The Plant Cell, Vol. 26: 2996–3009, July 2014, www.plantcell.org

The Structure of the Catalytic Domain of a Plant Cellulose Synthase and Its Assembly into Dimers[C][W][OPEN]

Anna T. Olek,[a,1] Catherine Rayon,[a,1,2] Lee Makowski,[b,c] Hyung Rae Kim,[d] Peter Ciesielski,[e] John Badger,[f] Lake N. Paul,[g] Subhangi Ghosh,[d] Daisuke Kihara,[d,h] Michael Crowley,[e] Michael E. Himmel,[e] Jeffrey T. Bolin,[d] and Nicholas C. Carpita[a,d,g,3]

[a] Department of Botany and Plant Pathology, Purdue University, West Lafayette, Indiana 47907-2054
[b] Department of Bioengineering, Northeastern University, Boston, Massachusetts 02115
[c] Department of Chemistry and Chemical Biology, Northeastern University, Boston, Massachusetts 02115
[d] Department of Biological Sciences, Purdue University, West Lafayette, Indiana 47907-1971
[e] National Renewable Energy Laboratory, Biomolecular Science Group, Golden, Colorado 80401-3305
[f] DeltaG Technologies, San Diego, California 92122
[g] Bindley Bioscience Center, Purdue University, West Lafayette, Indiana 47907-2057
[h] Department of Computer Science, Purdue University, West Lafayette, Indiana 47907-2107

ORCID ID: 0000-0003-0770-314X (N.C.C.)

Cellulose microfibrils are *para*-crystalline arrays of several dozen linear (1→4)-β-D-glucan chains synthesized at the surface of the cell membrane by large, multimeric complexes of synthase proteins. Recombinant catalytic domains of rice (*Oryza sativa*) CesA8 cellulose synthase form dimers reversibly as the fundamental scaffold units of architecture in the synthase complex. Specificity of binding to UDP and UDP-Glc indicates a properly folded protein, and binding kinetics indicate that each monomer independently synthesizes single glucan chains of cellulose, i.e., two chains per dimer pair. In contrast to structure modeling predictions, solution x-ray scattering studies demonstrate that the monomer is a two-domain, elongated structure, with the smaller domain coupling two monomers into a dimer. The catalytic core of the monomer is accommodated only near its center, with the plant-specific sequences occupying the small domain and an extension distal to the catalytic domain. This configuration is in stark contrast to the domain organization obtained in predicted structures of plant CesA. The arrangement of the catalytic domain within the CesA monomer and dimer provides a foundation for constructing structural models of the synthase complex and defining the relationship between the rosette structure and the cellulose microfibrils they synthesize.

INTRODUCTION

An estimated 2×10^{11} tons of cellulose are synthesized annually (Lieth, 1975), making it the primary substrate for biofuels and bio-based products. In plants, synthesis of cellulose microfibrils is performed by a large complex of many synthase (CesA) proteins arranged in six-membered hexagonal arrays called "particle rosettes" (Giddings et al., 1980; Mueller and Brown, 1980), with their catalytic domains clustered in 50-nm diameter structures within the cell (Bowling and Brown, 2008). Information about the structure of plant CesAs and the synthase complexes they form is needed to understand the relationship between the structure of the rosettes, (1→4)-β-D-glucan synthesized by each CesA, and the structure of the *para*-crystalline cellulose microfibrils extruded by these large complexes.

A bacterial gene encoding cellulose synthase was discovered in *Gluconacetobacter xylinus* (syn *Acetobacter xylinum*) (Wong et al., 1990; Saxena et al., 1990), and plant CesAs share homology with four catalytic sequences of bacterial CesA proteins containing D, DxD, and Q/RxxRW residues essential for substrate binding and catalysis (Saxena et al., 1995; Pear et al., 1996; Delmer, 1999). The structure of *Rhodobacter sphaeroides* cellulose synthase (BcsA) provided a basis for modeling of CesA structures from vascular plants (Morgan et al., 2013). Sethaphong et al. (2013) proposed a complete structure of the catalytic domain of a plant CesA synthase based on ab initio modeling of the BcsA and similar synthase structures.

In contrast to bacterial synthases, plant synthase complexes comprise an estimated two to three dozen CesA polypeptides of ~110 kD each. Plant CesAs have two regions adjacent to their catalytic domains not present in bacterial CesAs: the plant-conserved sequence (P-CR) between the first D and the DxD motif and a domain toward the C terminus of the catalytic domain originally termed the "hypervariable region" because so few consensus sequences were found (Pear et al., 1996). Subsequent work showed that this region constituted a class-specific region (CSR) for which high similarity across many species was observed among subclasses of orthologous isoforms (Vergara and

[1] These authors contributed equally to this work.
[2] Current address: EA 3900-BIOPI, Université de Picardie Jules Verne, 80039 Amiens, France.
[3] Address correspondence to carpita@purdue.edu.
The author responsible for distribution of materials integral to the findings presented in this article in accordance with the policy described in the Instructions for Authors (www.plantcell.org) is: Nicholas C. Carpita (carpita@purdue.edu).
[C] Some figures in this article are displayed in color online but in black and white in the print edition.
[W] Online version contains Web-only data.
[OPEN] Articles can be viewed online without a subscription.
www.plantcell.org/cgi/doi/10.1105/tpc.114.126862

Carpita, 2001). The cytoplasmic region also includes an N-terminal RING-finger or zinc-finger (ZnF) domain implicated in interactions among CesAs. We found that threading of a rice (*Oryza sativa*) CesA catalytic domain sequences homologous to the BcsA synthase gave structures in close agreement with those of Sethaphong et al. (2013). However, wide disagreement was found between the ab initio model and our threading with similar glycosyl transferases when the plant-specific P-CR and CSR were included. For this reason, a soluble 57-kD catalytic domain of rice CesA8 (Supplemental Figure 1) was expressed recombinantly to obtain structural information about conformation of plant CesAs, to gain an understanding of the biochemical mechanism of synthesis, and to determine the possible roles of the P-CR and CSR domains in assembly of the multimeric synthase complex. Size-exclusion chromatography and analytical ultracentrifugation (AUC) showed that the catalytic domains of CesA polypeptides form dimers as the fundamental units of construction of the synthase complex. Specificity of binding and saturation stoichiometries with UDP and UDP-Glc indicated a properly folded protein, and each monomer in the dimer was capable of synthesizing a (1→4)-β-D-glucan chain independently.

Small-angle x-ray scattering (SAXS) has become a valuable tool for generating solution structures of proteins otherwise inherently difficult to crystallize or larger than current NMR capabilities (Svergun and Koch, 2003; Hura et al., 2009; Grant et al., 2011; Daily et al., 2012). An advantage of solution scattering is its ability to characterize protein structures and interactions that exist in multiple states (Konarev et al., 2003; Williamson et al., 2008; Grant et al., 2011). This makes possible separation of contributions due to monomers and dimers and estimates of the relative abundances of distinct species within solutions (Minh and Makowski, 2013). Examples relevant to plants include the extracellular cellulosome scaffolding structures plant cell wall digesting bacteria (Linder and Teeri, 1997; Currie et al., 2013), homo- and heterodimerization related to metal binding detoxifying proteins (Bilecen et al., 2005), pectinases-inhibitor binding interactions (Benedetti et al., 2011), and a plant hormone abscisic acid receptor (Nishimura et al., 2009). SAXS has also been used to study time-resolved structural dynamics to illustrate differences between active versus inactive states of proteins and solution properties that differ from their respective crystal structures (Lamb et al., 2009; Kezuka et al., 2010).

Solution scattering of the rice CesA catalytic domain provided an image of the monomer as an elongated two-domain structure with a large central domain and a cap region distal to the small, extended domain that coupled into dimers through the smaller domains. Docking studies indicated that the catalytic core, modeled on the basis of the BcsA structure without the P-CR and CSR, must reside near the center of a monomer. In contrast to ab initio structural predictions by Sethaphong et al. (2013), we find that the positions of the P-CR and CSR flank the catalytic domain on opposite sides to one other and implicate the CSR in formation of the CesA dimers. We propose a model for how the P-CR and Zn-finger domains couple the CesA dimer scaffold units into the particles of the rosette complex. Knowledge of this basic unit of construction of the rosette complex provides insights to genetically manipulate the structure and organization of plant cell walls in a way that will produce improved feedstocks for production of biofuels and bioproducts.

RESULTS

Structural Similarity to a Bacterial Cellulose Synthase

Structure predictions of a rice *CesA8* (LOC_Os07g10770) catalytic domain (CatD) monomer and dimer were made using a platform of algorithms at LOMETS (Wu and Zhang, 2007), including MUSTER (Wu and Zhang, 2008), FUGUE (Shi et al., 2001), HH-Pred (Söding et al., 2005), SparksX (Yang et al., 2011), Phyre2 (Kelley and Sternberg, 2009), MUSCLE (Edgar, 2004), and I-TASSER (Roy et al., 2010). Threading with FUGUE provided similar structural information as those obtained with LOMETS or MUSTER. All algorithms gave a consensus template BcsA (Protein Data Bank [PDB] ID: 4HG6:A), and the structure versus structure alignment was confirmed by TM-Align (Zhang and Skolnick, 2005) and DALI consensus (Holm and Rosenström, 2010). We used the *R. sphaeroides* BcsA protein as the threading target, but the rice protein sequence was truncated to exclude the P-CR and CSR before threading through the cognate bacterial BcsA CatD (Supplemental Figure 2A). Comparison of sequences revealed strong conservation of the four catalytic motifs essential for binding to a UDP moiety, the diphosphate of UDP-Glc, and the nonreducing terminal cellobiosyl unit of the β-D-glucan chain that extended into the protein, all of which mapped directly to the catalytic amino acids and other essential motifs integral to the BcsA active site (Supplemental Figures 2B and 2C and Supplemental Movie 1). The structure presented a general nucleotide binding fold (Rossmann et al., 1974) found in many of the templates (Supplemental Table 1). Local global alignment (Zemla, 2003) displayed 155 amino acid residues well coaligned in similarity, mostly in the upper domain including the first D, DxD, and QxxRW motifs, giving a root mean square deviation (RMSD) value of 3.12 Å.

To determine the degree of similarity of the BcsA catalytic core to other structurally related glycosyl transferases (Supplemental Table 1), *R. sphaeroides* BcsACatD, two catalytic domains of *Escherichia coli* chondroitin polymerase, SpsA synthase from *Bacillus subtilis*, and an SpsA homolog from *Bacteroides fragilis* that crystallizes as a dimer were aligned in pairs using tools on the UniProt website (UniProt Consortium, 2013). All pairs showed good consensus when threaded with FUGUE (Shi et al., 2001; Supplemental Figure 3). The resulting catalytic core sequences were then extracted from the PDB coordinates of each glycosyl transferase. All of the TM scores (Zhang and Skolnick, 2005) exceeded 0.5, and the RMSD values between matched Cα atoms in the paired proteins varied only between 1.8 and 3.5 Å, indicating that the catalytic cores of all of these proteins adopted the same variation of the classical nucleotide binding fold (Figure 1).

Structural Modeling Is Problematic after Inclusion of P-CR and CSR Domains

The rice CatD threaded structure was severely distorted using the BcsA CatD template when the P-CR and CSR sequences were included, regardless of the algorithms used. Of all structure templates surveyed (Supplemental Table 1), the best threaded positional or conformational fit when the P-CR and CSR sequences were included was obtained with *E. coli* chondroitin polymerase (ChondP), a type 2 UDP-GalNAc transferase (Osawa et al., 2009).

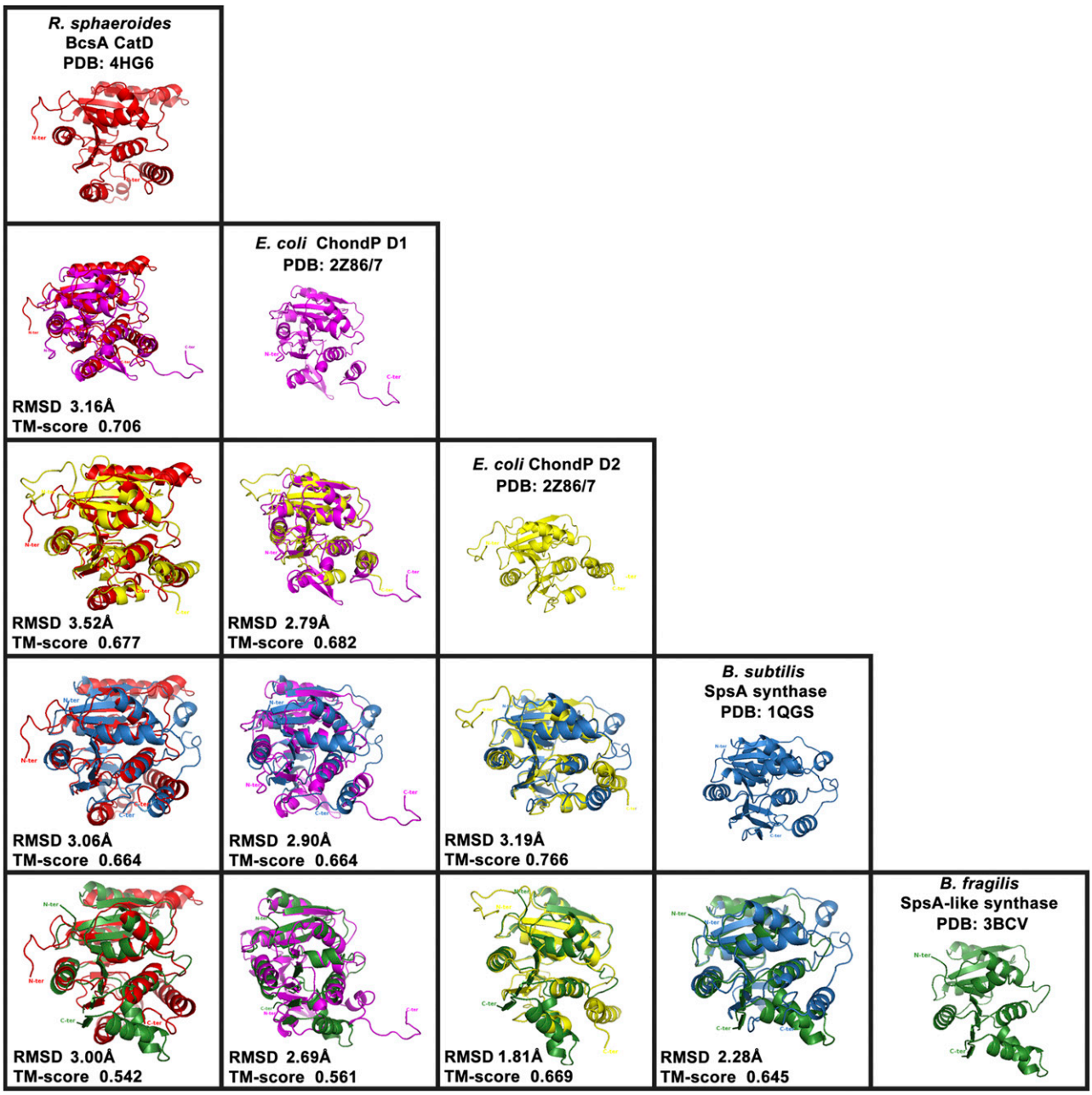

Figure 1. Structural Comparisons of the Common Catalytic Core Motifs of *Rhodobacter sphaeroides* BcsA, *E. coli* ChondP Nucleotide Binding Domains 1 and 2 (D1 and D2), *B. subtilis* SpsA Synthase, and *B. fragilis* SpsA Synthase Homolog.

The catalytic cores of the glycosyl transferases were aligned in pairs with Uniprot (UniProt Consortium, 2013). The secondary structure matching based on RMSD value calculation and on best equivalent residues of two proteins calculation (TM score) was determined by TM-align (Zhang and Skolnick, 2005), and superimposition was visualized with PyMol (DeLano, 2002). *R. sphaeroides* cellulose synthase catalytic domain (BcsA CatD), *E. coli* chondroitin polymerase (ChondP D1 and D2), *B. subtilis* SpsA synthase (SpsA), and *B. fragilis* SpsA like-synthase (Fragilis) are depicted in red, pink, yellow, blue, and green, respectively.

In contrast to the truncated sequence threaded to BcsA, the rice CatD had only 18.4% sequence identity to ChondP when the P-CR and CSR domains were included (Supplemental Figure 4A). The rice CatD and ChondP shared structural and catalytic features absent in the BcsA sequence (Supplemental Figures 4A and 4B), and the native ChondP dimerizes (Supplemental Figure 4C). However, inclusion of the CSR corrupted the catalytic core structure, indicating that this particular model was unlikely to represent the rice CatD.

The best fit (30.8% sequence identity to rice CatD) was obtained when a chimeric model with the BcsA catalytic

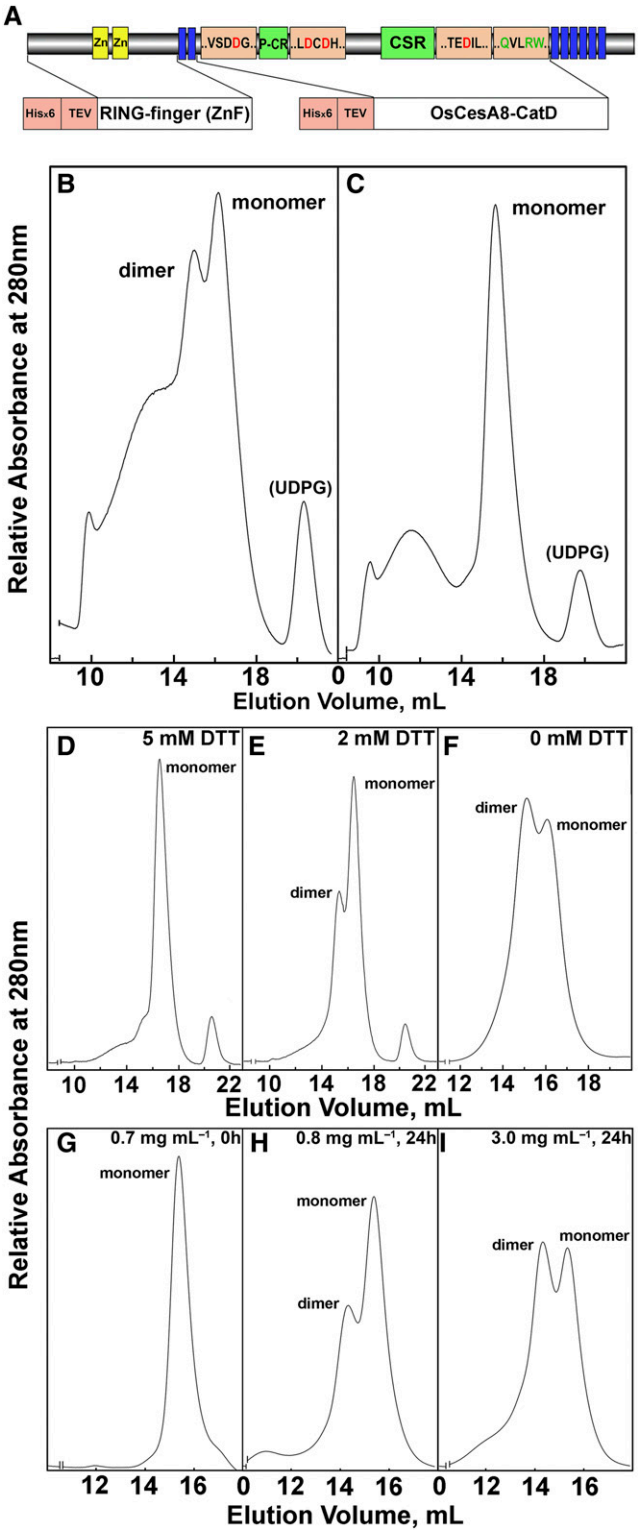

Figure 2. Recombinant CatD Domains Spontaneously Form Dimers upon Affinity Purification.

Protein model of CesA (rice CesA8) and the design of the ZnF and CatD affinity purification constructs **(A)**. The CatD domains yield monomers, dimers, and higher order aggregates upon affinity purification **(B)**, but only monomers in the presence of 10 mM DTT **(C)**. Dilution and removal of the DTT (**[D]** to **[F]**) or concentration of the protein in the presence of 10 mM DTT (**[G]** to **[I]**) results in reformation of the dimers.

[See online article for color version of this figure.]

domain integrated with sequences of a maize (*Zea mays*) oxidoreductase corresponding to the rice P-CR and CSR (Supplemental Table 1 and Supplemental Figures 4D and 4E). However, while these models showed good alignment of the catalytic core of the structure predicted by ab initio modeling (Sethaphong et al., 2013) and our threading results (Supplemental Figure 2C), they varied widely in modeling the P-CR and CSR domains, giving a large RMSD value of 24.4 Å. Clearly, the positions of the P-CR and CSR could not be unequivocally defined by computational modeling alone. For this reason, we began studies to determine the structure of the plant CesA catalytic domain empirically.

Reversible Dimerization of the Catalytic Domains of Rice CesA8

An *Arabidopsis thaliana CesA1* (At4g32410) ZnF and the rice CatD, excluding the membrane spans, were expressed in a high-efficiency prokaryotic expression system using a 6xHis fusion for affinity purification and a tobacco etch virus protease cleavage site to remove the His fusion sequence postpurification (Figure 2A). Consistent with the observations of Kurek et al. (2002), size-exclusion chromatography (SEC) showed that the ZnF domains spontaneously dimerized as they were purified directly from Ni^{+}-affinity matrices but separated to monomers in 10 mM DTT (Supplemental Figures 5A and 5B). Both monomer and dimer of the rice CatD were observed upon separation by SEC after purification from the Ni^{+}-affinity matrix (Figure 2B). When 10 mM DTT was included in the protein extraction buffer, little dimerization was observed (Figure 2C). However, when the monomer was isolated and DTT concentration diluted or the proteins concentrated on size-exclusion membrane filters, dimers spontaneously reformed and no higher molecular mass components were observed (Figures 2D to 2I).

Substrate Binding Stoichiometry

We tested whether the catalytic domains were properly folded by their specificity of binding of uridinyl substrates. Reactions at 10°C prolonged the stability of the recombinant monomer and dimer proteins for at least three days. The UDP and UDP-Glc were stable in the presence or absence of CatD protein, and radiolabel in UDP-Glc bound to CatD proteins remained in the nucleotide sugar. Binding assays with radiolabeled UDP-Glc and UDP showed that molar equivalents of each are bound per milligram of total protein for both monomer and dimer (Figure 3A). Curiously, binding stoichiometry at saturation for UDP-[6-^{3}H]-Glc indicated two substrate binding sites per monomer, i.e., four sites per dimer, whereas molar binding ratios were only 1:1 for [5,6-^{3}H]-UDP and monomer equivalents (Figure 3B). The

CatD monomer and dimer forms each bound UDP-[6-^{3}H]-Glc in a Mg^{2+}-dependent manner (Figures 3C and 3D). Binding of UDP-[6-^{3}H]-Glc was blocked strongly, but not completely, by UDP and to a certain extent by Glc-1-P, but not by GDP-Glc, ADP-Glc, or UDP-Xyl (Figures 3C and 3D).

Solution X-Ray Scattering and Analytical Ultracentrifugation

Solution x-ray scattering generated low-resolution images of their molecular shapes, and AUC gave the axial ratios of CatD monomer and dimer. We confirmed by AUC the dependency of dimer formation on thiol reductant concentration. Analysis of experimental sedimentation coefficients ($S_{20,w}$) and calculated frictional ratios ($f/f_{min} = S_{max}/S$) indicated moderately elongated structures (Erickson, 2009): $S_{max,monomer}/S_{20,w} = 1.44$ and $S_{max,dimer}/S_{20,w} = 1.59$, for the 57-kD monomer and 114-kD dimer, respectively (Figure 4, Table 1).

Analysis of solution scattering curves corresponding to soluble CatD monomers gave a radius of gyration (R_g) of 36.2 Å, which deviated 1.4-fold from the 25 Å predicted for a 57-kD globular protein. Analysis of data corresponding to the CatD dimer gave an R_g of 58.8 Å, a 1.8-fold deviation from the 33 Å predicted for a 114-kD globular protein. As was the case for the AUC data, the solution scattering results indicated a more extended structure for the dimer compared with the monomer (Figures 5A to 5C). Kratky plots of the scattering data showed characteristic peaks and troughs expected for monomers and dimers that were both well-folded and exhibited a similar degree of flexibility (Figure 5B). The peak for the dimer fell at lower q than the monomer, indicating that the spatial extent of the dimer was substantially larger than that of the monomer. A pair distribution function showed the interatomic distances for the two forms, with the dimer containing four times the number of interatomic vectors compared with the monomer, as reflected in the relative magnitude of the curves (Figure 5C). In contrast to the solution scattering patterns of the CatDs, those of the ZnF domains indicated that the proteins were largely unfolded in solution, precluding the possibility of carrying out reconstructions of their molecular envelope (Supplemental Figure 5C).

Structure reconstruction methods (Svergun et al., 2001) were used to determine low-resolution molecular envelopes for both the CatD monomer and dimer from the x-ray solution scattering data. Reconstruction of the three-dimensional molecular envelope for the monomer revealed two domains (Figure 5D; Supplemental Movie 2). A small domain protruded at a modest angle from the plane of a flattened saucer-shaped large domain. The larger domain contained a wide central region with dimensions adequate to accommodate the catalytic core predicted by modeling from the BcsA catalytic domain, but with an additional capping extension.

Solution scattering data for the CatD protein collected under conditions leading to dimerization were initially fit by a three-dimensional ellipsoid with the program SASHA (Svergun et al., 1996), indicating an elongated shape of ~162 Å in length. Three-dimensional reconstruction methods (Svergun et al., 2001) applied to these data provided a molecular envelope with a longest dimension that was consistent with this determination. The envelope had a particularly narrow cross section in the central section near the 2-fold symmetry axis and displayed a slight curvature along the principal axis (Figure 5E). A monomer pair generated by a 2-fold symmetry operation was fit into the dimer envelope to compare the shapes of the two reconstructions. Paired monomers must be placed with their small domains in side-by-side positions about the dimer 2-fold symmetry axis to accurately match the length of the dimer (Figure 5E, inset). The monomer and dimer reconstructions were performed independently, and the level of agreement between the shapes of paired monomers related by a 2-fold symmetry axis and the shape of the reconstructed symmetric dimer supported the integrity of both reconstructions. In particular, the monomer reconstruction was necessarily relatively elongated in order to match the long dimension of the dimer reconstruction and this elongation was not an artifact of the reconstruction algorithms.

Model-Based Interpretation of the CatD Reconstruction

Identification of a conserved catalytic core across several functionally related structures motivated adoption of atomic coordinate data from BcsA as a partial homology model for the catalytic core of CatD. The volume of the small domain in the SAXS reconstruction was 1.7×10^4 Å^3, which was similar to the calculated volume for the full length of either the CSR or P-CR sequence. This volume constraint, together with the shape of the catalytic core, required placement of the catalytic core in the central part of the SAXS-derived reconstruction envelope, leaving the small domain and a distal part of the large domain to be accounted for by the CSR and P-CR (Figures 6A to 6D).

The BcsA catalytic core was docked into the SAXS-derived monomer envelope obtained from solution scattering data by rotation through a set of three Euler angles, followed by a translation search to obtain the best fit to the envelope for each rotation setting. Evaluations were performed with both the Cα trace of the entire BcsA catalytic core and with a model restricted to the two longer and most homologous regions. This search indicated that the position of the catalytic core was constrained to central region of the reconstruction—the only volume where the cross section was wide enough to accommodate the atomic coordinates representing the catalytic core (Figure 6D; Supplemental Movie 3).

Of the top 16 fits to the reconstruction envelope, three solutions placed attachment points for the CSR close to a large unfilled volume. In calculations where the scoring was performed using the more restricted definition of the catalytic core, these topologically most plausible solutions included the best fit (Figure 6D) and second best fit (Supplemental Figure 6A) of the core into the reconstructed envelope. The third ranked solution (Supplemental Figure 6B) was a less optimal fit at a small subsidiary minimum ~60° from the orientation of the best solution. The ChondP model (Supplemental Figure 4B), the composite model (Supplemental Figure 4E), and the ab initio model of the cotton (*Gossypium hirsutum*) CesA catalytic domain (Sethaphong et al., 2013), which include the P-CR and CSR, were inconsistent with the CatD protein envelope derived from our solution scattering data (Supplemental Figures 7A to 7C). In all of those models, the small domain volume was incompletely occupied, and significant regions of protein model were outside the reconstruction envelope. These results underscored the need to rationalize modeling strategies with the actual solution structural envelope.

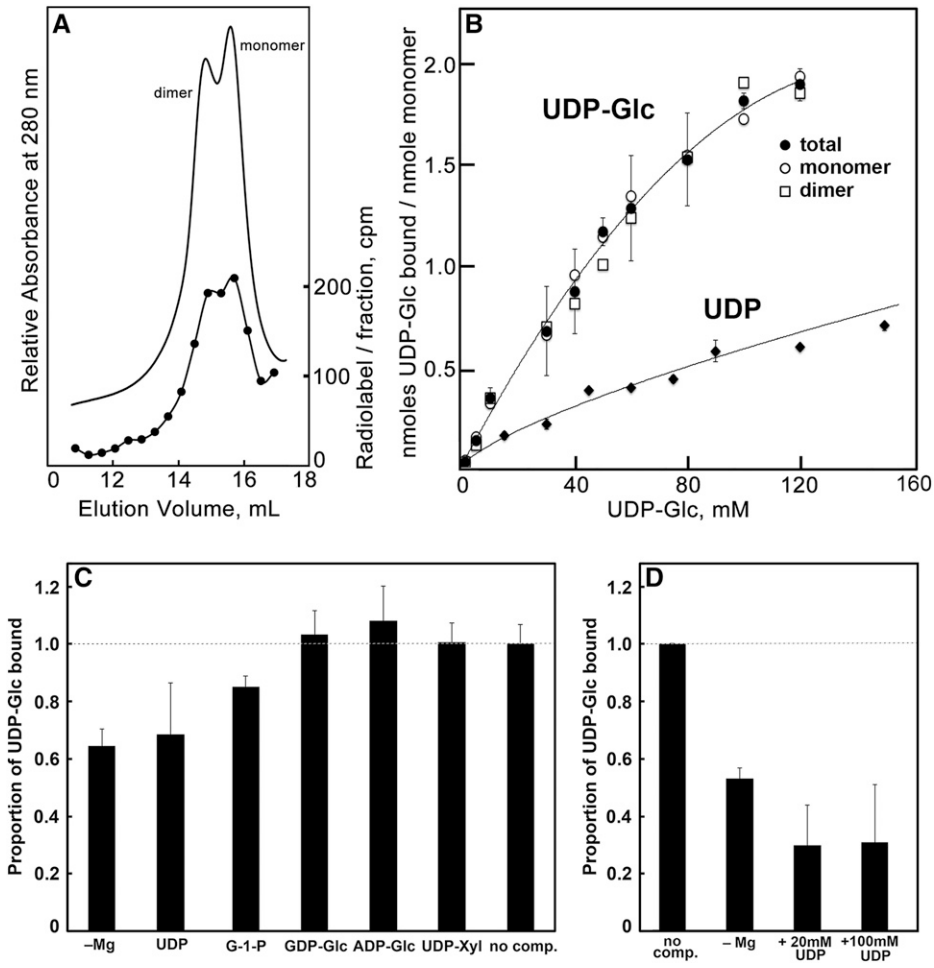

Figure 3. The Monomer and Dimer CatDs Bind Specifically UDP and UDP-Glc.

(A) Example of an incubation of purified monomer and dimer with UDP-[6-^{3}H]-glucose and [5,6-^{3}H]-UDP for up to 3 d at 10°C showing stabilized binding to both monomer and dimer in a concentration-dependent manner. Binding was near completion after 18 h and stable for the subsequent time measured. Radioactivity associated with each was quantified, and nanomoles were determined by ratio of the cpm of UDP-[6-^{3}H]-Glc bound and unbound and the concentration of the UDP-Glc in the reaction mixture.

(B) Stoichiometry combined from several binding experiments at concentrations of UDP-Glc up to 120 mM and UDP up to 160 mM. Error bars show SD of at least three samples.

(C) Stoichiometric binding of radiolabeled 2 mM UDP-Glc in the presence or absence of equimolar $MgCl_2$ and competitors at 5 mM.

(D) Stoichiometric binding of radiolabeled 20 mM UDP-Glc in the presence or absence of equimolar $MgCl_2$ and UDP at 20 mM or 100 mM with equimolar $MgCl_2$.

DISCUSSION

Solving the crystal structure of the BscA cellulose synthase (Morgan et al., 2013) was a seminal advance in understanding the structure of plant cellulose synthases. The homologous sequences between the bacterial and plant CesA catalytic domains are conserved in a variant of the classical nucleotide binding fold (Figure 1; Rossmann et al., 1974). This finding strongly indicates that the plant domain of catalysis conforms to this structure. However, when the plant-unique sequences corresponding to the P-CR and CSR are included in modeling algorithms, the significant differences in structures generated by ab initio modeling (Sethaphong et al., 2013) and our own threading (Supplemental Figure 4) prompted us to determine empirically how the P-CR and CSR are incorporated in a plant CesA CatD.

The Catalytic Domains of Plant CesAs Dimerize

Our results indicated that fusion of the catalytic domains into dimers constitutes the basic scaffolding unit of construction. This finding is in contrast to all reported studies to date, which propose that single CesA catalytic units occupy discrete locations within each particle of the rosette complex. While homodimers of different CesA isoforms might be recruited into the synthase complex, certain heterodimer combinations of the coexpressed isoforms might be favored over homodimers of the same isoform.

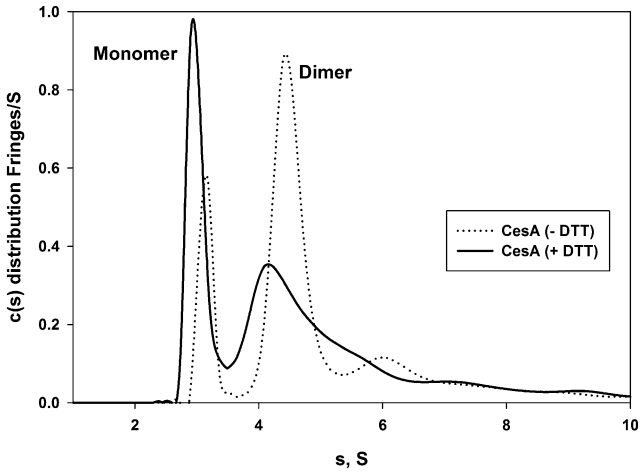

Figure 4. Determination by Analytical Centrifugation of the Sedimentation Coefficient Distributions of the CesA CatD in the Presence and Absence of 10 mM DTT.

A c(S) distribution is a representation of a particle's hydrodynamic properties, such a diffusion, sedimentation coefficient, buoyant molar mass, and shape (Schuck, 2000). CatD forms a significant amount of dimer when DTT is not present; however, the distribution shifts to mostly monomer in the presence of DTT.

Future work can be done to determine in vitro the relative interaction potential of different isoforms of CesA CatDs into dimers.

The best fit of the BcsA CatD into the SAXS-derived envelope places its connection to the membrane-spanning domains lateral to the long axis, with the P-CR and CSR flanking the CatD within the cytoplasm. This model also predicts the CSR to fill the small domain of the CesA CatD in a position where it will contribute to coupling of monomers into dimers. The range of flexibility of this interaction domain is unknown, so we are unable to predict if the large domains fold together and are locked into place by the P-CR domain, if they exist as flattened structures in vivo, or if monomers are free to "saucer" into flattened structures because of the absence of the *trans*-membrane domains (Figures 5D and 5E). Many of the target proteins that form dimers, including ChondP, give evidence of flexible interaction domains, or "hinges" (Gerstein et al., 1994), that fold the catalytic domains into a tight structure (Supplemental Table 1). When crystallized as monomers, the inherent flexibility often results in the inability to obtain high structural resolution in crystallography. Such is the case with the SpsA synthase, where the interaction domain near the C terminus is unassigned in crystals of monomers (Charnock and Davies, 1999), whereas the *B. fragilis* homolog crystallizes as a dimer, revealing the structures (Palani et al., 2011).

Proposed Role for Dimers in Particle Rosette Assembly

The formation of dimers of CesA catalytic units demonstrates that the organization of plant synthases is different from that of bacterial systems. Our finding that CesA CatD polypeptides dimerize to form catalytic units is fully consistent with many studies that show that different isoforms interact to form cellulose synthase complexes (Saxena et al., 2001; Taylor et al., 2004; Taylor, 2008; Guerriero et al., 2010). Close interactions of three distinct CesA polypeptides occur in vivo as imaged by bimolecular fluorescence complementation (Desprez et al., 2007). Mutations in any one of the three coexpressed isoforms result in impaired cellulose synthase activity (Taylor et al., 2004; Taylor, 2008), and lack of one isoform of CesA prevents incorporation of the other two into the plasma membrane (Gardiner et al., 2003). The principal differences in sequence structures among the different isoforms occur in the CSR (Vergara and Carpita, 2001), the structures we infer from SAXS modeling to participate in the dimerization of CesA monomers (Figures 5D and 5E).

Direct evidence for an interaction of different isoforms comes from copurification of CesAs after solubilization from isolated membranes (Taylor et al., 2003; Taylor, 2008; Wang et al., 2008). Consistent with these observations, Triton X-100–soluble microsomal preparations subjected to native PAGE gave an apparent 840-kD complex, whereas null mutants, but not missense mutations, gave smaller 420-kD complexes (Wang et al., 2008). Atanassov et al. (2009) affinity-trapped several isoforms of CesA corresponding to a hexameric 730-kD complex that, in the presence of DTT, separates into tetrameric, dimeric, and monomeric units, but not odd-numbered oligomers. These data give strong complementary evidence that CesA dimers are the fundamental scaffold unit of construction, with up to three dimers per particle. Kurek et al. (2002) demonstrated that the ZnF domains of CesA couple in a redox-dependent manner to link CesAs—linkages likely to couple dimers into the larger complex. On the basis of ab initio structure modeling the catalytic domain structure, Sethaphong et al. (2013) suggested that the monomers form a symmetrical hexamer that involves coupling of the CSR and P-CR regions. However, a symmetrical model does not account for how the ZnF domains can function to couple the CesAs together. By contrast, the SAXS-derived solution structure dimerized through the small domains (Figure 5E, inset; Supplemental Movie 1). This type of dimerization of the full-length CesAs places the ZnF domains on opposite sides of the dimer, providing a means to couple one or two additional dimers to form one particle of the six-membered particle rosette. Consistent with data of Atanassov et al. (2009) and given the width constraints of a rosette, no more than a triad of dimers could constitute one particle of the rosette. Given the maximum microfibril size of 36 chains, the suggestion

Table 1. DTT-Dependent Dimerization as Determined by Analytical Ultracentrifugation

Species	$S_{20,w\ monomer}$	$S_{20,w\ dimer}$	S_{max}	$S_{max}/S_{20,w}$
CesA8 CatD (+DTT)	3.6	5.6	5.2	1.44 (monomer)
CesA8 CatD (–DTT)	3.7	5.2	8.3	1.59 (dimer)

For the calculations of S ratio, the value for $S_{20,w\ monomer}$ was taken from the average of the +DTT and –DTT analytical ultracentrifugation studies. Because of concerns regarding possible reversible equilibria between species examined in the +DTT experiment, the $S_{20,w\ dimer}$ value was taken from the –DTT experiment. Values for $S_{max,monomer}$ and $S_{max,dimer}$ were calculated using the known molecular masses of 56.8 and 113.6 kD, respectively.

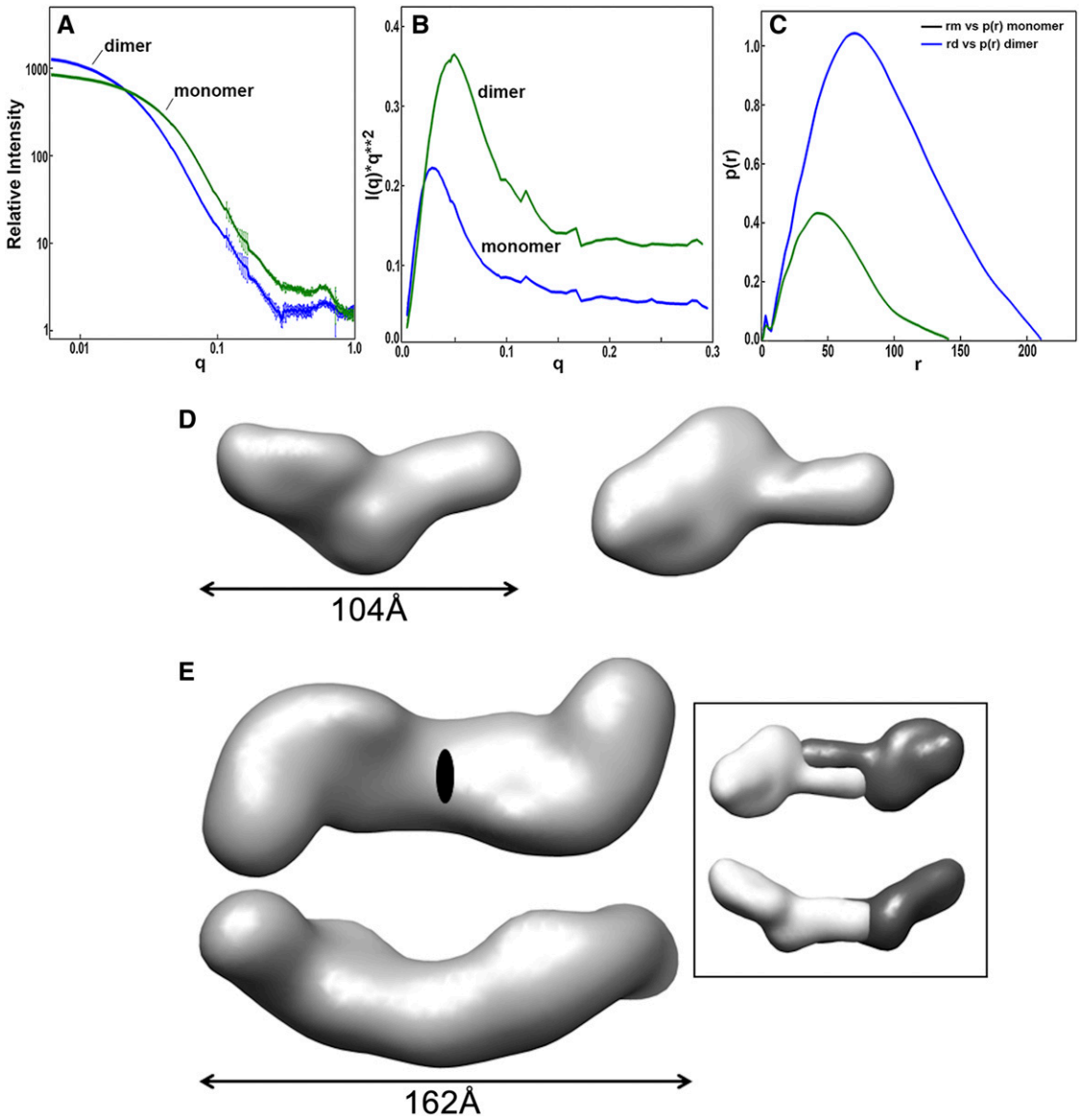

Figure 5. Solution X-Ray Scattering Experiments and Determination of 3-D Surface Contour Structure.

(A) Solution x-ray scattering curves from CesA monomers (green) and dimers (blue).
(B) Kratky plots of the data shown in **(A)**.
(C) P(R), the pair distribution function, for intensity data corresponding to monomer (green) and dimer (blue) as calculated by GNOM (Svergun, 1992).
(D) and **(E)** Three-dimensional molecular envelopes for monomer **(D)** and dimer **(E)** reconstructed from the solution scattering data using the ATSAS software package (Konarev et al., 2006) with GASBOR for the reconstruction runs (Svergun et al., 2001). Molecular surfaces generated from the grid objects that were obtained by averaging multiple reconstruction runs were rendered with Chimera (Pettersen et al., 2004). Inset: Delineation of each monomer in the dimer structure. For 3D view, see Supplemental Movie 2.

by Carpita (2011) that dimers act cooperatively to make a single (1→4)-β-D-glucan chain is not supported by these size restrictions. Binding stoichiometries (Figure 3) and the solution structure of the CatD dimers (Figure 5) argue that each monomer synthesizes a glucan chain independently. How the particles assemble into the hexagonal rosette is not understood, but domain-swap experiments implicate the C-terminal domains of CesAs in these interactions (Wang et al., 2006).

The formation of dimers appears to produce a common scaffolding feature among the synthases of other (1→4)-β-D-glycans besides CesAs (Carpita, 2011). With the exception of some members of the *CslD* group, other *Csls* that encode synthases for other (1→4)-glycans, such as mannans, xyloglucan, and mixed-linkage (1→3),(1→4)-β-D-glucans, lack ZnF domains and form complexes that synthesize single chain backbones. Of the Csl-associated synthases, the (1→3),

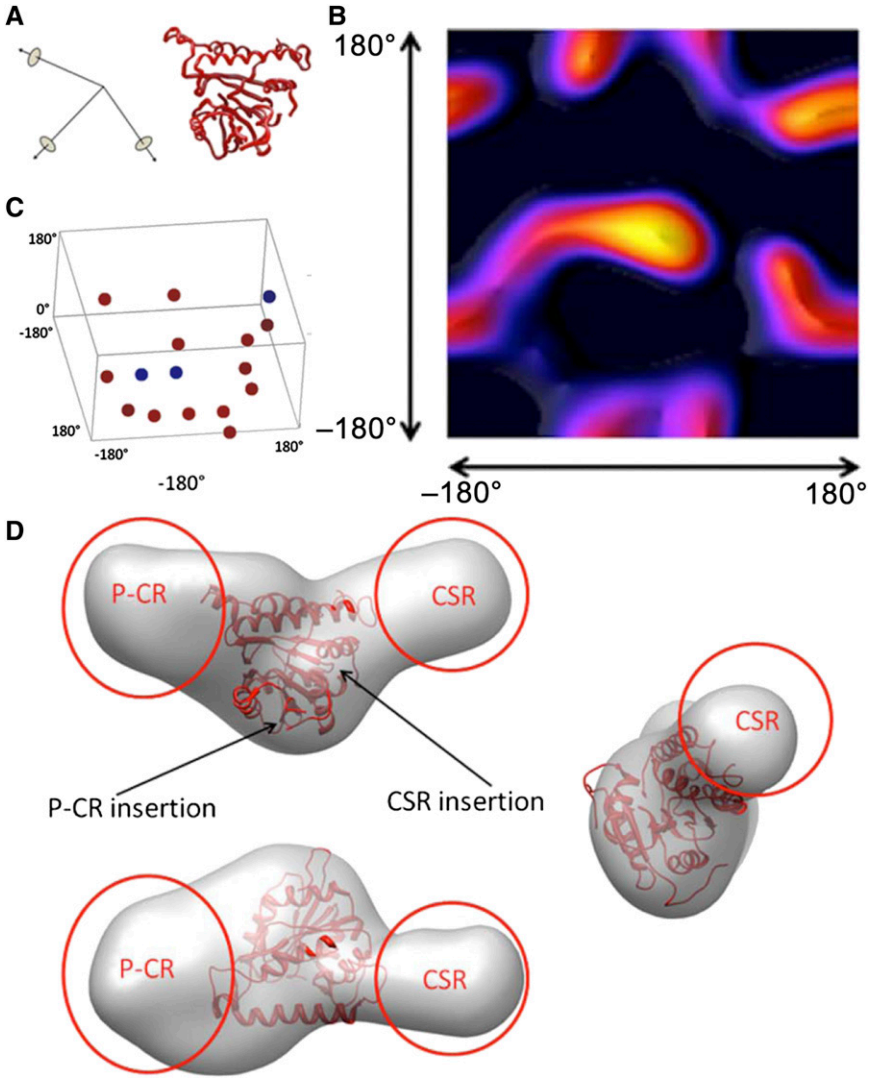

Figure 6. Interpretation of Reconstruction from Solution Scattering Data.

(A) A global search for feasible docking solutions for the catalytic core from BcsA into the envelope determined from solution scattering data was performed by systematically calculating models rotated over the full range of orientation angles and performing a local translation search to maximize fit to the envelope at each orientation. Solutions were scored with a function that minimized the sum of minimum distances from the set of Cα atoms in the model to grid points that define the volume contained within the reconstruction envelope, similar to that used in SUPCOMB (Kozin and Svergun, 2001).
(B) One section of a color-coded contour map of the target function, with the orientation that resulted in the highest overlap as the bright yellow area set at the center [(0,0,0)] position.
(C) Docked poses for sixteen other local maxima in this function, scattered across the search space and marked by red and blue dots positions were evaluated (Supplemental Movie 3). Only two of these secondary solutions (blue dots) appeared plausible in terms of placement of attachment points for the CSR and P-CR near to unfilled protein volumes. Both of these secondary solutions would also predict that the CSR corresponds to the small domain.
(D) Three nearly orthogonal views of the best solution within the low-resolution protein envelope determined from the solution scattering data.

(1→4)-β-D-glucans are synthesized synergistically by heterologously coexpressed CslF and CslH in *Arabidopsis* (Doblin et al., 2009), suggesting the two distinct isoforms interact to form an integral catalytic complex. No Csl appears to be involved directly in (1→4)-β-D-xylan synthesis, but two nonredundant glycosyl transferases are required for proper chain elongation (Lee et al., 2012). Although a UDP-galacturonsyl transferase1 (GAUT1) synthesizes homogalacturonan oligomers in vitro, coupling GAUT1 to a catalytically inactive isoform GAUT7 is required for pectin synthesis in vivo (Atmodjo et al., 2011). Together, these data indicate an emerging theme that dimerization into structural scaffolds or membrane anchors is a common feature of synthase complexes for cellulose and noncellulosic polysaccharides.

METHODS

3D Structure Prediction and Modeling of the CatD Monomer and Dimer

BcsA (Morgan et al., 2013) and 18 additional templates were selected to query the full-length rice (*Oryza sativa*) CatD (containing the P-CR and CSR) based on sequence identity, secondary structure comparison, protein family structure, and good query coverage (Supplemental Table 1). Tertiary structures were generated with Modeller 9v11 (Sali and Blundell, 1993) and visualized and labeled with the PyMol software (DeLano, 2002). The homology models of monomer or dimer models threaded with 2Z87 and built with Modeller were refined using a modified CABS ab initio modeling program (Kolinski, 2004; Koliński and Bujnicki, 2005). All the predicted quality models were evaluated by three protein model quality assessment methods, TM-align (Zhang and Skolnick, 2005), Protein Quality Predictor server (Wallner and Elofsson, 2006), ProQ, Model Quality Assessment Server (McGuffin, 2007), and QMEANscore6 (Benkert et al., 2009). The protein quality prediction of CatD was generally improved after refinement. The best model was selected based on the overall structure, the secondary structure, and position of the important motifs of the catalytic site.

CatD and ZnF Cloning

A pBluescript SKII vector (Stratagene) containing the rice *CELLULOSE SYNTHASE8* cDNA, *OsCesA8* (LOC_Os07g10770, formerly called *OsCesA13*), was used as a template to clone the catalytic site of *OsCesA8* (*CatD*) using the primer pair 5′-TCTACGCTCGAGGTATGACAGAGAAGGTGA-3′ (forward) and 5′-CAAACCGCTCGAGTTACTAAAGACGTCCTCCATAGC-3′ (reverse). The 1525-bp *CatD* fragment was ligated into a modified pETM11 (EMBL) vector, which contained an N-terminal 6x His-tag and a tobacco etch virus (TEV) protease cleavage site [ENLYF(Q/G)A], thus generating a fusion protein HIS-TEV-CatD. The native pETM11 plasmid was obtained from EMBL (http://www.embl-heidelberg.de). The RING-finger (ZnF) domain consisted of 807-bp fragment spanning from Glu-2 to Arg-270 of AtCesA1 (locus At4g32410) was generated using the primer pair 5′-TAAACTAGTGAGGCCAGTGC-3′ (forward) and 5′-TAACTCGAGGCGAGAAGATGGGATAGG-3′ (reverse) and ligated into a modified pETM11 vector, generating a fusion protein HIS-TEV-ZnF.

Transformation and Expression of CatD and ZnF Polypeptides

Plasmid constructs were transformed by heat shock into competent *Escherichia coli* [Rosetta2 (DE3); Novagen]. The 500-mL cultures in Luria-Bertani media were initiated with 5-mL overnight cultures, grown to OD_{600} of 0.6 to 0.8 (~4 h), and cooled to 16°C, and recombinant protein production was induced with 0.2 mM isopropyl β-D-1-thiogalactopyranoside (IB02100; IBI Scientific) for 13 h before harvesting.

Isolation and Purification of Protein

Frozen cells suspended in homogenization buffer were burst by sonication. Cell debris was pelleted, and a suspension of Ni^+ 6-Fast Flow resin microbeads (GE Healthcare) was added to supernatant liquids. After a 1-h incubation, the suspension was microfuged to collect resins with bound protein. After washes in low-imidazole buffer, the proteins were eluted in 0.5-mL batches a buffer containing 300 mM imidazole. In some experiments, the His-tag was excised from CatD by the TEV protease. Proteins in 10 mM DTT eluted from the Ni^+ column (~0.6 mg mL^{-1}) were digested with 30 units of AcTEV protease itself tagged with Hisx6 (Invitrogen) at ambient temperature for 6 h with gentle rocking. The Hisx6 tag had no influence on dimerization or UDP-Glc binding.

Proteins were separated on a column of Superdex 200, 10/300GL (GE Healthcare). Protein elution was monitored by absorbance at 280 nm, and 0.5-mL fractions were collected. In experiments to reduce DTT, fractions of protein from the column in 10 mM DTT were diluted to 5 mM DTT and 2 mM DTT and concentrated to ~0.5 mg mL^{-1} and rechromatographed in buffer supplemented with the respective DTT concentrations. For 0 mM DTT, protein was separated by SEC in buffer without DTT. Protein-containing fractions of dimer and monomer were concentrated using Amicon Ultra Centrifugal Filters with 30K membranes (Millipore; UFC503024).

UDP-Glc and UDP Binding Assays

UDP-[6-^{3}H]-glucose [20 Ci $mmol^{-1}$] was from American Radiolabeled Chemicals, and [5,6-^{3}H]-UTP [36 Ci $mmol^{-1}$] was from Perkin-Elmer. The labeled UTP was converted to UDP by action of 5′-nucleoside kinase (Sigma-Aldrich) with 10 mM ADP and 10 mM $MgCl_2$, and the labeled nucleotides were separated by anion-exchange HPLC in a gradient of 8 to 1.2 M potassium phosphate, pH 7.5. UDP-Glc, UTP, UDP, UDP-Xyl, GDP-glucose, ADP-glucose, and glucose-1-phosphate were from Sigma-Aldrich Chemical Co. Stock solutions of UDP-Glc and other nucleotides and other competitors were made in reaction buffer containing equivalent concentrations of $MgCl_2$. Stock reaction mixes (200 μL) were initiated by addition of 50 μL of reaction buffer containing 1.5 to 5.4 mg mL^{-1} (26.4 to 95.1 nmole mL^{-1}) of the 56.8-kD protein. Relative milligrams of protein was determined by a bicinchoninic acid assay (Sigma-Aldrich) using BSA as standard. Samples were withdrawn daily and chromatographed by SEC; radioactivity in 0.5-mL fractions representing monomer, dimer, and unbound UDP-Glc or UDP was determined by liquid scintillation counting. Equilibrium binding was achieved by 2 d reaction, and nanomoles bound were determined by the ratio of label bound to total label and the reaction concentration of nucleotide.

Analytical Ultracentrifugation

Sedimentation velocity experiments were conducted using both the Beckman-Coulter XLA and XLI analytical ultracentrifuges. The samples were centrifuged at 45,000 rpm using two-sector 1.2-cm path length carbon-filled epon centerpieces. The experiments were run on an An-50 Ti rotor at 20°C. The density and relative viscosity were calculated using SEDNTERP version 1.09, 1.0227 g mL^{-1}, and 0.01081 Poise (http://sednterp.unh.edu/). The partial specific volume (vbar) of the protein was also calculated from the protein sequence using SEDNTERP (0.734 mL g^{-1}). The samples were monitored at 655 nm (Rayleigh Interference) with a 4-min delay and 150 scans. The c(s) distributions were analyzed using SEDFIT version 12.43 (Schuck, 2000), with the buffer mismatch model to compensate for the oxidation of DTT in the cell.

SAXS Data Collection and Processing

SAXS data were collected using the undulator-based beam line X9 at the National Synchrotron Light Source at Brookhaven National Laboratory using two detectors to collect both SAXS and WAXS data simultaneously, to collect scattered intensity in the range of $0.006 < q < 2.0$ $Å^{-1}$ (Allaire and Yang, 2011), where q is the momentum transfer [$q = 4\pi \sin(\Theta)/\lambda$], 2Θ is the scattering angle, and λ is the wavelength of the incident x-rays. Data were collected at an x-ray wavelength of 0.9184 Å. A Photonic Science CCD detector operated as the WAXS detector and a Mar 165 CCD as the SAXS detector. The SAXS detector was located 3.4 m from the sample. Samples were loaded into a 96-well plate and aspirated into the 1.5-mm diameter, thin-walled sample tube using an automated system previously described (Allaire and Yang, 2011). The protein concentration for the sample treated with DTT was 4.9 mg/mL and 7.1 mg/mL for the untreated sample. All data were collected at 10°C. Preliminary data processing was performed using the X9 software package to produce circularly averaged intensity profiles combining data from the two detectors and extending over the entire q range.

Analysis of Solution Scattering Patterns

X-ray scattering patterns were checked, manipulated, and plotted for analysis using programs from the ATSAS suite (Konarev et al., 2006). SASPLOT was used to generate plots of log(I) versus log(q) (Figure 5A) and the I(q).q^2 versus q (the Kratky plot) (Figure 5B). The radii of gyration, Rg, corresponding to the data used to reconstruct molecular envelopes, were obtained with GNOM (Svergun, 1992) using standard settings with P(r) constrained to zero at the origin and at r_{max}.

Separation of Scattering of Dimers from Monomers

WAXS studies were with CatD prepared with and without 10 mM DTT to generate monomers and dimers, respectively, and each were purified subsequently by SEC. The SEC separations indicated monomer samples predominated the DTT-treated sample (~85%) and dimers more abundant (~88%) in the untreated sample. Respective monomer and dimer fractions from chromatography were pooled and concentrated on spin filters. WAXS patterns from the two samples were considered to be linear combinations of WAXS patterns from monomers and dimers with

$$S = aD + (1 - a)M$$
$$T = bD + (1 - b)M$$

where S and T are measured intensity and D and M are intensities from dimers and monomers, respectively. Assuming that we know a, the proportion of dimer in sample 1, and b, the proportion of dimer in sample 2, these equations can be solved for D and M according to

$$D = [S(1 - b) - T(1 - a)]/(a - b) \text{ and}$$
$$M = [aT - bS]/(a - b).$$

However, a and b are unknowns making the problem indeterminant. A strategy for overcoming this problem is to solve for D and M for all possible combinations of a and b and evaluate the physical plausibility of the answers. In particular, the zero angle scattering from a dimer will be 4 times the scattering from the monomer and combinations of D and M that do not satisfy this constraint can be eliminated from consideration. To determine the values of a and b consistent with this constraint, two programs, "separate" and "separate-grid," were written. These programs solved the system of linear equations specified above and evaluated the ratio of D and M as q approached 0.

When this separation was performed for the scattering patterns from CesA, it was found that a very limited range of solutions was consistent with this constraint. The allowed values for a and b were a linear function of one another, with the allowed values centered on a = 0.880 and b = 0.145, indicating that dimers made up ~88% of the scattering particles in the untreated sample, and only ~14.5% of the particles in the treated sample, consistent with the estimation from UV monitoring of SEC. Using deviation from the expected ratio of zero angle scattering as a guide, these numbers were accurate to within ±5% (i.e., a = 0.88 ± 0.05; b = 0.145 ± 0.05). Within the allowed range, use of different values of a and b led to only small changes in D and M making possible a good estimate of the scattering intensity from monomers and from dimers. Visually, the scattering curves corresponding to D and M are almost indistinguishable from the scattering curves measured from untreated and treated samples and, due to the dominating effect of an individual component in each of the two samples, represent only a small quantitative correction to the data.

When this separation was performed for the scattering patterns from CesA, only a very limited range of solutions were consistent with this constraint. The allowed values for a and b were a linear function of one another, with the allowed values centered on a = 0.880 and b = 0.145, indicating that dimers made up ~88% of the scattering particles in the "dimer" sample, and only ~14.5% of the particles in the "monomer" sample, consistent with the estimation from UV monitoring of SEC. Using deviation from the expected ratio of zero angle scattering as a guide, these numbers were accurate to within ±5% (i.e., a = 0.88 ± 0.05; b = 0.145 ± 0.05). Within the allowed range, use of different values of a and b led to only small changes in D and M making possible a good estimate of the scattering intensity from monomers and from dimers. Visually, the scattering curves corresponding to D and M were almost indistinguishable from the scattering curves measured from untreated and treated samples and, due to the dominating effect of an individual component in each of the two samples, represented only a small quantitative correction to the data.

Reconstruction of Molecular Envelopes

Three-dimensional molecular envelopes for CatD monomers and dimers were determined using programs from the ATSAS suite (Konarev et al., 2006) for the analysis of solution x-ray scattering data. The reconstruction of the protein monomer employed data from the DTT-treated sample after correction to remove the minor scattering contribution from the dimer. The GNOM program (Svergun, 1992) was used to calculate a particle distance distribution function, P(r) from x-ray scattering data in the range $0.007 < q < 1.0$ Å^{-1} with r_{max} = 120 Å. The GASBOR program (Svergun et al., 2001) was used to generate three-dimensional models of connected beads to fit this data, with the number of beads set approximately equal to the number of amino acids in the CatD construct. To assess the level of uniqueness and accuracy of these solutions 40 bead models were generated and the final corrected χ^2 values for all of these solutions were in the range 0.96 to 1.10. Additional calculations were performed to show that reducing the value of r_{max} in the calculation of P(r) to 100 Å, as would be the case for a more compact protein, results in significantly poorer χ^2 scores for the bead models generated by GASBOR.

Reconstruction averages were obtained by taking four subsets of 10-bead models and aligning them with DAMSEL and SUPCOMB (Kozin and Svergun, 2001), using the most representative model within each subset as reference. These averaged reconstructions were transformed into grid objects with volumes equal to the partial specific volume calculated from the CatD amino acid sequence. The grid objects were generated on a cubic 4-Å grid by counting the number of beads within 8Å of each grid point and choosing thresholds to give the required partial specific volume. Occupied grid points were represented by pseudo-atoms in PDB format to allow for convenient manipulation and display in molecular graphics software. Comparisons of the molecular envelopes represented by these four grid objects showed reproducible features that included a division of the protein volume into a large somewhat flattened domain, with some suggestion of a central depression, and a distinct small domain.

The molecular envelopes represented by these grid objects were aligned and averaged to create an improved low-resolution image of the protein for subsequent docking and display with the MIFit, PyMol (DeLano, 2002), and Chimera (Pettersen et al., 2004) molecular graphics programs. The reproducibility of the position of the protein surfaces were on the order of ~1 to 2 grid points (4 to 8 Å) and the average differences between these grid objects, as measured by the NSD score output by the SUPCOMB program, is comparable to this grid spacing.

For the three-dimensional reconstruction of the CatD dimer, scattering data in the range $0.022 < q < 1.0$ Å^{-1} from the sample untreated with DTT were used. The particle distance distribution function calculated with the program GNOM used r_{max} = 180 Å, and the bead models obtained with the GASBOR program were generated with P2 symmetry using approximately the number of amino acids in the CatD sequence. In this case, 20-bead models were generated and the final reduced χ^2 values for these models were in the range 1.21 to 1.32. These bead models were aligned subject to the constraint that the axes defining the 2-fold symmetry of the models should overlap so as to preserving the twofold symmetry in the aggregate model. Thus, the alignment procedure required only a rotation search about this symmetry axis together with checks to evaluate inverted and enantiomorphic solutions. Using the same procedure that was used for the monomer reconstruction, the set of 20 aligned bead models was then transformed so as to occupy contiguous points on a cubic grid with a volume set to twice the partial specific volume estimated from the CatD amino acid sequence.

Docking of the Model for the Conserved Catalytic Core into the Solution Scattering Reconstruction

Interactive docking of the chain trace for catalytic core regions of BcsA, identified as homologous to CatD, was performed with the molecular graphics program, MIFit (http://code.google.com/p/mifit/), and fits were also scored by a local standalone program incorporating a similar scoring function to that used by SUPCOMB (Kozin and Svergun, 2001). The chain trace atoms used for docking included amino acids 129 to 182, 216 to 329, and 333 to 408 from the BcsA coordinates set (PDB ID: 4HG6). Although this catalytic core region contains only ~48% of the total number of amino acids in the CatD construct, the range of feasible docking models is highly constrained by (1) the shape of the partial model and (2) the requirement that proximate, unoccupied volumes of the reconstruction are available for the introduction atoms corresponding to the sequences of the CSR and the P-CR. In particular, the positions where the CSR sequence joins the catalytic core are well defined in this partial homology model as between amino acids 329 and 333, indicating that a volume representing a distinct CSR domain in the reconstruction should be available at this position in the docked model.

Accession Numbers

Sequence data from this article can be found in the PDB databases under the following accession numbers: *Rhodobacter sphaeroides* BcsACatD (PDB ID:4HG6:A), two catalytic domains of *E. coli* chondroitin polymerase, ChondP D1 and ChondP D2 (PDB ID:2Z87), *Bacillus subtilis* SpsA synthase (PDB ID:1QGS), and *Bacteroides fragilis* SpsA homolog (Fragilis; PDB ID:3BCV). A complete list of the target sequences and PDB accession numbers is provided in Supplemental Table 1.

Supplemental Data

The following materials are available in the online version of this article.

Supplemental Figure 1. The Amino Acid Sequence of the Rice (*Oryza sativa*) CesA8 Catalytic Domain.

Supplemental Figure 2. Modeling Comparisons of the Rice CesA8 CatD and the *Rhodobacter sphaeroides* BcsA Catalytic Domain (BcsA CatD).

Supplemental Figure 3. Sequence Alignments of Five Catalytic Domains from Structurally Related Glycosyl Transferases.

Supplemental Figure 4. Structure-Based Sequence Alignment and Threaded Structure of the Rice CesA8 CatD and the *E. coli* Chondroitin Polymerase (ChondP) and a Composite Sequence-Structure Alignment and Model of the Rice CesA8 CatD with *Rhodobacter sphaeroides* BcsA Catalytic Domain (BcsA CatD) and the P-CR and CSR Domains with *Zea mays* Oxidoreductase (1B5Q).

Supplemental Figure 5. Recombinant ZnF Domains Spontaneously Form Dimers in a Redox-Dependent Manner but Do Not Fold Properly.

Supplemental Figure 6. Alternative Docking Models of the BcsA Catalytic Core into the Solution Scattering Envelope.

Supplemental Figure 7. Lack of Fit of CesA Theoretical Models into the Solution Scattering Envelope.

Supplemental Table 1. Templates Used for Threading.

Supplemental Movie 1. Predicted Structure of the Rice CesA8 CatD Threaded with the *Rhodobacter sphaeroides* BcsA Catalytic Domain (BcsA CatD).

Supplemental Movie 2. Determination 3D Surface Contour Structure from Small-Angle X-Ray Scattering Experiments.

Supplemental Movie 3. Docking Reconstruction of the BcsA Catalytic Domain within the Solution Scattering Envelope.

ACKNOWLEDGMENTS

We thank Robert F. Fischetti (APS, Argonne, IL) for helpful discussions on SAXS acquisition. We thank Marc Allaire (National Synchrotron Light Source, Brookhaven National Laboratory) for help collecting the SAXS data. We thank Debby Delmer for helpful comments on the nucleotide and nucleotide-sugar binding. This work was supported by the Center for Direct Catalytic Conversion of Biomass to Biofuels (C3Bio), an Energy Frontier Research Center funded by the U.S. Department of Energy, Office of Science, Office of Basic Energy Sciences, Award DE-SC0000997. Use of the National Synchrotron Light Source, Brookhaven National Laboratory, was supported by the U.S. Department of Energy, Office of Science, Office of Basic Energy Sciences, under Contract DE-AC02-98CH10886. SAXS method development was supported by National Science Foundation Grant 1158340 to L.M. D.K. acknowledges support from the National Institutes of Health (R01GM097528), the National Science Foundation (IIS1319551, DBI1262189, and IOS1127027), and the National Research Foundation of Korea (NRF-2011-220-C00004).

AUTHOR CONTRIBUTIONS

N.C.C. designed the experiments and assisted with substrate binding experiments. C.R. designed and cloned the ZnF and CatD domains and performed initial purifications. A.T.O. constructed additional clones and performed initial purifications, separations of monomer and dimer forms by SEC, and substrate binding assays. L.M. performed and interpreted SAXS and WAXS experiments. J.B. performed the docking analysis on the SAXS reconstructions. P.C. and M.C. performed surface envelope studies of SAXS results. H.R.K. and D.K. led studies of protein modeling with C.R. L.N.P. performed and analyzed the AUC experiments. M.E.H. assisted in interpretations. S.G. and J.T.B. performed early SEC experiments that revealed the CatD dimers. All authors assisted in writing the article.

Received April 18, 2014; revised June 8, 2014; accepted June 17, 2014; published July 10, 2014.

REFERENCES

Allaire, M., and Yang, L. (2011). Biomolecular solution x-ray scattering at the National Synchrotron Light Source. J. Synchrotron Radiat. **18:** 41–44.

Atanassov, I.I., Pittman, J.K., and Turner, S.R. (2009). Elucidating the mechanisms of assembly and subunit interaction of the cellulose synthase complex of *Arabidopsis* secondary cell walls. J. Biol. Chem. **284:** 3833–3841.

Atmodjo, M.A., Sakuragi, Y., Zhu, X., Burrell, A.J., Mohanty, S.S., Atwood, J.A., III., Orlando, R., Scheller, H.V., and Mohnen, D. (2011). Galacturonosyltransferase (GAUT)1 and GAUT7 are the core of a plant cell wall pectin biosynthetic homogalacturonan:galacturonosyltransferase complex. Proc. Natl. Acad. Sci. USA **108:** 20225–20230.

Benedetti, M., Leggio, C., Federici, L., De Lorenzo, G., Pavel, N.V., and Cervone, F. (2011). Structural resolution of the complex between a fungal polygalacturonase and a plant polygalacturonase-inhibiting protein by small-angle X-ray scattering. Plant Physiol. **157:** 599–607.

Benkert, P., Künzli, M., and Schwede, T. (2009). QMEAN server for protein model quality estimation. Nucleic Acids Res. **37:** W510–W514.

Bilecen, K., Ozturk, U.H., Duru, A.D., Sutlu, T., Petoukhov, M.V., Svergun, D.I., Koch, M.H., Sezerman, U.O., Cakmak, I., and Sayers, Z. (2005). *Triticum* durum metallothionein. Isolation of the gene and structural characterization of the protein using solution scattering and molecular modeling. J. Biol. Chem. **280:** 13701–13711.

Bowling, A.J., and Brown, R.M., Jr., (2008). The cytoplasmic domain of the cellulose-synthesizing complex in vascular plants. Protoplasma **233:** 115–127.

Carpita, N.C. (2011). Update on mechanisms of plant cell wall biosynthesis: how plants make cellulose and other (1->4)-β-D-glycans. Plant Physiol. **155:** 171–184.

Charnock, S.J., and Davies, G.J. (1999). Structure of the nucleotide-diphospho-sugar transferase, SpsA from *Bacillus subtilis*, in native and nucleotide-complexed forms. Biochemistry **38:** 6380–6385.

Currie, M.A., Cameron, K., Dias, F.M.V., Spencer, H.L., Bayer, E.A., Fontes, C.M.G.A., Smith, S.P., and Jia, Z. (2013). Small angle X-ray scattering analysis of *Clostridium thermocellum* cellulosome N-terminal complexes reveals a highly dynamic structure. J. Biol. Chem. **288:** 7978–7985.

Daily, M.D., Makowski, L., and Phillips, G.N., Jr.,., and Cui, Q. (2012). Large-scale motions in the adenylate kinase solution ensemble: coarse-grained simulations and comparison with solution X-ray scattering. Chem. Phys. **396:** 84–91.

DeLano, W.L. (2002). The PyMOL Molecular Graphics System. (San Carlos, CA: DeLano Scientific).

Delmer, D.P. (1999). Cellulose biosynthesis: exciting times for a difficult field of study. Annu. Rev. Plant Physiol. Plant Mol. Biol. **50:** 245–276.

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

Doblin, M.S., Pettolino, F.A., Wilson, S.M., Campbell, R., Burton, R.A., Fincher, G.B., Newbigin, E., and Bacic, A. (2009). A barley cellulose synthase-like CSLH gene mediates (1,3;1,4)-β-D-glucan synthesis in transgenic Arabidopsis. Proc. Natl. Acad. Sci. USA **106:** 5996–6001.

Edgar, R.C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. **32:** 1792–1797.

Erickson, H.P. (2009). Size and shape of protein molecules at the nanometer level determined by sedimentation, gel filtration, and electron microscopy. Biol. Proced. Online **11:** 32–51.

Gardiner, J.C., Taylor, N.G., and Turner, S.R. (2003). Control of cellulose synthase complex localization in developing xylem. Plant Cell **15:** 1740–1748.

Gerstein, M., Lesk, A.M., and Chothia, C. (1994). Structural mechanisms for domain movements in proteins. Biochemistry **33:** 6739–6749.

Giddings, T.H., Jr., Brower, D.L., and Staehelin, L.A. (1980). Visualization of particle complexes in the plasma membrane of *Micrasterias denticulata* associated with the formation of cellulose fibrils in primary and secondary cell walls. J. Cell Biol. **84:** 327–339.

Grant, T.D., Luft, J.R., Wolfley, J.R., Tsuruta, H., Martel, A., Montelione, G.T., and Snell, E.H. (2011). Small angle X-ray scattering as a complementary tool for high-throughput structural studies. Biopolymers **95:** 517–530.

Guerriero, G., Fugelstad, J., and Bulone, V. (2010). What do we really know about cellulose biosynthesis in higher plants? J. Integr. Plant Biol. **52:** 161–175.

Holm, L., and Rosenström, P. (2010). Dali server: conservation mapping in 3D. Nucleic Acids Res. **38: W545–W549.**

Hura, G.L., et al. (2009). Robust, high-throughput solution structural analyses by small angle X-ray scattering (SAXS). Nat. Methods **6:** 606–612.

Kelley, L.A., and Sternberg, M.J.E. (2009). Protein structure prediction on the Web: a case study using the Phyre server. Nat. Protoc. **4:** 363–371.

Kezuka, Y., Kojima, M., Mizuno, R., Suzuki, K., Watanabe, T., and Nonaka, T. (2010). Structure of full-length class I chitinase from rice revealed by X-ray crystallography and small-angle X-ray scattering. Proteins **78:** 2295–2305.

Kolinski, A. (2004). Protein modeling and structure prediction with a reduced representation. Acta Biochim. Pol. **51:** 349–371.

Koliński, A., and Bujnicki, J.M. (2005). Generalized protein structure prediction based on combination of fold-recognition with de novo folding and evaluation of models. Proteins **61** (suppl. 7)**:** 84–90.

Konarev, P.V., Volkov, V.V., Sokolova, A.V., Koch, M.H.J., and Svergun, D.I. (2003). PRIMUS: a Windows PC-based system for small-angle scattering data analysis. J. Appl. Crystallogr. **36:** 1277–1282.

Konarev, P.V., Petoukhov, M.V., Volkov, V.V., and Svergun, D.I. (2006). ATSAS 2.1, a program package for small-angle scattering data analysis. J. Appl. Crystallogr. **39:** 277–286.

Kozin, M.B., and Svergun, D.I. (2001). Automated matching of high- and low-resolution structural models. J. Appl. Crystallogr. **34:** 33–41.

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

Lamb, J.S., Zoltowski, B.D., Pabit, S.A., Li, L., Crane, B.R., and Pollack, L. (2009). Illuminating solution responses of a LOV domain protein with photocoupled small-angle X-ray scattering. J. Mol. Biol. **393:** 909–919.

Lee, C., Zhong, R., and Ye, Z.H. (2012). Arabidopsis family GT43 members are xylan xylosyltransferases required for the elongation of the xylan backbone. Plant Cell Physiol. **53:** 135–143.

Lieth, H. (1975). Primary production of the major vegetation units of the world. In Primary Productivity of the Biosphere, H. Lieth and R.H. Whittaker, eds (New York: Springer-Verlag), pp. 203–215.

Linder, M., and Teeri, T.T. (1997). The roles and function of cellulose-binding domains. J. Biotechnol. **57:** 15–28.

McGuffin, L.J. (2007). Benchmarking consensus model quality assessment for protein fold recognition. BMC Bioinformatics **8:** 345.

Minh, D.D., and Makowski, L. (2013). Wide-angle X-ray solution scattering for protein-ligand binding: multivariate curve resolution with Bayesian confidence intervals. Biophys. J. **104:** 873–883.

Morgan, J.L.W., Strumillo, J., and Zimmer, J. (2013). Crystallographic snapshot of cellulose synthesis and membrane translocation. Nature **493:** 181–186.

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

Nishimura, N., Hitomi, K., Arvai, A.S., Rambo, R.P., Hitomi, C., Cutler, S.R., Schroeder, J.I., and Getzoff, E.D. (2009). Structural mechanism of abscisic acid binding and signaling by dimeric PYR1. Science **326:** 1373–1379.

Osawa, T., Sugiura, N., Shimada, H., Hirooka, R., Tsuji, A., Shirakawa, T., Fukuyama, K., Kimura, M., Kimata, K., and Kakuta, Y. (2009). Crystal structure of chondroitin polymerase from *Escherichia coli* K4. Biochem. Biophys. Res. Commun. **378:** 10–14.

Palani, K., Kumaran, D., Burley, S.K., and Swaminathan, S. (2011). Crystal structure of a putative glycosyltransferase from *Bacteroides fragilis*. Protein Data Bank: 3BCV. (http://www.rcsb.org/pdb/explore/explore.do?structureId=3BCV).

Pear, J.R., Kawagoe, Y., Schreckengost, W.E., Delmer, D.P., and Stalker, D.M. (1996). Higher plants contain homologs of the bacterial *celA* genes encoding the catalytic subunit of cellulose synthase. Proc. Natl. Acad. Sci. USA **93:** 12637–12642.

Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C., and Ferrin, T.E. (2004). UCSF

Chimera—a visualization system for exploratory research and analysis. J. Comput. Chem. **25:** 1605–1612.

Rossmann, M.G., Moras, D., and Olsen, K.W. (1974). Chemical and biological evolution of nucleotide-binding protein. Nature **250:** 194–199.

Roy, A., Kucukural, A., and Zhang, Y. (2010). I-TASSER: a unified platform for automated protein structure and function prediction. Nat. Protoc. **5:** 725–738.

Sali, A., and Blundell, T.L. (1993). Comparative protein modelling by satisfaction of spatial restraints. J. Mol. Biol. **234:** 779–815.

Saxena, I.M., Lin, F.C., and Brown, R.M., Jr.,. (1990). Cloning and sequencing of the cellulose synthase catalytic subunit gene of *Acetobacter xylinum.* Plant Mol. Biol. **15:** 673–683.

Saxena, I.M., Brown, R.M., Jr., Fevre, M., Geremia, R.A., and Henrissat, B. (1995). Multidomain architecture of β-glycosyl transferases: implications for mechanism of action. J. Bacteriol. **177:** 1419–1424.

Saxena, I.M., and Brown, R.M., Jr.,., and Dandekar, T. (2001). Structure-function characterization of cellulose synthase: relationship to other glycosyltransferases. Phytochemistry **57:** 1135–1148.

Schuck, P. (2000). Size-distribution analysis of macromolecules by sedimentation velocity ultracentrifugation and lamm equation modeling. Biophys. J. **78:** 1606–1619.

Sethaphong, L., Haigler, C.H., Kubicki, J.D., Zimmer, J., Bonetta, D., DeBolt, S., and Yingling, Y.G. (2013). Tertiary model of a plant cellulose synthase. Proc. Natl. Acad. Sci. USA **110:** 7512–7517.

Shi, J., Blundell, T.L., and Mizuguchi, K. (2001). FUGUE: sequence-structure homology recognition using environment-specific substitution tables and structure-dependent gap penalties. J. Mol. Biol. **310:** 243–257.

Söding, J., Biegert, A., and Lupas, A.N. (2005). The HHpred interactive server for protein homology detection and structure prediction. Nucleic Acids Res. **33:** W244–W248.

Svergun, D.I. (1992). Determination of the regularization parameter in indirect-transform methods using perceptual criteria. J. Appl. Crystallogr. **25:** 495–503.

Svergun, D.I., and Koch, M.H.J. (2003). Small-angle scattering studies of biological macromolecules in solution. Rep. Prog. Phys. **66:** 1735–1782.

Svergun, D.I., Volkov, V.V., Kozin, M.B., and Stuhrmann, H.B. (1996). New developments in direct shape determination from small-angle scattering. 2. Uniqueness. Acta Crystallogr. A **52:** 419–426.

Svergun, D.I., Petoukhov, M.V., and Koch, M.H.J. (2001). Determination of domain structure of proteins from X-ray solution scattering. Biophys. J. **80:** 2946–2953.

Taylor, N.G. (2008). Cellulose biosynthesis and deposition in higher plants. New Phytol. **178:** 239–252.

Taylor, N.G., Howells, R.M., Huttly, A.K., Vickers, K., and Turner, S.R. (2003). Interactions among three distinct CesA proteins essential for cellulose synthesis. Proc. Natl. Acad. Sci. USA **100:** 1450–1455.

Taylor, N.G., Gardiner, J.C., Whiteman, R., and Turner, S.R. (2004). Cellulose synthesis in the Arabidopsis secondary cell wall. Cellulose **11:** 329–338.

UniProt Consortium (2013). Update on activities at the Universal Protein Resource (UniProt) in 2013. Nucleic Acids Res. **41:** D43–D47.

Vergara, C.E., and Carpita, N.C. (2001). β-D-glycan synthases and the CesA gene family: lessons to be learned from the mixed-linkage (1→3),(1→4)β-D-glucan synthase. Plant Mol. Biol. **47:** 145–160.

Wallner, B., and Elofsson, A. (2006). Identification of correct regions in protein models using structural, alignment, and consensus information. Protein Sci. **15:** 900–913.

Wang, J., Howles, P.A., Cork, A.H., Birch, R.J., and Williamson, R.E. (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.

Wang, J., Elliott, J.E., and Williamson, R.E. (2008). Features of the primary wall CESA complex in wild type and cellulose-deficient mutants of *Arabidopsis thaliana.* J. Exp. Bot. **59:** 2627–2637.

Williamson, T.E., Craig, B.A., Kondrashkina, E., Bailey-Kellogg, C., and Friedman, A.M. (2008). Analysis of self-associating proteins by singular value decomposition of solution scattering data. Biophys. J. **94:** 4906–4923.

Wong, H.C., et al. (1990). Genetic organization of the cellulose synthase operon in *Acetobacter xylinum.* Proc. Natl. Acad. Sci. USA **87:** 8130–8134.

Wu, S., and Zhang, Y. (2007). LOMETS: a local meta-threading-server for protein structure prediction. Nucleic Acids Res. **35:** 3375–3382.

Wu, S., and Zhang, Y. (2008). MUSTER: Improving protein sequence profile-profile alignments by using multiple sources of structure information. Proteins **72:** 547–556.

Yang, Y., Faraggi, E., Zhao, H., and Zhou, Y. (2011). Improving protein fold recognition and template-based modeling by employing probabilistic-based matching between predicted one-dimensional structural properties of query and corresponding native properties of templates. Bioinformatics **27:** 2076–2082.

Zhang, Y., and Skolnick, J. (2005). TM-align: a protein structure alignment algorithm based on the TM-score. Nucleic Acids Res. **33:** 2302–2309.

Zemla, A. (2003). LGA: A method for finding 3D similarities in protein structures. Nucleic Acids Res. **31:** 3370–3374.