021/ct400751u. [PubMed: 24659926]
52. Li P, Song LF, and Merz KM Jr. (2015). Systematic Parameterization of Monovalent Ions Employing the Nonbonded Model. J. Chem. Theory Comput 11, 1645–1657. 10.1021/ct500918t.
53. Li P, Roberts BP, Chakravorty DK, and Merz KM Jr. (2013). Rational design of particle mesh ewald compatible lennard-jones parameters for +2 metal cations in explicit solvent. J. Chem. Theory Comput 9, 2733–2748. 10.1021/ct400146w. [PubMed: 23914143]
54. Zgarbová M, Otyepka M, Sponer J, Mládek A., Banáš P, Cheatham TE III, and Jurečka P (2011). Refinement of the Cornell et al. nucleic acids force field based on reference quantum chemical calculations of glycosidic torsion profiles. J. Chem. Theory Comput 7, 2886–2902. 10.1021/ct200162x. [PubMed: 21921995]
55. Petrová P, Koča J, and Petrová P (1999). Potential energy hypersurfaces of nucleotide sugars: Ab initio calculations, force-field parametrization, and exploration of the flexibility. J. Am. Chem. Soc 121, 5535–5547.
56. Aduri R, Psciuk BT, Saro P, Taniga H, Schlegel HB, and SantaLucia J (2007). AMBER force field parameters for the naturally occurring modified nucleosides in RNA. J. Chem. Theory Comput 3, 1464–1475. 10.1021/ct600329w. [PubMed: 26633217]
57. Wang P, Izatt RM, Oscarson JL, and Gillespie SE (1996). 1H NMR study of protonation and Mg(II) coordination of AMP, ADP, and ATP at 25, 50, and 70 °C. J. Phys. Chem 100, 9556–9560.
58. Jancan I, and Macnaughtan MA (2012). Acid dissociation constants of uridine-5'-diphosphate compounds determined by 31phosphorus nuclear magnetic resonance spectroscopy and internal pH referencing. Anal. Chim. Acta 749, 63–69. 10.1016/j.aca.2012.08.052. [PubMed: 23036468]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

59. Dupradeau FY, Cézard C, Lelong R, Stanislawiak E, Pêcher J, Delepine JC, and Cieplak P (2008). R.E.DD.B: a database for RESP and ESP atomic charges, and force field libraries. Nucleic Acids Res 36, D360–D367. 10.1093/nar/gkm887. [PubMed: 17962302]

60. Singh A, Kwansa AL, Kim HS, Williams JT, Yang H, Li NK, Kubicki JD, Roberts AW, Haigler CH, and Yingling YG (2020). In silico structure prediction of full-length cotton cellulose synthase protein (GhCESA1) and its hierarchical complexes. Cellulose 27, 5597–5616.

61. Sethaphong L, Haigler CH, Kubicki JD, Zimmer J, Bonetta D, DeBolt S, and Yingling YG (2013). Tertiary model of a plant cellulose synthase. Proc. Natl. Acad. Sci. USA 110, 7512–7517. 10.1073/pnas.1301027110. [PubMed: 23592721]

62. Lee J, Cheng X, Jo S, MacKerell AD, Klauda JB, and Im W (2016). CHARMM-GUI input generator for NAMD, Gromacs, Amber, Openmm, and CHARMM/OpenMM simulations using the CHARMM36 additive force field. Biophys. J 110, 641a. 10.1016/j.bpj.2015.11.3431.

63. Darden T, York D, and Pedersen L (1993). Particle Mesh Ewald - an N.Log(N) method for ewald sums in large systems. J. Chem. Phys 98, 10089–10092. 10.1063/1.464397.

64. Ryckaert J-P, Ciccotti G, Berendsen HJ, van der Spoel D, and van Drunen R (1977). Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. J. Comp. Phys 23, 327–341.

65. Schneider T, and Stoll E (1978). Molecular-dynamics study of a three-dimensional one-component model for distortive phase transitions. Phys. Rev. B 17, 1302–1322.

66. Berendsen HJC, Postma JPM, Vangunsteren WF, Dinola A, and Haak JR (1984). Molecular-Dynamics with Coupling to an External Bath. J. Chem. Phys 81, 3684–3690. 10.1063/1.448118.

67. Le Grand S, Götz AW, and Walker RC (2013). SPFP: Speed without compromise—A mixed precision model for GPU accelerated molecular dynamics simulations. Comput. Phys. Commun 184, 374–380.

68. Salomon-Ferrer R, Götz AW, Poole D, Le Grand S, and Walker RC (2013). Routine microsecond molecular dynamics simulations with AMBER on GPUs. 2. Explicit solvent particle Mesh Ewald. J. Chem. Theory Comput 9, 3878–3888. 10.1021/ct400314y. [PubMed: 26592383]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Highlights

- Cryo-EM structure of UDP-glucose-bound poplar cellulose synthase-8
- Site-directed mutagenesis analyzes the function of a flexible gating loop
- Conserved gating loop residues are necessary for cellulose biosynthesis
- Molecular dynamics simulations support persistent gating loop-substrate interactions

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

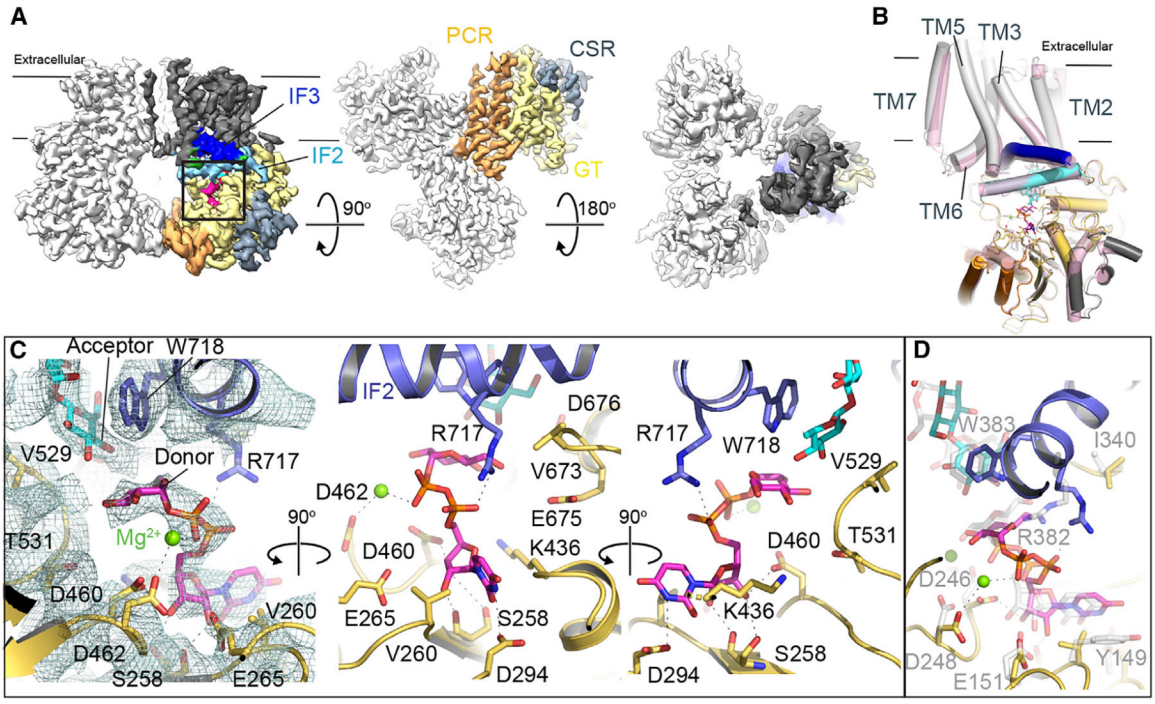


Figure 1. Structure of UDP-glucose-bound poplar CesA8

(A) Cryo-EM map of the substrate-bound poplar CesA8 trimer. Two subunits are shown in light gray and one subunit is colored according to its domains: dark gray: TM region, wheat: GT domain, orange: PCR, steel blue: CSR, and light blue and blue for IF helices 2 and 3, respectively. See also Figures S1, S2, and S3A.

(B) Overlay of one subunit of the apo (PDB: 6WLB, colored pink) and substrate (UDP-Glc)-bound CesA8 structures. The substrate-bound CesA8 structure is colored as in panel (A).

(C) Zoom-in view of the active site and substrate coordination. The cryo-EM map of the substrate and surrounding residues is shown as a mesh in the left panel.

(D) Superimposition of substrate-bound BcsA (PDB: 5EIY) and poplar CesA8. BcsA is colored light gray for its carbon atoms. See also Figure S3.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

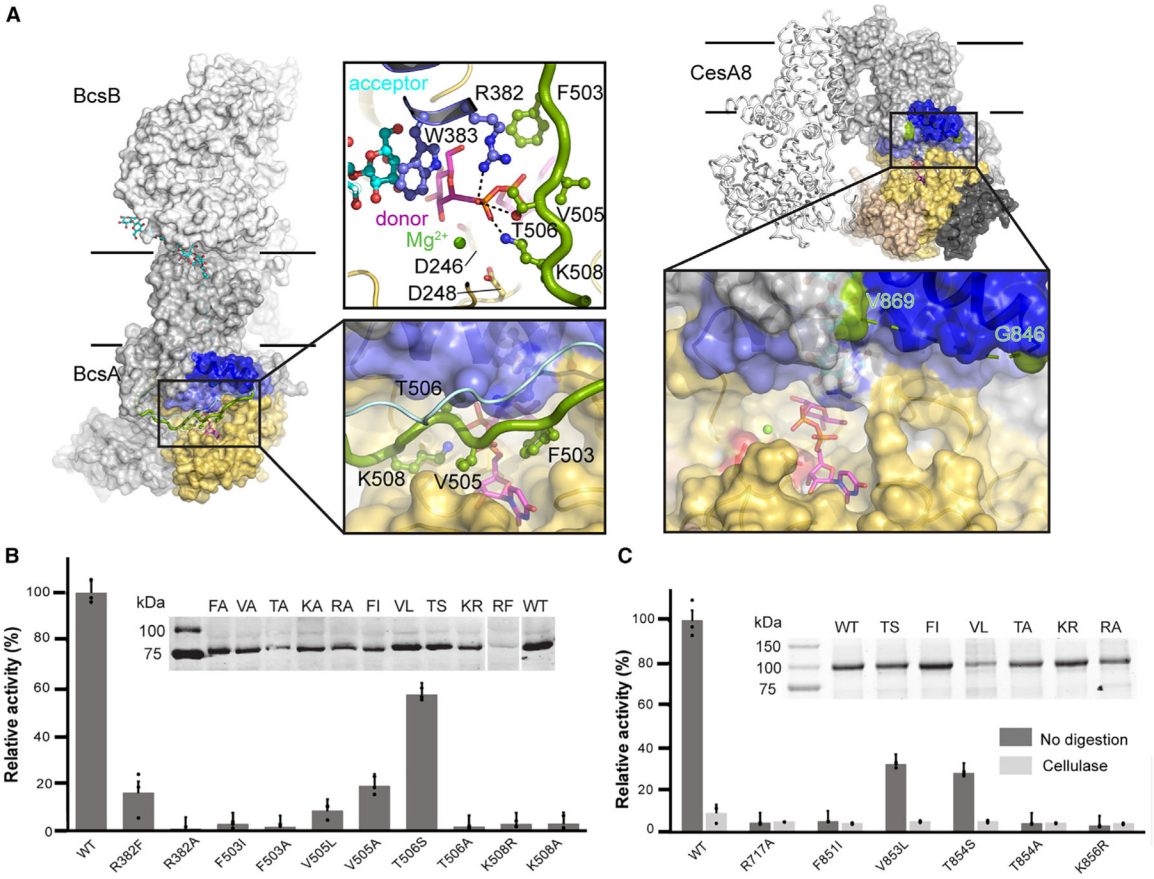


Figure 2. The gating loop is critical for catalytic activity

(A) Substrate-bound structures of bacterial (BcsA, PDB: 5EIY) and poplar cellulose synthase (this study). The substrate is shown as sticks colored magenta for carbon atoms. The gating loop positions observed in BcsA in the absence (BcsA, PDB: 4P02) and the presence of substrate are shown as light blue and green ribbons, respectively.

(B) Comparison of the catalytic activities of the bacterial BcsA-B wild-type (WT) complex and its variants containing the indicated BcsA mutants. Experiments were performed in IMVs and activities were measured upon incorporation of tritium-labeled glucose into cellulose. The product yield for the WT enzyme was set to 100%. Inset: Western blot of the BcsA-B containing IMVs used for activity assays detecting His-tagged BcsA. Blotting was done to normalize the concentration of mutants with respect to the WT enzyme. The lanes for the ‘‘RF’’ and ‘‘WT’’ constructs were moved from the left to the right side of the Western blot for clarity.

(C) Catalytic activity determined for WT CesA8 and its mutants. Reactions were performed in the presence or absence of cellulase using affinity-purified proteins. The product obtained for the undigested WT CesA8 was set as 100%. Inset: Coomassie stained 10% SDS-PAGE of the purified CesA8 constructs used for normalization and activity assays. In panels (B) and (C) error bars represent positive standard deviations from means of three replicas. See also Figure S4.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

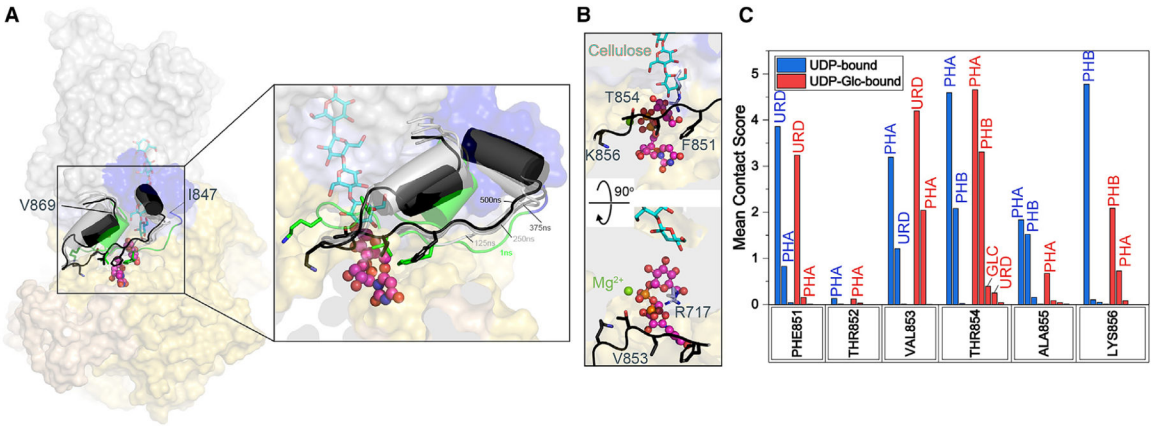


Figure 3. Molecular dynamics simulation analysis of gating loop-UDP interactions

(A) Overlay of UDP-Glc-bound CesA8 models with an inserted gating loop after 1, 125, 250, 375, and 500 ns of MD simulation. The original model is shown as a surface-colored structure as in Figure 1. The gating loop is shown as a cartoon for models obtained after the indicated simulation times. Right panel shows magnified view of the boxed area.

(B) Close-up view of interactions between the gating loop and the substrate's uracil moiety after 500 ns of MD simulation.

(C) A bar plot of mean contact scores for individual CesA gating loop residues interacting with UDP/UDP-Glc and/or Mg^{2+}. Mean contact score refers to a mean over 1,000 frames over time (500 ns), and each bar represents an average from three independent simulations. URD: Uridine, PHA/B: alpha and beta phosphate, GLC: alpha-D-glucosyl group. Generated with Origin.[30] See also Figure S5.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

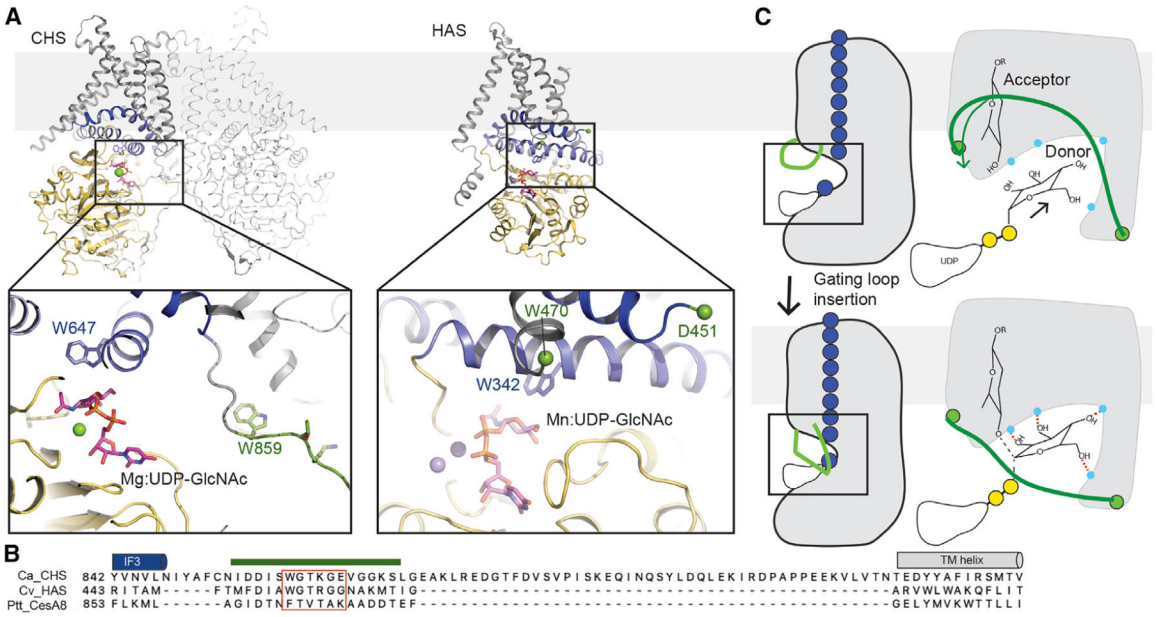


Figure 4. Localization and proposed function of the gating loop in membrane-integrated processive GT-2 glycosyltransferases

(A) Substrate-bound structures of *Candida albicans* chitin synthase (CHS, PDB: 7STM) and Chlorella virus hyaluronan synthase (HAS, PDB: 7SP8). Both enzymes place a conserved WGTR/KG motif near the catalytic pocket (unresolved in HAS).

(B) Alignment of the gating loop regions of chitin, hyaluronan, and cellulose synthases.

(C) Gating loop insertion likely positions the substrate molecule closer to the base catalyst and helps in accepting glycosyl unit. Left panel: Side view of entire enzyme, right panel: Close-up view of the active site. Conformational changes of the donor sugar may increase the reactivity of the electrophile.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Penta · His Antibody, BSA-free	Qiagen	Cat# 34660; RRID: AB_2619735
Anti-Mouse IgG (H&L) (Goat) Antibody DyLight 800 Conjugated	Rockland	Cat# 610-145-002; RRID:AB_10703265
Bacterial and virus strains		
Rosetta 2(DE3) Singles Competent Cells	Novagen	Cat# 71400
NEB® 5-alpha Competent *E. coli* (High Efficiency)	New England Biolabs	Cat# C2987
Chemicals, peptides, and recombinant proteins		
Uridine 5′-diphosphate disodium salt hydrate	Sigma-Aldrich	Cat# 94330; CAS: 27821-45-0
Uridine 5′-diphosphoglucose disodium salt hydrate from *Saccharomyces cerevisiae*	Sigma-Aldrich	Cat# U4625; CAS: 28053-08-9
Uridine diphospho-D-glucose, [6-^{3}H]-, 250 μCi	PerkinElmer	Cat# NET1163250UC
Lauryl Maltose Neopentyl Glycol	Anatrace	Cat# NG310
Cholesteryl Hemisuccinate Tris Salt	Anatrace	Cat# CH210
Glyco-diosgenin (GDN)	Anatrace	Cat# GDN101
Aprotinin, Bovine Lung	AG Scientific	Cat# A-1420; CAS: 9087-70-1
Leupeptin hemisulfate	Adooq	Cat# A14145; CAS: 103476-89-7
Bestatin hydrochloride	APExBIO	Cat# A8621; CAS: 65391-42-6
AEBSF HCl	AG Scientific	Cat# A-1018; CAS: 30827-99-7
Pepstatin A	Adooq	Cat# A14262; CAS: 26305-03-3
Benzamidine Hydrochloride Monohydrate	GoldBio	Cat# B-050; CAS: 1670-14-0
E-64	APExBIO	Cat# A2576; CAS: 66701-25-5
Phenylmethanesulphonyl fluoride	VWR	Cat# 97064; CAS: 329-98-6
FuGENE HD Transfection Reagent	Promega	Cat# E2311
ESF 921 Insect Cell Culture Medium, Protein Free	Expression Systems	Cat# 96-001-01
Fetal Bovine Serum, Heat Inactivated	Sigma Aldrich	Cat# 12306C
Critical commercial assays		
Cellulase (*endo*-1,4-β-D-glucanase) (*Trichoderma longibrachiatum*)	Megazyme/Neogen	Cat# E-CELTR
Deposited data		
Structure of homotrimeric poplar cellulose synthase isoform 8	Purushotham et al.[19]	PDB: 6WLB
Coordinates of UDP bound CesA8	This study	PDB: 8G27
Cryo-EM structure of UDP bound CesA8	This study	EMD: 29678
Coordinates of UDP-Glucose bound CesA8	This study	PDB: 8G2J
Cryo-EM structure of UDP-Glucose bound CesA8	This study	EMD: 29679

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Experimental models: Cell lines		
Spodoptera frugiperda Sf9	ATCC	Cat# CRL-1711; RRID: CVCL_0549
Oligonucleotides		
Shown in Table S5	This study	N/A
Recombinant DNA		
pETDuetBcsA_WT-His_{12}-BcsB	Morgan et al.[7]	N/A
pETDuetBcsA_F503A-His_{12}-BcsB	This study	N/A
pETDuetBcsA_F503I-His_{12}-BcsB	This study	N/A
pETDuetBcsA_V505A-His_{12}-BcsB	This study	N/A
pETDuetBcsA_V505L-His_{12}-BcsB	This study	N/A
pETDuetBcsA_T506A-His_{12}-BcsB	This study	N/A
pETDuetBcsA_T506S-His_{12}-BcsB	This study	N/A
pETDuetBcsA_K508A-His_{12}-BcsB	This study	N/A
pETDuetBcsA_K508R-His_{12}-BcsB	This study	N/A
pETDuetBcsA_R382A-His_{12}-BcsB	This study	N/A
pETDuetBcsA_R382F-His_{12}-BcsB	This study	N/A
pACEBac1-His_{12}-*Pt*CesA8_WT	Purushotham et al.[19]	N/A
pACEBac1-His_{12}-CesA8_R717A	This study	N/A
pACEBac-His_{12}-CesA8_F851I	This study	N/A
pACEBac-His_{12}-CesA8_V853L	This study	N/A
pACEBac-His_{12}-CesA8_T854S	This study	N/A
pACEBac-His_{12}-CesA8_T854A	This study	N/A
pACEBac-His_{12}-CesA8_K856R	This study	N/A
Software and algorithms		
Cryosparc v.2.8	Punjani et al.[31]	https://cryosparc.com/
Coot v0.9.6	Emsley and Cowtan[32]	https://www2.mrc-lmb.cam.ac.uk/personal/pemsley/coot/
Phenix	Adams et al.[33]	http://www.phenix-online.org/
PyMol v2.5.4	PyMol[34]	https://pymol.org/2/
Chimera 1.15	Pettersen et al.[35]	https://www.cgl.ucsf.edu/chimera/
Origin 2021b	OriginLab[30]	https://www.originlab.com/
GraphPad Prism 6.0	GraphPad	https://www.graphpad.com/scientific-software/prism/
Amber 2019	Case et al.[36]	https://ambermd.org/
PyContact 1.0.4	Scheurer et al.[37]	https://pycontact.github.io/
MDAnalysis 0.20.1	Michaud-Agrawal et al.[38]; Gowers et al. [39]	https://www.mdanalysis.org/

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Packmol 18.169	Martínez et al.[40]; Schott-Verdugo and Gohlke[46]	https://m3g.github.io/packmol/
Clustal Omega	Sievers et al.[41]	https://www.ebi.ac.uk/Tools/msa/clustalo/
Other		
Superose 6 Increase 10/300 GL (Bed dimensions 10 X 300)	Cytiva	Cat# 29091596
HisPur Ni-NTA Resin	ThermoFisher Scientific	Cat# 88222