Short Title

In vitro synthesis & assembly of cellulose fibrils

Corresponding authors

B. Tracy Nixon,
Department of Biochemistry and Molecular Biology, The Pennsylvania State University,
University Park, PA 16802, USA
Tel: 814-863-4904
FAX: 814-863-7024
e-mail: btn1@psu.edu

Manish Kumar
Department of Chemical Engineering, The Pennsylvania State University, University Park, PA
16802, USA
Tel: 814-865-7519
FAX:
e-mail: manish.kumar@psu.edu

Journal Research Area

Biochemistry and Metabolism
Synthesis and self-assembly of cellulose microfibrils from reconstituted cellulose synthase

Sung Hyun Cho1, Pallinti Purushotham2, Chao Fang3, Cassandra Maranas4, Sara M. Díaz-Moreno5, Vincent Bulone5,6, Jochen Zimmer2, Manish Kumar3*, B. Tracy Nixon1*

1Department of Biochemistry and Molecular Biology, The Pennsylvania State University, University Park, PA 16802, USA
2Department of Molecular Physiology and Biological Physics, University of Virginia School of Medicine, 480 Ray C. Hunt Dr., Charlottesville, VA 22908, USA
3Department of Chemical Engineering, The Pennsylvania State University, University Park, PA 16802, USA
4Department of Chemical Engineering, University of Washington, Seattle, WA 98105
5Division of Glycoscience, School of Biotechnology, Royal Institute of Technology (KTH), Stockholm, SE-10691, Sweden
6Australian Research Council Centre of Excellence in Plant Cell Walls, School of Agriculture, Food and Wine, University of Adelaide, Waite Campus, Urrbrae 5064, South Australia, Australia

One Sentence Summary

Liposome-reconstituted, heterologously expressed cellulose synthases that contribute to primary/secondary plant cell walls synthesized glucan chains that assembled into cellulose microfibrils.
Footnotes

- Author’s contributions

S.H.C., M.K. and B.T.N. designed the experiments, interpreted the data and wrote the paper. P.P and J.Z. constructed Pichia strains and performed radiolabeling assays. S.H.C. was involved in most experiments. S.H.C., C.F. and C.M conducted TEM analysis. S.M.D and V.B performed linkage analysis.

- Funding Information

This material is based upon work supported as part of The Center for LignoCellulose Structure and Formation, an Energy Frontier Research Center funded by the U.S. Department of Energy, Office of Science, and Office of Basic Energy Sciences under Award Number DE-SC0001090.

Correspondence: btn1@psu.edu / manish.kumar@psu.edu
Abstract

Cellulose, the major component of plant cell walls, can be converted to bioethanol and is thus highly studied. In plants, cellulose is produced by cellulose synthase, a processive family-2 glycosyltransferase. In plant cell walls, individual β-1,4-glucan chains polymerized by CesA are assembled into microfibrils that are frequently bundled into macrofibrils. An in vitro system in which cellulose is synthesized and assembled into fibrils would facilitate detailed study of this process. Here, we report the heterologous expression and partial purification of His-tagged CesA5 from Physcomitrella patens. Immunoblot analysis and mass spectrometry confirmed enrichment of PpCesA5. The recombinant protein was functional when reconstituted into liposomes made from yeast total lipid extract. The functional studies included incorporation of radiolabeled Glc, linkage analysis and imaging of cellulose microfibril formation using transmission electron microscopy. A number of microfibrils were observed either inside or on the outer surface of proteoliposomes, and strikingly, several thinner fibrils formed ordered bundles that either covered the surfaces of proteoliposomes or spawned from lipidosome surfaces. We also report this arrangement of fibrils made by proteoliposomes bearing CesA8 from hybrid aspen. These observations describe minimal systems of membrane-reconstituted CesAs that polymerize β-1,4-glucan chains that coalesce to form microfibrils and higher-ordered macrofibrils. How these micro- and macrofibrils relate to those found in primary and secondary plant cell walls is uncertain, but their presence enables further study of the mechanisms that govern the formation and assembly of fibrillar cellulotic structures and cell wall composites during or after the polymerization process controlled by CesA proteins.
**Introduction**

Cellulose is the most abundant biopolymer on earth (Pauly and Keegstra, 2008; McNamara et al., 2015). It consists of bundles of β-1,4-glucan chains typically organized as microfibrillar structures. It is mostly found in plants, bacteria, oomycetes and green algae (Fugelstad et al., 2009; John et al., 2011; Augimeri et al., 2015). Cellulose is the main structural component of plant cell walls, and is used for several technological applications including manufacturing of paper, textiles, and furniture (McFarlane et al., 2014). In recent years cellulose has become a focus of those pursuing the production of sustainable biofuels due to its potential to be converted to ethanol and other compounds (Pauly and Keegstra, 2008). Detailed knowledge regarding the structure and function of cellulose synthases (CesAs) and the mechanisms of cellulose polymerization and microfibril assembly is likely to facilitate downstream processing of cellulose (Cantarel et al., 2009; McNamara et al., 2015).

A simplified *in vitro* system that includes levels of CesA assembly sufficient to produce cellulose microfibrils would greatly facilitate the acquisition of such knowledge. There are many CesA genes in plants whose products exhibit complex interactions (Popper et al., 2011). For example, *Arabidopsis thaliana* has 10 CesA genes that are divided into two groups by the type of cell wall that they make: CesA1, -2, -3, -5, -6 and -9 are involved in primary cell wall synthesis (Arioli et al., 1998; Cano-Delgado et al., 2003; Desprez et al., 2007; Persson et al., 2007) and CesA4, -7 and -8 are involved in secondary cell wall synthesis (Taylor et al., 2000; Brown et al., 2005; Bosca et al., 2006; Mendu et al., 2011; McFarlane et al., 2014). These CesAs form a hexameric rosette structure in the plasma membrane, called Cellulose Synthase Complex (CSC), and it was believed until recently that each of the six lobes have six subunit CesA proteins (Kimura et al., 1999; Doblin et al., 2002; Cosgrove, 2005; Somerville, 2006). The latest developments based on imaging, structural and computational approaches suggest that each CSC lobe possesses three CesA molecules (Newman et al., 2013; Hill et al., 2014; Nixon et al., 2016; Vandavasi et al., 2016). A single CSC should thus produce up to 18 single glucan chains, presuming that all individual CesA proteins are active synthases; these chains subsequently crystallize to form cellulose microfibrils that are 3-5 nm in diameter (Ha et al., 1998; Cosgrove, 2005; Thomas et al., 2013). Higher order bundles of microfibrils are also seen in cell walls (Marga et al., 2005; Somerville, 2006; Zhang et al., 2016).
While assembly of CesAs and assembly of glucan chains involve different macromolecules, they are likely to be linked processes, but the extent of linkage remains unclear. One scenario is that the CesA enzymes assemble into trimeric lobes, and these into rosette CSCs. Subsequent glucan synthesis and extrusion from the organized rosette provides chains to form fibrillar cellulose. Alternatively, CesA might assemble into trimeric lobes that each synthesize three adjacent glucan chains which self-assemble into a sub-microfibril. The latter further assemble into complete microfibrils providing feedback onto individual lobes, moving them into a rosette distribution. The microfibrils could then further assemble with each other and wall matrix materials.

Some clues have been garnered regarding assembly of CesA proteins. The secondary structure of most plant CesA proteins consists of at least 6 transmembrane helices (TMHs) that anchor and separate a cytoplasmic glycosyltransferase (GT) domain from an N-terminal, intrinsically unfolded domain (Sethaphong et al., 2013). In addition to possessing the catalytic function, the GT domain includes a plant conserved region (P-CR) and a class specific region (CSR) (Slabaugh et al., 2014). The catalytic function is provided in part by three broadly conserved aspartate-containing motifs designated as DDG, DCD and TED, followed by a conserved QXXRW motif (Saxena and Brown Jr, 1997; Saxena et al., 2001; Sethaphong et al., 2013). \textit{In vitro} studies have shown that the catalytic domain of rice CesA1 forms redox-sensitive dimers (Olek et al., 2014), and that the same region of \textit{Arabidopsis thaliana} CesA1 forms redox-insensitive trimers (Vandavasi et al., 2016). Small-angle X-ray scattering (SAXS) and transmission electron microscopy (TEM) based low-resolution volumes of the trimer fit well to averaged lobe volumes from \textit{P. patens} rosettes imaged by freeze-fracture TEM (Nixon et al., 2016). A Zn$^{2+}$-binding RING domain is encoded in the N-terminal regions of CesAs, which may be responsible for dimerization of CesA proteins (Kurek et al., 2002).

Superimposed on this as yet poorly defined tendency for CesA subunits to assemble is an intrinsic propensity for nascent $\beta$-1,4-glucan chains to form microfibrils with varied packing of the chains (Ha et al., 1998; Cosgrove, 2005; Thomas et al., 2013). Cellulose microfibrils are deposited in transverse direction with respect to the cell growth (Szymanski and Cosgrove, 2009), and reoriented during cell wall expansion (Baskin, 2005; Anderson et al., 2010). Within a microfibril, the cellulose chains are held together by hydrogen bonds and Van-der-Waals forces (Nishiyama et al., 2002; Nishiyama et al., 2003). In nature, parallel chains are packed with low
order, hence called ‘amorphous’, or with high order, as found in crystalline cellulose. Because individual β-glucan chains are synthesized as a flat ribbon with 180-degree flipping of the orientation of successive glucose monomers, it is possible to pack them in crystalline sheets with the lateral register of the ‘up’ or ‘down’ glucose moieties in adjacent chains in different ways, giving rise to the Iα and Iβ crystalline forms. The Iα form is mostly found in bacterial and algal cellulose, and the Iβ form in higher plants and tunicate animals (Atalla and Vanderhart, 1984; Belton et al., 1989). A single microfibril may contain a mixture of these forms, and interactions between microfibrils contribute significantly to the mechanical properties of cellulose. The parallel alignment of glucan chains is consistent with simultaneous synthesis of cellulose chains in CSCs (Somerville, 2006). It is not known if CesA facilitates the close placement of cellulose chains to enable hydrogen bonding and hydrophobic interactions, if the close distance between nascent glucan chains is sufficient to induce the formation of the higher order cellulose microfibrils, or if other protein(s) is(are) required for this assembly (Somerville, 2006).

Recent work demonstrated that a single CesA isoform (CesA8 from hybrid aspen, PttCesA8) synthesizes cellulose in vitro and packs it into fibrils observable by TEM (Purushotham et al., 2016) (Purushotham et al., 2016). PttCesA8 contributes to the formation of secondary cell walls. Here we report similar synthesis of cellulose microfibrils by reconstituted CesA5 of Physcomitrella patens (PpCesA5), which is involved in primary cell wall formation (Goss et al., 2012). Previously, we had shown that overexpression of PpCesA5 in P. patens itself yielded partially purified protein that synthesized cellulose microfibrils, but the presence of other CesA isoforms and accessory proteins could not be ruled out (Cho et al., 2015). Now we eliminate the possible presence of additional isoforms and other proteins by purifying heterologously expressed PpCesA5 from Pichia pastoris, and reconstituting it into proteoliposomes. Reconstituted PpCesA5 produced cellulose microfibrils, demonstrated by incorporation of radioactive Glc from UDP-[3H]-Glc, cellulase specific degradation, TEM analysis and linkage analysis. Furthermore, TEM analysis identified cellulose microfibrils within and on the surface of proteoliposomes bearing PpCesA5 or PttCesA8. These microfibrils coalesced into higher order ‘macrofibrils’. These results confirm that single isoforms of CesAs for primary and secondary cell wall synthesis are sufficient to produce cellulose microfibrils, and suggest that glucan chains can form higher order cellulose structures by self-assembly without the need for accessory proteins.
Results

**Heterologous expression and purification of PpCesA5**

PpCesA5 is predicted to have 7 transmembrane domains, an N-terminal Zn-binding domain, and a long cytosolic region between the second and the third transmembrane helices (Supplemental Fig. S1). The cytosolic region contains P-CR and CSR regions as well as a TED motif, the latter believed to deprotonate the acceptor hydroxyl via its aspartic acid residue (Morgan et al., 2014). PpCesA5 conjugated with 12x His-tag at the C-terminus was heterologously expressed in *Pichia*. Expression in *Pichia* was directed by using methanol to induce transcription from the *alcohol oxidase 1* (*AOX1*) promoter. Membrane proteins were isolated and further purified with TALON (cobalt) resin. PpCesA5 proteins were co-purified with other membrane proteins, as shown by immunoblot using anti-PpCesA5 antibody #1 (Lanes 1-3, Fig. 1). The final elution fraction contained a highly enriched PpCesA5 protein of ~125 kDa (Lane 9, Fig. 1). Two other PpCesA5 specific antibodies #2 and #3 also detected similar band patterns, whereas anti-His antibody identified an additional band at around 55 kDa (Supplemental Fig. S2). While the epitopes of the PpCesA5 specific antibodies were in the cytosolic domain (Supplemental Fig. S1), the His-tag was at the C-terminus, suggesting that the N-terminal region is more prone to degradation. Similarly, heterologously expressed *PttCesA8* also showed N-terminal degradation that was speculated to be due to conformational flexibility or loose association of the N-terminal RING finger region (Purushotham et al., 2016).

To further confirm that PpCesA5 was enriched, proteins in the region between 110-140 kDa were excised from an SDS-PAGE gel and analyzed by mass spectrometry (MS). Three peptides of PpCesA5 including \(444^{\text{VQPTFVK}}_{450}538^{\text{AGAMNALVR}}_{546}\), and \(808^{\text{GSAPINLSDR}}_{817}\) were identified, together with fragments of 18 proteins from *Pichia*, including the catalytic subunit of \(\beta-1,3\)-glucan synthase (Supplemental Table S1).

**PpCesA5 is functional when reconstituted in proteoliposomes**

Since CesA proteins are integral membrane proteins, the enriched PpCesA5 was reconstituted into proteoliposomes using total lipid extract from yeast. These proteoliposomes were then subjected to biochemical characterization using radiolabeled UDP-[\(^{3}\text{H}\)]Glc as a tracer,
followed by detergent disruption and descending paper chromatography (Fig. 2), as described (Omadjela et al., 2013; Purushotham et al., 2016). Radiolabeled glucose was incorporated into insoluble material as shown by radioactivity that failed to leave the origin, indicating that PpCesA5s in proteoliposomes were functional.

CesA proteins require cations as cofactors for function. Activity of PpCesA5 proteoliposomes was tested with different cations including Ca$^{2+}$, Mg$^{2+}$, Mn$^{2+}$ and Zn$^{2+}$. We found that Mn$^{2+}$ was most effective, and that Zn$^{2+}$ in combination with Mn$^{2+}$ appeared to have no synergistic effect (Fig. 2A). These results were consistent with those from PttCesA8 (Purushotham et al., 2016). Mn$^{2+}$ alone was used in all subsequent tests.

The incorporation of radiolabeled Glc from UDP-$^{3}$H-Glc into synthesized product was monitored over time (Fig. 2B). The radioactive signal reached saturation at 150 min, which might be due to the depletion of substrate, loss of protein activity, or product inhibition. In contrast, it took 90 min for PttCesA8 to reach a plateau (Purushotham et al., 2016). Enzyme kinetics assays were conducted using a constant level of UDP-$^{3}$H-Glc with variation of UDP-Glc concentration. These assays show monophasic Michaelis-Menten kinetics with $K_M = 137 (102 - 179) \mu$M (Fig. 2C; 95% confidence interval in parenthesis).

To confirm that the in vitro product contained cellulose, it was treated with $\beta$-1,4-glucanase, solubilizing 73% of the radiolabeled product (Fig. 2D). The enriched preparation of PpCesA5 contained other proteins from Pichia that were present in the 110-140 kDa region of an SDS-PAGE gel (Supplemental Table S1), including the catalytic subunit of $\beta$-1,3-D-glucan synthase (callose synthase). To examine potential contribution of callose synthase to the incorporated radioactive product signal, in vitro products were subjected to treatment with $\beta$-1,3-glucanase, solubilizing 27% of the radiolabeled product (Fig. 2D).

To further confirm that the observed signal resulted from PpCesA5 activity, we introduced amino acid substitution D782N of the catalytically crucial TED motif of PpCesA5 (Supplemental Fig. S3). The analogous substitution inactivated PttCesA8 (Purushotham et al., 2016). As evident in immuno-blots, PpCesA5(D782N) was sensitive to proteolysis. Proteoliposomes prepared as for the wildtype PpCesA5 showed no capacity to incorporate $^{3}$H-Glc (Supplemental Fig. S3).
In vitro-produced glucan chains assemble into cellulose microfibrils

β-1,4-D-glucan chains produced from CesAs form microfibrils and these are sometimes bundled into larger arrays. To determine if β-1,4-glucan chains produced by PpCesA5 proteoliposomes form microfibrils or higher order bundles, proteoliposomes that had been incubated with UDP-Glc were examined over time by TEM (Fig. 3). Individual glucan chains are too small to be seen with this method, but microfibrils are large enough and they were first observed within 5 min of incubation with UDP-Glc. The initial amount of microfibrils increased upon incubation for 60 and 180 min. There were no microfibrils formed in a negative control reaction lacking UDP-Glc, nor any jagged-edge fibrils as reported for callose (Him et al., 2001). Repeated experiments with the TED mutant protein PpCesA5(D782N) yielded no observable microfibrils (Supplemental Fig. S5). Pre-treatment of proteoliposomes bearing wild-type PpCesA5 with Triton X-100 caused them to fail to yield microfibrils (Supplemental Fig. S6). Also, the addition of Zn$^{2+}$ to PpCesA5-proteoliposomes had no effect on the level of microfibril formation (Supplemental Fig. S7). The thickness of microfibrils made by PpCesA5 was estimated by cryo-TEM and found to be 4.3±0.8 nm (Supplemental Fig. S8). Most microfibrils disappeared after 15 min of incubation with cellulase, while a negative control without cellulase showed no loss of microfibrils (Fig. 4).

To further confirm that the fibrils were cellulose microfibrils with 1,4-glucose linkage, a glycosidic linkage analysis was performed by permethylation followed by gas chromatography coupled to electron-impact mass spectrometry (GC/EI-MS). In vitro products from a large scale reaction were pretreated with α-amylase and amyloglucosidase to remove Pichia-derived glycogen-like α-1,4-glucan chains. Thereafter, the sample was permethylated to alditol acetates, and then subjected to GC/EI-MS analysis. Comparison with reference peaks confirmed that the sample mainly contained 1,4-D-glucose, with no 1,3-D-glucose being observed (Fig. 5; Supplemental Fig. S9A). The sample also contained a small level of terminal glucose, as further identified by electron-impact mass spectrometry (Supplemental Fig. S9B). Analogously treated material that was produced by proteoliposomes bearing the inactive PpCesA5(D782N) showed no 1,4-D-glucose (Supplemental Fig. S10A). However, 1,4-D-glucose was present prior to eliminating α-1,4-glucans with amylase (Supplemental Fig. S10B).
Higher order self-assembly of cellulose microfibrils

It is known that cellulose microfibrils are often bundled into larger fibrils in plant cell walls. To observe possible higher order assembly that might reflect a bundling process in the PpCesA5 and PttCesA8/proteoliposome systems, we thoroughly explored negatively stained TEM grids with in vitro products to find unbroken or partially broken proteoliposomes. This revealed cellulose microfibrils near and apparently attached to the surface of proteoliposomes with PpCesA5, as if they were spawned from the proteoliposomes (Fig. 6A and Fig. 6B). The thicknesses of such fibrils varied, with thinner fibrils merging to form thicker fibrils. We also found PpCesA5-proteoliposomes that contained numerous cellulose microfibrils inside and a few microfibrils exiting out through the membranes, which also formed higher order assemblies outside (Fig. 6C). Some cellulose microfibrils wrapped around the proteoliposomes (Fig. 6D). In many images, thin cellulose fibrils coalesced into higher order micro- or macrofibrils (Fig. 6, panels C, E and F) which in some cases might be considered bundles. This coalescence was also observed in cellulose microfibrils produced by proteoliposomes bearing PttCesA8 (Fig. 7). Such ‘bundled’ microfibrils were also observed in other grid areas (Fig. 4, panels A and G, marked by arrowheads). In total, the area of ‘bundled’ microfibrils accounted for 36%-43% of the microfibrillar area in each image. These ‘bundled’ microfibrils were also eliminated by cellulase digestion prior to imaging (Fig. 4H), indicating that they were cellulose microfibrils.

Discussion

In this study proteoliposomes bearing single isoforms of CesAs known to synthesize primary cell walls was able to synthesize β-1,4-glucan chains that were assembled into cellulose microfibrils. The cellulose microfibrils were often seen to merge into higher order forms. It is not yet clear if these in vitro products relate to the microfibrils, macrofibrils or bundles seen in plant cell walls, so in the following discussion we will explicitly denote the in vitro cellulose assemblies with superscript ‘iv’ (thus, microfibrils\textsuperscript{iv}, macrofibrils\textsuperscript{iv} or bundles\textsuperscript{iv}).

Both biochemical and TEM data showed the synthesis of β-1,4-glucan chains that assembled into microfibrils\textsuperscript{iv}. Incorporation of [\textsuperscript{3}H]-Glc from UDP-[\textsuperscript{3}H]-Glc into insoluble material that was sensitive to cellulase, the presence of cellulose-like fibers with diameters of 4.3 nm in negative stain and cryo TEM images, and the presence of β-1,4-D-Glc in linkage analysis in the in vitro product confirm the synthesis of cellulose microfibrils\textsuperscript{iv}. In addition to the
synthesis of cellulose, there was some evidence for the production of \( \beta \)-1,3-glucan chains. The presence of a 110-140 kDa fragment of the 220 kDa \( \beta \)-1,3-glucan synthase of \emph{Pichia} and observed partial digestion of \emph{in vitro} synthesized insoluble polymers by \( \beta \)-1,3-glucanase are consistent with callose synthesis. However, we observed no fibers in negative stain images that had jagged edges as reported for callose and the linkage analysis showed no evidence of \( \beta \)-1,3-D-glucose. These contradictory observations could be explained by variable presence of functional \emph{Pichia} \( \beta \)-1,3-glucan synthase in the PpCesA5 preparations of this study. Such variation was seen for preparations of PttCesA8 of hybrid aspen (Purushotham et al., 2016).

A key aspect of our previous report on PttCesA8 and this report on PpCesA5 is that a single isoform of plant CesA can alone produce \( \beta \)-1,4-glucan chains of cellulose. These observations force new considerations regarding the prior models of CSC composition. We note that the CSC is a complex of CesA proteins, arranged in tightly bound ‘lobes’ that are more loosely contained within mostly hexamers (sometimes pentamers) of lobes (Kimura et al., 1999; Desprez et al., 2007; Nixon et al., 2016). Recent evidence strongly suggests that each lobe is a trimer of CesA, and until recently these trimers were assumed to be heterotrimers of different CesA isoforms, the exact isoforms depending on whether the CSC is involved in making cellulose for primary versus secondary cell walls (Cosgrove, 2005). The genetic redundancy has been interpreted as a way plants provide differential regulation of cellulose synthesis in specific tissues and developmental events (McFarlane et al., 2014). The results presented here for CesA5 from the non-vascular plant \emph{P. patens} and previously for CesA8 from the vascular plant hybrid aspen undeniably show that a single isoform is capable of forming cellulose microfibrils\(^\text{iv}\) in \emph{in-vitro} systems. This is thus true for CesAs known for contributing \emph{in vivo} to the production of primary (PpCesA5) or secondary (PttCesA8) cell walls. These observations of single isoforms making cellulose microfibrils\(^\text{iv}\) are consistent with reported 1:1:1 stoichiometric presence of CesA isoforms needed for normal primary and secondary cell wall formation (Gonneau et al., 2014; Hill et al., 2014), but they do raise the possibility of homomeric complexes.

If one only considers the moss \emph{P. patens}, it could be argued that homomeric lobes may be restricted to non-vascular plants. This would be reasonable because the \emph{P. patens}’ genome has seven nearly identical CesA genes (Roberts and Bushoven, 2007) and also has CSCs in the plasma membranes similar to the rosettes seen in vascular plants (Roberts et al., 2012; Nixon et al., 2016). Based at least in part on such observations, it was predicted that a single CesA
isoform might be able to substitute for other CesAs in the CSCs of *P. patens*. Consistent with that prediction, single knock-out of *PpCesA6* or *PpCesA7* showed no phenotype, only the double knock-out resulted in reduction of stem lengths (Wise et al., 2011). However, our concurrent study of hybrid aspen PttCesA8 that was also heterologously expressed in *Pichia* and reconstituted into proteoliposomes makes it clear that only one isoform of CesA from vascular plants is sufficient to yield cellulose microfibrils (Purushotham et al., 2016). In both the moss and poplar studies, CesA in its detergent-solubilized form failed to incorporate Glc from UDP-Glc into glucan chains, regaining this function only when incorporated into proteoliposomes. The amino acid sequences of CesA proteins in hybrid aspen are diverse like those in Arabidopsis (Djerbi et al., 2004). Thus, we propose that CesA proteins will generally need to be in a lipid bilayer to adopt functional conformation(s), and that they can assemble into homotrimer as well as heterotrimer lobes (the latter not yet having been demonstrated *in vitro*).

In plants, the individual β-1,4 glucan chains made by CesA proteins are assembled into microfibrils, and these into macrofibrils or higher order bundles, such as those noted in atomic force microscopic images of onion epidermal cell walls (Zhang et al., 2016). The relationship between trimeric lobes in rosettes and assembly of crystalline cellulose microfibrils is not yet understood. It is very intriguing that we observe micro- and macrofibrils in both the CesA studies, including the one reported here for CesA5 and in the CesA8 studies presented earlier (Purushotham et al., 2016). Such microfibrils are not made by the bacterial cellulose synthase BcsA-BcsB, even when present at high concentrations (Purushotham et al., 2016). However, when BscA-BcsB was immobilized on a nickel-film, followed by reaction with UDP-Glc, fibrillar cellulose was produced (Basu et al., 2016; Basu et al., 2017). This suggests that close proximity of cellulose synthase proteins is required and sufficient for formation of microfibrils, and we infer that such close proximity is present in the membrane-reconstituted moss and hybrid aspen synthases CesA5 and CesA8.

The lateral dimension of an individual cellulose glucan chain is lower than 0.5 nm, which makes isolated chains invisible in negative stain TEM images. As shown in Fig. 6, and Fig. 7, variably thick fibers are seen, and these tend to coalesce into larger fibrils. While negative stain images do not give reliable size estimates, these relative differences in size are clearly present. We think it possible that a homotrimer of CesA5 or CesA8 is capable of making an elemental fibril of three glucan chains, and that these can further assemble to form more typical...
microfibrils of about 4.3 to 4.8 nm in diameter. The estimated diameter presented here from cryo-TEM images is reliable, given the absence of any staining, and such sizes are consistent with diameters reported for plant cell walls (Ha et al., 1998; Cosgrove, 2005; Thomas et al., 2013; Zhang et al., 2016). This raises the possibility that homotrimer lobes could be organized into higher order rosettes as the microfibrils\textsuperscript{iv} are formed. It is simultaneously possible that the dimerization propensity of N-terminal domains of CesA could help organize rosette formation by bringing together lobes (Kurek et al., 2002). Indeed, truncation of N-terminal Zn\textsuperscript{2+}-binding domains from PttCesA8 resulted in the formation of few or no microfibrils despite significant production of β-1,4-glucan chains (Purushotham et al., 2016). Future use of these in vitro systems may reveal the membrane distributions of wild-type and mutant CesAs before, during and after catalysis with UDP-Glc and thus allow testing these hypotheses.

**Conclusions**

PpCesA5 of the moss *P. patens* was successfully expressed heterologously in *Pichia* and partially purified, followed by reconstitution into proteoliposomes. Proteoliposomes bearing PpCesA5 synthesize glucan chains that assembled into higher order cellulose microfibrils\textsuperscript{iv} and macrofibrils\textsuperscript{iv} or bundles\textsuperscript{iv} comparable to what is seen for a vascular plant CesA. We discuss how formation of cellulose microfibrils from single isoforms of CesA could be attributable to close proximity of CesA proteins in lipid bilayers, formation of higher order complexes among CesAs, or a combination of that and feedback from glucan chain crystallization – these assembly processes remain to be investigated. These systems for reconstitution of functional cellulose synthases in vitro serve as foundations for further studies on the synthesis of cellulose by CesA proteins and the assembly of nascent glucan chains into elementary fibrils, microfibrils and microfibril bundles in the absence or presence of matrix polysaccharides. Access to purified, membrane-embedded, functional CesAs may also yield structures of CesA from non-vascular and vascular plants.

**Materials and Methods**

**Cloning and transformation of *PpCesA5***
The *PpCesA5* gene was amplified by PCR using a primer set designed to add 12 x His tag at the C-terminus: 5’-ACTAATCAAGGTACCATGGAGGCTAATGCAGGCCTTATT-3’ / 5’-ACAATAGCCGCCGCTCAGTATGGTGATGGTGATGGTGATGGTGATGGTGATGGGAC-3’. Once synthesized, the PCR product was digested with *Kpn*I and *Not*I restriction enzymes. The resulting fragment was ligated into *Kpn*I and *Not*I-linearized pPICZ A vector. Five µg of pPICZ A plasmid containing *PpCesA5* was used to transform into the *P. pastoris* strain SMD1168H. For preparation of the SMD1168H cells for transformation, a colony grown on YPDS plate containing 1% yeast extract, 2% peptone and 2% glucose was inoculated into 50 mL YPDS liquid medium, followed by culture at 30°C by shaking until growth reached early stationary phase (~0.6-2 x 10^8 cells/mL). Cells were harvested by centrifugation at 3000 rpm for 5 min at 4°C, followed by washing with 40 mL ice-cold sterile water and centrifugation at 2500 rpm for 5 min at 4°C. After washing with 20 mL ice-cold water, cells were resuspended in 5 mL ice-cold sorbitol solution (2% sorbitol in water) and pelleted at 2000 rpm for 5 min at 4°C. The cells were resuspended in 150 µL ice-cold sorbitol solution, from which 40 µL of cells were mixed with 5 µg *PmEl*-linearized DNA in a pre-chilled electroporation cuvette. Electroporation was performed at 1.5 kV, 25 µF and 200 Ohms for 5 sec, which was immediately followed by addition of 1 mL 1 M ice-cold sorbitol solution. Transformed cells were screened on YPDS medium containing 100 µg/ml of Zeocin.

The catalytically inactive *PpCesA5* construct was generated using the QuikChange mutagenesis kit (ThermoFisher Scientific) following the manufacturer’s protocol. Primers used for mutagenesis are: 5’-CTGTCACCGAGAATATTCTGACGG-3’ and 5’-CCGTCAGAATATTCTCGGT GACAG-3’

**Enrichment of PpCesA5 and PttCesA8**

*PpCesA5* and *PttCesA8* were partially purified and reconstituted to proteoliposomes following the method previously described (Purushotham et al., 2016). The *Pichia pastoris* strain expressing *PpCesA5* or *PttCesA8* was grown on YPDS medium at 30°C overnight. A colony was inoculated into a 500 mL pre-culture medium containing 1% yeast extract, 2% peptone, 1% glycerol, 0.34% yeast nitrogen base, 1% ammonium sulfate and 0.00004% biotin, followed by shaking incubation at 30°C overnight. Cells were collected by centrifugation at 2600 x g at 4°C for 15 min and resuspended in 1 L of induction medium (1% yeast extract, 2% peptone, 0.7%
methanol, 0.34% yeast nitrogen base, 1% ammonium sulfate and 0.00004% biotin) to an OD$_{600}$ of 0.5, followed by shaking incubation at 30°C for 24 h. Grown cells were collected by centrifugation and frozen at -80°C until use. Frozen cells were thawed in a Cell Resuspension Buffer containing 20 mM Tris-HCl pH 7.5, 1 M sorbitol, 1x EDTA-free protease inhibitor tablet (Thermo Scientific) and 1 mM phenylmethylsulfonyl fluoride (PMSF) and lysed by three passes using a microfluidizer at 20,000 psi. The lysate was centrifuged at 10,000 g at 4°C for 10 min and its supernatant was subjected to ultracentrifugation at 200,000 x g at 4°C for 2 h. The pellet corresponding to vesicle fraction was resuspended in 50 mL Membrane Resuspension Buffer containing 150 mM sodium phosphate pH 7.5, 100 mM sodium chloride, 40 mM n-dodecyl-β-D-maltoside (DDM), 10% glycerol, 1x EDTA-free Protease inhibitor tablet (Thermo Scientific) and 1 mM PMSF, followed by gentle agitation at 4°C for 1 h. Insoluble materials removed by ultracentrifugation at 200,000 x g at 4°C for 40 min. The obtained membrane protein fraction was incubated with 5 mL pre-equilibrated TALON Superflow Resin (GE Healthcare) with gentle agitation at 4°C overnight. The mix was applied to an empty column and washed with in a series of 15 mL washing buffer (150 mM sodium phosphate pH 7.5, 100 mM sodium chloride, and 1 mM LysoFos Choline Ether 14) containing 20, 40, 60 or 80 mM imidazole. Protein was eluted by 15 mL washing buffer containing 350 mM imidazole. Imidazole was reduced to less than 0.5 mM by repeated concentration/dilution cycles through a Vivaspin 20 desalting column (30 kDa cutoff, GE Healthcare). All fractions were examined by western blot analysis.

**Reconstitution of proteoliposomes**

Chloroform was evaporated from a pre-dissolved yeast total lipid extracts (Avanti Polar Lipids) by a stream of nitrogen gas, which was then completely dried by overnight incubation in a vacuum chamber. The lipid (100 mg) was resuspended in 1 mL of 120 mM lauryldimethylamine oxide (LDAO), solubilized by heating at 60°C for 1 h, and stored at -20°C until use. Bio-Beads SM-2 Adsorbent Media (Bio-Rad) was washed in deionized water at room temperature for 10 min and dried on a filter paper. Six hundred μL of partially purified PpCesA5 or PttCesA8 was mixed with 400 μL yeast lipid, to which dried Bio-Beads SM-2 Adsorbent Media was added to 75% of the total volume. This mix was incubated overnight with gentle rotation at 4°C. Reconstitution was determined by visual turbidity.
Immunoblot analysis

Protein fractions were separated on a 4-12% gradient PAGE gel (GenScript) and transferred to a nitrocellulose membrane (GE Healthcare), which was hybridized with one of three monoclonal PpCesA5-specific antibodies raised against the synthetic peptides CKFSRKKTAPTRSDS, CGHSGGHDTDGNELP, or CAQKVPEEGWTMQDG (GenScript). The hybridized blot was incubated with secondary antibody conjugated with horseradish peroxidase (HRP). Chemiluminescent signals generated by Super Signal West Dura Extended Duration Substrate (Thermos Scientific) were detected by a GelLogic 4000 Pro System (Carestream Health).

Activity assay

In general, reconstituted proteoliposome (PL) was mixed with a reaction buffer containing 20 mM Tris-HCl, pH 7.5, 100 mM sodium chloride, 10% glycerol, 20 mM manganese chloride, 3 mM UDP-Glc and 0.25 μCi UDP-[3H]-Glc, followed by incubation at 37°C for 3 h. However, assays with alternate components and conditions were performed as indicated in each experiment. Reactions were stopped by adding triton X-100 to a final concentration of 0.1%. The reaction mix was then centrifuged at 15,000 rpm for 20 min to collect the product, which was resuspended in 20 μL of cellulose resuspension buffer containing 20 mM Tris-HCl, pH 7.5, 100 mM sodium chloride and 10% glycerol and applied to a descending chromatography paper (Whatman 2MM) using 60% ethanol. Following air-dry, the chromatography paper was submerged in 5 mL ScintiVerse BD Cocktail (Fisher Scientific), and radioactivity was measured in a Beckman Coulter LS6500 Scintillation counter. To prepare samples for electron microscopy, proteoliposomes were mixed with a reaction buffer containing 20 mM Tris-HCl, pH 7.5, 100 mM sodium chloride, 10% glycerol, 20 mM manganese chloride and 3 mM UDP-Glc and incubated at 37°C for 3 h, which was followed by treatment with 0.1% triton X-100.

Kinetic analysis

Reactions were performed with 20 mM MnCl₂ and 0 to 3.5 mM UDP-Glc. A fourfold concentrated stock solution of 20 mM UDP-Glc and 1.0 μCi UDP-[3H]-Glc were diluted to the final substrate concentration required in each reaction. After incubation for 30 min, the reactions
were treated as described above. Untransformed data were fit to the Michaelis-Menton equation to extract parameter values.

**Transmission Electron Microscopy (TEM)**

400 mesh Glider Copper grids (Ted Pela) were coated by evaporation of carbon using Denton Vacuum 502B carbon coater. 3.5 μL of prepared sample was loaded onto a glow-discharged grid, incubated for 1 min and negatively stained with 10 serial drops of 0.75% uranyl formate on a parafilm. TEM images were taken using an FEI Tecnai 12 Spirit BioTWIN TEM [FEI; 120 kV; 6.3 mm spherical aberration (Cs); 56 K magnification; 4 K × 4 K Eagle CCD camera; located at the Huck Institutes of the Life Sciences, Penn State University].

**Cryo-EM analysis**

To measure the width of cellulose microfibrils, *in vitro* cellulose microfibril products were applied to Cryo EM analysis as described (Grassucci et al., 2008). Never-dried sample was loaded on a QUANTIFOIL holey carbon grid (300 mesh; Ted Pella), blotted for 3 sec and vitrified with a Vitrobot (FEI; at the Huck Institute of the Life Sciences, Penn State University).

**Cellulase sensitivity assay**

For assays of incorporation of radioactive UDP glucose, the *in vitro* products were incubated with 10 U of β-1,4-glucanases or β-1,3-glucanases (Megazyme) at 37 °C for 1 h. For TEM observation, *in vitro* produced cellulose microfibrils were incubated with a total of 150 ng highly purified cellulase mix containing Cel6A, Cel7A and Cel7B proteins (kindly provided by Giridhar Poosarla, Penn State University), which was incubated at 50 °C. Sample was taken for negative staining.

**Linkage analysis**

*In vitro* products were pre-treated for linkage analysis as described (Purushotham et al., 2016). Briefly, 2% SDS and 1 mg/ml proteinase K were used to eliminate proteins, followed by washing twice with water, treatment with hexane, and then freeze-drying. Dried samples were treated with chloroform/methanol followed by washing with 70% ethanol, and then a solution containing 50 mM Tris-HCl, 2% SDS, 10 mM Na-EDTA, 40 mM β-mercaptoethanol was used.
to remove residual proteins, followed by washing 3 times with 70% ethanol. To eliminate
*Pichia*-derived glycogen-like polysaccharide, the samples were subjected to α-amylase and
amyloglucosidase, and then washed with 70% ethanol. Subsequently, samples were freeze-dried.
These treatments greatly reduced the mass (from 10 to 2 mg for wild-type and 33 to 4.2 mg for
PpCesA5(D782N). The prepared samples were permethylated to alditol acetates and analyzed by
GC/EI-MS following the method previously described (Omadjela et al., 2013).

**Image analysis**

All image analyses were performed using ImageJ (National Institute of Health). In order to
measure areas occupied by microfibrils, the Set Scale option was used to set the real length, the
Polygon selection tool was used to define areas along the microfibrils and the Measure tool was
used to measure area. Measured areas were analyzed by mean and standard error. For
measurement of thickness of individual microfibrils, the ‘Straight Line’ tool was used to define
widths and the ‘Measure’ tool was used for measurement of thickness.

**Mass spectrometry**

For mass spectrometry analysis, partially purified protein sample was separated by SDS-
PAGE, followed by coomassie staining. The gel region corresponding to 110-140 kDa was cut
out and incubated with 5 mM Tris(2-carboxyethyl)phosphine hydrochloride (TCEP)/0.8% (w/v)
ammonium bicarbonate at 60°C for 10 min, followed by washing with 100 mM iodoacetamide
(IAA)/0.8% (w/v) ammonium bicarbonate at 37°C for 15 min. MS spectra were taken from 60
minute gradient from an Eksigent NanoLC-Ultra-2D Plus and Eksigent LC through a 200 µm x
0.5 mm Chrom XP C18-CL 3 µm 120 Å Trap Column and elution through a 75 µm x 15 cm
NanoLCMS C18-CL 2.6 µm 120 Å Nano LC Column. Sciex 5600 TripleTOF settings were:
Parent scan acquired for 250 msec, and then up to 50 MS/MS spectra were acquired over 2.5
seconds for a total cycle time of 2.8 sec. Gas 1 (Nitrogen) = 7 and Gas 3 (Nitrogen)= 25. Mass
spectrometry analysis was performed by the Proteomics and Mass Spectrometry Core Facility of
Penn State Hershey.

**Supplemental materials**

Figure S1 – Diagram of PpCesA5
Acknowledgments

We thank Missy Hazen and John Cantolina (Huck Institutes of the Life Sciences, Penn State University) for technical help with TEM and Giridhar Poosarla (Penn State University) for providing Cel6A, Cel7A and Cel7B proteins.

Figure legends

Figure 1. Partial purification of heterologously expressed PpCesA5 from Pichia. Membrane proteins were isolated from P. pastoris expressing PpCesA5::12xHis and purified using TALON (cobalt) resin. A, SDS-PAGE gel stained with coomassie. Lane 1, total protein extract; Lane 2, non-membrane proteins; Lane 3, membrane proteins; Lane 4, flow through; Lane 5, washing fraction 1 (20 mM imidazole); Lane 6, washing fraction 2 (40 mM imidazole); Lane 7, washing fraction 3 (60 mM imidazole); Lane 8, washing fraction 4 (80 mM imidazole); Lane 9, Elution (350 mM imidazole). B, Immunoblot of gel as in (A) using PpCesA5 specific antibody #1. The arrows point to bands of mass expected for full length PpCesA5. See Supplemental Figure S1 for epitope location.

Figure 2. Functional characterization of PpCesA5 reconstituted into liposomes. A, Partially purified PpCesA5 protein was reconstituted into liposomes using yeast total lipids and detergent
removal by BioBeads®, and then assayed for function by incubation with UDP-Glc and UDP-[3H]-Glc in the presence of the indicated cations. After stopping reactions by the addition of Triton X-100, the product was applied to descending paper chromatography, from which insoluble material at the origin was evaluated in a scintillation counter. Error bars represent standard deviation (n=3). B, Time course. C, Kinetic analysis. Michaelis-Menten parameter estimates and 95% confidence intervals: $K_M$=137 (102,179) µM; $V_{max}$=103 (96,111) relative units. D, Treatments with β-1,4-glucanase (cellulase) or β-1,3-glucanase (callase).

Figure 3. PpCesA5 in proteoliposomes produces microfibrils. A-F, Proteoliposomes bearing PpCesA5 were incubated with UDP-Glc, during which samples were taken at designated time points for imaging by negative stain TEM. Representative random images are shown. Arrows indicate microfibrils, magnified in insets for panels B and C. G, Negative control – representative image for a reaction incubated for 180 min without UDP-Glc. Scale bars - 100 nm. H, Plots of the mean and standard error for the areas occupied by microfibrils in 40 random images. NC, negative control.

Figure 4. Microfibrils are sensitive to cellulose-specific glucosidases. To examine if the microfibrils were made of cellulose, in vitro products were subjected to cellulase treatment and sampled at the designated time points by negative staining. Randomly taken TEM images were analyzed for areas occupied by fibrils. A-F, Representative images for reaction times of 0, 5, 10, 30, 45 and 60 min. G, Negative control – representative image for a reaction incubated for 60 min without cellulase. Arrows indicate microfibrils. Arrowheads in A and G indicate bundled microfibrils. Scale bars - 100 nm. H, Plots of the mean values of the areas occupied by microfibrils. Gray bars and white bars represent total microfibrils and bundled microfibrils, respectively. Error bars indicate standard errors (n=40). NC, negative control.

Figure 5. Microfibrils contain 1-4 linked glucose. In vitro products were pretreated with 2% SDS and proteinase K, followed by α-amylase and amyloglucosidase to remove Pichia-derived α-1,4-glucan chains. Thereafter, the sample was permethylated to alditol acetates, and then subjected to GC/MS analysis. The resulting spectra are compared with reference peaks. Identities
of each peak are marked. In particular, terminal-Glc (t-Glc) and 4-linked Glc (4-Glc) were detected at 9.348 min and 19.059 min, respectively.

**Figure 6.** Cellulose microfibrils are produced from PpCesA5 in proteoliposomes and coalesce into higher order macrofibrils. Proteoliposomes bearing PpCesA5 were incubated with UDP-Glc, followed by negative staining and TEM analysis. Arrows indicate where cellulose microfibrils were attached to proteoliposomes (A and B). Arrowheads indicate where microfibrils coalesced into macrofibrils (C-F). Scale bars - 100 nm.

**Figure 7.** Cellulose microfibrils from PttCesA8-proteoliposomes coalesce into higher order macrofibrils. Proteoliposomes bearing PttCesA8 were incubated with UDP-Glc, followed by negative staining and TEM analysis. Arrows indicate where cellulose microfibrils coalesced into macrofibrils. Scale bars - 100 nm.

**Literature Cited**


Saxena IM, Brown Jr RM (1997) Identification of cellulose synthase(s) in higher plants: Sequence analysis of processive β-glycosyltransferases with the common motif "D, D, D35Q(R,Q)XRW". Cellulose 4: 33-49


Figure 1. Partial purification of heterologously expressed PpCesA5 from *Pichia*. Membrane proteins were isolated from *P. pastoris* expressing PpCesA5::12xHis and purified using TALON (cobalt) resin. A, SDS-PAGE gel stained with coomassie. Lane 1, total protein extract; Lane 2, non-membrane proteins; Lane 3, membrane proteins; Lane 4, flow through; Lane 5, washing fraction 1 (20 mM imidazole); Lane 6, washing fraction 2 (40 mM imidazole); Lane 7, washing fraction 3 (60 mM imidazole); Lane 8, washing fraction 4 (80 mM imidazole); Lane 9, Elution (350 mM imidazole). B, Immunoblot of gel as in (A) using PpCesA5 specific antibody #1. The arrow indicates the band corresponding to PpCesA5. See Supplemental Figure S1 for epitope location.
Figure 2. Functional characterization of PpCesA5 reconstituted into liposomes. A, Partially purified PpCesA5 protein was reconstituted into liposomes using yeast total lipids and detergent removal by BioBeads®, and then assayed for function by incubation with UDP-Glc and UDP-[\(^{3}H\)]-Glc in the presence of the indicated cations. After stopping reactions by the addition of Triton X-100, the product was applied to descending paper chromatography, from which insoluble material at the origin was evaluated in a scintillation counter. Error bars represent standard deviation (n=3). B, Time course. C, Kinetic analysis. Michaelis-Menten parameter estimates and 95% confidence intervals: KM=137 (102,179) \(\mu\)M; Vmax=103 (96,111) relative units. D, Treatments with \(\beta\)-1,4-glucanase (cellulase) or \(\beta\)-1,3-glucanase (callase).
Figure 3. PpCesA5 in proteoliposomes produces microfibrils. A-F, Proteoliposomes bearing PpCesA5 were incubated with UDP-Glc, during which samples were taken at designated time points for imaging by negative stain TEM. Representative random images are shown. G, Negative control – representative image for a reaction incubated for 180 min without UDP-Glc. Scale bars - 100 nm. H, Plots of the mean and standard error for the areas occupied by microfibrils in 40 random images. NC, negative control.
Figure 4. Microfibrils are sensitive to cellulose-specific glucosidases. To examine if the microfibrils were made of cellulose, in vitro products were subjected to cellulase treatment and sampled at the designated time points by negative staining. Randomly taken TEM images were analyzed for areas occupied by fibrils. A-F, Representative images for reaction times of 0, 5, 10, 30, 45 and 60 min. G, Negative control – representative image for a reaction incubated for 60 min without cellulase. Scale bars - 100 nm. H, Plots of the mean values of the areas occupied by microfibrils. Error bars indicate standard errors (n=40). NC, negative control.
Figure 5. Microfibrils contain 1-4 linked glucose. In vitro products were pretreated with 2% SDS and proteinase K, followed by α-amylase and amyloglucosidase to remove Pichia-derived α-1,4-glucan chains. Thereafter, the sample was permethylated to alditol acetates, and then subjected to GC/MS analysis. The resulting spectra are compared with reference peaks. Identities of each peak are marked. In particular, terminal-Glc (t-Glc) and 4-linked Glc (4-Glc) were detected at 9.348 min and 19.059 min, respectively.
Figure 6. Cellulose microfibrils are produced from PpCesA5 in proteoliposomes. Proteoliposomes bearing PpCesA5 were incubated with UDP-Glc, followed by negative staining and TEM analysis. Arrows indicate where cellulose microfibrils were attached to proteoliposomes. Scale bars - 100 nm.
Figure 7. Cellulose microfibrils from PpCesA5 proteoliposomes coalesce into higher order macrofibrils. Proteoliposomes bearing PpCesA5 were incubated with UDP-Glc, followed by negative staining and TEM analysis. Arrows indicate where cellulose microfibrils coalesced into macrofibrils. Scale bars - 100 nm.


Saxena IM, Brown Jr RM (1997) Identification of cellulose synthase(s) in higher plants: Sequence analysis of processive β-glycosyltransferases with the common motif "D, D, D35Q(R,Q)XRW". Cellulose 4: 33-49


