

MECHANISMS IN CELLULOSE BIOSYNTHESIS

Inder M. Saxena¹, T. Dandekar², R. Malcolm Brown, Jr.¹

School of Biological Sciences, University of Texas at Austin, Austin, TX 78712¹; European Molecular Biology Laboratory, Postfach 102209, D-69012, Heidelberg, Germany²

Cellulose is a major industrial biopolymer in the forest products, textile, and chemical industries. It also forms a large portion of the biomass useful in the generation of energy. Moreover, cellulose-based biomass is a renewable energy source that can be used for the generation of ethanol as a fuel. Cellulose is synthesized by a variety of living organisms, including plants, algae, bacteria, and animals. It is the major component of plant cell walls with secondary cell walls having a much higher content of cellulose. The biosynthesis of cellulose essentially proceeds by the polymerization of glucose residues using an activated substrate UDP-glucose. In plants, cellulose is synthesized on the plasma membrane by the enzyme cellulose synthase that is present in the membrane. In the bacterium *Acetobacter xylinum*, the enzyme cellulose synthase is present on the cytoplasmic membrane, and the cellulose product is obtained extracellularly. However, in other organisms, cellulose is found to be synthesized in other regions of the cell. In the alga *Pleurochrysis*, cellulose scales are formed in the Golgi apparatus and then deposited on the cell surface.

Cellulose is an aggregate of glucan chains that are arranged in a specific manner to give rise to a crystalline state. Although cellulose produced by different organisms has the same chemical composition (polymer of β -1,4-linked glucose residues), there are remarkable differences in the physical properties of the cellulose product, mainly in the length of the glucan chains (as represented by degree of polymerization) and the crystallinity and crystalline form of the cellulose product. Depending upon the specific organism, this crystalline state is different, and it defines the physical properties of the product such as its strength, solubility in various solvents, and accessibility to various modifying reagents.

Cellulose biosynthesis proceeds in at least two stages – polymerization and crystallization. The first stage is catalyzed by the enzyme cellulose synthase, and the second stage is dependent on the organization of the cellulose synthases possibly with other proteins such that the glucan chains are assembled in a crystalline form.

Cellulose biosynthesis in *Acetobacter xylinum*.

The bacterium *A. xylinum* produces pure cellulose. A single cell may polymerize up to 200,000 glucose residues per second into β -1,4-glucan chains. Advantages of using a bacterial system for production of cellulose is that the bacterium grows rapidly under controlled conditions and produces cellulose from a variety of carbon sources including glucose, ethanol, sucrose, and glycerol. Moreover, genetic analysis is aided by the isolation of a large number of mutants affected in cellulose biosynthesis, and these mutants have allowed identification of specific genes involved in this process. The sites of synthesis of cellulose have been visualized using freeze-fracture techniques. Assembly of the ribbon takes place in several steps. The cellulose-synthesizing sites on the cell surface are made of 3.5 nm pores that are arranged in a linear row. Each pore covers a 10 nm-particle that consists of the cellulose-synthesizing enzymes involved in the polymerization reaction and accessory proteins possibly involved in other functions. Each 10 nm particle produces a number of glucan chains that form a 1.5 nm sub-elementary fibril. The sub-elementary fibrils together form the microfibril (1).

The biosynthetic pathway for cellulose biosynthesis is very well understood in *A. xylinum*. The pathway from the substrate glucose to cellulose involves a number of reactions in which glucose is first converted to glucose-6-phosphate by the enzyme glucokinase. In the second step, glucose-6-phosphate is converted to glucose-1-phosphate by the enzyme phosphoglucomutase. In the next step, glucose-1-phosphate is converted to UDP-glucose in the presence of UTP and the enzyme UDPG pyrophosphorylase. The UDP-glucose produced is used as a substrate by the enzyme cellulose synthase. In *A. xylinum*, this enzyme is activated by the cyclic nucleotide, c-di-GMP.

The cellulose synthase activator c-di-GMP is synthesized in *A. xylinum* by the enzyme diguanylate cyclase, and the concentration of this

activator is regulated by the action of phosphodiesterases. Genes for the diguanylate cyclase and phosphodiesterases have been identified and they are organized together in more than one operon in the *A. xylinum* chromosome (2).

Cloning of the genes for cellulose synthesis from *A. xylinum*

In our laboratory, purification of cellulose synthase activity in *A. xylinum* led to the identification of 2 polypeptides of molecular weight 83-kD and 93-kD in the purified fraction (3). When the purified fraction was incubated with a radioactively labeled substrate, the 83-kD polypeptide bound to the substrate suggesting that it was the catalytic subunit (4). Amino acid sequences obtained from the 83-kD polypeptide and 93-kD polypeptide were used to clone the genes for cellulose synthase and other proteins (5). A similar set of genes was also identified by analysis of *A. xylinum* mutants affected in cellulose biosynthesis (6). We believe that the proteins coded by these genes form the cellulose-synthesizing complex that is present in the cytoplasmic membrane. Three genes *acsAB*, *acsC*, and *acsD* are organized in an operon with the *acsAB* gene coding for a 168-kD polypeptide that functions as the cellulose synthase (7). The *acsC* and *acsD* genes code for polypeptides of molecular size 138-kD and 17-kD respectively. The AcsC and AcsD polypeptides are probably organized with the cellulose synthase (AcsAB polypeptide) in the cellulose-synthesizing complex. No enzymatic function has been assigned to the AcsC and AcsD polypeptides, and their role in cellulose biosynthesis is believed to be structural. Upon isolation of cellulose synthase activity, the 168-kD AcsAB polypeptide is observed as two separate polypeptides of 83-kD and 93-kD with the 83-kD polypeptide functioning as the catalytic subunit and the 93-kD polypeptide functioning as the c-di-GMP binding subunit. Analysis of mutants in which the *acsAB* gene was inactivated led to the identification of *acsAII*, a second gene for cellulose synthase (8). The role of this gene in cellulose biosynthesis is not well understood. Apart from genes organized in the *acs* operon, two genes that may function in cellulose biosynthesis have been identified by analysis of mutants (9). One of these genes codes for an endoglucanase. The role of endoglucanase in cellulose biosynthesis is not very clear, although a similar gene has also been observed in *Agrobacterium tumefaciens*, another bacterium that produces cellulose (10).

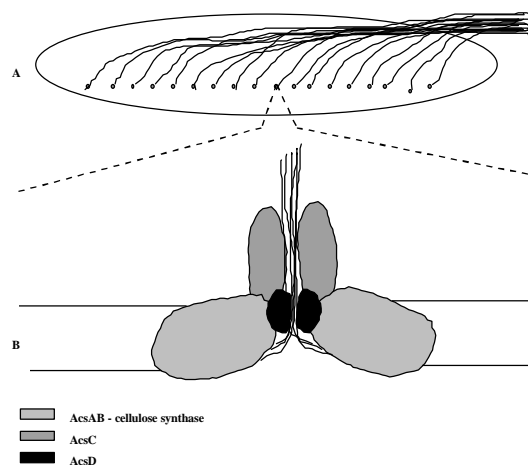


Figure 1. Site of cellulose synthesis in *Acetobacter xylinum*. (A) A single bacterial cell contains a number of cellulose-synthesizing sites that are observed as a linear row of particles on the membrane surface. Each particle synthesizes a sub-elementary fibril predicted to consist of 16 glucan chains. The sub-elementary fibrils from different particles join to give rise to the ribbon of microfibrils. (B) Each particle contains a number of cellulose synthases (AcsAB protein) and associated proteins (AcsC and AcsD). The arrangement of the cellulose synthase and associated proteins is shown in a section of the cellulose synthesizing complex in the cytoplasmic membrane of *A. xylinum*. Cellulose synthase is a membrane protein present in the cytoplasmic membrane. The globular region of this protein faces the cytoplasm and it is this region that contains the conserved amino acid residues involved in the polymerization step.

Cellulose synthase

The enzyme cellulose synthase uses α -linked UDP-glucose as the substrate. It forms a polymer with β -1,4-glycosidic linkage by inversion of configuration at the anomeric carbon. Every glucose residue in the polymer chain is rotated 180° with respect to its neighbor thereby displaying a twofold screw axis. Interestingly, c-di-GMP, the activator of cellulose synthase in *A. xylinum* has a twofold degree of rotational symmetry (11). The glucan chains are synthesized by cellulose synthase in a processive manner in which the enzyme remains tightly bound to the polymeric product during chain elongation.

Cellulose synthase is a membrane protein with a number of transmembrane regions. It is believed that the globular region of the protein is involved in the catalytic function. When the amino acid sequence of this region was compared with the sequence of other proteins in the database, it was found that cellulose synthase showed weak similarity to a few proteins. These proteins were identified as β -glycosyltransferases and when these proteins were aligned with cellulose synthase, it was found that certain amino acid residues were conserved in all these sequences (12). Since the glycosyl transfer reaction proceeds by an acid-base catalytic mechanism, we predicted that the residues involved in catalysis were either glutamic acid or aspartic acid residues. At least three aspartic acid residues and a QXXRW motif were found to be conserved in cellulose synthase and other proteins such as the chitin synthase, hyaluronan synthase, NodC, and others. All these proteins function as processive enzymes since they transfer more than a single sugar residue. On the other hand, only two conserved aspartic acid residues were found in proteins that transferred only a single sugar residue (non-processive enzymes). Based upon these observations we proposed a model for polymerization in which two UDP-glucose residues were predicted to bind in the catalytic pocket such that the glucose moiety of each substrate molecule was positioned 180° with respect to the other residue. During polymerization, two β -1,4 glycosidic bonds are formed either sequentially or simultaneously and the growing chain is elongated two glucose residues at a time (12). Experiments using *A. xylinum* have shown that the glucan chain is elongated from the non-reducing end (13).

Structural features of the globular region of *A. xylinum* cellulose synthase

Cellulose synthase has not been purified from any source, and as a result it has not been possible to determine its structure. Because it is a membrane protein, it is difficult to purify it and obtain crystals for determination of its structure. Since the catalytic activity of the cellulose synthase is present in the globular region of this protein, we have predicted the structure of this region by folding a 285 amino acid segment of the *A. xylinum* cellulose synthase using the genetic algorithm (14). The predicted structure shows that the globular region of the cellulose synthase folds into a single large domain with dimensions of roughly $60 \times 60 \times 70 \text{ \AA}$. The folded structure has a mixed topology of β -strands and α -helices (10 strands and 8 helices). The helices and strands do not alternate regularly and are interrupted by long loop regions. Importantly, the predicted structure shows the presence of a central, elongated cavity that can accommodate two UDP-glucose residues. The residues in the QXXRW motif are present in a region close to the central cavity. Using the genetic algorithm, we also predicted the structure of the globular region of the cotton cellulose synthase and notice that it shows an arrangement of strands and helices similar to the globular region of the *A. xylinum* cellulose synthase. The predicted structure of the globular region of the *A. xylinum* cellulose synthase was found to be similar with recently determined crystal structures of other glycosyltransferases including the SpsA protein from *Bacillus subtilis* (15) and bovine β 4Gal-T1 catalytic domain (16).

Mechanism of cellulose biosynthesis

In light of the available evidence, we have proposed a general model for cellulose synthase and other processive β -glycosyltransferases. Presently, there is no evidence for the requirement of a primer by cellulose synthase. We believe that initiation of cellulose synthesis probably takes place by the formation of a cellobiose unit in the catalytic pocket of cellulose synthase. In the next step, the cellobiose unit moves from the catalytic pocket into the glucan-binding region of cellulose synthase. During chain elongation, two UDP-glucose molecules enter the catalytic pocket, and two new β -1,4 linkages are formed either sequentially or simultaneously. This model

predicts the presence of two catalytic centers in the enzyme active site to take care of the 180° rotation of alternate sugar residues in the glycan chain. As a result neither the enzyme nor the glycan chain needs to rotate for addition of every sugar residue. However, is there a need for two catalytic centers in cellulose synthase and possibly other processive β -glycosyltransferases to account for the 180° rotation of alternate sugar residues in the glycan chain? For example, the glucose residues can be added at a single catalytic site and the alternating residues could relax into opposing orientations after they have exited the catalytic site. This and other possibilities in support of a single catalytic site have been proposed and are being debated in the absence of a crystal structure of the active site of cellulose synthase. Apart from the crystal structure of the cellulose synthase, we believe that the structure of the glucan chain in the active site may also be important in understanding the mechanism of cellulose biosynthesis. For example, do the glucose residues in the glucan chain in the active site have intramolecular hydrogen bonds or do these hydrogen bonds form after the glucan chain has exited the active site and are formed at a later stage?

Identification of genes for cellulose biosynthesis in plants

Although the gene for cellulose synthase was first identified in *A. xylinum* in 1990, it was not until 1996 that the first cellulose synthase gene was identified from a higher plant (17). Unexpectedly, the plant gene was not identified by DNA-DNA hybridization experiments using the bacterial gene as a probe. This, in spite of the fact that one would have predicted sequence similarity between the bacterial and the plant cellulose synthase genes. Instead, the plant gene was identified by sequencing random cDNAs from a cotton fiber library and analyzing the derived amino acid sequences for presence of the conserved features identified in the bacterial cellulose synthase and other processive β -glycosyltransferases. Mutant analysis, specifically in *Arabidopsis thaliana*, and large scale efforts to sequence the expressed genes and complete genomes in a number of plants has led to the identification of a large glycosyltransferase gene family. Members of this gene family code for proteins that are predicted to function as processive β -glycosyltransferases based on the presence of the 'D,D,D35QXXRW' motif (18). Some of these genes code for cellulose synthases (referred to as CesAs), while others code for proteins with an as yet to be determined function (referred to as cellulose synthase-like, Csl) (19). The Csl proteins show sequence similarities to the cellulose synthases and are believed to function in the biosynthesis of non-cellulosic polysaccharides. In *A. thaliana*, over 40 protein sequences have been identified to constitute the CesA superfamily. Based on sequence comparisons, these proteins are classified into 7 families. The CesA family contains the cellulose synthases (about a dozen members), while the remaining protein members are grouped into the CslA, CslB, CslC, CslD, CslE, and CslG families (<http://cellwall.stanford.edu/cellwall/>).

The identification of almost a dozen genes coding for cellulose synthases in *A. thaliana* was surprising. Studies with other plants (cotton, maize, rice, pine, poplar, etc) have also shown the presence of a large number of cellulose synthase genes in them. Why do plants have so many cellulose synthase genes? The proteins coded by the different cellulose synthase genes may have different catalytic properties, they may be tissue-specific, or they may provide functional redundancy. In fact, cellulose synthases (Rsw-1 and Ath-A) involved in primary cell wall formation can be differentiated by sequence comparisons from cellulose synthases (Ce1A1, Irx3, and Ath-B) involved in secondary cell wall formation (20). These observations suggest that there is a differential expression of the cellulose synthase genes with the cellulose synthases having catalytic properties that distinguishes them from one another. For example, cellulose synthases active during secondary cell wall formation probably synthesize a cellulose product with a higher degree of polymerization.

Cellulose-synthesizing complexes and cellulose microfibril structure

Different organisms produce cellulose with properties that show differences in molecular weight, microfibril size and shape, crystallinity, and in the ratio of the I α and I β sub-allomorphs of cellulose I. Why do different organisms produce cellulose with different physical properties? Although some of these unique properties may be due to differences in the cellulose synthases present in different organisms, we believe that the organization of the cellulose synthesizing complexes (also referred to as terminal complexes or TCs) in the plasma membrane probably plays a major role in the physical properties of the cellulose product (21). So far two types of organizations

have been observed in the various cellulose producing organisms – rosette type or linear type. In rosette TCs, usually a hexagonal array of particles is observed by electron microscopy following freeze-fracture. In some cases, as in the case of the alga *Coleochaete*, the rosette is found to have an eightfold symmetry. The rosette TCs may be present individually as in the case of land plants, *Chara*, and *Nitella*, or they may be further organized into larger complexes as observed in *Micrasterias*, *Closterium*, and *Spirogyra*. Linear TCs have been observed as single rows in the bacterium *A. xylinum* or as multiple rows found in the algae *Valonia*, *Boergesenia*, *Oocystis*, and others. Sufficient information is now available from a number of organisms regarding TC structure and geometry (22). How does TC structure relate to microfibril shape and size, and the ratio of the cellulose I sub-allomorphs in cellulose I? For example, cellulose synthesized by rosette TCs in vascular plants is a mixture of cellulose I α and cellulose I β , with cellulose I α forming a smaller proportion as compared to cellulose synthesized in organisms with linear TCs. It will be important to determine if there is a correlation between TC geometry and the relative proportion of cellulose I sub-allomorphs found in different organisms (22).

Future directions

In the last decade, dramatic progress has been made in our understanding of cellulose biosynthesis in a variety of organisms. Genes for cellulose synthases have now been identified in bacteria, plants, and more recently in the amoeba *Dictyostelium discoideum* (23). The structure of cellulose synthesizing complexes (TCs) has been observed in bacteria, algae, plants, and animals. These complexes contain cellulose synthases and possibly other proteins that function during cellulose biogenesis. One of the challenges now is to understand the nature of the proteins in these complexes, and determine how they are regulated and organized to form a structure that synthesizes cellulose so efficiently.

Since cellulose is a major component of the biomass used in a number of industries, one of our goals is to utilize our knowledge of cellulose biosynthesis for improved cellulose production in the bacterium *A. xylinum* and selected plants including forest trees. In certain organisms we propose to produce cellulose with modified physical and chemical properties to be useful in a number of industrial applications. We are also interested in screening for cellulose biosynthesis in a variety of nitrogen-fixing blue-green algae. Cellulose production has been reported in certain blue-green algae (24) and one of our goals is to transfer genes for cellulose biosynthesis from other organisms into blue-green algae for large-scale economic production of cellulose.

We acknowledge support from the Department of Energy (DE-FG03-94ER20145).

References

- (1) Haigler, C.H.; Benziman, M. In: *Cellulose and other natural polymer systems* (Ed. R.M.Brown, Jr.), Plenum press, New York. **1982**, 273.
- (2) Tal, R.; Wong, H.C.; Calhoon, R.; Gelfand, D.; Fear, A.L.; Volman, G.; Mayer, R.; Ross, P.; Amikam, D.; Weinhouse, H.; Cohen, A.; Sapir, S.; Ohana, P.; Benziman, M. *J. Bacteriol.* **1998**, *180*, 4416.
- (3) Lin, F.C.; Brown, R.M., Jr. In: *Cellulose and Wood – Chemistry and Technology* (Ed. C. Schuerch), John Wiley and Sons, New York. **1989**, 473.
- (4) Lin, F.C.; Brown, R.M., Jr.; Drake, R.R. Jr.; Haley, B.E. *J. Biol. Chem.* **1990**, 4782.
- (5) Saxena, I.M.; Lin, F.C.; Brown, R.M., Jr. *Plant Mol. Biol.* **1990**, *15*, 673.
- (6) Wong, H.C.; Fear, A.L.; Calhoon, R.D.; Eichinger, G.H.; Mayer, R.; Amikam, D.; Benziman, M.; Gelfand, D.H.; Meade, J.H.; Emerick, A.W.; Bruner, R.; Ben-Bassat, A.; Tal, R. *Proc. Natl. Acad. Sci. USA.* **1990**, *87*, 8130.
- (7) Saxena, I.M.; Kudlicka, K.; Okuda, K.; Brown, R.M., Jr. *J. Bacteriol.* **1994**, *176*, 5735.
- (8) Saxena, I.M.; Brown, R.M., Jr. *J. Bacteriol.* **1995**, *177*, 5276.
- (9) Standal, R.; Iversen, T.-G.; Coucheron, D.H.; Fj arvik, E.; Blatny, J.M.; Valla, S. *J. Bacteriol.* **1994**, *176*, 665.
- (10) Matthyse, A.; White, S.; Lightfoot, R. *J. Bacteriol.* **1995**, *177*, 1069.
- (11) Egli, M.; Gessner, R.V.; Williams, L.D.; Quigley, G.J.; van der Marel, G.A.; van Boom, J.H.; Rich, A.; Frederick, C.A. *Proc. Natl. Acad. Sci. USA.* **1990**, *87*, 3235.
- (12) Saxena, I.M.; Brown, R.M., Jr.; Fevre, M.; Geremia, R.A.; Henrissat, B. *J. Bacteriol.* **1995**, *177*, 1419.
- (13) Koyama, M.; Helbert, W.; Imai, T.; Sugiyama, J.; Henrissat, B. *Proc. Natl. Acad. Sci. USA.* **1997**, *94*, 9091.
- (14) Dandekar, T.; Argos, P. *J. Mol. Biol.* **1994**, *236*, 844.
- (15) Charnock, S.J.; Davies, G.J. *Biochemistry* **1999**, *38*, 6380.
- (16) Gastinel, L.N.; Cambillau, C.; Bourne, Y. *EMBO J.* **1999**, *18*, 3546.
- (17) Pear, J.R.; Kawagoe, Y.; Schreckengost, W.E.; Delmer, D.P.; Stalker, D.M. *Proc. Natl. Acad. Sci. USA.* **1996**, *93*, 12637.
- (18) Saxena, I.M.; Brown, R.M., Jr. *Cellulose* **1997**, *4*, 33.
- (19) Delmer, D.P. *Ann. Rev. Plant Physiol. Plant Mol. Biol.* **1999**, *50*, 245.
- (20) Taylor, N.G.; Scheible, W.-R.; Cutler, S.; Somerville, C.R.; Turner, S.R. *Plant Cell* **1999**, *11*, 769.
- (21) Brown, R.M. Jr.; Saxena, I.M. *Plant Physiol. Biochem.* **2000**, *38*, 57.
- (22) Brown, R.M. Jr. *J. Macro. Sci.: Pure Appl. Chem.* **1996**, *A33*, 1345.
- (23) Blanton, R.L.; Fuller, D.; Iranfar, N.; Grimson, M.J.; Loomis, W.F. *Proc. Natl. Acad. Sci. USA.* **2000**, *97*, 2391.
- (24) De Winder, B.; Stal, L.J.; Mur, L.R. *J. Gen. Microbiol.* **1990**, *136*, 1645.