Commentary

New hope for old dreams: Evidence that plant cellulose synthase genes have finally been identified

Candace H. Haigler and Richard L. Blanton
Department of Biological Sciences, Texas Tech University, Lubbock, TX 79409

It is common in the cellulose literature to read that cellulose is the most abundant macromolecule on earth. Authors perhaps hoped that the abundance of the polymer would compensate for the lack of progress in understanding the biochemistry and molecular biology of its synthesis in higher plants! Despite substantial effort, cellulose biogenesis has remained for decades one of the most enigmatic aspects of plant biochemistry, with only low activity observable in vitro, no synthase purified to homogeneity, and no relevant genes identified. Many scientists who tried to make progress eventually quit to concentrate on more productive and publishable research. However, the subject is of such critical importance to plant biology and global productivity that a few research groups worldwide persevered.

Cellulose is important because it is an almost ubiquitous component of plant cell walls. The high molecular weight chains of the (1 → 4)-β-D-glucan homopolymer form crystalline microfibrils of high tensile strength (Fig. 1a), and, in growing walls, the microfibril orientation constrains the direction of plant cell expansion. Therefore, cellulose synthesis is a fundamental determinant of plant growth rate (without it, plants cannot grow substantially) and morphogenesis (because plant cells do not migrate during development, direction of cell expansion determines plant form). Furthermore, cellulose is a substantial component of the biomass produced and decomposed in the biosphere because it is highly abundant in thick secondary walls of plant cells that conduct water and provide support, finally forming wood in trees. Cellulose also composes the secondary cell walls of cotton fibers, our most important natural fiber. Finally, cellulose biogenesis likely reflects many novel regulatory processes, since the cytoplasmic precursor UDP-glucose is converted at the plasma membrane into insoluble extracytoplasmic microfibrils, which are often laid down in precise patterns. For all these reasons, it is essential to understand how the synthesis of cellulose is regulated.

One of the research groups that persevered in the study of cellulose biogenesis was that of Deborah P. Delmer, who, along with her associate Yasushi Kawagoe, is a coauthor on the paper in this issue of the Proceedings identifying the first strong candidates for higher plant cellulose synthase genes (1). Other authors on the paper, Julie R. Pear, William E. Schreckengost, and David M. Stalker are from Calgene, a biotechnology company interested in the control and modification of cotton fiber development. Their approach was not biochemical, relying instead on random sequencing of plant cDNAs combined with insightful sequence analysis. This finding is noteworthy because it provides the first likely direct route to unraveling the cellular and molecular details of plant cellulose biogenesis.

Problems with the Biochemical Approach to Cellulose Biogenesis in Higher Plants

The past biochemical approach to studying cellulose synthesis was largely stymied by the failure of isolated membranes or permeabilized plant cells fed radiolabeled UDP-glucose to synthesize radiolabeled cellulose at high percentages and high rates as would occur in vivo. Instead, another homopolymer of glucose, callose, or (1 → 3)-β-linked glucose was synthesized preferentially (2, 3). In many ways this was not surprising, since callose synthesis is a wound response in plant cells (in addition to smaller scale developmental roles) (4), and this synthase likely remains poised to work in many plasma membranes. In a hypothesis born out of adversity, it was even suggested that the same enzyme might accomplish both cellulose and callose synthesis under different regulatory conditions, with the callose-specific conditions occurring in disrupted cells (5, 6).

Progress in the Cell Biology of Cellulose Biogenesis

Steady progress has been made through the years in identifying cell biological phenomena related to cellulose synthesis. In higher plants and some algae, cellulose-synthesizing cells were observed by freeze-fracture electron microscopy to have distinct aggregates of proteins in their plasma membranes (Fig. 1b). The number, arrangement, and location of these aggregates (called rosettes because they are circles of six particles) was proportional to the rate, pattern, and location of cellulose synthesis (3, 7, 8). The aggregation of proteins in association with cellulose biogenesis is necessary to generate about 40 chains in close proximity so that the natural allomorph, cellulose I, will crystallize. Further modification of the extent and pattern of aggregation of plasma membrane proteins provides cellular control over microfibril size (9). Although not identified biochemically, these aggregated proteins provided evidence that the cellular organization of cellulose biogenesis would prove to be a fascinating and possibly unique cellular process. Similarly, evidence accumulated that the location and orientation of microtubules affected the pattern and orientation of cellulose microfibril deposition (Fig. 2a and b), possibly through direct tracking of rosettes on microtubules (now considered less likely) or microtubule-mediation of special membrane fluidity domains (a hypothesis awaiting proof) (7).

Identification of Bacterial Cellulose Synthase Genes

A major breakthrough occurred when it became possible to obtain high-activity in vitro cellulose synthesis with solubilized membrane proteins from the Gram-negative bacterium, Acetobacter xylinum (11). Two independent efforts led to the identification of the genes involved directly in cellulose synthesis (12, 13). In one, complementation of cellulose-minus strains led to the identification of an operon of four genes (13). In the other, microsequencing of purified polypeptides identified by photoaffinity labeling with a UDP-glucose analogue allowed the isolation of the gene for the catalytic subunit of the cellulose synthase (12). The four genes of the operon coded for the cellulose synthase catalytic subunit, a regulatory protein, and two other proteins that likely mediate biophysical aspects of cellulose biogenesis (13, 14).

For 4 years, the cellulose synthesis community was tantalized by the Acetobacter genes. Despite the attempts of several research groups, use of probes prepared from the A. xylinum catalytic subunit gene did not result in the identification of
nucleic acid sequences from plants or the slime mold *Dictyostelium discoideum*. With only two gene sequences, meaningful regions of sequence conservation could not be identified. In fact, the two sequences (from different strains of the same species) diverged at the nucleotide level by 28%, which suggested that conservation across larger taxonomic groups would not be high. Other probes that would have proved useful, such as antibodies to fusion proteins prepared from the genes (15), were not generally available because of commercial considerations.

In 1995, the activity that led to the discovery by Pear *et al.* (1) began. A second *Acetobacter* cellulose synthase gene (*acsAII*) was identified (16). Studies of *Agrobacterium tumefaciens* yielded several genes involved in cellulose synthesis and presented the prospect of a significantly different biochemical pathway for this process (17, 18). However, one of the genes, *celA*, had significant sequence homology to the *Acetobacter* cellulose synthase catalytic subunit. Surprisingly, the *Escherichia coli* genome sequencing project yielded another gene with sequence homology to the *Acetobacter* synthase (19), an observation that remains a curiosity, since *E. coli* is not known to synthesize cellulose. Sequence comparisons of the five related genes then available revealed regions of amino acid conservation. But once again, these observations did not lead to the isolation of relevant genes from plants or *Dictyostelium*. Approaches based on photoaffinity labeling (both with UDP-gluc analogues and the novel regulator of cellulose synthesis in *Acetobacter*), mutational analysis, and purification of polypeptides from *in vitro* synthesis reactions also did not lead to the genes for the plant cellulose synthase (2, 3). Instead, the key was to pursue sequence comparisons of UDP-gluc binding proteins in general.

**Conserved Sequence Motifs Provide the Critical Key to the Higher Plant Problem**

Linear sequence alignments of a number of UDP-gluc and UDP-glcnAc-binding proteins led to the identification of a region containing conserved residues. Delmer and Amor (2) identified a relatively small motif, which was found in the larger region identified by Saxena *et al.* (14). The scattered conserved residues in this region meant that it was of limited utility for designing probes for screening cDNA libraries. However, the identification of conserved residues inspired an approach that combined a sensitive method of sequence analysis called hydrophobic cluster analysis (HCA) (20, 21) with a consideration of the reaction mechanism involved in different types of glycosyltransferase reactions (22).

In simple terms, HCA combines secondary structure prediction and sequence alignment; in effect, it reveals threedimensional similarities in polypeptides in a manner that once awaited three-dimensional reconstructions. It has proven to be useful in comparing proteins of similar function that have low sequence homology, it can reveal regions of similar structure that are separated by regions of high variability, and, significantly, it can predict the catalytic residues that would be expected to be conserved in sequences of proteins with related functions (20–22). A variety of glycosyltransferases that use nucleotide diphospho sugars as substrates were subjected to HCA (22). They were found to fall into two groups, those that contained domain A only (transferases that add a single sugar) and those that contained domains A and B (transferases that add several sugars). A common pattern in domain A was the presence of conserved aspartate residues in loops at the C-terminal ends of β-strands. Domain B also featured a conserved aspartate and a conserved sequence motif of QXXRW. Saxena *et al.* (22) postulated a reaction mechanism based upon three UDP-gluc binding sites to explain how the cellulose synthase can add glucoses that alternate spatially by approximately 180° (a characteristic of (1→4)-β-gluc linkages). In the model, each binding site included a conserved aspartate and the third binding site was followed by the QXXRW motif. The spacing and location of these conserved residues relative to secondary structure revealed by HCA were important clues for Pear *et al.* (1).

Random data base searches using the bacterial cellulose synthase sequences or the conserved bacterial motifs do not identify any plant sequences with significant sequence similarity. However, many expressed sequence tags, which are partial cDNA sequences generated by various genome projects, are noted in their data base file to have sequence similarity to the bacterial cellulose synthase, but this similarity is usually insignificant and meaningless. Establishing a se-

---

**Fig. 1.** Biophysical aspects of cellulose biogenesis. (a) Cellulose microfibrils overlaying plasma membrane (Right) in the secondary cell wall of a *Zinnia elegans* tracheary element. (Bar = 100 nm.) (b) Rosettes in the plasma membrane underlying the cellulose-rich secondary wall thickening of a differentiating *Z. elegans* tracheary element. (Bar = 30 nm.) (Micrographs by M. J. Grimson and C.H.H.)
to be enriched in sequences related to cellulose synthesis. In addition, careful sequence comparisons and analyses were feasible for the relatively limited number of sequences emerging from the cotton fiber cDNA sequencing project. Two clones with limited homology to celA were identified and subjected to refined functional domain comparisons with the bacterial synthase genes, leading to their identification as plant homologues for the bacterial cellulose synthase, as described by Pear et al. (1).

In these rather large sequences, only four small amino acid subdomains are highly conserved within three regions of lower overall conservation. The U-1 subdomain contains the conserved motif identified by Delmer and Amor (2). The conserved region identified by Saxena et al. (14) begins at the middle of the H-1 region and includes the U-1 subdomain, H-2 region, and U-2 subdomain. The pattern predicted by Saxena et al. (22) is found; conserved aspartates are found in the U-1, U-2, and U-3 subdomains, and the QXXR motive is found in the U-4 subdomain. Their prediction that these would be involved in UDP-glc binding and catalysis (22) is supported by experiments of Pear et al. (1) in which a partial recombinant polypeptide with all four conserved subdomains can bind UDP-glc, whereas binding is destroyed when the U-1 subdomain is deleted.

The more highly conserved subdomains, together with the more loosely conserved H1–H3 regions, must be sufficient to confer cellulose synthase activity, and these have been relatively highly conserved at the amino acid level during evolutionary time since the separation of bacteria from higher plants. The small U1–U4 subdomains emphasize that conservation of the catalytic function of enzymes requires only very small motifs, a finding that also applies to other enzyme families (21). The substantial sequence divergence at the N and C termini and in an internal hypervariable region between the higher plant genes undoubtedly points to multiple ways that the synthase can be regulated or interact with other cellular components (1). This point is further emphasized by the existence of a multi-gene family of the putative higher plant cellulose synthase (1).

Challenges for Future Work

The conserved subdomains are obvious candidates for primers to be used in PCR-based methods to isolate cellulose synthase homologues from other organisms. Pitfalls in this approach include the low conservation at the nucleotide level in the subdomains, the difficulty of predicting expected fragment sizes based upon the intervening plant-conserved and hypervariable regions, and the need to have long-range sequence information of sufficient accuracy to provide a translated sequence that can be subjected to HCA. Candidate sequences must then be shown to be cellulose synthases rather than other UDP-glc binding proteins.

The paper of Pear et al. (1) does not prove that the identified genes code for higher plant cellulose synthases. Proof would require synthesis of cellulose by a recombinant protein in vitro, complementation of a cellulose-negative mutant (which cannot exist in higher plants), or tissue- and stage-specific suppression of gene expression coupled with a cellulose-negative phenotype (e.g., in the secondary wall of cotton fibers). The complementing molecular and biochemical data do however make a strong case that the correct genes have been identified (1). The temporal expression of the genes (as revealed by Northern blot analysis of cotton fiber RNAs) from 17–35 days after flowering in the developing cotton fiber is consistent with the well-characterized development of this unusual cell, which synthesizes a thick secondary wall composed of almost pure cellulose at this stage. In addition, binding of UDP-glc by the recombinant polypeptide is favored by ionic conditions that
have been well-established to favor cellulose over callose synthesis and disfavored by the opposite conditions.

Exploration of regulatory and cellular phenomena associated with high plant cellulose synthesis is sure be one of the most productive and interesting areas of plant biology in future years. The model proposed by Saxena et al. (22) can be tested by site-directed mutagenesis experiments. If the catalytic domain of the cellulose synthase is cytoplasmic as predicted by the model of Pear et al. (1), then there must either be another glycosyltransferase to transfer a low molecular weight glucan intermediate into the extracytoplasmic space for microfibril formation [analogously to the proposed mechanism in A. tumefaciens (18), but not A. xilum (24)] or there must be substantial transport capacity in the plasma membrane through which polymerized glucan chains can exit. Already a family of proteins that may interact with the cellulose synthase in cotton is being identified. For example, Amor et al. (25) recently provided evidence that cellulose synthesis in cotton fibers requires a membrane-bound sucrose synthase, which channels UDP-gluc to the cellulose synthase. Because permeabilized cotton fibers will synthesize cellulose more efficiently from sucrose than UDP-gluc, this finding may also illuminate one reason for the previous difficulties with in vitro synthesis of cellulose. Sucrose synthase often appears colocalized with the pattern of cellulose microfibrils in cotton (Fig. 2 c and d), and further work may demonstrate directly that it is closely coupled to the cellulose synthase. Proteins that interact with the cellulose synthase may also include a homologue to rac, a GTP-binding protein that may modulate glucan synthesis activity by regulation of the cytoskeleton, activation of NADPH oxidase to up-regulate reactive oxygen derivatives, or other mechanisms (26). Similarly, an annexin-like protein that can affect callose synthase activity has been identified (27).

Inhibitor studies, mutant analysis, and biochemical experiments have indicated that primary and secondary wall synthesis may be accomplished by different cellulose synthases (2, 3, 28), a finding that is consistent with the existence of a small gene family in cotton described by Pear et al. (1). Researchers should now be able to use a variety of molecular and microscopic tools to elucidate the details of the cellulose synthase complex and its assembly, regulation, and interaction with the cytoskeleton, signal transduction mechanisms, and membranes. For example, it should not be much longer until the biochemical identity of the aggregated proteins in the rosette associated with microfibril formation (Fig. 1b) is determined. These remarkable complexes will provide a fascinating study of assembly of multiple subunit membrane proteins.

One thing seems certain—the identification of probable cellulose synthase genes in higher plants will, as predicted by D. P. Delmer in 1987 (6), stimulate additional scientists to join the field and contribute to future progress. We can be sure that the elucidation of the molecular and cellular details of cellulose synthesis will be so interesting in its own right that it will no longer be necessary to remind readers of the abundance of the polymer! Therefore, the paper of Pear et al. (1) represents new hope for old dreams in one of the most fundamental aspects of plant regulatory mechanisms.