2018

Cellulose synthase 'class specific regions' are intrinsically disordered and functionally undifferentiated

Tess R. Scavuzzo-Duggan

Arielle M. Chaves
University of Rhode Island, ariellec@uri.edu

See next page for additional authors

Citation/Publisher Attribution
Scavuzzo-Duggan TR, Chaves AM, Singh A, Sethaphong L, Slabaugh E, Yingling YG, Haigler CH, Roberts AW (2018) Cellulose synthase 'class specific regions' are intrinsically disordered and functionally undifferentiated. *J Integr Plant Biol* 60: 481–497. doi: 10.1111/jipb.12637 Available at: https://doi.org/10.1111/jipb.12637

Authors
Tess R. Scavuzzo-Duggan, Arielle M. Chaves, Abhishek Singh, Latsavongsakda Sethaphong, Erin Slabaugh, Yaroslava G. Yingling, Candace H. Haigler, and Alison Roberts

Title: Cellulose synthase 'class specific regions' are intrinsically disordered and functionally undifferentiated

The 'class specific regions' are intrinsically disordered and interchangeable between different classes of *Physcomitrella patens* cellulose synthases

Running title: Intrinsic disorder in cellulose synthases

Highlight:

Tess R. Scavuzzo-Duggan[1*], Arielle M. Chaves[1*], Abhishek Singh[3*], Latsavongsakda Sethaphong[3], Erin Slabaugh[2], Yaroslava G. Yingling[3], Candace H. Haigler[2], and Alison W. Roberts[1**]

[*]The first three authors contributed equally to this work.

[1]Department of Biological Sciences, University of Rhode Island, 120 Flagg Road, Kingston, RI 02881 USA

[2]Department of Crop and Soil Sciences and Department of Plant and Microbial Biology, North Carolina State University, Raleigh, NC 27695 USA

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

[**]Correspondence: Alison Roberts (aroberts@uri.edu; +1 (401) 874-4098)

Email addresses:
TSD trsduggan@berkeley.edu
AMC ariellec@uri.edu
AS asingh7@ncsu.edu
SS sethapho@usna.edu
ES erindavis257@gmail.com
YGL yara_yingling@ncsu.edu
CHH candace_haigler@ncsu.edu
AWR aroberts@uri.edu

ABSTRACT

Cellulose synthases (CESAs) are glycosyltransferases that catalyze formation of cellulose microfibrils in plant cell walls. Seed plant CESA isoforms cluster in six phylogenetic clades, whose non-interchangeable members play distinct roles within Cellulose Synthesis Complexes (CSCs). A 'class specific region' (CSR) with higher sequence similarity within vs. between functional CESA classes has been suggested to contribute to specific activities or interactions of different isoforms. We investigated CESA isoform specificity in the moss *Physcomitrella patens* (Hedw.) B. S. G. to gain evolutionary insights into CESA structure/function relationships. Like seed plants, *P. patens* has oligomeric rosette-type CSCs, but the PpCESAs diverged independently and form a separate CESA clade. We showed that *P. patens* has two functionally distinct CESAs classes based on the ability to complement the gametophore-negative phenotype of a *ppcesa5* knockout line. Thus, non-interchangeable CESA classes evolved separately in mosses and seed plants. However, testing of chimeric moss *CESA* genes for complementation demonstrated that functional class-specificity is not determined by the CSR. Sequence analysis and computational modeling showed that the CSR is intrinsically disordered and contains predicted Molecular Recognition Features, consistent with a possible role in CESA oligomerization and explaining the evolution of class-specific sequences without selection for class-specific function.

INTRODUCTION

Cellulose is a major component of plant cell walls and is also produced by widely divergent eukaryotic and prokaryotic organisms. In all of these organisms, the β-1,4-linked glucan chains of cellulose are polymerized by family 2 glycosyl transferases known collectively as cellulose synthase catalytic subunits. These membrane-spanning proteins share a similar central catalytic core, although other aspects of the sequences and structures vary among taxa (Pear et al. 1996; Morgan et al. 2013; Sethaphong et al. 2013). In land plants and their closest algal relatives, multiple cellulose synthases (called CESAs) are organized into six-lobed oligomeric complexes known as rosettes (Mueller and Brown 1980; Kimura et al. 1999). The organization of rosettes and other types of Cellulose Synthesis Complexes (CSCs) is correlated with microfibril structure in diverse organisms (Tsekos 1999).

In rosette-forming CESAs, the large central cytosolic region includes the catalytic domain and two plant-specific domains called the plant-conserved region (P-CR; Pear et al. 1996) and class-specific region (CSR; Vergara and Carpita 2001). Although we do not yet know the roles of the P-CR or CSR, their location on the periphery of the catalytic domain (Sethaphong et al. 2013; Olek et al. 2014) and presence

only in rosette-forming CESAs suggest that they could participate in CESA-CESA interactions that are important for rosette CSC assembly and stability (Somerville 2006; Olek et al. 2014; Slabaugh et al. 2014). The 126 amino acid P-CRs are conserved both in sequence and structure across diverse CESA isoforms (Pear et al. 1996; Sethaphong et al. 2016; Rushton et al. 2017). In contrast, the CSRs have variable lengths (e.g. 86-106 amino acids in Arabidopsis) and poorly conserved sequences (Pear et al. 1996; Vergara and Carpita 2001; Carroll and Specht 2011; Kumar et al. 2016). Correspondingly, the predicted CSR structures vary substantially between AtCESA isoforms, although they are consistently rich in randomly oriented alpha helices (Sethaphong et al. 2016). In addition to the P-CR and CSR, CESAs have two additional structural elements that are absent from other cellulose synthases. These include a conserved Zn-binding RING domain near the N-terminus and a variable region between the RING domain and the first transmembrane helix referred to as hypervariable region I (HVRI) (Pear et al. 1996). The RING domain has been implicated in CESA dimerization in vitro (Kurek et al. 2002; Xu and Joshi 2010).

Arabidopsis has six functionally distinct classes of CESAs, classes 1, 3 and 6-like required for primary cell wall deposition (called hereafter primary CESAs) and classes 4, 7 and 8 required for secondary cell wall deposition (secondary CESAs) (Taylor et al. 2003; Desprez et al. 2007; Persson et al. 2007; McFarlane et al. 2014). The strong phenotypes that result from mutating *AtCESA1* (Arioli et al. 1998), *AtCESA3* (Fagard et al. 2000), or any one of the secondary *AtCESA*s (Taylor et al. 2003) are consistent with the possibility that each isoform occupies one of three unique positions within each lobe of its respective CSC. Complementation experiments have confirmed very limited interchangeability of AtCESAs from different classes (Carroll et al. 2012). Thus, the functions of the CESA proteins include their biosynthetic output, as well as their interactions with their required partners. Phylogenetic and expression analyses support conservation of these distinct functional classes within six CESA clades shared by angiosperms (Holland et al. 2000; Samuga and Joshi 2002; Tanaka et al. 2003; Burton et al. 2004; Djerbi et al. 2004; Djerbi et al. 2005; Ranik and Myburg 2006; Suzuki et al. 2006; Kumar et al. 2009; Burton et al. 2010; Song et al. 2010; Carroll and Specht 2011; Handakumbura et al. 2013; Zhang et al. 2014; Kaur et al. 2016; Kumar et al. 2016), conifers (Nairn and Haselkorn 2005; Yin et al. 2014), and ferns (Yin et al. 2014), but not lycophytes (Harholt et al. 2012).

Because rosettes composed of three non-interchangeable CESA isoforms seem to be conserved, efforts to identify protein domains important for CSC assembly have focused on class-specific regions with higher sequence similarity within functional CESA classes than between them (Carroll and Specht 2011; Kumar et al. 2016; Sethaphong et al. 2016). The CSR has been discussed in terms of potential interactions

between different isomers (Somerville 2006; Sethaphong et al. 2016) based on within-class sequence similarity (Vergara and Carpita 2001). Alternatively, CSCs may be homo-oligomeric as supported by in vitro microfibril synthesis by a single poplar CESA (Purushotham et al. 2016) and evidence that a single CESA isoform and rosette CSCs existed in the common ancestor of mosses and seed plants (Roberts and Bushoven 2007). Experimental modeling of CESA homotrimers has shown different possibilities, with CSRs either forming CESA-CESA contacts within lobes or residing on the periphery of lobes where they could generate CESA-CESA contacts between lobes or interface with other partner proteins (Sethaphong et al. 2013; Nixon et al. 2016; Sethaphong et al. 2016; Vandavasi et al. 2016; Rushton et al. 2017), However, none of these models are consistent with all available data (Vandavasi et al. 2016; Rushton et al. 2017).

The CESA family of the moss *Physcomitrella patens* (Hedw.) B. S. G. diversified independently and does not include members of the functionally distinct seed plant CESA clades (Roberts and Bushoven 2007; Roberts et al. 2012). The seven PpCESA isoforms cluster in two clades with conserved intron position. PpCESA3, PpCESA5, and PpCESA8 (hereafter referred to as clade A) share 7 introns and 81.7-95.0% sequence identity. PpCESA4, PpCESA6, PpCESA7, and PpCESA10 (clade B) share 12 introns and 89.7-99.8% sequence identity (Roberts and Bushoven 2007; Yin et al. 2009; Wise et al. 2011; Roberts et al. 2012). Little was previously known about the functional specialization of *P. patens* CESAs. Knockout mutants of *PpCESA5* fail to produce gametophores (Goss et al. 2012). However, neither double knockouts of *PpCESA3* and *PpCESA8* (Norris et al. 2017) nor quadruple knockouts of the clade B *PpCESAs* (Li et al., unpublished) share this phenotype. This is in contrast to Arabidopsis in which loss of any of the secondary CESAs results in an irregular xylem phenotype (Taylor et al. 2003) and loss of any class of primary CESAs is gametophytic or embryo lethal (Desprez et al. 2007; Persson et al. 2007). Also, no *PpCESA*s are strongly co-expressed (Tran and Roberts 2016), in contrast to the secondary *CESA*s in many vascular plants (Ruprecht et al. 2011). These observations lead us to test the functional interchangeability of the PpCESAs.

The easily scored phenotype of the *ppcesa*5 knockout (KO) mutant enables functional testing of CESAs by complementation analysis (Scavuzzo-Duggan et al. 2015; Slabaugh et al. 2015). In wild type *P. patens*, gametophores develop from bud initial cells that form from asymmetric division of caulonemal cells. Bud initial cells divide several times to produce buds, each with a single pyramidal apical initial cell that divides to produce leaf primordia (Harrison et al. 2009). The gametophore buds of *ppcesa5*KO are cellulose deficient and unable to divide and expand normally after the four-cell stage, which leads to formation of irregular tissue clumps instead of leafy gametophores (Goss et al. 2012). When expressed

under the control of a strong constitutive promoter, PpCESA5 complements the *ppcesa5*KO phenotype, restoring cellulose synthesis and normal gametophore development (Goss et al. 2012; Scavuzzo-Duggan et al. 2015). An assay based on this observation provides a readout for the efficacy of CESAs carrying engineered mutations (Scavuzzo-Duggan et al. 2015)

Here we show that clade A and clade B constitute functionally distinct classes of PpCESAs. Although their sequences are class specific, complementation assays showed that the CSRs of PpCESAs are interchangeable between classes. A resolution to this paradox is suggested by the results of structural modeling and sequence analysis, which show that the CSR is intrinsically disordered, thus providing an explanation for the evolution of sequence class-specificity in a region that can be exchanged between classes without impairing function.

RESULTS

P. patens has two functionally distinct CESA classes:

When expressed under the control of the strong constitutive *Act1* promoter, PpCESA5 complements the *ppcesa5*KO phenotype, restoring normal gametophore development (Goss et al. 2012; Scavuzzo-Duggan et al. 2015). When tested in the same way, both of the other clade A PpCESAs (PpCESA3 and PpCESA8) also restored gametophore production in a high percentage of stably transformed colonies (Fig. 1). In contrast, all members of Clade B (PpCESA4, PpCESA6, PpCESA7 and PpCESA10) were unable to rescue the gametophore-deficient mutant phenotype (Fig. 1). Western blot analysis confirmed that the clade B CESA proteins were expressed in the stably transformed lines (Fig. S1). These results indicate that Clade A and Clade B each comprise a functionally distinct CESA class, and that the Clade A CESAs are functionally interchangeable.

Divergent CSR sequences correlate with intrinsic protein disorder

As observed in seed plants (Vergara and Carpita 2001), the CSRs from the PpCESAs show more similarity within than between functional classes (Fig. 2), especially in the N-termini where the clade A CESAs are richer in cysteine and the basic amino acids lysine and arginine. The C-termini are more conserved, including an E(K/M)xFGxS motif that is also shared with Arabidopsis CESAs (Sethaphong et al. 2016). The CSRs of PpCESA6 and PpCESA7 are identical and only one was included in further analysis.

Structural analysis was performed on the CSRs to characterize structural elements potentially responsible for the functional differences between clade A and clade B PpCESAs. Initially, Rosetta was used to generate decoy structures (Fig. S2) as described previously (Sethaphong et al. 2016). For each sequence, an optimal predicted structure was identified by k-means clustering of the top 10 percent of the decoys in each population. The selected structure was further refined using all atom Molecular Dynamics (MD) to determine high resolution protein conformation states and their energy landscapes. To avoid trapping in local minima, we enhanced sampling using a hyperdynamics approach, which incorporates a harmonic boost to the potential energy function (protein force field) to smooth the potential energy surface (Miao et al. 2015). This accelerates transitions between low energy states and produces an accurate free energy profile. Based on their two-dimensional energy landscapes (Fig. 3) of root mean square distance (RMSD) vs. Radius of gyration (R_g), the CSRs from the PpCESAs (designated as CSRX where X is the PpCESA isoform) are predicted to be highly flexible. The energy landscapes include local minima (metastable states) separated by small energy barriers with one or more lowest energy regions (yellow, Fig. 3). Among all *P. patens* isoforms, CSR5 from clade A and CSR6/7 from clade B shared the broadest size distribution (R_g = 14 Å to 21 Å) and the most structural deviation (RMSD = 3 Å to 20Å) from initial conformation determined by ab initio modeling. This is indicative of diverse globular-extended conformations. In contrast, CSR3 and CSR8 displayed less conformational variation (R_g = 14 Å to 16 Å). Although the CSRs of clade B PpCESAs were predicted to be more flexible overall, CSR5 in clade A and CSR6/7 in clade B are both predicted to be highly flexible. Representative lowest energy structures consisting of short helical regions separated by flexible regions are shown as insets in Fig. 3.

High flexibility of the CSRs suggests the presence of disordered regions, which lack a fixed tertiary structure. Disordered regions are integral components of protein structure/function relationships (Tompa 2012) and can be identified through use of the Predictor of Natural Disordered Regions (PONDR) algorithm. PONDR VL-XT scores are based on three feed forward neural network algorithms including one trained on variously characterized long disordered regions (VL) and two trained on X-ray characterized terminal disordered regions (XT), using amino acid coordination number, hydropathy, and net charge as input attributes. The PONDR calculations, scaled 0 to 1 reflecting ideal order to ideal disorder with a threshold transition at 0.5 (Oldfield et al. 2005), show that all *P. patens* CSRs are predicted to be largely disordered, but each also contains a predicted ordered region corresponding to the conserved E(K/M)xFGxS motif (Fig. S3). Next we investigated whether the predicted disordered regions contain Molecular Recognition Features (MoRF), which can undergo disorder to order transition upon binding to a partner. Use of the Molecular Recognition Feature Predictor [MORFPRED; http://biomine.cs.vcu.edu/servers/MORFPRED/ (Disfani et al. 2012)] predicted two MoRFs for CSR5 (Fig.

4B, red blocks) as well the CSRs of other PpCESAs (Fig. S4). The predicted MoRF regions were 8-10 amino acids long, within the typical range of 5-25 amino acids (Disfani et al. 2012), and were separated by 20-25 amino acids. Based on our simulations, CSR5 is predicted to exist in several globular-extended conformations, but we did not observe disorder-to-order transitions in the absence of a binding partner (Fig. 4A). In a homology model of PpCESA5 using GhCESA1 as a template [excluding the 160 amino acid N-terminus and short C-terminus (Nixon et al. 2016)], we observed that the predicted MoRF region (amino acids 427-434) closest to the conserved E(K/M)xFGxS motif is in the form an α helix (α-MORF; Fig. 4A). The other predicted MoRF in CSR5 (amino acids 392-401) had only partial helical structure in a trimeric in silico assembly of PpCESA5 (Fig. 5).

We also measured the Root Mean Squared Fluctuations (RMSF) of amino acids determined during the MD simulations to investigate positional variance in mobility of CSR amino acids averaged over time. The normalized RMSFs for CSR5 isolated in solution, as part of a CESA monomer, and within a trimeric CESA assembly (Fig. 5A-B) are shown in Fig. 5C. Possible locations of MoRFs and the conserved E(K/M)xFGxS motif are highlighted in side and top views of a trimeric assembly of PpCESA5 (Fig. 5A-B). The α-MoRF region displays lower predicted mobility than the conserved E(K/M)xFGxS motif when the CSR is in isolation, part of the monomer, and in a trimeric assembly of monomers (Fig. 5C). The positional variance in mobility for the other CSRs is shown in Fig. S5. A Ramachandran plot of α-MORF residues (Fig. S6; MD trajectory data blue green) comparing one of the disordered states from MD simulation (black) and the ordered state from a trimeric assembly (red) is consistent with a complex energetic pathway for the transition.

The CSR is not responsible for isoform-specific function of PpCESAs:

To test the hypothesis that the CSR confers isoform specificity to the PpCESAs, we tested vectors in which the CSR of PpCESA5 was replaced with the CSR from the other PpCESAs (Fig. 6). As expected, chimeras of PpCESA5 containing the CSR of PpCESA3 (CESA5-CSR3) or the CSR of PpCESA8 (CESA5-CSR8) fully rescued the mutant phenotype (Fig. 7). However, PpCESA5 chimeras containing a CSR from any one of the Clade B PpCESAs (CESA5-CSR4, CESA5-CSR6/7, CESA5-CSR10) were also able to fully rescue the *ppcesa5*KO mutant phenotype (Fig. 7). Since the sequences of the CSRs of PpCESA6 and PpCESA7 are identical, these results indicate that the CSRs of all seven PpCESAs are functionally interchangeable with CSR5. Given that most of the sequence divergence within central cytosolic portions of CESA proteins occurs in the CSR, we replaced the entire central cytosolic region of PpCESA5 with the corresponding region from PpCESA4 in Clade B (CESA5-CAT4; Fig. 6). As shown in Fig. 7, the mutant phenotype was fully rescued.

To test whether cellulose content was completely restored by complementation with PpCESA chimeras, we used fluorescence microscopy to examine gametophores buds stained with Pontamine Fast Scarlet 4B (S4B; Anderson et al. 2010) in wild type, *ppcesa5*KO, and complementation lines. In wild type, positive control, and CESA5-CSR4 and CESA5-CSR7 complementation lines, gametophore buds appeared brighter than protonemal filaments. In contrast, the fluorescence intensity of buds and filaments was similar in *ppcesa5*KO and negative control lines (Fig. 8A). Quantitative analysis showed that the CESA5-CSR4 and CESA5-CSR7 complementation lines were significantly brighter than *ppcesa5*KO and negative controls, while being similar to wild type and positive controls (Fig. 8B). This confirms that CSRs from Clade B PpCESAs do not impair PpCESA5 activity.

The N-terminus allowed a chimeric clade B CESA to function in place of clade A PpCESA5:
Finally, we tested whether specific regions of PpCESA5 are sufficient to confer clade A-specific function on a PpCESAs from Clade B. Constructs that did not rescue the *ppcesa5*KO phenotype included the CSR, the P-CR and CSR together, the central cytosolic region, and the C-terminus of PpCESA5 (CESA4-CSR5, CESA7-CSR5, CESA4-PCRCSR5, CESA4-cat5 and CESA4-Cterm5; Fig. 9). Western blot analysis confirmed protein expression for non-rescuing vectors (Fig. S7). However, partial rescue was achieved when the N-terminus of PpCESA5 was swapped for the equivalent regions of PpCESA4 (CESA4-Nterm5). In this case, the number of colonies with gametophores was significantly different from both the PpCESA5 positive control and the empty vector negative controls. The CESA4-Nterm5, non-rescued lines had lower protein expression than rescued lines (Fig. S7).

DISCUSSION

Identifying CESA-CESA interfaces is important for understanding CSC assembly and developing strategies to modify cellulose microfibril structure. Although past efforts have focused on class-specific regions in the primary sequence of angiosperm CESAs, the first rosettes were apparently homo-oligomeric (Roberts et al. 2012) with all interfaces within the CSC occurring between identical subunits. Our results and analysis provide a unified view of rosette CSC structure over long evolutionary time. The computational and experimental data show that the CSR is intrinsically disordered and not responsible for class-specific function. As described below, the intrinsic disorder of the CSR : 1) explains evolution of sequence class specificity in the absence of selection for functional class specificity; and 2) implicates the CSR in CESA-CESA interactions that are not class-specific.

More than one functionally distinct CESA class is a common feature in extant land plants

The *P. patens* CESAs fall into two non-interchangeable classes as shown by complementation of *ppcesa5*KO by clade A (PpCESA3 and PpCESA8), but not by clade B PpCESAs (Fig. 1). More extensive diversification and specialization has produced six non-interchangeable CESA classes in seed plants (Taylor 2008). Nonetheless, both mosses and seed plants have evolved multiple CESA isoforms that differ in their ability to co-function with other CESA isoforms. On its face, this seems remarkable because the CESA families in the two lineages diversified independently and followed independent paths of functional differentiation (Roberts and Bushoven 2007; Yin et al. 2009; Roberts et al. 2012). However, according to the constructive neutral evolution hypothesis this outcome is a likely consequence of the replication of genes encoding proteins within multimeric complexes like CSCs. After gene replication, random mutations result in loss of interfaces between identical subunits and concomitant reliance on interfaces between non-identical subunits (Doolittle 2012; Finnigan et al. 2012). The two types of Arabidopsis CSCs, each composed of three non-interchangeable CESA isoforms, match the obligate hetero-oligomeric end state of this evolutionary pathway (Roberts et al. 2012). Because genome replications occurred more recently in the moss lineage (Rensing et al. 2007, 2016), *P. patens* CSCs may represent an intermediate state with only two functionally distinct types of subunits. Our ongoing efforts to determine the isoform composition of the *P. patens* CSCs may provide insight into the processes through which the obligate hetero-oligomeric CSCs of seed plants evolved.

Intrinsically disordered CSRs are functionally interchangeable despite class-specific sequences

There are two important consequences of the constructive neutral evolution scenario (Finnigan et al. 2012) with respect to understanding the functions of CESA isoforms within CSCs. First, it is expected that the differences among CESA isoforms impact only the position that they are able to occupy within the CSC, with no anticipated consequence for biosynthetic output. Second, interfaces between non-identical subunits can be expected to be class-specific at the sequence level. The CSR is class-specific based on analysis of sequence alignments in *P. patens* (Fig. 2) and seed plants (Vergara and Carpita 2001; Carroll and Specht 2011). This, along with evidence that CSRs may form interfaces between CESA subunits (Olek et al. 2014; Kumar et al. 2016; Sethaphong et al. 2016; Vandavasi et al. 2016), suggests a potential role for CSRs in isoform-specific CESA-CESA interactions (Somerville 2006). To the contrary, the CSRs of AtCESA1 and AtCESA3 (Sethaphong et al. 2016), the secondary AtCESAs (Kumar et al. 2016), and the PpCESAs (Fig. 5) are interchangeable in domain-swap experiments. Although we cannot rule out the possibility that CSRs contribute to class-specific function in some as yet undescribed cases, they are not universal determinants.

Evidence that the *P. patens* CSRs are intrinsically disordered include their broad energy landscapes containing several local minima (Fig. 3) and high PONDR scores (Figs. 3, S2). Intrinsically disordered regions evolve rapidly and are poorly conserved due to selection for maintenance of overall disorder (Brown et al. 2011; Schlessinger et al. 2011). Analysis of a previously published alignment of angiosperm sequences (Carroll and Specht 2011) shows that within-class identities in the structured P-CR are greater than 82%, whereas within-class identities in the CSR are as low as 29% with high similarities only between paralogs or sequences from close relatives (see Supplementary Table 2). Differences are due to insertions, deletions, and expansion of single amino acid repeats, as expected for intrinsically disordered regions due to selection for maintenance of disorder (Brown et al. 2011; Schlessinger et al. 2011). Given rapid evolution, class-specific sequences could have evolved in the absence of selection for class-specific function because the CESAs within each class shared longer evolutionary histories with each other than with CESAs from other classes (Carroll and Specht 2011; Kumar et al. 2016), allowing class members more time to accumulate shared mutations before they diverged. In contrast, structured CESA regions that are under purifying selection for shared function (e.g. glucan polymerization or stability within the membrane) are expected to have lower class specificity.

Although domain swap experiments in *P. patens* (Fig. 5) and Arabidopsis (Kumar et al. 2016; Sethaphong et al. 2016) show that the CSRs are not universal determinants of CESA isoform specificity, a role of the CSR in CESA-CESA interactions remains possible and consistent with known functions of intrinsically disordered regions (Radivojac et al. 2007). Binding characteristics of disordered proteins include low affinity, high specificity, and the ability to bind to multiple partners based on different conformations (Mohan et al. 2006). The "flycasting mechanism" suggests that disordered regions initiate binding interactions owing to the smaller entropic barrier, form limited short lived contacts with other subunits, and in the process assist large proteins in approaching closer and binding (Shoemaker et al. 2000). The conserved E(K/M)xFGxS motif, with its low VL-XT score and high mobility, may initiate protein-protein interactions (Radivojac et al. 2007), whereas the α-MoRFs with low mobility both in solution and as part of the CESA monomer, may serve as an ad hoc weakly coupled binding site (Verkhivker et al. 2003). Currently, the MoRF binding partners of PpCESA5 and other PpCESAs are unknown, but they may reside on the CESA surface and form contacts either within or between lobes. Our docked model (Fig. 5A) is similar to prior structural models based on small angle X-ray scattering data from a homotrimer of recombinant AtCESA1 cytosolic domains (Vandavasi et al. 2016) and an in silico homotrimer of a nearly complete GhCESA1 protein (Nixon et al. 2016). A third trimer structure assembled from the SAXS molecular envelopes of native recombinant monomeric CesA8 catalytic domains (Olek et al. 2014), with P-CR contacts that conform to a three-fold contact within the crystal

structure, is not possible when the membrane spanning domains are added (Rushton et al. 2017). CESA catalytic subunits have also been shown to dimerize (Olek et al. 2014). In this same study, the positions of the P-CR and CSR were predicted by docking the catalytic core of BcsA into the SAXS-derived molecular envelope of the monomer (Olek et al. 2014). Because none of these assemblies include the entire CESA protein, the position of the CSR in an assembly representing one lobe of the rosette CSC must currently be viewed as unresolved. If located on the periphery of individual lobes, the conformational variability of the CSRs could be under regulatory control in vivo, corresponding to different activity states of the CSC that remain to be characterized. Flexible interfaces between the lobes through the CSRs could help to explain the variable diameter of entire rosette CSCs and variation in inter-lobe spacing, as viewed where the transmembrane helices cross the membrane (Nixon et al. 2016). Further understanding of MoRF binding partners and their dynamic interactions and impacts on the overall structure will require analysis of atomistic models of entire CSCs.

Class specificity comparisons between whole AtCESA sequences have been used to explain the results of extensive domain swaps experiments (Kumar et al. 2016). These authors argue that the high class specificity score of AtCESA7 accounts for its inability to accept small regions from AtCESA4 and AtCESA8 while remaining active (Kumar et al. 2016). However, when phylogeny is taken into account, a different interpretation emerges. Based on consensus of published phylogenies that include wide species representation (Carroll and Specht 2011; Kumar et al. 2016), the AtCESA7 class diverged prior to the divergence of the AtCESA4 and AtCESA8 classes. Thus, AtCESA4 and AtCESA8 are expected to share higher sequence identity, and therefore a lower class specificity score, due to shared ancestry. AtCESA7 is also more similar to the PpCESAs (65.1-67.3% identity) than either AtCESA4 or AtCESA8 (59.5-61.3% and 57.6-58.8% identity, respectively). On these bases, AtCESA7 is appropriately described as ancestral and less specialized, whereas AtCESA4 and AtCESA8 are derived and more specialized. In this light, Kumar and Taylor's results (2016) can be interpreted as follows: AtCESA4 and AtCESA8 tolerate small portions of AtCESA7 because it is less specialized, whereas AtCESA7 is impaired by small portions of AtCESA4 or AtCESA8 because they are more specialized. This is also consistent with the observation that AtCESA7 is the only secondary AtCESA that can substitute for a primary AtCESA (Carroll et al. 2012).

The N-terminus contributes to class-specific function in *P. patens* CESAs.

Additional domain swap experiments revealed that the PpCESA5 N-terminus confers Clade A-specific function to PpCESA4 from clade B (Fig. 6). The CESA N-terminus includes the Zn-binding RING domain and HVRI (Pear et al. 1996). Although the RING domain is potentially involved in CESA-CESA

interactions (Kurek et al. 2002), it is highly conserved among all CESAs and thus unlikely to be responsible for functional differences between the isoforms. Therefore, the structural elements responsible for PpCESA5-specific function may reside within HVRI. This is in contrast to results from domain swap experiments in Arabidopsis, which showed that the region C terminal of the second transmembrane helix determines whether AtCESA1 and AtCESA3 can occupy their respective positions within the primary CSC (Wang et al. 2006). Results of HVRI swaps in secondary AtCESA were mixed. An HVRI from AtCESA4 or AtCESA8 abolished AtCESA7 function, whereas both AtCESA4 and AtCESA8 were still functional with the HVRI from AtCESA7 (Kumar et al. 2016). Given that the Arabidopsis and *P. patens* CESA families diversified independently, it is not surprising that different isoform specific interfaces evolved in each lineage.

Although it is not proven that the CSR plays a role in CESA-CESA interaction, its peripheral location, presence only in rosette-forming CESAs, and flexible and disordered nature are consistent with this hypothesis. The interactions between CESA subunits are likely to be complex, with both class-specific and non-class-specific interfaces. The lack of functional class-specificity for the CSRs of *P. patens* and at least some Arabidopsis CESAs indicates that this region may have formed CESA-CESA interfaces in ancestral homo-oligomeric rosettes. Selection for maintenance of intrinsic disorder and conserved binding motifs could have preserved these interfaces over evolutionary time, while rapid evolution generated sequence class-specificity as the isoform lineages diverged. Other interfaces, yet to be defined and perhaps differing by only a few amino acids between isoforms, may be responsible for the requirement for three non-redundant CESA isoform in both primary and secondary wall CSCs in seed plants. Isoform-specific interfaces have been shown to arise in other protein complexes through gene duplication followed by asymmetric and complementary degeneration of interfaces between like isoforms, resulting in restricted spatial roles, but no change in catalytic function (Finnigan et al. 2012). Our results do not preclude the possibility that other distinct roles of the CSR (or other parts of the protein) could have evolved in individual CESA lineages, although any such isoform-specific functional differences are currently uncharacterized in the CESA family.

MATERIALS AND METHODS

Construction of expression vectors

Primer pairs and templates used for vector construction are listed in Table 1. Amplification programs for Phusion Polymerase (New England Biolabs, Ipswich, MA, USA) consisted of a 30 s denaturation at 98°C; 35 cycles of 7 s at 98°C, 7 s at 68ºC (unless noted otherwise), and 30 s/kbp at 72°C.

To construct entry clones containing *PpCESA4, PpCESA7,* and *PpCESA8* cDNA clones pdp 21409, pdp38142 and pdp39044 (RIKEN Bioresource Center, Tsukuba, Ibaraki, Japan) were amplified using the primers listed in Table S1. The resulting PCR products were cloned in pDONR 221 P5-P2 according to the manufacturer's instructions (Life Technologies, Grand Island, NY, USA). To construct an entry clone containing *PpCESA6,* RIKEN clone pdp16421, which contains two base substitutions compared to the genomic sequence (Wise et al. 2011), was repaired by PCR fusion. Two fragments amplified using primers and templates shown in Table 1 were fused in a single overlap extension reaction and then cloned into pDONR 221 P5-P2 as described previously (Scavuzzo-Duggan et al. 2015). For the *PpCESA3* entry clone, RIKEN clone pdp10281, which contains a deletion and a splicing error that introduce frame shifts, was repaired by PCR fusion. Three fragments amplified from pdp10281 using primer pairs CESA8attB5/CES3repairR1, CES3repairF1/CES3repairR2, and CES3repairF2/CESA3attB2 (64ºC annealing temperature) were fused, cloned into pDONR 221 P5-P2 as described previously (Scavuzzo-Duggan et al. 2015), and sequence verified. An entry clone containing the *PpCESA10* coding sequence was constructed by amplifying cDNA prepared from wild type *P. patens* (Tran and Roberts 2016) with primer pair CesA10CDSattB5/CesA10CDSattB2 and cloning the product into pDONR 221 P5-P2. Methods used to construct entry clones containing *PpCESA5* were described previously (Scavuzzo-Duggan et al. 2015). To construct entry clones containing chimeric *PpCESA* genes, gene fragments were amplified using primer pairs and templates listed in Table 1. The fragments were fused in single overlap extension reactions and cloned into pDONR 221 P5-P2 (Life Technologies) as described previously (Scavuzzo-Duggan et al. 2015). To construct expression vectors, sequence verified entry clones were transferred, along with an entry clone containing a 3XHA tag in pDONR 221 P1-P5r, to the pTHAct1Gate destination vector as described previously (Scavuzzo-Duggan et al. 2015).

Complementation assays

The *Physcomitrella patens* (Hedw.) B. S. G. *ppcesa5*KO2-line (Goss et al. 2012) was cultured on basal medium supplemented with ammonium tartrate (BCDAT) as described previously (Roberts et al. 2011). Following transformation of protoplasts with test, positive control, and negative control expression vectors, stable antibiotic resistant colonies, each representing an independent transformation event, were arrayed and scored for complementation of the mutant phenotype as described previously (Scavuzzo-Duggan et al. 2015). The number of independent lines (=colonies) scored per treatment ranged from 29-123. The Wilson Score method (Wilson 1927; Newcombe 1998) was used to calculate 95% confidence intervals of the proportions. A two-tailed Fisher's Exact Test of Independence (Sokal and Rohlf 1981) was used for statistical analysis as described previously (Scavuzzo-Duggan et al. 2015). Transgene

expression was verified by western blot analysis for selected lines from transformations with expression vectors that did not rescue the mutant phenotype (Scavuzzo-Duggan et al. 2015).

For quantification of gametophore bud cellulose content, explants were cultured on solid BCD medium (Roberts et al. 2011) for 7 days and samples from each explant were incubated in phosphate buffered saline (PBS) containing 0.1mg mL^{-1} S4B (Anderson et al. 2010) for 30 min followed by rinsing in PBS. Fluorescence images of gametophore buds with 2 to ~16 cells and no leaf primordia or rhizoids (5-12 buds per explant) were captured with manual exposure under identical conditions (Zeiss Axio Imager M2 with 43HE DsRed filter set, Plan-Neofluar 203/0.5 objective, AxioCam MR R3 camera, and Zen Blue software, version 1.1.2.0; Carl Zeiss Microscopy, Jena, Germany). Buds were outlined manually in each image and average pixel intensities were calculated using the Fiji version of ImageJ, (Schindelin et al. 2012). Three independent transformed lines per genotype were analyzed for transformations of *ppcesa5*KO-2 with test, positive control and negative control vectors, and three independent explants were analyzed for wild type and the *ppcesa5*KO-2 background line. The experiment was repeated twice and the results were combined (n=6) after determining that there was no significant variation between experiments. Means were calculated for the six biological replicates for each genotype and analyzed by one-way ANOVA with posthoc Tukey's Honestly Significant Difference test (astatsa.com/OneWay_Anova_with_TukeyHSD/).

Ab initio and molecular dynamic simulation modeling of PpCESA CSR regions

The Rosetta ab initio modeling algorithm (Rohl et al. 2004; Kaufmann et al. 2010) was used as described previously (Sethaphong et al. 2016) to generate a minimum of 20,000 decoys for each sequence and up to 40,000 decoys for those proving difficult to fold, as was the case for CSR8. The top ten percent of each decoy population was isolated and subjected to k-means clustering to select the optimal predicted structure. These Rosetta predicted structures were further refined by all atom MD using AMBER 16 software suite (Case et al. 2017) with FF14 protein variant force field and TIP3P water model. The simulation protocol included conventional MD (cMD) stages of 1000 step minimization using the conjugate gradient and the steepest descent, solute constrained, isothermal-isobaric (NPT) ensemble simulations. The Gaussian accelerated MD module performs 200,000 steps of cMD for equilibration followed by 100,000 additional steps of cMD to obtain statistics on potential energy, which are required for determining boost potential. After adding boost potential, the system was equilibrated for 200,000 steps. Next 100,000 steps were used to obtain Gaussian acceleration parameters, the threshold potential, and the scaling factor. The upper limit on the standard deviation of the total potential boost and dihedral boost was set to the recommended value of 6.0 kcal/mol. The Particle Mesh Ewald summation method

was used to calculate the electrostatic potential under periodic boundary conditions applied in all directions. The non-bonded interactions were cut at 9 Å with 0.00001 tolerance of Ewald convergence. The temperature was maintained at 300 K using a Langevin thermostat. The simulations were run for 300 ns for each CSR region analyzed. The cpptraj module from AMBER16 and in-house scripts were used to perform post processing of simulation data. Two-dimensional energy landscapes were constructed by order parameters root mean square deviation (RMSD) and radius of gyration (R_g) in which $\Delta G=-k_BT$ [ln P(order parameter)-ln P_{max}], such that $\Delta G=0$ for the lowest free energy minimum. PyMOL (Molecular Graphics System, Version 1.5.0.4, Schrödinger, LLC) was used to capture and render protein structures.

Using PONDR (http://www.pondr.com/), disordered domains were predicted from their VL-XT scores. The disorder predictions are averaged and expressed over a sliding window of nine amino acids (Oldfield et al. 2005). MORFPRED (Disfani et al. 2012) was used to identify MoRFs within the disordered regions. The monomer homology model was used to build a trimer assembly using SymmDock (Schneidman-Duhovny et al. 2005). The docking protocol and simulation details for conventional MD were described previously (Nixon et al. 2016).

ACKNOWLEDGEMENTS

This research was 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, Office of Basic Energy Sciences under Award Number DE-SC0001090. Sequencing conducted by the Rhode Island Genomics and Sequencing Center, a Rhode Island NSF EPSCoR research facility, was supported in part by the National Science Foundation EPSCoR Cooperative Agreement EPS-1004057. The contribution of CHH was aided in part by the USDA National Institute of Food and Agriculture Hatch project 1000932. We thank Pierre-Francois Perroud for vector pTHAct1Gate and RIKEN BRC for cDNA clones cDNA pdp16421, pdp38142, pdp24095, pdp10281, pdp21409, and pdp39044.

AUTHOR CONTRIBUTIONS

A.W.R., T.S.D., A.M.C., A.S., and Y.G.L. designed the experiments. T.S.D., A.M.C., A.S. and S.S. performed the experiments and analyzed the data. A.W.R. wrote the manuscript with help from T.S.D., A.S., E.S. and C.H.H.

REFERENCES

Anderson CT, Carroll A, Akhmetova L, Somerville C (2010) Real-time imaging of cellulose reorientation during cell wall expansion in Arabidopsis roots. **Plant Physiol** 152: 787-796

Arioli T, Peng L, Betzner AS, Burn J, Wittke W, Herth W, Camilleri C, Höfte H, Plazinski J, Birch R, Cork A, Glover J, Redmond J, Williamson RE (1998) Molecular analysis of cellulose biosynthesis in *Arabidopsis*. **Science** 279: 717-720

Brown CJ, Johnson AK, Dunker AK, Daughdrill GW (2011) Evolution and disorder. **Curr Opin Struct Biol** 21: 441-446

Burton RA, Ma G, Baumann U, Harvey AJ, Shirley NJ, Taylor J, Pettolino F, Bacic A, Beatty M, Simmons CR, Dhugga KS, Rafalski JA, Tingey SV, Fincher GB (2010) A customized gene expression microarray reveals that the brittle stem phenotype *fs2* of barley is attributable to a retroelement in the *HvCesA4* cellulose synthase gene. **Plant Physiol** 153: 1716-1728

Burton RA, Shirley NJ, King BJ, Harvey AJ, Fincher GB (2004) The *CesA* gene family of barley. Quantitative analysis of transcripts reveals two groups of co-expressed genes. **Plant Physiol** 134: 224-236

Carroll A, Mansoori N, Li S, Lei L, Vernhettes S, Visser RG, Somerville C, Gu Y, Trindade LM (2012) Complexes with mixed primary and secondary cellulose synthases are functional in Arabidopsis plants. **Plant Physiol** 160: 726-737

Carroll A, Specht CD (2011) Understanding plant cellulose synthases through a comprehensive investigation of the cellulose synthase family sequences. **Front Plant Sci** 2: 5

Case DA, Cheatham TE, III, Darden TA, Duke RE, Giese TJ, Gohlke H, Goetz AW, Greene D, Homeyer N, Izadi S, Kovalenko A, Lee TS, LeGrand S, Li P, Lin C, Liu J, Luchko T, Luo R, Mermelstein D, Merz KM, Jr., Monard G, Nguyen H, Omelyan I, Onufriev A, Pan F, R. Q, Roe DR, Roitberg A, Sagui C, Simmerling C, Botello-Smith WM, Swails J, Walker RC, Wang B, Wolf RM, Wu X, Xiao L, York DM, Kollman PA (2017) AMBER 2017. University of California, San Francisco.

Desprez T, Juraniec M, Crowell EF, Jouy H, Pochylova Z, Parcy F, Hofte H, Gonneau M, Vernhettes S (2007) Organization of cellulose synthase complexes involved in primary cell wall synthesis in *Arabidopsis thaliana*. **Proc Natl Acad Sci U S A** 104: 15572-15577

Disfani FM, Hsu WL, Mizianty MJ, Oldfield CJ, Xue B, Dunker AK, Uversky VN, Kurgan L (2012) MoRFpred, a computational tool for sequence-based prediction and characterization of short disorder-to-order transitioning binding regions in proteins. **Bioinformatics** 28: i75-83

Djerbi S, Aspeborg H, Nilsson P, Sundberg B, Mellerowicz E, Blomqvist K, Teeri TT (2004) Identification and expression analysis of genes encoding putative cellulose synthases (CesA) in the hybrid aspen, *Populus tremula* (L.) X *P. tremuloides* (Michx.). **Cellulose** 11: 301-312

Djerbi S, Lindskog M, Arvestad L, Sterky F, Teeri TT (2005) The genome sequence of black cottonwood (*Populus trichocarpa*) reveals 18 conserved cellulose synthase (*CesA*) genes. **Planta** 221: 739-746

Doolittle WF (2012) Evolutionary biology: A ratchet for protein complexity. **Nature** 481: 270-271

Fagard M, Desnos T, Desprez T, Goubet F, Refregier G, Mouille G, McCann M, Rayon C, Vernhettes S, Höfte H (2000) *PROCUSTE1* encodes a cellulose synthase required for normal cell elongation specifically in roots and dark-grown hypocotyls of Arabidopsis. **Plant Cell** 12: 2409-2423

Finnigan GC, Hanson-Smith V, Stevens TH, Thornton JW (2012) Evolution of increased complexity in a molecular machine. **Nature** 481: 360-364

Goss CA, Brockmann DJ, Bushoven JT, Roberts AW (2012) A *CELLULOSE SYNTHASE* (*CESA*) gene essential for gametophore morphogenesis in the moss *Physcomitrella patens*. **Planta** 235: 1355-1367

Handakumbura PP, Matos DA, Osmont KS, Harrington MJ, Heo K, Kafle K, Kim SH, Baskin TI, Hazen SP (2013) Perturbation of Brachypodium distachyon *CELLULOSE SYNTHASE A*4 or *7* results in abnormal cell walls. **BMC Plant Biol** 13: 131

Harholt J, Sørensen I, Fangel J, Roberts A, Willats WGT, Scheller HV, Petersen BL, Banks JA, Ulvskov P (2012) The glycosyltransferase repertoire of the spikemoss *Selaginella moellendorffii* and a comparative study of its cell wall. **PLoS One** 7: e35846

Harrison CJ, Roeder AH, Meyerowitz EM, Langdale JA (2009) Local cues and asymmetric cell divisions underpin body plan transitions in the moss *Physcomitrella patens*. **Curr Biol** 19: 461-471

Holland N, Holland D, Helentjaris T, Dhugga KS, Xoconostle-Cazares B, Delmer DP (2000) A comparative analysis of the plant cellulose synthase (*CesA)* gene family. **Plant Physiol** 123: 1313-1323

Kaufmann KW, Lemmon GH, DeLuca SL, Sheehan JH, Meiler J (2010) Practically useful: what the ROSETTA protein modeling suite can do for you. **Biochemistry** 49: 2987-2998

Kaur S, Dhugga KS, Gill K, Singh J (2016) Novel structural and functional motifs in *Cellulose Synthase* (*CesA*) genes of bread wheat (*Triticum aestivum*, L.). **PLoS One** 11: e0147046

Kimura S, Laosinchai W, Itoh T, Cui X, Linder CR, Brown RM, Jr. (1999) Immunogold labeling of rosette terminal cellulose-synthesizing complexes in the vascular plant *Vigna angularis*. **Plant Cell** 11: 2075-2085

Kumar M, Atanassov I, Turner S (2016) Functional analysis of cellulose synthase CESA protein class-specificity. **Plant Physiol** 173: 970-983

Kumar M, Thammannagowda S, Bulone V, Chiang V, Han KH, Joshi CP, Mansfield SD, Mellerowicz E, Sundberg B, Teeri T, Ellis BE (2009) An update on the nomenclature for the cellulose synthase genes in *Populus*. **Trends Plant Sci** 14: 248-254

Kurek I, Kawagoe Y, Jacob-Wilk D, Doblin M, Delmer D (2002) Dimerization of cotton fiber cellulose synthase catalytic subunits occurs via oxidation of the zinc-binding domains. **Proc Natl Acad Sci U S A** 99: 11109-11114

McFarlane HE, Doring A, Persson S (2014) The cell biology of cellulose synthesis. **Annu Rev Plant Biol** 65: 69-94

Miao Y, Feher VA, McCammon JA (2015) Gaussian accelerated molecular dynamics: unconstrained enhanced sampling and free energy calculation. **J Chem Theory Comput** 11: 3584-3595

Mohan A, Oldfield CJ, Radivojac P, Vacic V, Cortese MS, Dunker AK, Uversky VN (2006) Analysis of molecular recognition features (MoRFs). **J Mol Biol** 362: 1043-1059

Morgan JLW, Strumillo J, Zimmer J (2013) Crystallographic snapshot of cellulose synthesis and membrane translocation. **Nature** 493: 181-U192

Mueller SC, Brown RM, Jr. (1980) Evidence for an intramembrane component associated with a cellulose microfibril-synthesizing complex in higher plants. **J Cell Biol** 84: 315-326

Nairn CJ, Haselkorn T (2005) Three loblolly pine *CesA* genes expressed in developing xylem are orthologous to secondary cell wall *CesA* genes of angiosperms. **New Phytol** 166: 907-915

Newcombe RG (1998) Two-sided confidence intervals for the single proportion: comparison of seven methods. **Statistics in medicine** 17: 857-872

Nixon BT, Mansouri K, Singh A, Du J, Davis JK, Lee JG, Slabaugh E, Vandavasi VG, O'Neill H, Roberts EM, Roberts AW, Yingling YG, Haigler CH (2016) Comparative structural and computational analysis supports eighteen cellulose synthases in the plant cellulose synthesis complex. **Sci Rep** 6: 28696

Norris JH, Li X, Huang S, Van de Meene AML, Tran ML, Killeavy E, Chaves AM, Mallon B, Mercure D, Tan H-T, Burton RA, Doblin MS, Kim SH, Roberts AW (2017) Functional specialization of cellulose synthase isoforms in a moss shows parallels with seed plants. **Plant Physiol** 175: 210-222

Oldfield CJ, Cheng Y, Cortese MS, Romero P, Uversky VN, Dunker AK (2005) Coupled folding and binding with alpha-helix-forming molecular recognition elements. **Biochemistry** 44: 12454-12470

Olek AT, Rayon C, Makowski L, Kim HR, Ciesielski P, Badger J, Paul LN, Ghosh S, Kihara D, Crowley M, Himmel ME, Bolin JT, Carpita NC (2014) The structure of the catalytic domain of a plant cellulose synthase and its assembly into dimers. **Plant Cell** 26: 2996-3009

Pear JR, Kawagoe Y, Schreckengost WE, Delmer DP, Stalker DM (1996) Higher plants contain homologs of the bacterial *celA* genes encoding the catalytic subunit of cellulose synthase. **Proc Natl Acad Sci U S A** 93: 12637-12642

Persson S, Paredez A, Carroll A, Palsdottir H, Doblin M, Poindexter P, Khitrov N, Auer M, Somerville CR (2007) Genetic evidence for three unique components in primary cell-wall cellulose synthase complexes in Arabidopsis. **Proc Natl Acad Sci U S A** 104: 15566-15571

Purushotham P, Cho SH, Díaz-Moreno SM, Kumar M, Nixon BT, Bulone V, Zimmer J (2016) A single heterologously expressed plant cellulose synthase isoform is sufficient for cellulose microfibril formation in vitro. **Proc Natl Acad Sci U S A** 113: 11360–11365

Radivojac P, Iakoucheva LM, Oldfield CJ, Obradovic Z, Uversky VN, Dunker AK (2007) Intrinsic disorder and functional proteomics. **Biophys J** 92: 1439-1456

Ranik M, Myburg AA (2006) Six new cellulose synthase genes from Eucalyptus are associated with primary and secondary cell wall biosynthesis. **Tree Physiol** 26: 545-556

Rensing SA, Ick J, Fawcett JA, Lang D, Zimmer A, Van de Peer Y, Reski R (2007) An ancient genome duplication contributed to the abundance of metabolic genes in the moss *Physcomitrella patens*. **BMC Evol Biol** 7: 130

Rensing SA, Ick J, Fawcett JA, Lang D, Zimmer A, Van de Peer Y, Reski R (2016) Erratum to: An ancient genome duplication contributed to the abundance of metabolic genes in the moss Physcomitrella patens. **BMC Evol Biol** 16: 184

Roberts AW, Bushoven JT (2007) The cellulose synthase (*CESA*) gene superfamily of the moss *Physcomitrella patens*. **Plant Mol Biol** 63: 207-219

Roberts AW, Dimos C, Budziszek MJ, Goss CA, Lai V (2011) Knocking out the wall: protocols for gene targeting in *Physcomitrella patens*. **Methods Mol Biol** 715: 273-290

Roberts AW, Roberts EM, Haigler CH (2012) Moss cell walls: structure and biosynthesis. **Front Plant Sci** 3: 166

Rohl CA, Strauss CE, Misura KM, Baker D (2004) Protein structure prediction using Rosetta. **Methods Enzymol** 383: 66-93

Ruprecht C, Mutwil M, Saxe F, Eder M, Nikoloski Z, Persson S (2011) Large-scale co-expression approach to dissect secondary cell wall formation across plant species. **Front Plant Sci** 2: 23

Rushton PS, Olek AT, Makowski L, Badger J, Steussy CN, Carpita NC, Stauffacher CV (2017) Rice *Cellulose SynthaseA8* plant-conserved region is a coiled-coil at the catalytic core entrance. **Plant Physiol** 173: 482-494

Samuga A, Joshi CP (2002) A new cellulose synthase gene (*PtrCesA2*) from aspen xylem is orthologous to *Arabidopsis AtCesA7* (*irx3*) gene associated with secondary cell wall synthesis. **Gene** 296: 37-44

Scavuzzo-Duggan TR, Chaves AM, Roberts AW (2015) A complementation assay for in vivo protein structure/function analysis in *Physcomitrella patens* (Funariaceae). **App Plant Sci** 3: 1500023

Schindelin J, Arganda-Carreras I, Frise E, Kaynig V, Longair M, Pietzsch T, Preibisch S, Rueden C, Saalfeld S, Schmid B, Tinevez JY, White DJ, Hartenstein V, Eliceiri K, Tomancak P, Cardona A (2012) Fiji: an open-source platform for biological-image analysis. **Nat Methods** 9: 676-682

Schlessinger A, Schaefer C, Vicedo E, Schmidberger M, Punta M, Rost B (2011) Protein disorder--a breakthrough invention of evolution? **Curr Opin Struct Biol** 21: 412-418

Schneidman-Duhovny D, Inbar Y, Nussinov R, Wolfson HJ (2005) PatchDock and SymmDock: servers for rigid and symmetric docking. **Nucleic Acids Res** 33: W363-367

Sethaphong L, Davis JK, Slabaugh E, Singh A, Haigler CH, Yingling YG (2016) Prediction of the structures of the plant-specific regions of vascular plant cellulose synthases and correlated functional analysis. **Cellulose** 23: 145-161
Sethaphong L, Haigler CH, Kubicki JD, Zimmer J, Bonetta D, DeBolt S, Yingling YG (2013) Tertiary model of a plant cellulose synthase. **Proc Natl Acad Sci U S A** 110: 7512-7517
Shoemaker BA, Portman JJ, Wolynes PG (2000) Speeding molecular recognition by using the folding funnel: the fly-casting mechanism. **Proc Natl Acad Sci U S A** 97: 8868-8873
Slabaugh E, Davis JK, Haigler CH, Yingling YG, Zimmer J (2014) Cellulose synthases: new insights from crystallography and modeling. **Trends Plant Sci** 19: 99-106
Slabaugh E, Scavuzzo-Duggan TR, Chaves AM, Wilson L, Wilson C, Davis JK, Cosgrove DJ, Anderson CT, Roberts AW, Haigler CH (2015) The valine and lysine residues in the conserved FxVTxK motif are important for the function of phylogenetically distant plant cellulose synthases. **Glycobiology** 65: 6645-6653
Sokal RR, Rohlf FJ (1981) *Biometry*. Second. W. H. Freeman and Company, New York
Somerville C (2006) Cellulose synthesis in higher plants. **Annu Rev Cell Dev Biol** 22: 53-78
Song D, Shen J, Li L (2010) Characterization of cellulose synthase complexes in Populus xylem differentiation. **New Phytol** 187: 777-790
Suzuki S, Li L, Sun Y-H, Chiang VL (2006) The cellulose synthase gene superfamily and biochemical functions of xylem-specific cellulose synthase-like genes in *Populus trichocarpa*. **Plant Physiol.** 142: 1233-1245
Tanaka K, Murata K, Yamazaki M, Onosato K, Miyao A, Hirochika H (2003) Three distinct rice cellulose synthase catalytic subunit genes required for cellulose synthesis in the secondary wall. **Plant Physiol.** 133: 73-83
Taylor NG (2008) Cellulose biosynthesis and deposition in higher plants. **New Phytol** 178: 239-252
Taylor NG, Howells RM, Huttly AK, Vickers K, Turner SR (2003) Interactions among three distinct CesA proteins essential for cellulose synthesis. **Proc Natl Acad Sci U S A** 100: 1450-1455
Tompa P (2012) Intrinsically disordered proteins: a 10-year recap. **Trends Biochem Sci** 37: 509-516
Tran ML, Roberts AW (2016) *Cellulose synthase* (*CESA*) gene expression profiling of *Physcomitrella patens*. **Plant Biol** 18: 362–368
Tsekos I (1999) The sites of cellulose synthesis in algae: Diversity and evolution of cellulose-synthesizing enzyme complexes. **J Phycol** 35: 635-655
Vandavasi VG, Putnam DK, Zhang Q, Petridis L, Heller WT, Nixon BT, Haigler CH, Kalluri U, Coates L, Langan P, Smith JC, Meiler J, O'Neill H (2016) A structural study of CESA1 catalytic domain of Arabidopsis cellulose synthesis complex: evidence for CESA trimers. **Plant Physiol** 170: 123-135
Vergara CE, Carpita NC (2001) b-D-glycan synthases and the *CesA* gene family: lessons to be learned from the mixed-linkage (1→3),(1→4)b-D-glucan synthase. **Plant Mol Biol** 47: 145-160
Verkhivker GM, Bouzida D, Gehlhaar DK, Rejto PA, Freer ST, Rose PW (2003) Simulating disorder-order transitions in molecular recognition of unstructured proteins: where folding meets binding. **Proc Natl Acad Sci U S A** 100: 5148-5153
Wang J, Howles PA, Cork AH, Birch RJ, Williamson RE (2006) Chimeric proteins suggest that the catalytic and/or C-terminal domains give CesA1 and CesA3 access to their specific sites in the cellulose synthase of primary walls. **Plant Physiol** 142: 685-695
Wilson EB (1927) Probable Inference, the Law of Succession, and Statistical Inference. **Journal of the American Statistical Association** 22: 209-212
Wise HZ, Saxena IM, Brown RM, Jr. (2011) Isolation and characterization of the cellulose synthase genes *PpCesA6* and *PpCesA7* in *Physcomitrella patens*. **Cellulose** 18: 371-384
Xu F, Joshi C (2010) *In vitro* demonstration of interactions among zinc-binding domains of cellulose synthases in *Arabidopsis* and aspen. **Adv Biosci Biotech** 1: 152-161
Yin Y, Huang J, Xu Y (2009) The cellulose synthase superfamily in fully sequenced plants and algae. **BMC Plant Biol** 9: 99

Yin Y, Johns MA, Cao H, Rupani M (2014) A survey of plant and algal genomes and transcriptomes reveals new insights into the evolution and function of the cellulose synthase superfamily. **BMC Genomics** 15: 260

Zhang Q, Cheetamun R, Dhugga KS, Rafalski JA, Tingey SV, Shirley NJ, Taylor J, Hayes K, Beatty M, Bacic A, Burton RA, Fincher GB (2014) Spatial gradients in cell wall composition and transcriptional profiles along elongating maize internodes. **BMC Plant Biol** 14: 27

FIGURES AND FIGURE LEGENDS

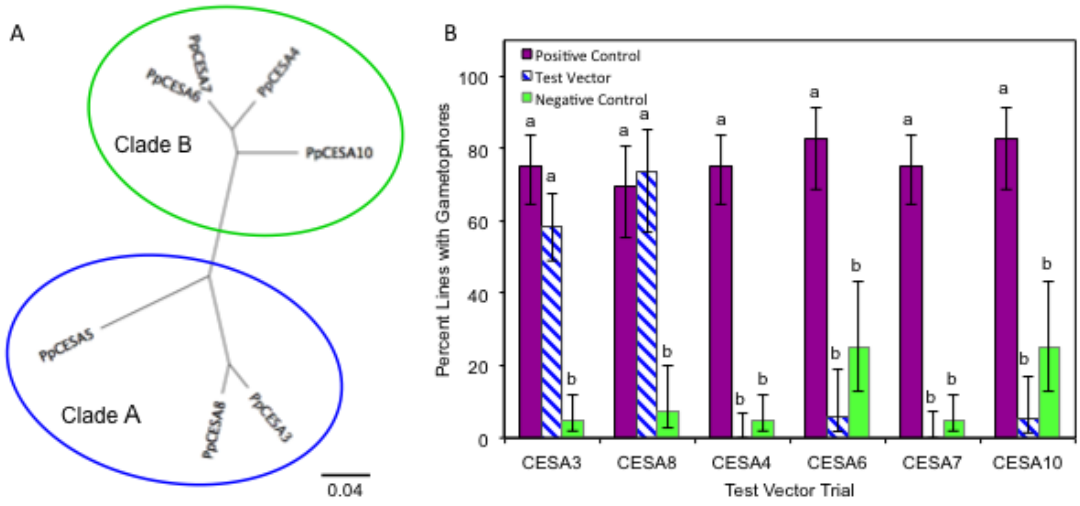

Fig. 1. *Physcomitrella patens* has functionally distinct CESA classes. A) Unrooted phylogram of PpCESAs showing clade A and clade B. B) Percentages of lines that produced gametophores when *ppcesa5*KO-2 was complemented with vectors driving expression of each of the other PpCESAs. In each experiment, complementation with PpCESA5 provided a positive control, and an empty vector was tested as a negative control. The two other PpCESAs from clade A complemented the phenotype and members of clade B did not. Percentages include the results of two trials for each vector and error bars show 95% confidence intervals. Bars with different letters designate statistically different means within each trial (p<0.01, Fisher's exact test).

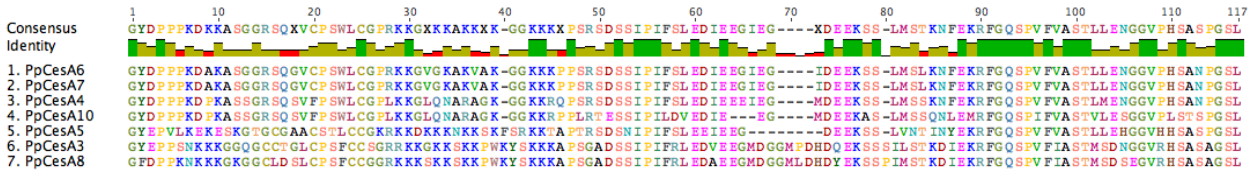

Fig. 2. The clade A and clade B CESAs differ in the CSR region. CLUSTALW alignment of the CSR regions of the seven PpCESAs. The N-terminal CSR regions of the clade A isoforms (PpCESA3, PpCESA5, PpCESA8) are rich in cysteine (C) and basic amino acids lysine (K) and arginine (R). The CSR regions of PpCESA6 and PpCESA7 are identical. In the identity bar, green=100% identity, brown= 30% to 99% identity, and red=less than 30% identity.

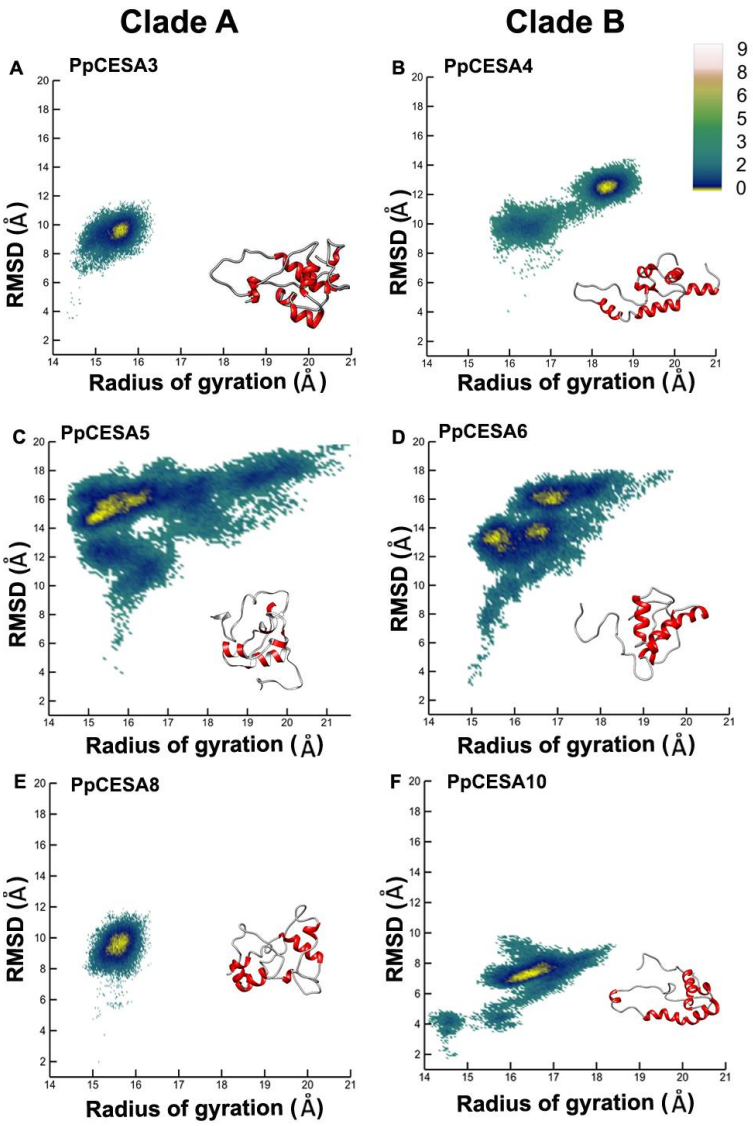

Fig. 3. The CSR sequences of moss CESAs are predicted to form flexible structures. Conformational sampling of CSRs from A) PpCESA3, B) PpCESA4, C) PpCESA5, D) PpCESA7, E) PpCESA8, and F) PpCESA10. The conformational space is defined by size (R_g) and overall structural deviation (RMSD) where $\Delta G=-k_BT$ [ln P (order parameter) - ln P_{max}], such that $\Delta G=0$ for the lowest free energy minimum (see color scale with yellow for $\Delta G=0$). Insets show representative low energy structures for each CSR with helical regions shown in red and flexible regions shown in gray.

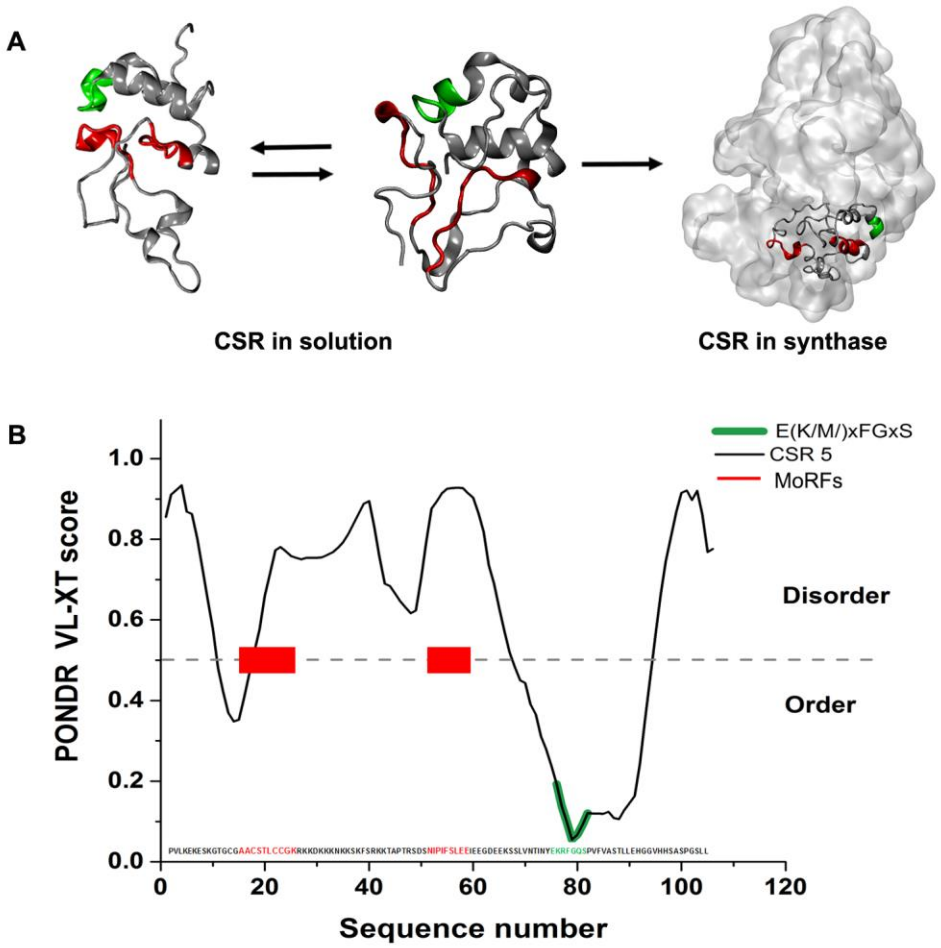

Fig. 4. CSR5 is an intrinsically disordered region. A) CSR5 can exist in several globular-extended conformations in solution and is more stable within the modeled protein. B) Profiling the CSR5 sequence revealed regional variability in predicted order (PONDR VL-XT score). The sequence shows the positions of MoRFs predicted with MORFPRED (red), and the conserved E(K/M)xFGxS motif (green).

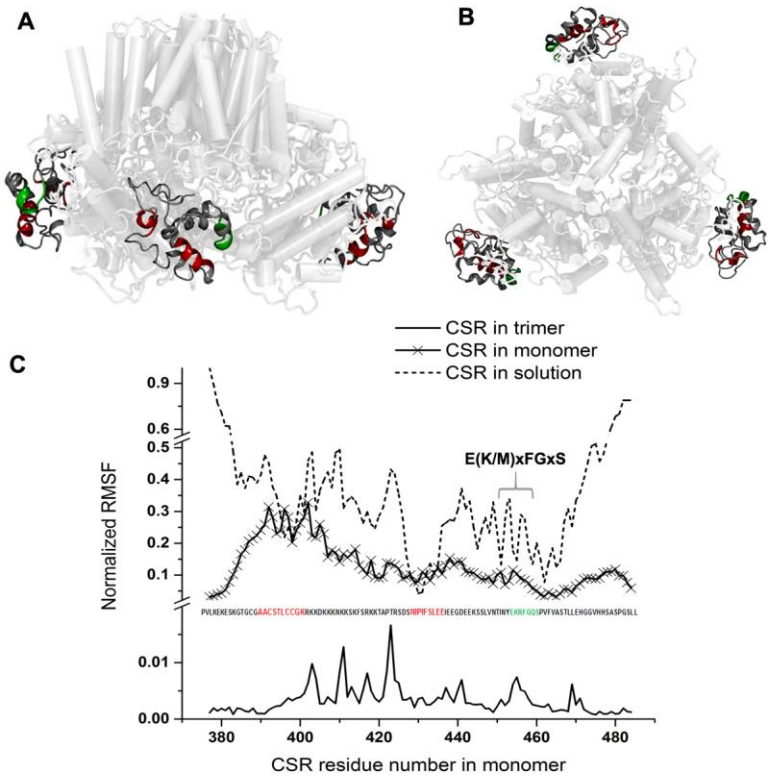

Fig. 5. Predicted MoRFs within the CSR have low mobility. Three dimensional model showing a possible spatial arrangement of the CSR in a trimer assembly of PpCESA5 is shown A) from the side or B) from the top. Highlights indicate the MoRFs (red), the conserved E(K/M)xFGxS motif (green), and other regions of the CSR (dark grey). C) Normalized fluctuation per residue in the CSR region: within the monomer; within the trimeric assembly of monomers; and alone in solution. Normalization was performed based on highest fluctuating amino acid.

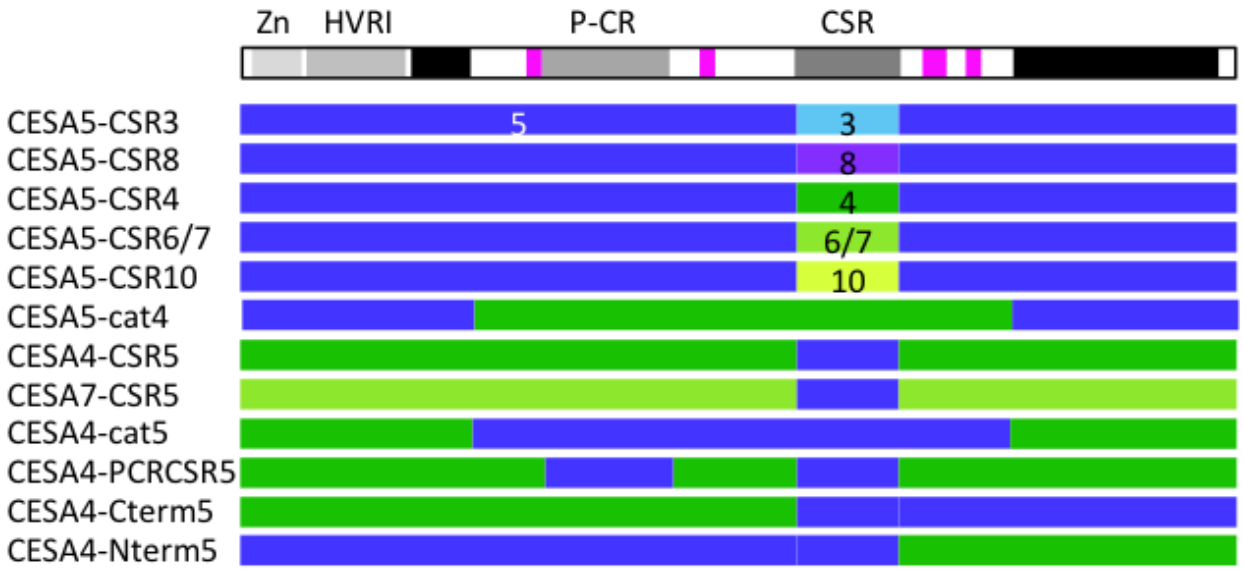

Fig. 6. Chimeric PpCESA expression vectors tested for complementation of the *ppcesa5*KO-2. The top bar shows a scale diagram of PpCESA5. Labels indicate the positions of the Zn-binding RING domain (Zn), HVRI, P-CR and CSR, with predicted transmembrane regions and conserved catalytic motifs shown in black and magenta, respectively. The twelve color-coded bars show the composition of each vector (clade A in shades of blue and clade B in shades of green), corresponding to vector names consisting of the host gene name followed by the region substituted in the chimera.

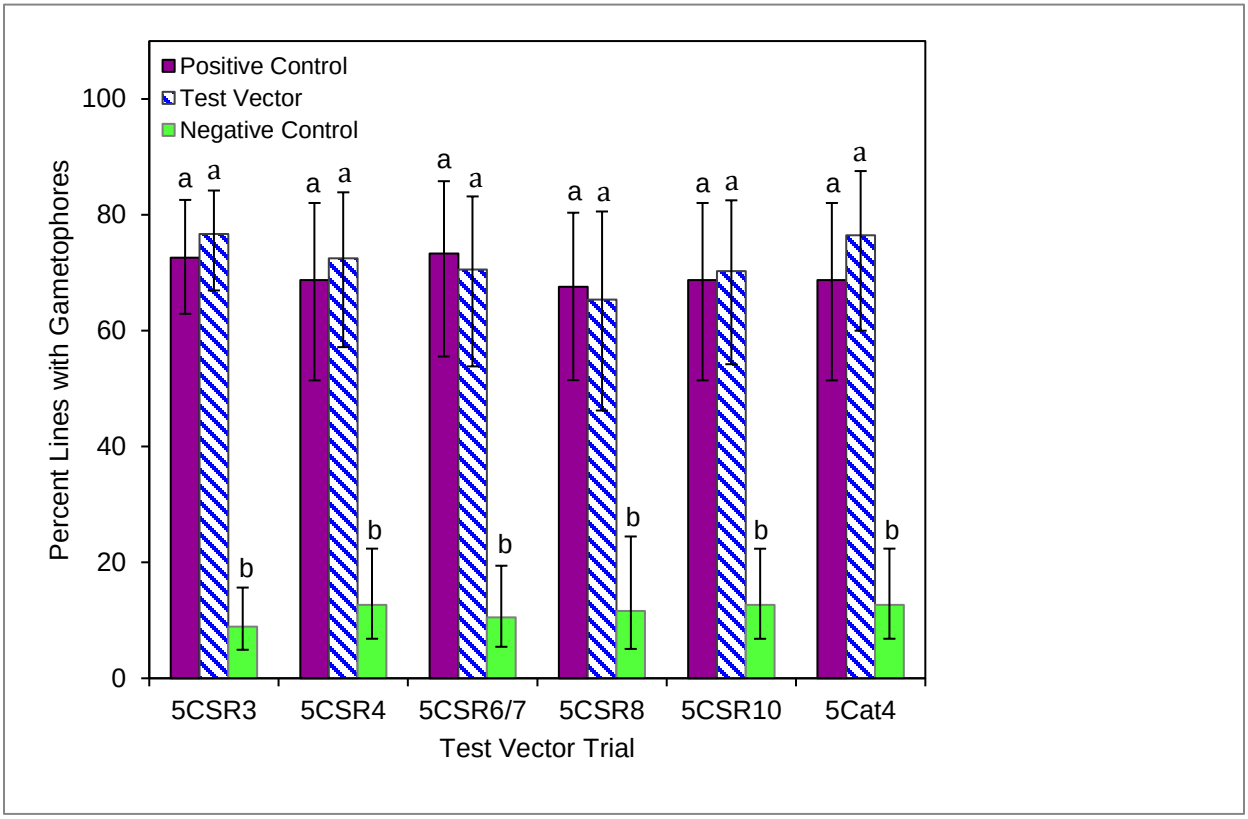

Fig. 7. Swapping out the CSR or the entire central cytosolic region does not interfere with PpCESA5 function. *ppcesa5*KO-2 complementation percentages for vectors driving expression of PpCESA5 modified with an alternative CSR or the central cytosolic region (Cat) from PpCESA4. For clarity, the X-axis labels here and in Fig. 8 omit 'CESA' prior to the host gene number. Controls were as described for Fig. 1. Percentages include the results for two (5CSR3, 5CSR8) or three (5CSR4, 5CSR6/7, 5CSR10, 5cat4) trials and error bars show 95% confidence intervals. Bars with different letters designate statistically different means within each trial ($P<0.01$, Fisher's exact test).

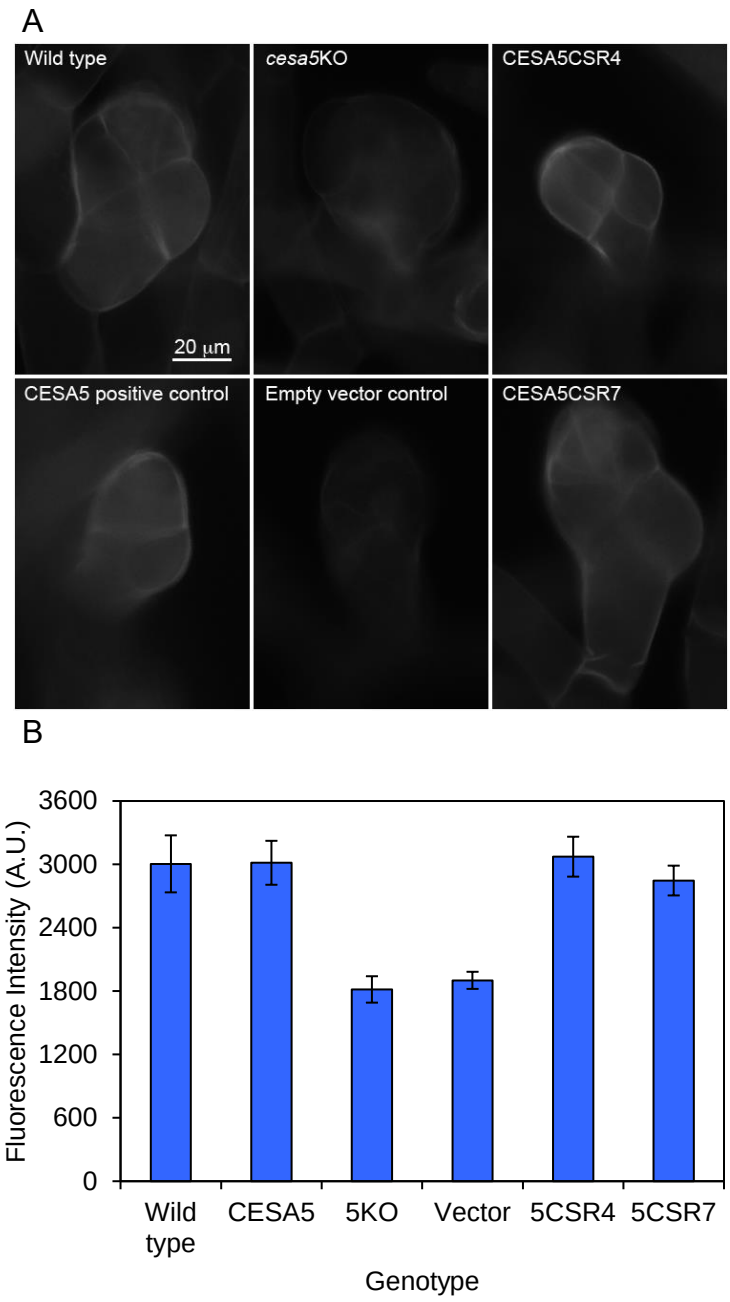

Fig. 8. Cellulose content of gametophore buds is restored by chimeric vectors containing *CSR*s from clade B *PpCESA*s. A) Fluorescence micrographs of gametophore buds stained with S4B from six lines: wild type, the *ppcesa5*KO background line, and *ppcesa5*KO transformed with unmodified CESA5 (positive control), the empty vector; the CESA5CSR4 chimera; or the CESA5CSR7 chimera. B) Average fluorescence intensity of gametophores buds of each line stained with S4B. Error bars show 95% confidence intervals. Bars with different letters show statistically different means (n=6, $p<0.01$, ANOVA with Tukey HSD post-hoc test).

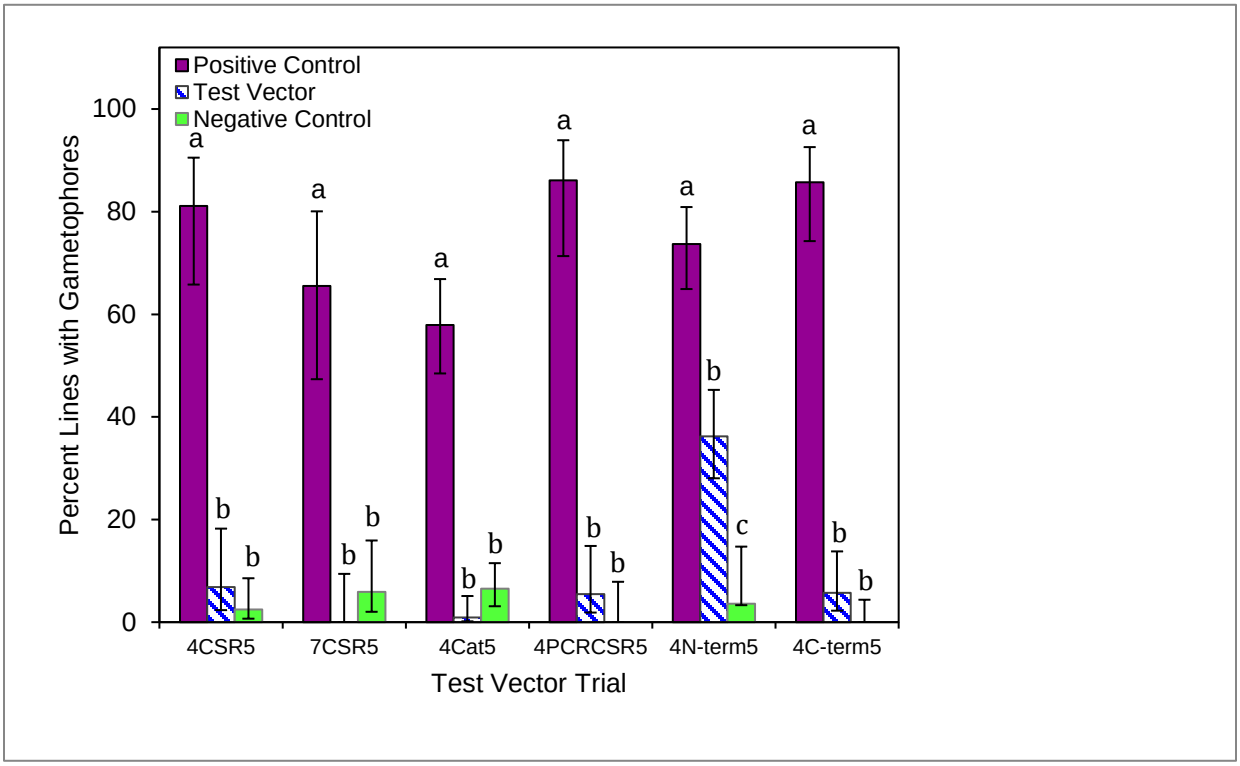

Fig. 9. The N-terminus of PpCESA5 enhances the ability of a clade B PpCESA to complement *ppcesa5*KO-2. *ppcesa5*KO-2 complementation percentages for vectors driving expression of PpCESA4 or PpCESA7 modified with various regions from PpCESA5. Controls were as described for Fig. 1. Percentages include the results for two (4CSR5, 7CSR5, 4PCRCSR5), three (4N-term5, 4C-term5), or four (4cat5) trials and error bars show 95% confidence intervals. Bars with different letters designate statistically different means within each trial ($p<0.01$, Fisher's exact test).

SUPPPLEMENTARY MATERIAL

Supplementary Table S1. Primers used for vector construction.

Primer Name	Primer Sequence	Template	Tm (°C)/Amplicon size (bp)	Description
CESA4CDSattB5	GGGGACAACTTTGTATACAAAAGTTGTCATGAAGGCGAATGCGGGGCTGTT	pdp21409	3358 bp	*PpCESA4* CDS
CESA4CDSattB2	GGGGACCACTTTGTACAAGAAAGCTGGGTACTATCGACAGTTGATCCCACACTG			
CESA7attB5	GGGGACAACTTTGTATACAAAAGTTGGCATGGAGGCGAATGCAGGGCTGCT	pdp38142	3349 bp	*PpCESA7* CDS
CESA6_7attB2	GGGGACCACTTTGTACAAGAAAGCTGGGTATCAACAGTTTATCCCGCACTGCGA			
CSEA10CDSattB5	GGGGACAACTTTGTATACAAAAGTTGTCATGGAGTCGAGTCCAGGGCTTCT	cDNA from wild type *P. patens*	3346 bp	*PpCESA10* CDS
CESA10CDSattB2	GGGGACCACTTTGTACAAGAAAGCTGGGTACTATCAGCAGTTGATCCCGCACTC			
CESA6CDSattB5	GGGGACAACTTTGTATACAAAAGTTGGCATGGAGGCCAATGCGGGGTTGGT	pdp16421	254 bp	Fragment 1 *PpCESA6* CDS
CESA6repairR	ATCCTTCCGTTCATACTCGTAGCA			
CESA6repairF	TGCTACGAGTATGAACGGAAGGATG	pdp38142	3120 bp	Fragment 2 *PpCESA6* CDS
CESA6_7attB2	See above			
CESA5attB5	GGGGACAACTTTGTATACAAAAGTTGCGATGGAGGCTAATGCAGGCCTTAT	pdp24095	1945 bp	Fragment 1 CESA5-CSR3
CESA5CSR3_SR1	TTGCCTCCTGAACACGCATCCCGTGCCTACATATAC			
CESA5CSR3_SF1	GTATATGTAGGCACGGGATGCGTGTTCAGGAGGCAA	pdp10281	417 bp	Fragment 2 CESA5-CSR3
CESA5CSR3_SR2	GACGTGAATTGCCTCCTTGAGGAGTGAGCCCGCGCT			
CESA5CSR3_SF2	AGCGCGGGCTCACTCCTCAAGGAGGCAATTCACGTC	pdp24095	1041 bp	Fragment 3 CESA5-CSR3
CESAA5attB2	GGGGACCACTTTGTACAAGAAAGCTGGGTACTAACAGCTAAGCCCGCACTCGAC			
CESA5attB5	See above	pdp24095	1947 bp	Fragment 1 CESA5-CSR4
CESA5CSR4_10_SR1	TCGCCTCTTGAAACAGCATCCCGTGCCTACATATAC			
CESA5CSR4_10B_SF1	GTATATGTAGGCACGGGATGCTGTTTCAAGAGGCGAGC	pdp21409	381 bp	Fragment 2 CESA5-CSR4
CESA5CSR4B_ SR2	GACGTGAATTGCCTCCTTCAACAATGACCCAGGATTGG			
CESA5CSR4_SF2	AATCCTGGGTCATTGTTGAAGGAGGCAATTCACGTC	pdp24095	1041 bp	Fragment 3 CESA5-CSR4
CESA5attB2	See above			
CESA5attB5	See above	pdp24095	1945 bp	Fragment 1 CESA5-CSR6/7
CESA5CSR6_SR1	TTGCCTCTTGAAACAACATCCCGTGCCTACATATAC			
CESA5CSR6_SF1	GTATATGTAGGCACGGGATGTTGTTTCAAGAGGCAA	pdp38142	399 bp	Fragment 2 CESA5-CSR6/7
CESA5CSR6_SR2-2	CACGTGGATTGCCTCCTTCAACAGCGACCCCGGATT			
CESA5CSR6_SF2-2	AATCCGGGGTCGCTGTTGAAGGAGGCAATCCACGTG	pdp24095	1041 bp	Fragment 3 CESA5-CSR6/7
CESA5attB2	See above			
CESA5attB5	See above	pdp24095	1945 bp	Fragment 1 CESA5-CSR8
CESA5CSR8_SR1	TTGCCTCCTAAAGACACATCCCGTGCCTACATATAC			
CESA5CSR8_SF1	GTATATGTAGGCACGGGATGTGTCTTTAGGAGGCAA	Pdp39044	417 bp	Fragment 2 CESA5-CSR8
CESA5CSR8_SR2	GACGTGAATTGCCTCCTTGAGGAGGGAGCCCGCGCT			
CESA5CSR8_SF2	AGCGCGGGCTCCCTCCTCAAGGAGGCAATTCACGTC	pdp24095	1041 bp	Fragment 3 CESA5-CSR8
CESAA5attB2	See above			

CESA5attB5 CESA5CSR4_10_ SR1	See above See above	pdp24095	1945 bp	Fragment 1 CESA5-CSR10
CESA5CSR4_10B_SF1 CESA5CSR10B_ SR2	See above GACGTGAATTGCCTCCTTCAACAATGACCCTGGACTGG	CESA10	388 bp	Fragment 2 CESA5-CSR10
CESA5CSR10_ SF2 CESAA5attB2	AGTCCAGGGTCATTGTTGAAGGAGGCAATTCACGTC See above	pdp24095	1041 bp	Fragment 3 CESA5-CSR10
CESA5attB5 CESA5CAT4_SR1-2	See above GCCACTTCGGGAACTGATCCAAAATCCATGAGA	pdp24095	1007 bp	Fragment 1 CESA5-CAT4
CESA5CAT4_SF1-2 CESA5CAT4_SR2	TCTCATGGATTTTGGATCAGTTCCCGAAGTGGCTACGCAACAAGAGGCAACGACGTCAGTGGATAGATC	pdp21409	1687 bp	Fragment 2 CESA5-CAT4
CESA5CAT4_SF2 CESAA5attB2	GATCTATCCACTGACGTCGTTGCCTCTTGTTGCGTA See above	pdp24095	700 bp	Fragment 3 CESA5-CAT4
CESA4CDSattB5 CESA4CSR5_SR1	See above CTTCCTGTTGAACACAGTCCCCGTACCCACATAAAC	pdp21409	1975 bp	Fragment 1 CESA4-CSR5
CESA4CSR5_SF1 CESA4CSR5_SR2	GTTTATGTGGGTACGGGGACTGTGTTCAACAGGAAG CACATGAATTGCCTCCTTGAGAAGAGATCCCGGGCT	pdp24095	390 bp	Fragment 2 CESA4-CSR5
CESA4CSR5_SF2 CESA4CDSattB2	AGCCCGGGATCTCTTCTCAAGGAGGCAATTCATGTG See above	pdp21409	1056 bp	Fragment 3 CESA4-CSR5
CESA7attB5 CESA7CSR5_SR1	See above GTGTACGTGGGTACTGGAACTGTGTTCAACAGGAAG	pdp38142	1132 bp	Fragment 1 CESA7-CSR5
CESA7CSR5_SF1 CESA7CSR5_SR2	GTGTACGTGGGTACTGGAACTGTGTTCAACAGGAAG CACGTGGATTGCCTCCTTGAGAAGAGATCCCGGGCT	pdp24095	390 bp	Fragment 2 CESA7-CSR5
CESA7CSR5_SF2 CESA6_7attB2	AGCCCGGGATCTCTTCTCAAGGAGGCAATCCACGTG See above	pdp38142	1053 bp	Fragment 3 CESA7-CSR5
CESA4CDSattB5 CESA4PCRCSR5_SR1	See above CTCAAATGTCAGCATGGACGCTCCATC	pdp21409	1267 bp	Fragment 1 CESA4-PCR,CSR5
CESA4PCRCSR5_SF1 CESA4PCRCSR5_SR2	GATGGAGCTTCGATGCTGACCTTTGAG CGACACATACACCAGCCGGGGGAGCTC	pdp24095	393 bp	Fragment 2 CESA4-PCR,CSR5
CESA4PCRCSR5_SF2 CESA4CDSattB2	GAGCTCCCCCGGCTGGTGTATGTGTCG See above	LR clone CESA4CSR5	1743 bp	Fragment 3 CESA4-PCR,CSR5
CESA4CDSattB5 CESA4CAT5_SR1	See above ACCACTTCGGGAACTGATCCAGGATCCACGATA	pdp21409	1037 bp	Fragment 1 CESA4-CAT5
CESA4CAT5_SF1 CESA4CAT5_SR2	TATCGTGGATCCTGGATCAGTTCCCGAAGTGGTTACGCCAGCAGTGGCAAAGACGTCAACGGATAGATT	pdp24095	1666 bp	Fragment 2 CESA4-CAT5
CESA4CAT5_SF2 CESA4CDSattB2	AATCTATCCGTTGACGTCTTTGCCACTGCTGGCGTA See above	pdp21409	703 bp	Fragment 3 CESA4-CAT5
CESA4CDSattB5 CESA4CSR5_SR1	See above See above	pdp21409	1975 bp	Fragment 1 CESA4-Cterm5
CESA4CSR5_SF1 CESAA5attB2	See above See above	pdp24095	1395 bp	Fragment 2 CESA4-Cterm5
CESA5attB5 CESA4CSR5_SR2	See above See above	pdp24095	2299 bp	Fragment 1 CESA4-Nterm5
CESA4CSR5_SF2 CESA4CDSattB2	See above See above	pdp21409	1056 bp	Fragment 2 CESA4-Nterm5

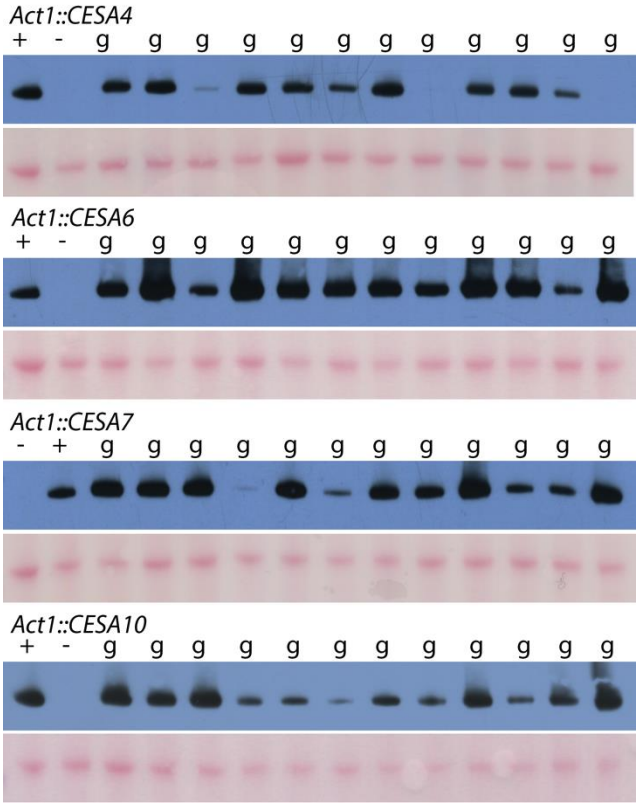

Figure S1. Western blot analysis of protein expression for *P. patens* lines derived from transformation of *ppcesa5*KO-2 protoplasts with vectors driving expression of clade B CESAs. Western blots probed with anti-HA are shown above the same blot stained with Ponceau S as a loading control. Protein loading per lane was 4.9 µg (*Act::CESA4*), 5.8 µg (*Act::CESA6*), 5.0 µg (Act::CESA7), and 3.3 µg (Act::CESA10). None of these lines produced gametophores as indicated by lower case "g". Positive (+) and negative (-) control lines are included with lines from each test transformation.

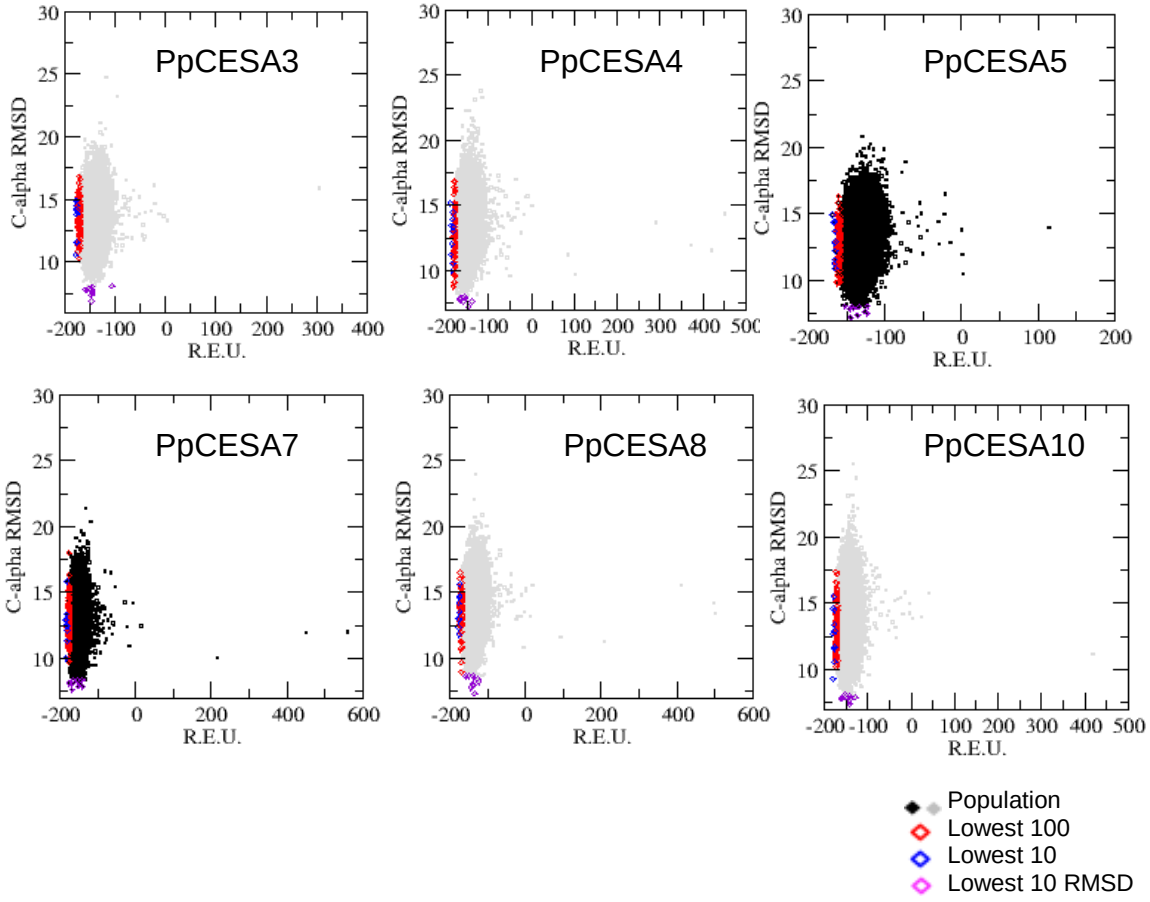

Fig. S2. Fold plots of CSRs modeled using the Rosetta ab initio modeling. RMSD values in Angstroms are plotted against Rosetta Energy Unit (REU).

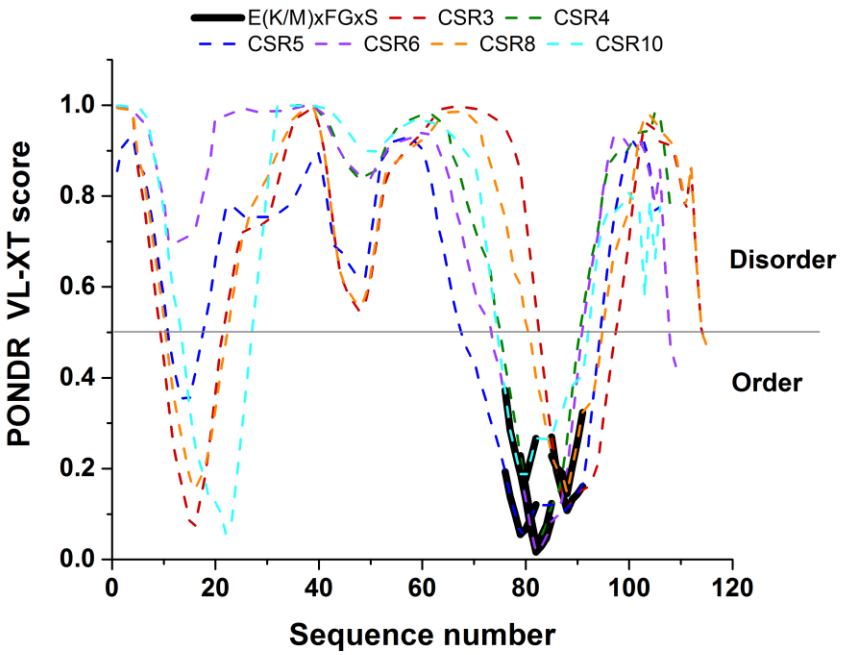

Figure S3. Predicted order (PONDR VLXT score) of the CSRs from six PpCESA isoforms. The locations of the conserved E(K/M)xFGxS motifs, within an ordered domain in each isoform, are shown by the solid black lines.

```
CSR5        PVLKEKESKGTGCGAACSTLCCGKRKKDKKKNKKSKFSRKKTAPTRSDSNIPIFSLEEIEEG--------DEEKSS-LVNTINYEKRFGQSPVFVASTLLEHGGVHHSASPGSLL
MoRFpred    nnnnnnnnnnnnnnMMMMMMMMMMnnnnnnnnnnnnnnnnnnnnnnnnnnnnMMMMMMMMMnnnn--------nnnnnn-nnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnM
CSR3        PPSNKKKGGQGCCTGLCPSFCCSGRRKKGKKSKKPWKYSKKKAPSGADSSIPIFRLEDVEEGMDGGMPDHDQEKSSSILSTKDIEKRFGQSPVFIASTMSDNGGVRHSASAGSLL
MoRFpred    nnnnnnnnnnnMnMMMMMMMMMnnnnnnnnnnnnnnnnnnnnnnnnnnnnMMMMMMMMMnnnnnnnnnnnnnnnnnnnnnnnnnnnnnMnMMnnnnnnnnnnnnnnnnnMnnMM
CSR8        PPKNKKKGKGGCLDSLCPSFCCGGRKKKSKKSKKPWKYSKKKAPSGADSSIPIFRLEDAEEGMDGGMLDHDYEKSSPIMSTKDIEKRFGQSPVFIASTMSDSEGVRHSASAGSLL
MoRFpred    nnnnnnnnnnMMnMMMMMMMMnnnnnnnnnnnnnnnnnnnnnnnnnnnnMMMMMMMMMnnnnnnnnnnnnnnnnnnnnnnnnnnnnnMnMMnnnnnnnnnnnnnnnnnMnnnn
CSR4        PPPKDPKASSGRSQSVFPSWLCGPLKKGLQNARAGK-GGKKRQPSRSDSSIPIFSLEDIEEEIEG----MDEEKSS-LMSSKNFEKRFGQSPVFVASTLMENGGVPHSANPGSLL
MoRFpred    nnnnnnnnnnnnnnMMMMMMMMMMnnnnnnnnnn-nnnnnnnnnnMMMMMMMMMMnnnn----nnnnnnn-nnnnnnnnnnnnnnnMnnnnnnnnnnnnnnnnnnnnnM
CSR6/7      PPPKDAKASGGRSQGVCPSWLCGPRKKGVGKAKVAK-GGKKKPPSRSDSSIPIFSLEDIEEGIEG----IDEEKSS-LMSLKNFEKRFGQSPVFVASTLLENGGVPHSANPGSLL
MoRFpred    nnnnnnnnnnnnnnMMMMMMMMMMMMnnnnnnnnnn-nnnnnnnnnnnnMMMMMMMMMMnnnnn----nnnnnnnnnnn-nnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnM
CSR10       PPPKDPKASSGRSQSVFPSWLCGPLKKGLQNARAGK-GGKKRPPLRTESSIPILDVEDIE---EG----MDEEKAS-LMSSQNLEMRFGQSPIFVASTVLESGGVPLSTSPGSLL
MoRFpred    nnnnnnnnnnnnnnMMMMMMMMMMMMnnnnnnnnnn nnnnnnnnnnnnMMMMMMMMMMn---nn----nnnnnnnnnnnnnn-nnnnnMnMMMnnnnnnnnnnnnnnnnnnMnM
```

Figure S4. Alignment of CSRs from PpCESA isoforms. Colors highlight the predicted MoRFs (red) and the conserved motif, E(K/M)xFGxS (green).

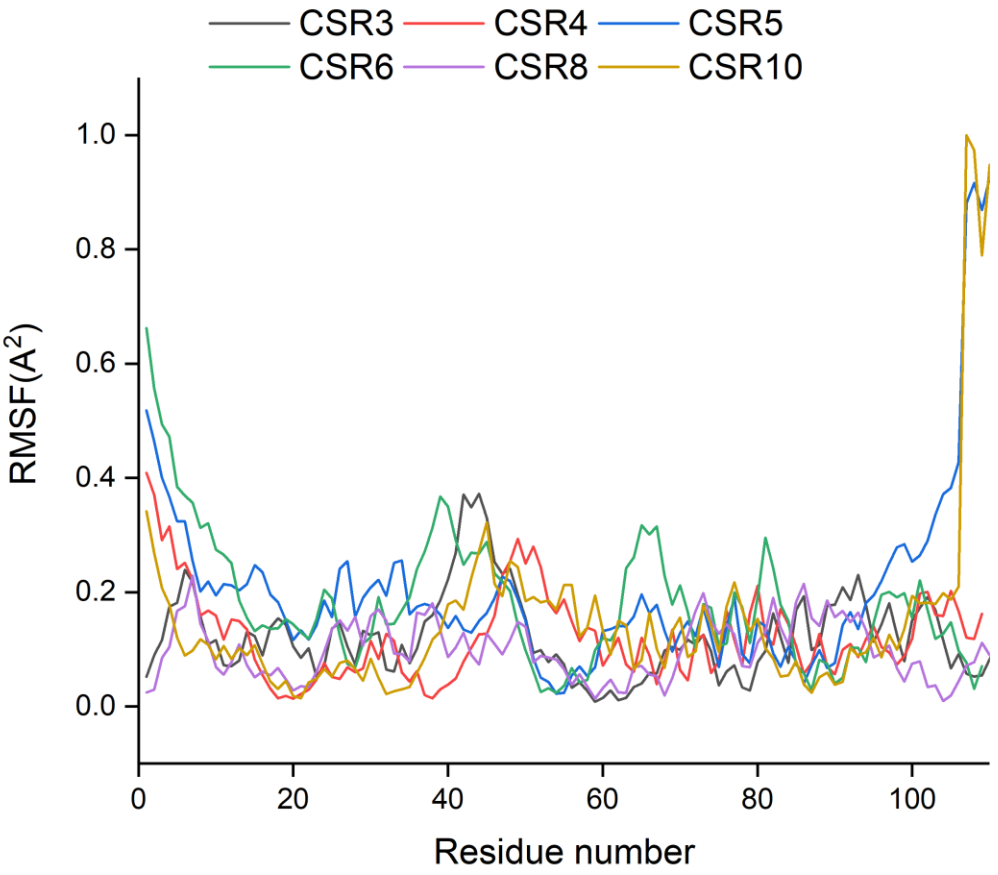

Figure S5. Normalized fluctuation per residue in the CSR regions of PpCESA3, PpCESA4, PpCESA7, PpCESA8 and PpCESA10. Normalization was performed based on highest fluctuating amino acid among all CSRs.

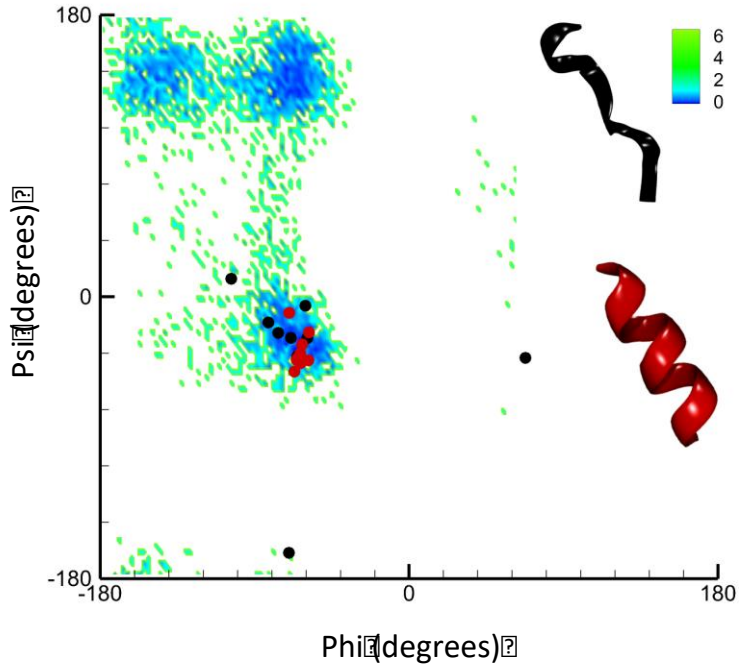

Figure S6. Ramachandran plot of α-MoRF in the CSR region of PpCESA5 (blue-green) from MD simulations. Insets show the helical form α-MoRF (red) from higher order assembly and the disordered form (black) from MD.

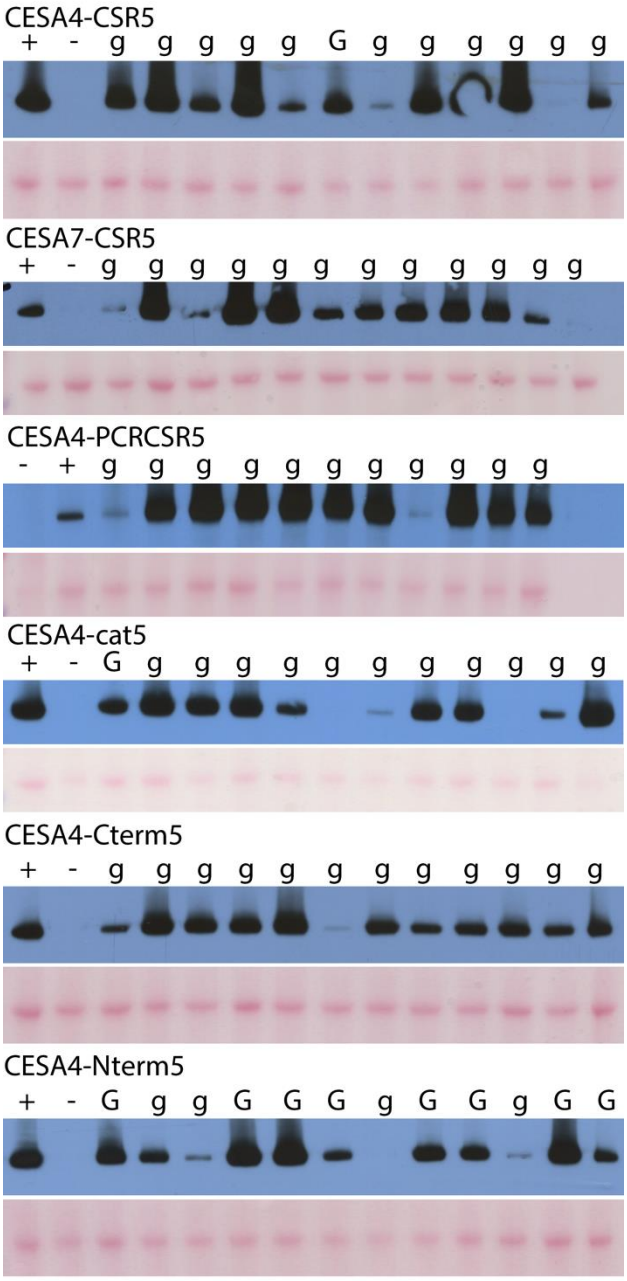

Figure S7. Western blot analysis of protein expression for *P. patens* lines derived from transformation of *ppcesa5*KO-2 protoplasts with vectors driving expression of chimeric PpCESAs. Western blots probed with anti-HA are shown above the same blot stained with Ponceau S as a loading control. Protein loading per lane was 4.9 µg (CESA4-CSR5), 3.9 µg (CESA7-CSR5), 5.0 µg (CESA4-PCRCSR5), 3.9 µg (CESA4-cat5), 7.2 µg (CESA4-Cterm5), and 5.1 µg (CESA4-Nterm5). The 'G' or 'g' labels indicate lines that did or did not produce gametophores, respectively. Positive (+) and negative (-) control lines are included with lines from each test transformation.