Review

Cellulose synthases: new insights from crystallography and modeling

Erin Slabaugh[1*], Jonathan K. Davis[1*], Candace H. Haigler[1,2], Yaroslava G. Yingling[3], and Jochen Zimmer[4]

[1] Department of Crop Science, North Carolina State University, Raleigh, NC, USA
[2] Department of Plant and Microbial Biology, North Carolina State University, Raleigh, NC, USA
[3] Department of Materials Science and Engineering, North Carolina State University, Raleigh, NC, USA
[4] Center for Membrane Biology, Department of Molecular Physiology and Biological Physics, University of Virginia, Charlottesville, VA, USA

Detailed information about the structure and biochemical mechanisms of cellulose synthase (CelS) proteins remained elusive until a complex containing the catalytic subunit (BcsA) of CelS from *Rhodobacter sphaeroides* was crystalized. Additionally, a 3D structure of most of the cytosolic domain of a plant CelS (GhCESA1 from cotton, *Gossypium hirsutum*) was produced by computational modeling. This predicted structure contributes to our understanding of how plant CelS proteins may be similar and different as compared with BcsA. In this review, we highlight how these structures impact our understanding of the synthesis of cellulose and other extracellular polysaccharides. We show how the structures can be used to generate hypotheses for experiments testing mechanisms of glucan synthesis and translocation in plant CelS.

Cellulose synthase proteins in plants and bacteria

Cellulose is a major structural component of plant cell walls and some bacterial biofilms, giving it importance as a renewable biomaterial and in human health [1–5]. Cellulose biosynthesis in plants is a complex process that involves the formation of cellulose synthase complexes (CSCs), polymerization of glucose into a β-(1→4)-linked polysaccharide, translocation of glucan across the plasma membrane, and coalescence of multiple glucan chains to form paracrystalline microfibrils. Describing cellulose synthesis in 1995, Deborah Delmer and Yehudit Amor wrote 'There is probably no major biochemical process in plants that is both so important and so poorly understood at the molecular level. . .' [6]. Although cellulose synthase (CelS) genes from the bacterium *Acetobacter xylinum* (now *Gluconoacetobacter hansenii*) were described in 1990, the plant homologs (called CESA proteins) were not described until 1996 because low sequence similarity (<30%) made screening plant cDNA libraries with bacterial nucleic acid probes ineffective. Ultimately, hydropathy plot analysis of translated cotton (*Gossypium hirsutum*) fiber expressed sequence tags implicated GhCESA1 as a putative plant cellulose synthase with some structural relatedness to the bacterial cellulose synthases [7]. Following the identification of plant CESA proteins, their function to produce cellulose in the plant cell wall has been confirmed with genetic approaches. Subsequent work defined the essential D, D, D, QxxRW amino acid motifs necessary for CelS function [8], and provided additional analyses of missense mutations that impact CelS activity (reviewed in [9] and [10]).

All known CelS proteins are classified as Glycosyltransferase Family2 (GT2) proteins in the CAZy database [11], based on sequence similarity. The GT2 family also includes cellulose synthase-like proteins involved in plant cell wall polysaccharide synthesis, as well as enzymes synthesizing some extracellular polysaccharides in other kingdoms, including chitin synthase in fungi and insects and hyaluronan synthase in animals. All GT2 proteins are predicted to be inverting enzymes, meaning that the configuration of the anomeric sugar carbon is inverted during the transfer reaction. For example, CelS proteins utilize UDP-α-D-glucose as the substrate in the formation of β-(1→4)-linked cellulose. GT2 proteins are also thought to share a common catalytic fold, designated GT-A, first described in 1999 from the crystal structure of the GT2 protein SpsA from *Bacillus subtilis* [12]. The GT-A fold has been observed seven times in GT2 family proteins that have so far been crystallized (CAZy database [11]) and is characterized by distinct acceptor substrate and sugar–nucleotide donor substrate binding sites within one catalytic site. The GT-A fold is composed of two juxtaposed β/α/β folds, which form a β-sheet surrounded by α-helices. For a detailed review of GT structures and mechanisms, see [13].

Recent advances in cellulose synthase structure

Despite previous work, detailed information about the atomic structure and biochemical mechanisms of CelS proteins remained elusive until the publication of two studies in 2013. The first study presents the structure of the A and B subunits of a bacterial CelS complex from *Rhodobacter sphaeroides* (BcsA and BcsB) in a heterodimeric complex that included one uridine diphosphate (UDP) molecule and a translocating glucan [14]. In

Corresponding author: Davis, J.K. (jkdavis2@ncsu.edu).
* These authors contributed equally to this manuscript.

1360-1385/$ – see front matter
 http://dx.doi.org/10.1016/j.tplants.2013.09.009

addition to the large central domain of BcsA containing a single UDP-glucose binding site, the 3.25 Å resolution crystal structure revealed the organization of eight transmembrane helices (TMHs) from BcsA and one TMH from BcsB. Cellulose synthesis occurs within the cytosolic domain of BcsA, whereas BcsB resides in the periplasm and may help to guide the chain in the extrusion process. Plant cells do not have a second outer membrane through which glucan must be translocated, and thus plants may or may not have a functional equivalent of BcsB. The BcsA structure clarified the role of the classical conserved amino acids in CelS proteins (D, D, D, QxxRW) as well as the role of additional functional motifs [14].

Complementing the solved crystal structure of BcsA, another study resulted in a 3D computational model of most of the predicted cytosolic domain (506 amino acids) of cotton GhCESA1 (Figure 1A) [15]. The authors used bioinformatics-based techniques for protein structure prediction, *de novo* protein folding, and molecular dynamics simulations to generate an atomistic model structure, which contained one apparent catalytic site formed by a six-stranded β-sheet surrounded by α-helices as predicted for the GT-A fold. The GT-A folds of BcsA and GhCESA1, which have limited primary sequence similarity, show structural congruence, even though the BcsA structure was not used for prediction of the plant CESA model (Figure 1B) [15].

In both the crystal and predicted structures, the catalytic residues positioned by the GT-A fold include the invariant DDG, DCD, and TED motifs. As shown by the BcsA structure, the DDG and DCD motifs coordinate UDP and the essential divalent cation, whereas the D of the TED motif probably acts as the catalytic base [12,14]. Nearby these conserved motifs, an interfacial helix (IF) at the plasma membrane (Figure 1B) contains a QRGRW motif, which interacts with the cellulose acceptor substrate. In GhCESA1, the QVLRW motif probably carries out the same function and is also likely to be positioned near the membrane, as shown by structural alignment of its modeled cytosolic region to BcsA (Figure 1B). Although the BcsA crystal structure does not unequivocally confirm whether the reducing or the non-reducing end of the elongating glucan is at the acceptor site, the latter is supported by the position of UDP in the structure. If glucose addition occurred at the reducing end, we would expect UDP to be bound to the 1-hydroxyl of the terminal glucose residue of the acceptor glucan. Instead, UDP is not bound to the polysaccharide and is located in a binding pocket that is large enough to accommodate UDP-Glc [14] This structural evidence is consistent with previous experiments supporting cellulose elongation from the non-reducing end [9,16,17].

It is clear from the BcsA crystal structure and predicted by the GhCESA1 model that cellulose elongation occurs at a single active site. Also, from the perspective of the catalytic mechanism, a single CelS protein is capable of synthesizing a β-(1→4)-linked polymer from UDP-glucose even though each successive glucose residue is rotated 180° relative to its neighbors [14]. Speculation about two active sites possibly being present within one peptide [8,18] is no longer necessary to explain cellulose synthesis. A model of how this occurs is presented in [14]. Briefly, the glucose moiety of UDP-glucose is presented to the elongating glucan acceptor at an angle relative to the polymer axis. Each glucose added to the glucan acceptor can freely rotate around the newly formed glycosidic bond to adopt the sterically favored inverted orientation relative to the preceding glucose residue as the glucan simultaneously translocates out of the active site. The direction of terminal glucose rotation reverses with every glucose addition due to the steric environment presented by the preceding glucose and the β-(1→4)-linkage, resulting in the characteristic cellulose polymer structure.

In the GhCESA1 model, the homologous locations of most of the known CESA missense mutations in the cytosolic domain that cause cellulose-deficient phenotypes in *Arabidopsis* converge near the catalytic site [15]. In this review, we extend the analysis to all CESA missense mutations that we were able to identify in rice (*Oryza sativa*) and *Arabidopsis* in both the cytosolic and membrane regions (Table S1 in the supplementary material online). Most CESA missense mutations lie within ~20 Å of the modeled GhCESA1 catalytic site (Table S1 in the supplementary material online), and their approximate locations on the secondary structural elements of CESA, as inferred from the computational model and predicted TMH topology, are shown in Figure 2.

Comparing the two structures, we can make predictions about the mechanisms by which CESA mutations might result in cellulose synthesis phenotypes. Although speculative, such predictions can inform experimental design to gain a better understanding of CelS structure and function. For example, CESA proteins share a conserved HxKAG motif (Figure S1 in the supplementary material online) with BcsA. In the BcsA crystal structure, the HxKAG motif is positioned so that it contributes to the formation of a pocket, along with the glutamate of the conserved TED motif that probably accommodates the glucose of UDP-Glc [14]. In the modeled GhCESA1 structure, the HxKAG motif is found in a similar location near TED and may play a similar role in forming the substrate-binding pocket. An *Arabidopsis* CESA3 mutant, *eli1-2* (A522V in HxKAG; see Table S1 in the supplementary material online and Figure 3), has stunted growth and ectopic lignin deposition phenotypes associated with a defect in cellulose synthesis [19,20]. We would predict that disruption of CESA activity in this mutant occurs because the part of the substrate-binding pocket that usually accommodates the glucose moiety of UDP-glucose is disrupted. In the BcsA crystal structure, the conserved alanine of HxKAG points towards a small pocket adjacent to the TED motif. It is possible that a change from alanine to the more bulky valine residue at this position would prevent the HxKAG and TED motifs from juxtaposing in a way that properly coordinates UDP-Glc.

Another *Arabidopsis* mutant in AtCESA1 (*rsw1-1*, A594V) is positioned ten amino acids C terminal of the HxKAG motif. In the modeled structure, the GhCESA1 residue aligning to AtCESA1 A594 lies on the same α-helix as HxKAG but on the opposite end, approximately 17 Å from the active site (Table S1 in the supplementary material online, Figure 3). The *rsw1-1* mutant is