Chapter 1

The Plant Biomass

COPYRIGHTED MATERIAL

1.1 The Structure of Plant Cell Wall

When we talk of plant biomass in terms of its use for biofuel and other biorefineries, we mostly mean the plant cell wall that makes up for more than 50% of the plants dry weight. This most outside located structure of the plant cell is also its most distinguishing feature, and because of its rigidity an essential component for their sedentary lifestyle. This rigidity also provides the strength to withstand mechanical stress and forms and maintains the plants shape. Despite this rigidity, nevertheless, the cell wall is a dynamic and metabolically active entity that plays crucial roles in growth, differentiation, and cell-to-cell communication and acts as a pressure vessel that prevents overexpansion when water enters the cell (Raven et al., 1999).

Plant cell walls typically consist of three layers: the "primary cell wall" (a rather thin but continuously extending layer that is produced by growing cells), the "secondary cell wall" (a thick layer that is formed inside the primary cell wall after termination of cell growth), and the "middle lamella" (the outermost layer that forms an interface between secondary walls of adjacent plant cells and glues them together) (Figure 1.1).

The primary cell wall consists of the polysaccharides cellulose, hemicellulose, and pectin (Rose et al., 2004). The cellulose thereby aggregates to microfibrils that are covalently linked to hemicellulosic chains and form a cellulose—hemicellulose network that is embedded in the pectin matrix. The secondary wall is formed in some plants between the plant cell and primary wall when either a maximum size or a critical point in development has been reached and makes the plant cells rigid. It is made up from cellulose, hemicelluloses (mostly xylan), and lignin. The latter is a complex polymer of aromatic aldehydes that fills the spaces between cellulose, hemicellulose, and pectin components of the cell wall. Because of its hydrophobic nature, it drives out water and so strengthens the wall. In wood, three layers of the secondary cell wall, referred to as the S_1, S_2, and S_3 lamellae, are found that result from different arrangements of the cellulose microfibrils (Mauseth, 1988; Figure 1.1). The first outermost layer—the S_1 lamella—has both left- and right-handed microfibril helices; in contrast, the S_2 (middle) and

Fungi and Lignocellulosic Biomass, First Edition. Christian P. Kubicek.
© 2013 John Wiley & Sons, Inc. Published 2013 by John Wiley & Sons, Inc.

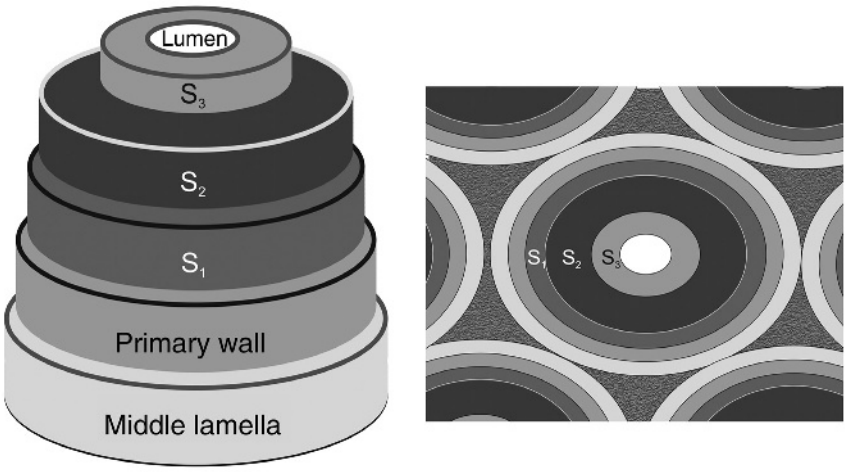


Figure 1.1. Organization of the different layers of the plant cell wall.

S_3 (innermost) lamellae only comprise a single helix of microfibrils, although with opposite handedness to each other. During formation of the secondary cell wall, lignification takes place in the S_1 and S_2 but not S3 lamellae and also in the primary wall and middle lamella (Levy and Staehlin, 1992; Reiter, 2002; Popper, 2008). This arrangement allows the cellulose microfibrils to become embedded and fixed within the lignin, similar to steel rods that become embedded in concrete (Figure 1.2).

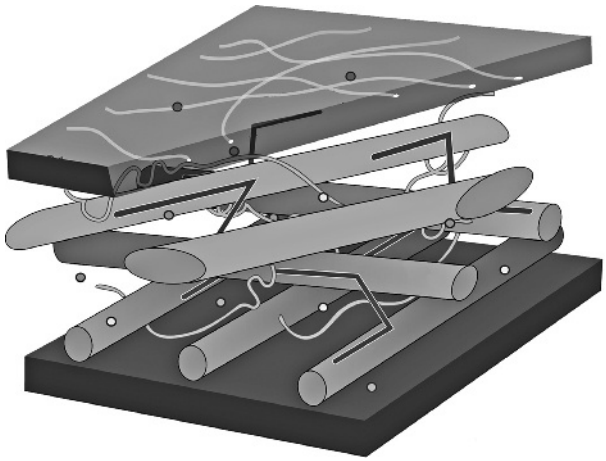


Figure 1.2. Schematic diagram of the three-dimensional arrangement of the main polymers in the primary plant cell wall. The top sheet represents the middle lamella, the bottom sheet represents the plasma membrane; and the area in between represents the primary cell wall. Bright gray threads symbolize pectin, dark gray rectangular lines indicate hemicelluloses, small globes indicate soluble proteins, and the gray tubes indicate the cellulose microfibrils.

In addition, structural proteins (1–5%) are found in most plant cell walls; they are usually classified as hydroxyproline-rich glycoproteins (HRGPs), arabinogalactan proteins (AGPs), glycine-rich proteins (GRPs), and proline-rich proteins (PRPs) (Albenne et al., 2009). The function of these proteins is not well understood. However, it is likely that the glycan moieties in these proteins can form hydrogen bonds and salt bridges to the cell wall polysaccharides, and thereby contribute to the mechanical strength to the wall. The relative composition of carbohydrates, secondary compounds, and proteins varies between plants and between the cell type and age (Levy and Staehelin, 1992; Reiter, 2002; Popper, 2008).

The secondary cell wall may also contain additional layers of lignin in xylem cell walls, and suberin in cork cell walls, that confer rigidity and contribute to the exclusion of water.

1.2 Chemical and Physicochemical Properties of the Major Plant Cell Wall Constituents

1.2.1 Cellulose

As noted earlier, cellulose is one of the principal components of both primary and secondary plant cell walls and reaches its highest abundance (40%) in the secondary cell walls. Cellulose consists of unbranched, unsubstituted 1,4-β-D-glucan chains that can reach degrees of polymerization of 2,000–6,000 and 2,000–10,000 residues in primary and secondary walls, respectively. The CH_2OH side group is arranged in a *trans-gauche* position (a term that described the separation of two vicinal groups by a 60° torsion angle) relative to the O5–C5 and C4–C5 bonds. Because of the absence of coiling or branching, the molecule adopts an extended, rod-like conformation, aided by the equatorial conformation of the glucose residues. The multiple hydroxyl groups on the glucose from one chain can form hydrogen bonds with oxygen molecules on the same or on a neighboring chain (Figure 1.3), and so hold the chains firmly together side by side and form microfibrils with high tensile strength. This strength is one of the major sources of rigidity to the plant cell wall (Klemm et al., 2004; O'Sullivan, 1997).

Carl Naegeli suggested in 1858 that cellulose has a crystalline structure (reviewed by Wilkie, 1961), and this was experimentally verified 80 years later by Meyer and Misch (1937). It consists of two parallel glucan chains that are bound into sheets by hydrogen bonding and

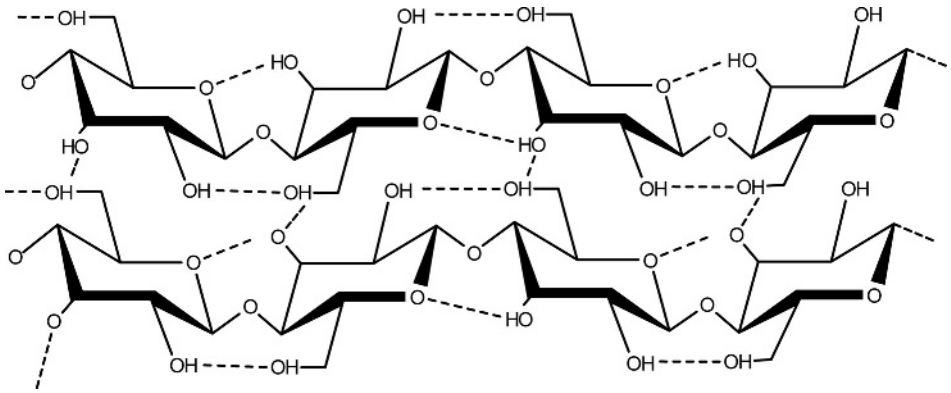


Figure 1.3. The chemical structure of cellulose. Dotted lines represent hydrogen bonds.

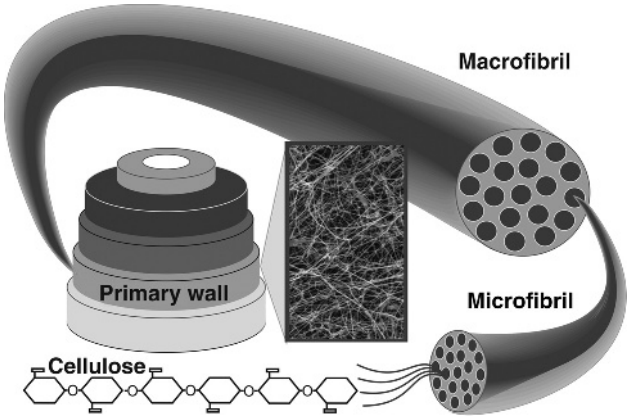


Figure 1.4. Generation of cellulose micro- and macrofibrils.

Van der Waals forces, and which are then stacked to needle-shaped fibers. The morphological hierarchy is defined by elementary fibrils, microfibrils, and microfibrillar bands. The length of these structural units is between 1.5 and 3.5 nm (elementary fibrils), 10 and 30 nm (microfibrils), and on the order of 100 nm (microfibrillar bands) (Figure 1.4). The length of the microfibrils is on the order of several hundred nanometers, which typically contain 30–40 parallel chains that form a diameter of 3.5–4 nm. The microfibrils form characteristic helices that differ as a function of the composition of the cell wall layer (see Section 1.1 above) and according to the plant type as well. As an example, cotton fibers have a lower orientation of the cellulose microfibrils (helix angle ≈18°) as bast fibers (helix angle ≈4°–5°), which results in decreased elasticity, a higher elongation at breakage, and less fiber strength. This adaptation of the mechanical properties of wood to environmental conditions through corresponding helix angles is fascinating and has yet to be rivaled in technical composite materials (O'Sullivan, 1997).

Despite the uniform chemical structure, however, cellulose occurs in the form of seven polymorphs (I_α, I_β, II, III_1, III_{11}, IV_1, and IV_{11}) that can be interconverted, as shown in Figure 1.5 (O'Sullivan, 1997). Among them, cellulose I is the form that is found in nature and therefore called "native cellulose." The polymorphic forms II, III_1, and III_{11} arise from artificial treatments, which are, however, relevant to the process of biomass pretreatment

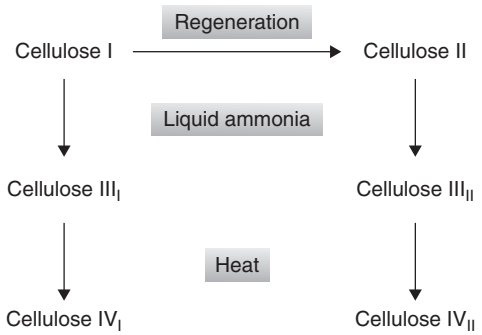


Figure 1.5. Interconversion of polymorphs of cellulose. L and g mean liquid and gaseous aggregate state, respectively.

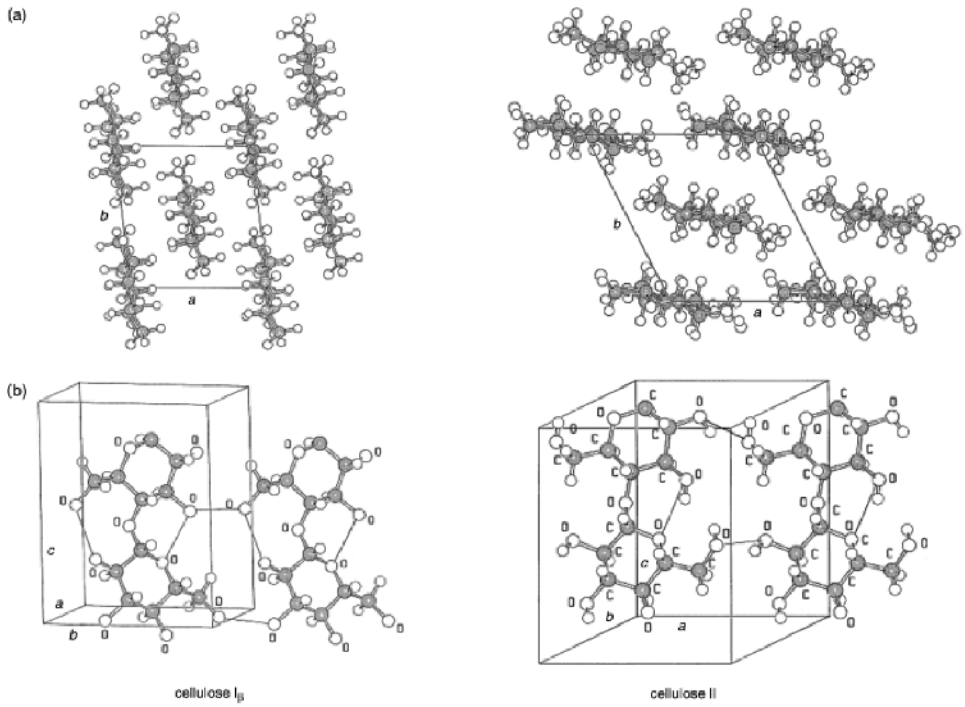


Figure 1.6. Arrangements of cellulose chains in cellulose Iβ and cellulose II. (Reprinted with permission from Klemm D, Heublein B, Fink HP, Bohn A. Cellulose: Fascinating biopolymer and sustainable raw material. Angew Chem Int, 2005; 44(22): 3358–3393. Copyright 2005, with permission from John Wiley & Sons, Inc.)

(see Chapter 10) and shall therefore be explained here as well. Polymorphs IV_1 and IV_{11} arise by heating celluloses III_1 and III_{11}, respectively, to 206°C in glycerol (Hess and Kissig, 1941; Gardiner and Sarko, 1985) and will thus not be treated here.

Cellulose I consists of two distinct crystalline forms (I_α and I_β), which differ from each other in their intramolecular bonding patterns (Wada et al., 1993). The crystal structure of cellulose I can be described by a monoclinic unit cell (space group $P2_1$), which contains two cellulose chains in a parallel orientation with a twofold screw axis. The two forms I_α and I_β can be found alongside each other, their ratio depending on the origin of the cellulose. I_α has triclinic and I_β has monoclinic unit cells, in which two new intramolecular, chain-stiffening hydrogen bonds inside neighboring molecular layers have also been described (O'Sullivan, 1997). The two crystalline forms occur in different proportions depending on the source, I_α being the dominant form in bacteria and lower organisms, whereas the more stable I_β form dominates in higher plants (Yamamoto and Horii, 1994).

Apart from the thermodynamically less stable cellulose I, cellulose II (Figure 1.6) is the most stable structure of technical relevance. It can be formed from cellulose I by treatment with aqueous sodium hydroxide or by dissolution of the cellulose and subsequent precipitation/regeneration. Its monoclinic crystal structure is characterized by the specific unit cell geometry with a modified H-bonding system. It is not yet understood how the parallel chain

arrangement of cellulose I undergoes transition into the antiparallel orientation of cellulose II without an intermediate dispersion of cellulose molecules (O'Sullivan, 1997).

1.2.2 Pectin

Pectin is the most structurally complex family of polysaccharides in nature, making up to 35% of primary walls in dicots and nongraminaceous monocots, 2–10% of grass and other commelinoid primary walls, and up to 5% of walls in woody tissue (Mohnen, 2008; Caffall and Mohnen, 2009; Harholt et al., 2010). It has been suggested to serve a fundamental role in the function of the plant primary and secondary cell walls, because the appearance as plants on land and their adaptation to upright growth correlate with an increase in pectin in their cell walls (Matsunaga et al., 2004).

Pectin generally consists of a backbone of α-1,4-linked D-galacturonic acid residues, of which about 20% can also be replaced by other residues. According to the nature of the monomer composition of this backbone and also of that of the side chains, pectins are distinguished into homogalacturonans (HGs), rhamnogalacturonans I and II (RG-I and RG-II), xylogalacturonan (XGA), and apiogalacturonan (AP). The rhamnose residues can bear long side chains consisting of L-arabinose and D-galactose residues, and small amounts of D-fucose and D-mannose resulting in a "hairy" appearance of the pectin (Mohnen, 2008). Some pectins (e.g., from sugar beet and apple) can also bear terminal ferulic acid residues that are linked to either *O*-5 of the arabinose or *O*-2 of the galactose residues (Oosterveld et al., 2000). Some of the galacturonic acids in pectin are methyl esterified or acetylated. The nonmethylated D-galacturonic acid sequences interact with Ca^{2+} and thereby cross-link different pectic acid chains, which contributes to the firmness of the plant tissue (Cafall and Mohnen, 2009).

HG is the most abundant pectic polysaccharide. It is a linear homopolymer of about 100 α-1,4-linked galacturonic acid residues that makes up for about 65% of all pectin in plant cell walls (Figure 1.7a). It can bear methyl ester groups at the C-6 carboxyl group, and—albeit less—acetyl groups at *O*-2 or *O*-3.

RG-I represents 20–35% of pectin. It contains a backbone of alternating α-D-galacturonic and α-L-rhamnose residues (Figure 1.7b), and a varying number of different sugars and oligosaccharides as side chains. Twenty to eighty percent of the rhamnosyl residues in the RG-I backbone can contain side chains of α-1,5-linked L-arabinan with C2- and C3-linked arabinan side chains, β-1,4-linked D-galactans with a degree of polymerization of up to 47, β-1,4-linked D-galactans with C3-linked L-arabinose or arabinan side chains, and β-1,3-linked D-galactan with β-6-linked galactan or arabinogalactan side chains (Schols et al., 1990). The side chains may further contain α-L-fucose, β-D-glucuronic acid, and 4-*O*-methyl-β-D-glururonic acid residues.

RG-II is the structurally most complex pectin and makes up for up to 10% of it (Figure 1.7c). Its structure consists of an HG backbone of at least eight α-1,4-linked α-D-galacturonic acid residues, which further contain side branches that can involve up to 12 different types of sugars in more than 20 different linkages. RG-II usually exists in plant walls as RG-II dimers cross-linked by a 1:2 borate diol ester between the apiosyl residues in side chain A of two RG-II monomers. RG-II dimerization cross-links HG domains resulting in a macromolecular pectin network. Mutations that result in even small modifications of the structure of RG-II lead to severe growth defects, suggesting that RG-II in the wall is crucial for normal plant growth and development (Mohnen, 2008).

Two other substituted galacturonans, XGA and AP, are only minor pectin components. XGA, which has been found mainly in reproductive plant tissues, is an HG substituted at *O*-3 with a β-linked xylose, which can sometimes bear additional β-linked xylose residues at *O*-4. AP is an HG substituted at *O*-2 or *O*-3 with D-apiofuraose and has been found in aquatic monocots (Cafall and Mohnen, 2009).

The way by which the pectic polysaccharides are linked to each other is still in debate. Most of the available data support the assumption that HG, RG-I, and RG-II are linked via their backbones. However, it has also been suggested that pectins may be covalently linked to, or tightly associated with, other types of wall polysaccharides such as xyloglucans and xylans. This hypothesis is supported by the finding of xylose residues in some pectins and that a mutation in genes of pectic biosynthesis influences also the xylan content of the cell walls (Orfila et al., 2005). This suggests that pectins may serve to hold at least some hemicelluloses in the wall.

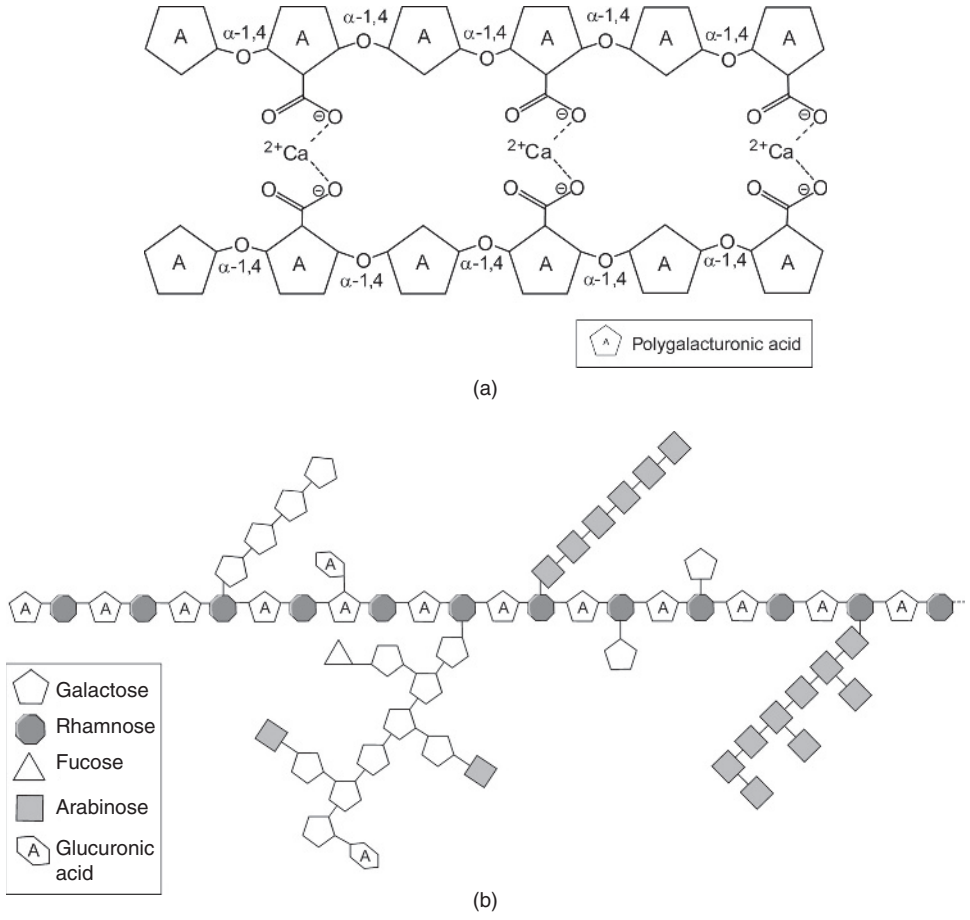
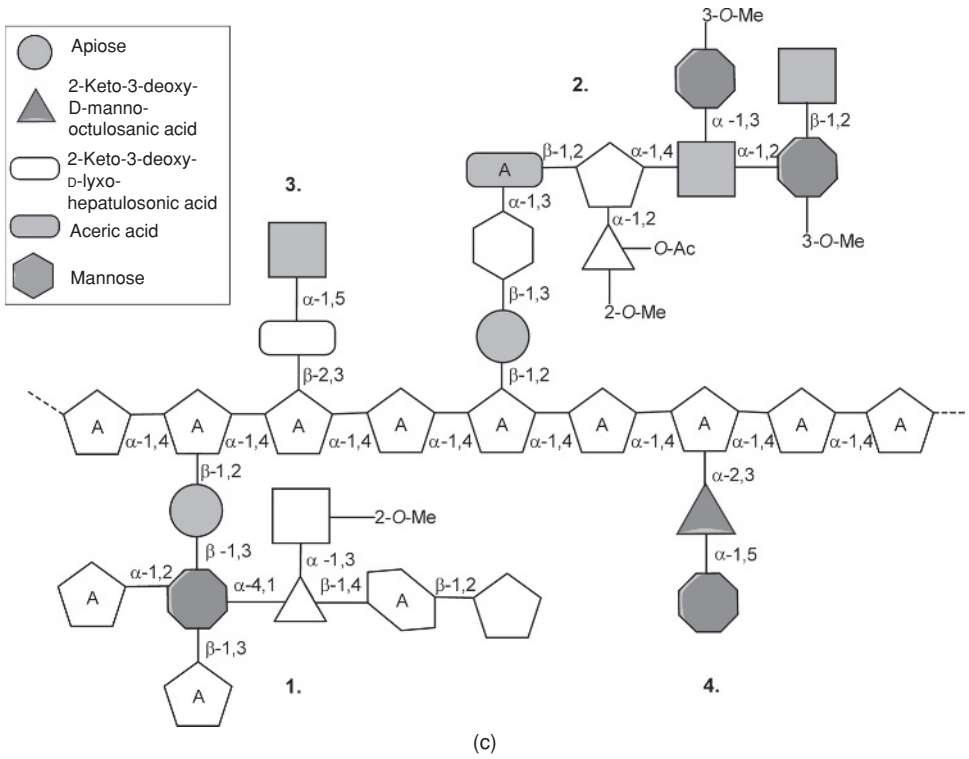


Figure 1.7. (a) The chemical structure of pectin and its complexation with Ca^{2+} ions. (b) The structure of rhamnogalacturonan I. (c) The structure of rhamnogalacturonan II. *O*-Ac and *O*-Met specify acetylation and methylation, respectively.



(c)

Figure 1.7. (*Continued*)

1.2.3 Hemicelluloses

Xyloglucan

The predominant hemicellulose in the primary cell wall of dicots and nongraminaceous monocots is xyloglucan, which may account for up to 20% of the dry weight of the primary wall. Xyloglucan has a backbone composed of 1,4-linked β-D-glucose residues, of which up to 75% are substituted at O6 with mono-, di-, or triglycosyl side chains. Xyloglucans are strongly associated with cellulose and thus add to the structural integrity of the cell wall. They are also believed to play an important role in regulating cell wall extension. The length of the xyloglucan polymers enables them to cross-link several cellulose microfibrils, thus creating a rigid network structure (Hayashi and Kaida, 2011).

Xyloglucans have been identified to occur in two types, that is, type XXXG and type XXGG (Figure 1.8): XXXG consists of repeating units of three β-1,4-linked D-glucopyranose residues, substituted with D-xylopyranose via an α-1,6-linkage, which are separated by an unsubstituted glucose residue. In xyloglucan type XXGG, two xylose-substituted glucose residues are separated by two unsubstituted glucose residues. The structural features of these, as well as some other types of xyloglucans, have been discussed in detail by Vincken et al. (1997). The xylose residues in xyloglucan can further be substituted with α-1,2-L-fucopyranose-β-1,2-D-galactopyranose and α-1,2-L-galactopyranose-β-1,2-D-galactopyranose disaccharides and *O*-linked acetyl groups (Maruyama et al., 1996; Hantus et al., 1997; Vincken et al., 1997).

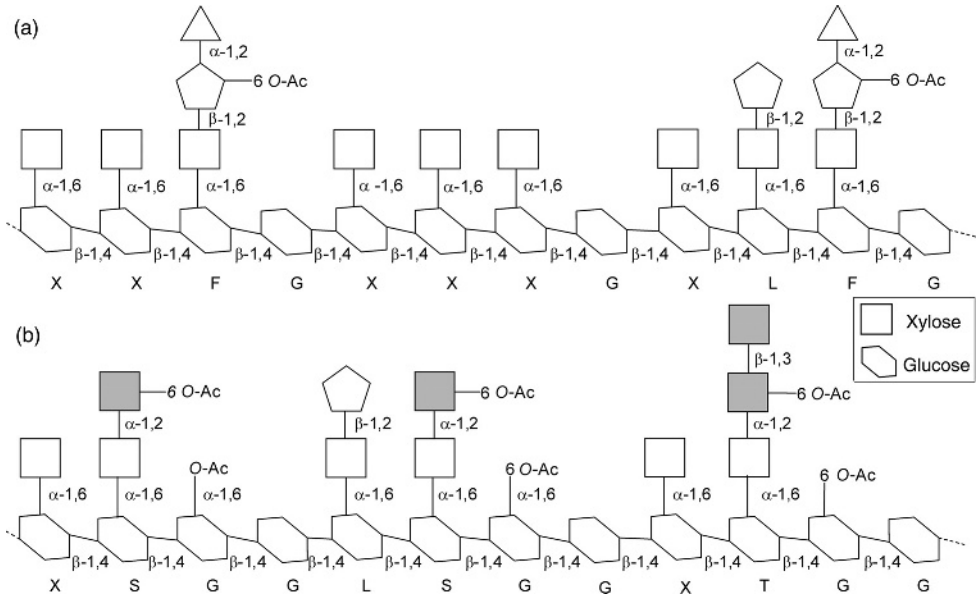


Figure 1.8. The chemical structure of xyloglucan. (a) XXXG-type; (b) XXGG-type. Symbols not shown are used as in Figure 1.7.

In cell walls of dicotyledons, the xyloglucans are partially replaced by glucuronoarabinoxylan, which has a linear β-1,4-linked D-xylopyranosyl backbone with both neutral and acidic side chains attached at intervals along its length. The acidic side chains are terminated with glucuronosyl or 4-*O*-methyl glucuronosyl residues, whereas the neutral side chains are composed of arabinosyl and/or xylosyl residues (Darvill et al., 1980). The structures of xyloglucans from several plants in the subclass Asteridae were characterized by nuclear magnetic resonance (NMR) spectroscopy and mass spectrometry to determine how their structures vary in different taxonomic orders (Hoffman et al., 2005). The structure of xyloglucan has been shown to differ also in a tissue-specific manner, for example, fucosyl residues are typically absent from seed xyloglucans but present on the xyloglucans in the vegetative portions of the same plant.

A single letter nomenclature is used to simplify the naming of xyloglucan side chain structures (Table 1.1). For example, a capital **G** represents an unbranched glucopyranose residue. A capital **F** represents a glucopyranose residue that is substituted with a fucose-containing

Table 1.1. Xyloglucan one letter code.

G	-4)-β-D-Glc_p-(1-
X	α-D-Xyl_p-(1–6)-β-D-Glc_p-(1-
L	β-D-Gal_p-(1–2)-α-D-Xyl_p-(1–6)-β-D-Glc_p-(1-
F	α-L-Fuc_p-(1–2)-β-D-Gal_p-(1–2)-α-D-Xyl_p-(1–6)-β-D-Glc_p-(1-
S	α-L-Ara_f-(1–2)-α-D-Xyl_p-(1–6)-β-D-Glc_p-(1-
T	α-L-Ara_f-(1–3)-α-L-Ara_f-(1–2)-α-D-Xyl_p-(1–6)-β-D-Glc_p-(1-
J	α-L-Gal_p-(1–2)-β-D-Gal_p-(1–2)-α-D-Xyl_p-(1–6)-β-D-Glc_p-(1-

From http://www.ccrc.uga.edu.

trisaccharide. The Complex Carbohydrate Research Center at the university of Georgia has developed a searchable ^{1}H NMR database (http://cell.ccrc.uga.edu/world/xgnmr/index.html) to facilitate the rapid identification of enzymatically generated xyloglucan subunit structures.

In the plant cell wall, the xyloglucans are arranged in such a way that they coat the surface of the cellulose microfibrils (some regions binding directly to the cellulose surface, other regions are not in direct contact with the cellulose but form cross-linking tethers, and still some regions of xyloglucan are entrapped within the cellulose microfibrils; Mellerowicz et al., 2008). This results in a limited aggregation of the cellulose chains and especially the tethers impact the mechanical properties of the wall. The binding of cellulose is likely a complex topological process, because the xyloglucan backbone must, to this end, adopt a "flat ribbon" conformation whose surface is complimentary to that of cellulose (Umemura and Yuguchi, 2005). Xyloglucans, however, normally tend to adopt a "twisted" conformation in solution. Bootten et al. (2009) demonstrated that binding of xyloglucan to cellulose may untwists the xyloglucan backbone, which—if both "ends" of the xyloglucan are attached in this way—may lead to the formation of coiled structures, that would form duplex antiparallel coils that are energetically stable.

Xylan

Xylan is the major hemicellulose polymer in cereals and hardwood. It always contains a β-1,4-linked D-xylose backbone, to which differently structured side chains can be attached, thus resulting in a high variety of xylan structures. Although most xylans are branched structures, linear polysaccharides have also been isolated. The xylans of cereals often contain large quantities of L-arabinose and are consequently termed arabinoxylans. In contrast, hardwood xylans contain large amount of D-glucuronic acid linked to the backbone and are named glucuronoxylans (Scheller and Ulvskov, 2010; Figure 1.9).

L-Arabinose (either single residues or short chains) is connected to the xylan backbone via either α-1,2- or α-1,3-linkage. These side chains can also contain D-xylose in a β-1,2-linkage to L-arabinose and D-galactose, which can be either β-1,5-linked to L-arabinose or β-1,4-linked to D-xylose. Acetyl residues can be attached to *O*-2 or *O*-3 of the D-xylan backbone, but the

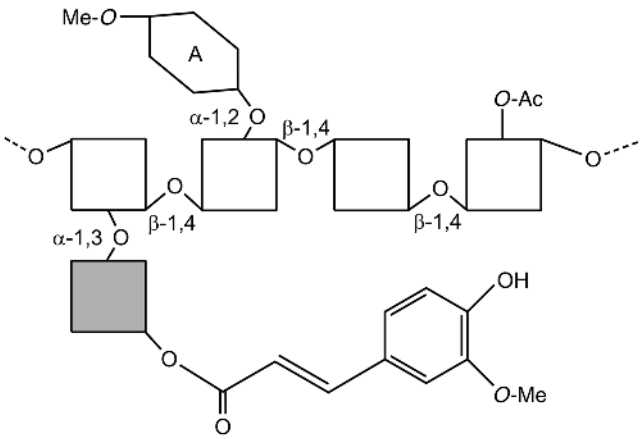


Figure 1.9. The chemical structure of xylan. Symbols not shown are used as in Figure 1.7.

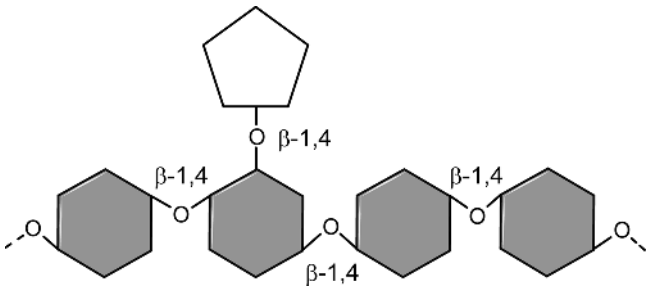


Figure 1.10. The chemical structure of galactomannan. Symbols not shown are used as in Figure 1.7.

degree of acetylation differs strongly depending on the origin. Glucuronic acid and its 4-*O*-methyl ether are attached to the xylan backbone via an α-1,2-linkage, whereas feruloyl and *p*-coumaroyl residues can be attached at the *O*-5 of terminal L-arabinose residues (Ebringerova and Heinze, 2000).

Galactomannans

Galactomannans and galactoglucomannans comprise a second group of hemicellulolytic structures, which form the major hemicellulose fraction of the gymnosperms cell walls (12–15%). They consist of a backbone of β-1,4-linked D-mannose residues, which can bear α-1,6-linked D-galactose residues (Figure 1.10) in ratios between 1:1 and 5:1 depending on the source (Dey, 1978). Therein the galactosyl side chain hydrogen interacts to the mannan backbone intramolecularly and so stabilizes the structure. Acetyl groups can be present but are irregularly distributed in glucomannan. Also some of the mannosyl units of galactoglucomannan are partially substituted by *O*-acetyl groups (Moreira and Filho, 2008).

1.2.4 Lignin

Lignin is found in all vascular plants, and is—after cellulose—the most abundant carbon source on earth. Lignin is characterized by a complex structure derived from oxidative coupling of three primary hydroxycinnamyl alcohols, that is, *p*-coumaryl, coniferyl, and sinapyl alcohol, which render it extremely recalcitrant to degradation. The corresponding phenylpropanoid units in the lignin polymer are usually denoted as *p*-hydrophenyl (H), guaiacyl (G), and syringyl (S) units, respectively, based on the methoxy substitution on the aromatic rings (Figure 1.11). Although gymnosperm lignin (=softwood) contains mostly G units and very low levels of H units (G/S/H = 96:*t*:4), angiosperm lignin (=hardwood) is composed of similar levels of G and S units with traces of H units (G/S/H = 50:50:*t*). Monocotyledons (e.g., grasses) contain all the three units in a ratio of G/S/H = 70:25:5 (Davin and Lewis, 2005; Calvo-Flores and Dobado, 2010).

Lignification is achieved by cross-linking reactions of the monomer with the growing polymer or by polymer–polymer coupling via radicals generated by oxidase enzymes (see Chapter 5). Endwise reactions coupling a lignin monomer to the growing polymer result in the formation of β-linked structures. Further, coupling between two preformed lignin oligomers or polymers results in 5–5 and 5-*O*-4 linked structures. Finally, end groups arise from coupling

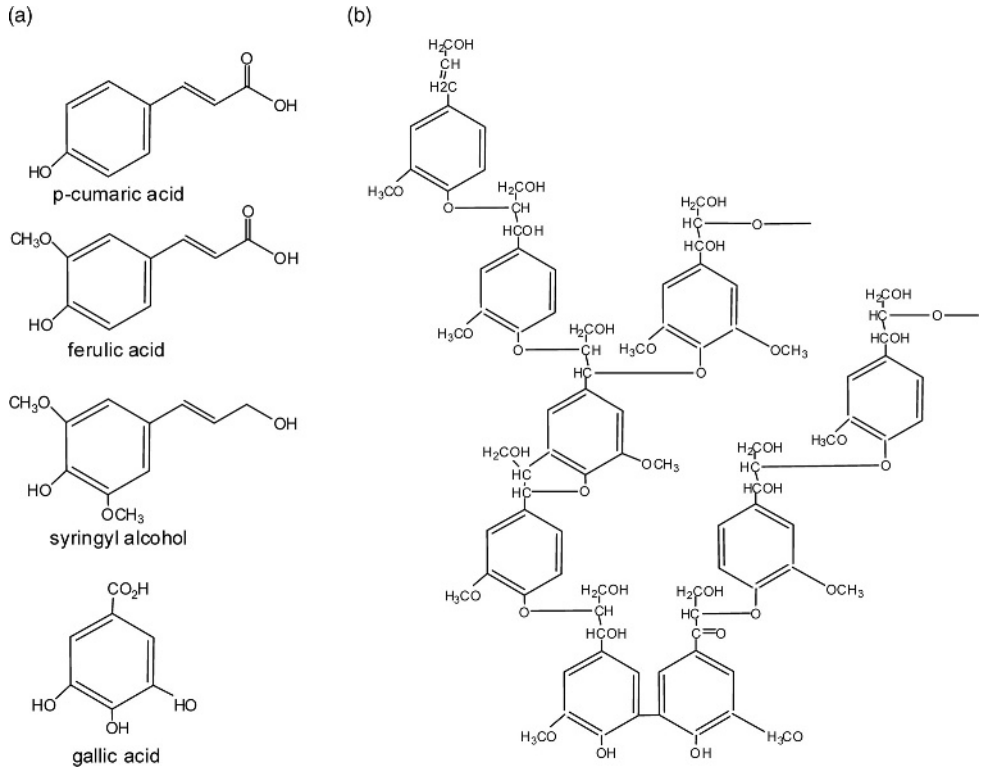


Figure 1.11. The chemical structure of lignin. (a) The monomers, from which lignin has been formed, are given on the left side of the figure. (b) Only a small part, yet indicating all occurring bonds between different aromatic rings, is shown.

reactions that are not at the side chain β-position of the monomer. The relative abundance of the different linkages largely depends on the relative contribution of the monomers to the polymerization process during lignin biosynthesis; for example, the β-*O*-4 (arylglycerol-β-aryl ether) coupling of a monolignol with the growing lignin oligomer/polymer creates the most abundant structural unit, involving generally about 50% and 80% of the phenylpropanoid units in softwood and in hardwood lignin, respectively (Alder, 1977). The β-5, 5-5, and 4-*O*-5 structures account for roughly 10%, 25%, and 4%, respectively, in softwood lignin (Brunow, 2001). Acylated structural units, such as 4-propoxy-sinapyl-γ-acetate, are found at high levels in some lignins, and in grass lignins, the hydroxycinnamic acids can be esterified at the γ-position of the propyl side chains. The amount of β-1 structures in softwood lignin is about 2%. The 5-5 unit is frequently etherified with additional monolignol via intramolecular reaction of quinone methide intermediates. The resulting dibenzodioxocin, an eight-member cyclic ether unit, as well as the 5-5 and 4-*O*-5 units may serve as branching points in softwood lignin. Similarly, some β-1-linkages seem to be part of spirodienone substructures (Boerjan et al., 2003; Wong, 2009). The lignins from the herbaceous plants sisal (*Agave sisalana*), kenaf (*Hibiscus cannabinus*), abaca (*Musa textilis*), and curaua (*Ananas erectifolius*) are extensively acylated at the gamma-carbon of the lignin side chain (up to 80% acylation) with acetate and/or *p*-coumarate groups and preferentially over S units. The structures of all these highly

acylated lignins are characterized by a very high S:G ratio, a large predominance of β-*O*-4′ linkages (up to 94% of all linkages), and a strikingly low proportion of traditional beta-beta′ linkages, which indeed are completely absent in the lignins from abaca and curaua (del Río et al., 2008).

Lignin interacts with the cellulose fibrils, creating a rigid structure strengthening the plant cell wall. They also form several types of covalent cross-links to hemicelluloses (for review, see Fry, 1986): one is formed by diferulic acid bridges between lignin and arabinoxylans, pectin polymers, or xylan and lignin. Another type is an ester linkage between lignin and the glucuronic acid residues in xylan, which has been observed, for example, in beech wood. A third type (Rizk et al., 2000) involves a protein- and pH-dependent binding of pectin and glucuronoarabinoxylan to xyloglucan, and it is dependent on the presence of fucose on the xyloglucan (for details see de Vries and Visser, 2001).

1.3 Abundant Sources of Carbohydrate Polymers and Their Monomer Composition

With the aim of producing ethanol and other biorefinery products from biomass without interfering with the food and feed chains, plant biomass that cannot be used for this purpose must be used ("second generation biofuels"). Materials that are currently considered as potentially useful include agricultural residues (currently considered the most likely feedstock to be adopted), forest harvest residues, and finally the dedicated breeding of energy crops. Their content in cellulose, hemicelluloses, and lignin is given in Table 1.2.

1.3.1 Agricultural Wastes

Cellulosic wastes, including waste products from agriculture (straw, stalks, leaves) and forestry, wastes generated from processing (nut shells, sugarcane bagasse, sawdust) and organic parts of municipal waste, could all be potential sources. However, it is also important to consider the crucial role that decomposing biomass plays in maintaining soil fertility and texture; excessive withdrawals for bioenergy use could have negative effects.

Kim and Dale (2004) presented a calculation about how much bioethanol could be produced from agricultural residues worldwide: to avoid conflicts between human food use and industrial use of crops, they considered only the wasted crop, which is defined as crop lost in distribution,

Table 1.2. Cell wall compositions of different plant lignocellulose sources.

		Cellulose (%)	Hemicellulose (%)	Lignin (%)
Hard wood	Birch	40	23	21
	Willow	37	23	21
	Aspen	51	29	16
Soft wood	Spruce	43	26	29
	Pine	46.4	8.8	29.4
Agro residues	Wheat straw	38.2	21.2	23.4
	Corn stover	37.5	22.4	17.6
Crop	Switch grass	31	20.4	17.6
	Miscanthus	40	18	25

Data taken from Chandel and Singh (2011).

as feedstock. Overall, they arrived at about 73.9 Tg of dry wasted crops in the world that could potentially produce 49.1 GL of bioethanol. Asia is the largest potential producer of bioethanol from crop residues and wasted crops, mainly rice straw, wheat straw, and corn stover, and could produce up to 291 GL/year of bioethanol. The next highest potential region is Europe (69.2 GL of bioethanol), in which most bioethanol comes from wheat straw. Corn stover is the main feedstock in North America, from which about 38.4 GL/year of bioethanol can potentially be produced. Options for dealing with the straw debris so far are (a) leaving them on the field, (b) plowing it back into the soil, (c) burning, or (d) removing from the land depending on the decision made by landowner. In recent years, however, burning has become discredited due to increased concern over the health effects of smoke from burning fields (Kerstetter and Lyons, 2001). In the case of wheat straw, full removal may lead to soil erosion, and the fraction of wheat straw that must be left on the field depends on the weather, crop rotation, existing soil fertility, slop of the land, and tillage practices. According to Kim and Dale (2004), 60% ground cover could be maintained in order to ensure prevention of soil erosion that roughly requires 1.7 Mg wheat residue per hectare. Globally considered rice straw would be the leader by producing 205 GL/year of bioethanol, and the next highest potential feedstock is wheat straw, which can produce 104 GL/year of bioethanol (Kim and Dale, 2004).

Another, albeit, less abundant biomass source could be leaf litter: Simmons et al. (2008) emphasized that leaf biomass can represent 5–15% of the total aboveground biomass in a year. While this figure is small, however, this process occurs annually and represents 25–60% equivalent of total biomass at harvest. Leaves also present the additional advantage compared with stemwood that they are easier to process.

1.3.2 Forest Product Residues

Current forest product manufacturing techniques produce large amounts of residues–chips and particles—from milling processes. They are currently used for the manufacture of particle board and paper products. In addition, wood residues accumulate that arise during forest products processing. Mabee and Saddler (2010) estimated that they represent approximately 50% of the biomass energy consumption in the United States. According to their calculations, 331 million dry tons could be dedicated to biofuel production on a sustainable basis.

The average forest residues generation rate for temperate northern countries is estimated to be about 12%. In order to be truly renewable, the removal of forest biomass must be carried out in a fashion that limits impacts on local ecosystems, in accordance with the principles of sustainable forest management. Estimates for Canada have calculated that the dry residues that may be sustainably removed from forest harvest operations range between 9.8 (Mabee et al., 2006) and 46 (Wood and Layzell, 2003) million dry tons per annum. In addition, residues from forest products manufacture have been conservatively estimated to range between 2 and 5.4 million dry tons per annum (Mabee et al., 2006; Wood and Layzell, 2003). However, use of this material is in strong competition with new ventures such as pelletization, and the perspectives for using a large proportion of this for bioethanol production are small.

An additional, nonrenewable source of forest biomass is the wood killed by plant pathogenic insects and microbes (“disturbance wood,” Schmitz and Gibson, 2003). Because these sources of forest biomass are nonrenewable, they are not included in the estimates of renewable biomass supply, but could constitute an additional significant amount of biomass for biofuel production (Mabee and Saddler, 2010).

1.3.3 Energy Crops

As lignocellulosic biomass becomes more widely established, forest and agricultural residues will no more be available in the desired quantities, and other feedstocks be required. Three energy crops, widely considered as feasible options in this direction, are *Miscanthus giganteus*, *Panicum virgatum* elephant grass, and hybrid poplar (Chandel and Singh, 2011).

M. giganteus is a grass native to Asia and can grow up to 12 ft (3.7 m) tall without significant addition of water or fertilizers. It is a sterile hybrid between *M. sinensis* and *M. sacchariflorus*, and because of its rapid growth (more than 3.5 m in one growth season), low mineral content, and high biomass yield (up to 25 tons per hectare), it has been proposed as a renewable biomass in Europe since the early 1980s.

P. virgatum, commonly known as switchgrass, is a perennial warm season bunchgrass native to North America, where it occurs naturally from 55°N latitude southward into the United States and Mexico. It is one of the dominant species of the central North American tallgrass prairie, and it is also found in remnant prairies, in native grass pastures, and naturalized along roadsides. It is similar to *M. giganteus* with respect to cold and drought tolerance and water use efficiency (Chandel and Singh, 2011). Switchgrass is a diverse species, with striking differences between different plants, thus providing a range of valuable traits for breeding programs. It has two distinct forms: the lowland cultivars, which tend to produce more biomass (they may grow up to $\geq$2.7 m), and the upland cultivars that are generally shorter ($\leq$2.4 m tall), more cold tolerant and of northern origin. In 1992, the US Department of Energy started a research program focused on developing switchgrass as a sustainable bioenergy feedstock (Sanderson et al., 1996). Switchgrass was chosen as a biomass-based renewable energy source crop because it had high forage yield and seed production at different regional cultivar testing fields in several states in the United States. Switchgrass cv. Alamo has been ranked as the highest yielding cultivar in most yield trials conducted and is relatively amenable to genetic transformation and subsequent regeneration into mature plants (McLaughlin and Kszos, 2005; Burris et al., 2009). Additionally, it can be grown and cultivated on marginal lands, thereby alleviating concerns about competing with land for cultivation of food crops. The best-case land-use scenario for switchgrass is its cultivation on agriculturally depleted soils that no longer support agriculturally important row crops (for review see Keshwani and Cheng, 2009; and Moon et al., 2010).

Hybrid poplar, a member of the willow family, is a cross between North America's cottonwoods, aspens, and Europe's poplars. Poplar is considered as a model example of a short-rotation woody crop and can produce 9–15.7 dry tons per hectare (4–7 dry tons per acre) annually over a 6–10-year rotation (Alig et al., 2000). Other trees that could be grown specifically for biofuel production are eucalyptus, loblolly pine, willow, and silver maple. Eucalyptus, native to Australia but grown throughout the world, has been studied extensively in California and Florida, and appears to be amenable to high-density cultivation in plantation farms (Rockwood et al., 2008).

1.3.4 Weedy Lignocellulosic Substrates

Weedy lignocelluloses have also been discussed as promising future biomass feedstocks (Huber and Dale, 2009). These include wild sugarcane (*Saccharum spontaneum*), a perennial weedy grass with worldwide distribution but which is especially abundant in Asian countries like India, or water hyacinth (*Eichhornia crassipes*), which is a free-floating perennial aquatic plant native to tropical South America. For a full account on these and many more plants, see

Huber and Dale (2009). A drawback of the targeted cultivation of these plants, however, is that they all are weeds that may be different to control and thus threaten other crop production.

1.4 Biosynthesis of Plant Cell Wall Polymers

1.4.1 Cellulose

Cellulose biosynthesis has been subject of many excellent and recent reviews (Somerville, 2006; Crowell et al., 2010; Guerriero et al., 2010; Harris and de Bolt, 2010; Wightman and Turner, 2010, Endler and Persson, 2010). It occurs by a membrane-bound cellulose synthase (CESA) protein complex consisting of a CESA hexamer called "rosette" that can be visualized in the inner leaflet of freeze-fractured plasma membranes (Mueller and Brown, 1980). A CESA gene has been first cloned from cotton, and subsequent analysis of the *Arabidopsis thaliana* genome sequence revealed the presence of 29 genes (Richmond and Somerville, 2000). The CESAs are in fact an ancient gene family and conserved from algae to higher plants (Crowell et al., 2010).

The rosette hexamer contains multiple CESA catalytic subunits, of which three (CESA1, CASE3, and CESA6) are involved in the formation of the primary cell wall, whereas three other isoforms (CESA4, CESA7, and CESA8) synthesize cellulose for the secondary walls. The occurrence of putative zinc-finger domains at the N-terminus of the CESA proteins and the analysis of the redox states of the cysteine residues present in them have suggests that they are responsible for the dimerization of the CESA subunits (Doblin et al., 2002). The requirement for three isoenzymes to assemble a functional CESA complex has been a matter of discussion and several models have been proposed: one of them suggests that they might have distinct binding sites for specific intra- and interparticle interaction for assembling the hexameric rosette (Doblin et al., 2002; Timmers et al., 2009). Alternatively, it has been discussed that the three subunits may catalyze distinct chain initiation and elongation reactions or add adjacent glucose residues in opposite orientations (Simmons et al., 2006; Guerriero et al., 2010).

The rosette complex synthesizes cellulose at the plasma membrane, to which it is delivered via actin filaments and cortical microtubules (reviewed in detail by Crowell et al., 2010). However, as explained by Guerriero et al. (2010), the possibility that cellulose biosynthesis already begins in the Golgi apparatus in higher plants has not been strictly ruled out as yet. Indeed, noncrystalline chains of β-(1→4)-glucan may be synthesized in the Golgi by catalytic subunits of the CESA that have not yet been assembled in the form of rosettes. Alternatively, the oligosaccharides that are formed in the Golgi could be too short to form microfibrillar strings, or interact in the Golgi with carbohydrates or an aglycone that would prevent microfibril formation (Guerriero et al., 2010). Following delivery, the CESA catalytic subunits appear to be inserted into the plasma membrane by means of eight transmembrane helices. As a consequence of this even number of helices, both the N-terminus and the C-terminus of the catalytic subunits face the same side of the plasma membrane. It is generally (although not univocally) assumed that cellulose synthesis occurs on the cytoplasmic side of the plasma membrane (Guerriero et al., 2010).

Cellulose synthesis proceeds by the attachment of glucosyl residues to the nonreducing terminus of the acceptor glucan chain. The substrate for this attachment is uridine-diphosphate (UDP)-glucose, which can be provided by the cytosolic enzyme UDP-glucose pyrophosphorylase (Figure 1.12). However, it is also possible that the provision of UDP-glucose comes from sucrose synthase, which converts sucrose and UDP into fructose and UDP-glucose

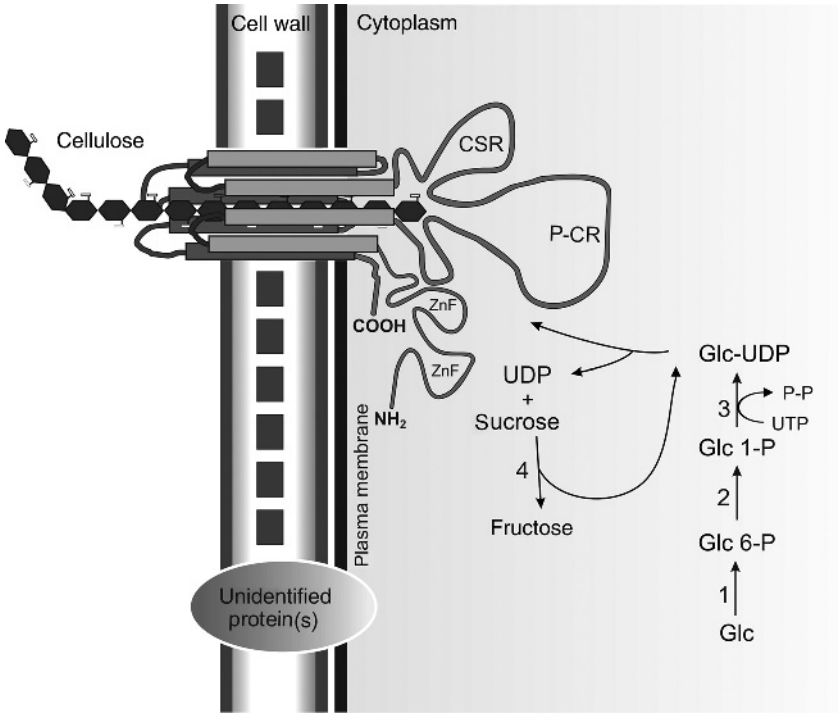


Figure 1.12. Biosynthesis of cellulose by the catalytic dimer hypothesis for cellulose synthase. Precursor biosynthesis in the cytosol from glucose and fructose involves five enzymes: 1, hexokinase/glucokinase; 2, phosphoglucomutase; 3, UDP-glucose pyrophosphorylase; 4, sucrose synthase. Two CESA proteins form a complex that synthesizes a single β-1,4-D-glucan chain. This dimerization also results in a channel composed of 16 membrane-spanning domains. Two zinc fingers (plant-specific conserved region, P-CR; class-specific region, CSR) are potential interaction sites. The catalytic site resides within the loop structures on the cytoplasmic side close to the membrane. (Modified and complemented from Carpita, 2011.)

(Amor et al., 2005). This theory is supported by the fact that a membrane-bound sucrose synthase can be found at sites of very active cellulose synthesis, but must still be viewed with caution because UDP-glucose also serves as a substrate for the various glycosyltransferases involved in the biosynthesis of noncellulosic cell wall carbohydrates (Simmons, 2006; Guerriero et al., 2010).

Initiation of cellulose biosynthesis is assumed to require a primer (although evidence for this *in vivo* is still lacking), which most likely is sitosterol-β-glucoside (Peng et al., 2002). The glycosyltransferase that initiates cellulose chain synthesis then uses UDP-glucose as a donor and sitosterol-β-glucoside as an acceptor. Further polymerization is believed to proceed by the direct addition of the monomers to the carbohydrate chains. However, the theory of Albersheim et al. (1997) that the sugars are first transferred to an S or T residue of the enzyme cannot be excluded at the moment. Guerriero et al. (2010) also presented a third alternative, that is, the polymerization of cellulose chains in higher plants could be initiated on lipid acceptors. Cello-oligosaccharides would then be transferred to the nonreducing end of the elongating glucan chains by CESA or a transglycosylase rather than a glycosyltransferase. This hypothesis

bears the attractivity that it is compatible with the proposed role of sitosterol-β-glucoside in cellulose biosynthesis (Peng et al., 2002).

Whatever the exact mechanism is, the elongating cellulose chains are then translocated across the membrane into the cell wall, either via a pore that is formed by the eight transmembrane helices of the CESAs or through adjacently located porins.

Assembly of the microfibrils from the arosen cellulose chains then occurs at some distance from the enzyme complex itself. This process is in fact still poorly understood: it may be occurring spontaneously through the formation of interchain hydrogen bonds when a number of chains sufficient to form an elementary microfibril such as those observed in primary walls (ca 3 nm diameter) are in close proximity. Alternatively, the formation of microfibrils may require the assembly of complete rosettes with their sixfold symmetry, or the involvement of accessory proteins that are not functional or present in the Golgi apparatus. The use of *Arabidopsis* CESA mutants strongly supports the hypothesis that the rosette structures are required for the formation of microfibrils (Guerriero et al., 2010). The CESA complexes have been observed to move in the plasma membrane, and a biophysical model suggests that this movement of the enzyme complex is driven by the polymerization and crystallization events (Diotallevi and Mulder, 2007). In this model, the aggregates of microfibrils form simultaneously as they are spun from groups of rosettes that are spatially close in a confined environment. Bessueille et al. (2009) showed that CESA is located in lateral patches of the plasma membrane that bear properties resembling lipid rafts, cholesterol, and sphingolipid-rich parts of the membrane that lower the free energy between the membrane and the adjacent aqueous phase (Rietveld and Simons, 1998). Guerriero et al. (2010) have thus speculated "that groups of rosettes co-localized in raft-like structures are responsible for the coordinated synthesis of multiple microfibrils that coalesce shortly after having been extruded from individual enzyme complexes to form fibrils or ribbons of cellulose."

The actual stoichiometry of the CESA catalytic subunits in the rosettes has not been experimentally demonstrated as yet, but it is assumed that the rosette would be able to synthesize 36 β-(1→4) glucan chains that would co-crystallize to form a microfibril (Delmer, 1999), which corresponds with the size of microfibrils isolated from most primary walls (∼3.5–4 nm) (Delmer, 1999). This model is based on the rationale that six globules of the rosette each have six CESA catalytic subunits. However, the number of active catalytic subunits per rosette has never been experimentally demonstrated. Thus, the above rationale that individual cellulose microfibrils consist of 36 chains is exclusively based on the observed sixfold symmetry of the rosette structures. It is as well possible that the different rosettes contain a lower number of catalytically active subunits and consequently form thinner microfibrils. In fact, the minimum number of chains to form a crystalline structure is a total of 16 chains (i.e., 4 × 4 package). This number of chains is compatible with a lateral size of 2–2.5 nm for an individual microfibril and consistent with the width observed for microfibrils from primary walls (Guerriero et al., 2010).

In this regard, one must also consider the fact that the synthesis of a β-1-4-D-glucan chain implies a big steric problem: to make a (1→4)-β-D linkage means that each glucosyl residue is turned 180° with respect to each neighbor. Thus, the *O*-4 position of nonreducing terminal sugar of the acceptor chain is displaced several angstroms upon addition of each successive unit, and the site of catalysis must thus also move several angstroms within the protein. In addition, the acceptor chain must swivel 180°, or the catalytic or acid–base amino acids must toggle between two forms to account for the displacement (Carpita, 2011). To overcome this problem, models have proposed that two CESAs associate to form two opposing catalytic sites (Capita, 2011). The requirement of two CESAs to synthesize one chain would in fact reduce

the stoichiometry to 18 chains, which is close to what is observed experimentally (*vide supra*). Regardless of the precise number of chains that form individual microfibrils, their packing into a crystalline structure is believed to occur spontaneously, driven by hydrogen and van der Waals bonding (Crowell et al., 2010).

There is strong evidence from mutants affected in cellulose biosynthesis that the rosette complex also contains other proteins than the CESAs. Examples are a membrane-bound sucrose synthase, a membrane-bound endo-β-(1→4)-glucanase (Korrigan in *Arabidopsis*), annexins, actin, tubulin, a putative lipid transfer protein, and others (for a detailed review, see Mølhøj et al., 2002). However, the precise roles of these proteins have not yet been determined and we cannot exclude that the observed effects of the respective mutations were indirect. In fact, purification of the rosette complex from *Arabidopsis* to physical homogeneity disproved the association of other proteins than AtCesA4, 7, and 8 (Atanassov et al., 2009). It is possible that the above named proteins would only be transiently associated with the complex, that is, be present only at a specific stage of the cellulose biosynthesis where they are needed (Guerriero et al., 2010).

1.4.2 Hemicellulose Biosynthesis

Structural similarities between the β-1,4-linked glucan chains of cellulose and the backbones of the various β-linked hemicellulosic polysaccharides led to the prediction that "CESA like" (CSL) genes might encode glycan synthases that are involved in the biosynthesis of these polysaccharides (Lerouxel et al., 2006). In contrast to cellulose synthase, the hemicelluloses synthases are located in the Golgi. Similar to the CESAs, however, the catalytic domains of the Golgi polysaccharide synthetases are also believed to face the cytosol. This then raises the problem of how the glycan chains are extruded. There is a general belief that this could occur in a similar way as cellulose synthesis (see earlier) through the formation of pores in the membrane. However, it appears that some members of the CslA, B, and C groups do not possess enough TM helices to allow the formation of such a pore. It has therefore been discussed that the catalytic sites of these enzymes are located in the Golgi lumen, which leaves the mechanism of further transport of the glycans open (Sandhu et al., 2009).

The *CSLs* are a family of genes that have sequence similarity to the *CESA* genes and appear to be present in all plant genomes (Figure 1.13). Certain *CSL* subfamilies are common to all plants, whereas other subfamilies are present only in specific groups of plants (Keegstra and Walton, 2006). Availability of a plant CESA sequence allowed annotation of a large number of genes with relatively low sequence similarity as CESA-*like* (*Csl*). Sequences in the *Csl* class are further divided into subgroups A–H and J. Groups B and G are specific to nongrass species, whereas groups F, H, and J are found only in grasses (Sandhu et al., 2009). In comparison with the biosynthesis of cellulose, however, that of hemicelluloses is far less well understood. Candidates for xylan synthase, feruloyl transferase, and arabinosyl transferase were tentatively identified by bioinformatic approaches, using a comparison of gene expression patterns of rice and *Arabidopsis* with a number of tissue expression libraries (Mitchell et al., 2007), and by the use of reversed genetics. All the β-glycan chain-forming enzymes identified thus far by these approaches belong to the Csl family (Sandhu et al., 2009).

Two groups of *Csl* genes, *CslF* and *CslH*, which are found only in grasses, have been shown to catalyze (1→3),(1→4)-β-D-glucan biosynthesis (Carpita, 2011). Heterologous expression of a rice *CslF* in *Arabidopsis*, a species that does not make (1→3),(1→4)-β-D-glucan, results in small amounts of the β-D-glucan in the cell walls (Burton et al., 2006). However, considerably greater amounts of the (1→3),(1→4)-β-D-glucan result when a *CslH* is coexpressed with *CslF*,

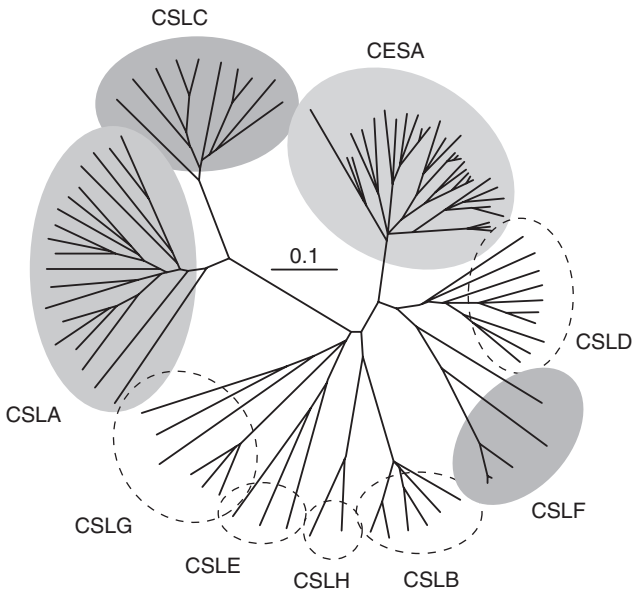


Figure 1.13. Phylogenetic identification of function of plant β-glucan synthases. CESA and CSLC are cellulose synthases; CSLA, β-mannan synthases; and CSLF, a β-1,3/1-4 glucan synthase. (Reprinted from Lerouxel O, Cavalier DM, Liepman AH, Keegstra K. Biosynthesis of plant cell wall polysaccharides—a complex process. Curr Opin Plant Biol, 2006; 9(6): 621–630. Copyright 2006, with permission from Elsevier.)

suggesting a synergistic role for both CslH and CslF in the synthesis of the polysaccharide and that a catalytic heterodimer enhances the activity (Carpita, 2011).

In addition to the backbones discussed earlier, most hemicelluloses also contain side chains, which are added by glycosyltransferases that are usually type II integral membrane proteins that have a catalytic domain facing the Golgi lumen (Lerouxel et al., 2006). The enzymes that add side chains to XyG and glucomannans are the best characterized, so we briefly consider them here.

Among the glycosyltransferases that add sugars onto the backbone galactosyltransferase, fucosyltransferase and glucuronoarabinoxylan have been identified (Zeng et al., 2010; Carpita, 2011).

1.4.3 Pectin Biosynthesis

As with hemicellulose synthases, pectin is also synthesized in the Golgi and transported to the wall in membrane vesicles. Pectin synthesis occurs simultaneously in numerous Golgi stacks in the cell in a process that appears to include a compartmentalization of specific biosynthetic enzymes to drive the construction of increasingly complex pectin polysaccharides through the *cis*, medial, and *trans*-Golgi cisternae (Mohnen, 2008).

Several reviews on pectin biosynthesis have recently been published. For a more detailed summary of the enzymology of pectin synthesis, readers are directed to Mohnen (2008) and Cafall and Mohnen (2009).

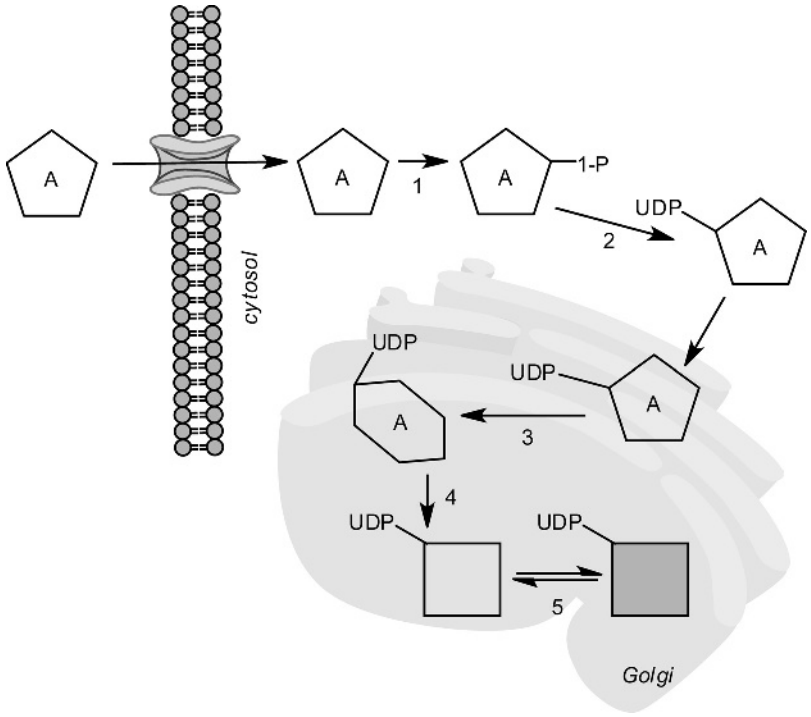


Figure 1.14. A model for galacturonic acid (GalA) recycling during plant cell wall assembly and restructuring from storage tissues: GalA (indicated by a the pentaeders with an A in the center), released from pectins, is transferred to the cytosol, where it is subsequently phosphorylated by galacturonate kinase and then converted to UDP-GalA in the cytosol by the UDP-sugar PPase "Sloppy." UDP-GalA can then be transferred into the Golgi, and can be either incorporated into glycans (not shown) or enter the nucleotide-sugar interconversion pathway, that is, being epimerized to UDP-GlcA (hexaeders with A in center) by UDP-GlcA 4-epimerase. UDP-GlcA is then further converted to UDP-xylose (*empty squares*) by UDP-GlcA decarboxylase. UDP-xylose 4-epimerase interconverts UDP-xylose and UDP-arabinose (*full squares*). (Modified after Yang et al., 2009.)

All evidence to date suggests that pectin is synthesized in the Golgi lumen by membrane-bound, or associated, Golgi-localized glycosyltransferases (GTs) that transfer glycosyl residues from nucleotide-sugars onto oligosaccharide or polysaccharide acceptors. It is not clear, however, how synthesis of any of the pectic polysaccharides is initiated or whether lipid or protein donors are required. Modification of the pectic glycosyl residues occurs by methyltransferases, acetyltransferases, and feruloyl transferases, using *S*-adenosylmethionine (SAM), acetyl-CoA, and feruloyl-CoA as substrates, respectively. Approximately 67 glycosyltransferase, methyltransferase, and acetyltransferase activities are predicted to be required for pectin synthesis (for review, see Caffall and Mohnen, 2009). Yet, conclusive identification of these genes has so far been accomplished only for a few of them.

Diverse biosynthetic pathways lead to the synthesis of the specific nucleotide-sugars required for plant pectin biosynthesis. Nucleotide-sugars may be formed via salvage pathways from sugars recycled from the wall polysaccharides or from sugars supplied to cultured cells (Bar-Peled and O'Neill, 2011; Figure 1.14). Such nucleotide-sugars, or primary sugar

Table 1.3. Predicted and proven *Arabidopsis thaliana* nucleotide-sugar interconverting enzymes required for pectin biosynthesis.

Enzyme	NDP-Sugar[a]	Enzymatic Reaction[b]
CMP-Kdo synthetase (CKS)	CMP-Kdo	E.C.2.7.7.38
GDP-mannose dehydrogenase (GMD)	GDP-4-keto-6-deoxy-Man	E.C.4.2.1.47
GDP-mannose epimerase (GME)	GDP-L-Gal	E.C.5.1.3.18
GDP-mannose epimerase/reductase (GER)	GDP-Fuc	NK
Kdo synthase (KDS)	CMP-Kdo	E.C.2.5.1.55
Myo-inositol oxygenase (MIOX)	D-GlcA	E.C.1.13.99.1
UDP-apiose/UDP-xylose synthase (AXS)	UDP-Api/UDP-Xyl	NK
UDP-glucose dehydrogenase (UGD)	UDP-Gal	E.C.1.1.1.22
UDP-glucose epimerase (UGE)	UDP-Glc	E.C.5.1.3.2
UDP-glucuronic acid epimerase (GAE)	UDP-GalA	E.C.5.1.3.6
UDP-rhamnose synthase (RHM)	UDP-L-Rha	E.C.4.2.1.76
UDP-xylose epimerase (UXE)	UDP-Ara	E.C.5.1.3.5
UDP-xylose synthase (UXS)	UDP-Xyl	E.C.4.1.1.35

NK, not known.
Data adapted from Caffall and Mohnen (2009), enzyme abbreviations are given in parentheses.
[a]The nucleotide-sugar synthesized.
[b]The enzyme commission number of each enzyme activity is listed based on the chemical reaction carried out.

phosphates derived directly from photosynthesis metabolism, are converted into a diverse array of sugar donor molecules by the nucleotide-sugar interconverting enzymes (NIEs) (Table 1.3). Nucleotide-sugars are supplied to wall biosynthetic glycosyltransferases by NIEs that regulate wall polysaccharide biosynthesis and are themselves frequently regulated by elements of the NIE pathway (Seifert et al., 2004). Evidence of NIE regulation of nucleotide-sugar availability is observed in the biological shift from primary wall to secondary wall synthesis: the abundance of nucleotide-sugars and their precursors is coordinately shifted to reflect an upregulation in hemicellulose and cellulose nucleotide-sugar substrates and a downregulation in pectic polysaccharide nucleotide-sugar substrates (Caffall and Mohnen, 2009).

GDP-D-mannose-4,6-dehydrogenase gene (GMD1) is a key component in the synthesis of GDP-fucose that is necessary for the correct synthesis of cell wall RG-II and of fucosylated xyloglucans. UDP-D-4-glucose epimerase (UGE) catalyzes the epimerization of UDP-D-Glc to UDP-D-Gal, which has an effect on the specific incorporation of Gal onto XG side chains, type-II arabinogalactan (β-1,6-galactan) and to a lesser extent RG-I (Seifert et al., 2002). UDP-D-Xyl is synthesized by the decarboxylation of UDP-D-GlcA by UDP-D-Xyl synthase (UXS), and UDP-D-Xyl is in turn utilized for the synthesis of UDP-D-Ara (Harper and Bar-Peled, 2009). Rhamnose biosynthesis (RHM) is responsible for the synthesis of UDP-L-Rha from UDP-D-Glc by three distinct activities: UDP-D-glucose-4,6-dehydratase, UDP-4-keto-6-deoxy-D-glucose-3,5-epimerase, and UDP-4-keto-L-rhamnose 4-keto-reductase (Oka et al., 2007).

The HG backbone is a polymer of α1,4-linked GalA residues and thus likely requires multiple HG:α1,4-GalATs to synthesize the entire complement of HG required throughout plant development. These enzymes (E.C.2.4.1.43) catalyze the transfer of D-GalA from UDP-D-GalA onto a growing stretch of HG via a catalytic domain facing the lumen of the Golgi. Unmethylated UDP-D-GalA was the preferred nucleotide-sugar substrate for elongation of endogenous acceptors in particulate membrane fractions of mung bean (*Vigna radiata*). A

subfraction of HG is composed of XGA; HG decorated with xylose residues at the *O*-3 of backbone GalA residues. A xylosyltransferase in *Arabidopsis* was identified that is reported to be an XGA xylosyltransferase, which would be the first specific glycosyltransferase for XGA synthesis (Caffall and Mohnen, 2009).

The modification of pectic polysaccharides by addition of methyl groups at the C-6 carboxyl group or acetyl groups at the *O*-2 or *O*-3 of GalA residues occurs during synthesis in the Golgi apparatus. Pectins are first secreted in a highly methyl esterified form, but after the deposition of pectins in the apoplast, the methyl esters are removed by pectic methylesterases. Methylation occurs by multiple pectin methyltransferases that catalyze the transfer of a methyl group from the donor *S*-adenosyl-methionine (SAM) to the pectic polysaccharide (Caffall and Mohnen, 2009).

Pectin *O*-acetyltransferase (PAT) activity catalyzes the transfer of an acetyl group from acetyl-CoA to a pectic polysaccharide acceptor. Acetyl groups decorate the GalA residues of pectic polysaccharides at the *O*-2 or *O*-3 positions. The functional consequences of acetylation are not clear but may play a role in preventing pectin breakdown by microbial hydrolases.

The biosynthesis of RG-I requires multiple glycosyltransferase activities to synthesize a backbone of [1→2)-α-L-Rha*p*-(1,4)-α-D-Gal*p*A-(1→4] disaccharide repeats that are branched at the C-4 of approximately half of the rhamnose residues by 5-linked and 3,5-linked arabinan, 4-linked and 4,6-linked galactan, as well as type-I and type-II AG. Potentially 34 specific activities may be required to synthesize RG-I backbone (for details, see Cafall and Mohnen, 2009).

The galactan side chains of RG-I are composed primarily of β-1,4-D-galactan with some branches of β-1,6- and β-1,3-Gal. Galactosyltransferases (GalTs) catalyze transfer of D-Gal*p* residues from UDP-D-Gal*p* to an acceptor molecule. Pectin GalTs are hypothesized to catalyze the initiation of the galactan side chains directly onto the backbone rhamnose residues of RG-I, elongate the galactan chains, initiate branch points onto the galactan chains, and elongate the galactan side chains. Synthesis of these structures may require up to ten or more different pectin-specific GalTs (Mohnen et al., 2008). The synthesis of α-1,5-linked arabinan by α-1,5-arabionosyltransferases (AraTs) occurs by the transfer of arabinose residues from UDP-L-arabinose to acceptor molecules in an α-(1,5)-configuration. Up to 18 AraTs have been estimated to be required for the synthesis of the complex branched arabinans in pectin (Mohnen et al., 2008).

The backbone of RG-II is α-1,4-D-GalA-linked HG. GAUT1 has recently been discovered, which catalyzes the addition of GalA residues onto HG oligomers (Sterling et al., 2006), and could, conceivably, synthesize the backbone of RG-II. There are 15 GAUT genes in the GAUT1-related gene family, and it is also possible that one of these may function to specifically synthesize the backbone of RG-II. Further work in this area will be needed to explore this possibility, because the acceptor specificities for the GAUTs have not been determined (Caffall and Mohnen, 2009).

1.4.4 Lignin Biosynthesis

Lignins are synthesized via oxidative coupling between monomers of *p*-hydroxycinnamyl alcohols, namely, the monolignols *p*-coumaryl (*p*-CA), coniferyl (CA), and sinapyl alcohols (SA) (Boerjan et al., 2003; Vanholme et al., 2008). In addition to these building blocks, lignin also incorporates many more compounds, such as products from incomplete monolignol biosynthesis (5-hydroxyconiferyl alcohol, hydroxycinnamaldehydes, and hydroxycinnamic

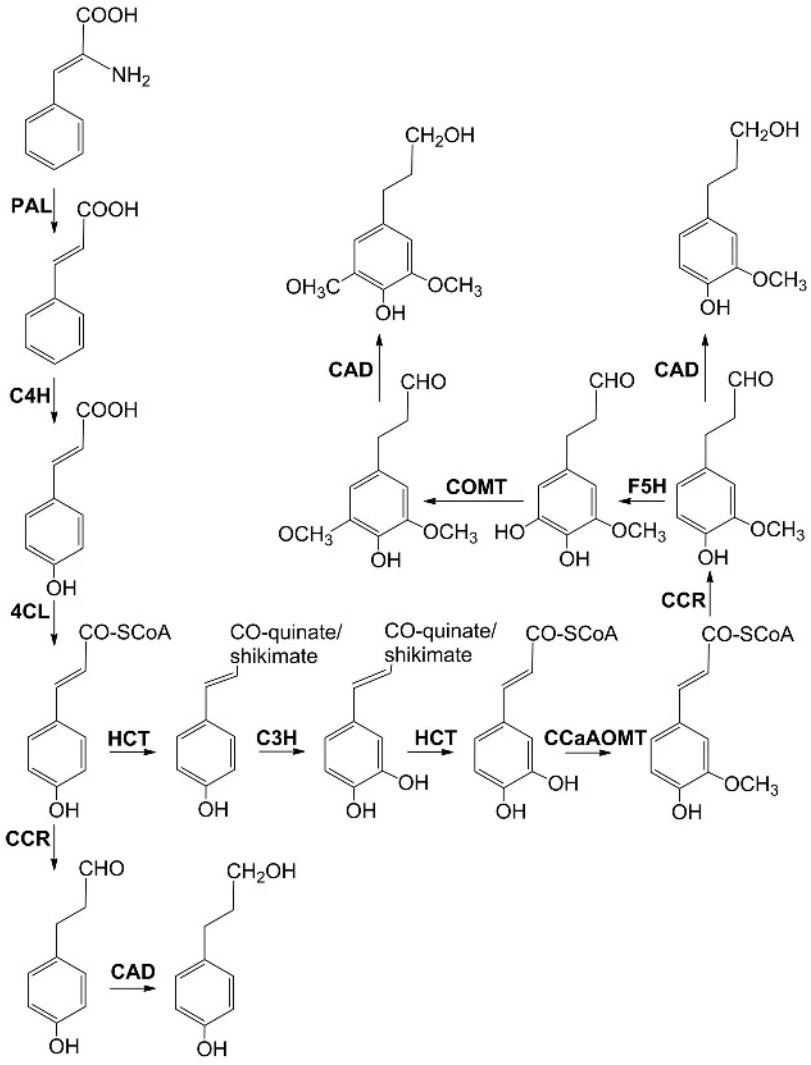


Figure 1.15. Biosynthesis of lignin via the phenylpropanoid pathway. PAL, phenylalanine ammonia-lyase; C4H, cinnamate-4-hydroxylase; 4CL, 4-coumarate:CoA ligase; HCT, hydroxycinnamoyl:CoA transferase; C3H, 5-*O*-(4-coumaroyl)shikimate 3′-hydroxylase; CCoAOMT, caffeoyl-CoA *O*-methyl transferase; CCR, cinnamoyl-CoA reductase; CAD, cinnamyl alcohol dehydrogenase; F5H, ferulate/coniferaldehyde 5-hydroxylase; COMT, caffeate/5-hydroxyconiferaldehyde *O*-methyltransferase.

acids), or derivatives of the classical monolignols (e.g., sinapyl *p*-hydroxybenzoate, coniferyl and sinapyl *p*-coumarate, and coniferyl and sinapyl acetate) (Vanholme et al., 2008).

A general reaction scheme for lignin biosynthesis is shown in Figure 1.15: the precursors are synthesized through the phenylpropanoid pathway, starting from phenylalanine ammonia-lyase (PAL) and leading to the three monolignols through a series of hydroxylations, methylations, and reductions (Simmons et al., 2010). The enzymes catalyzing the lignin biosynthetic pathway

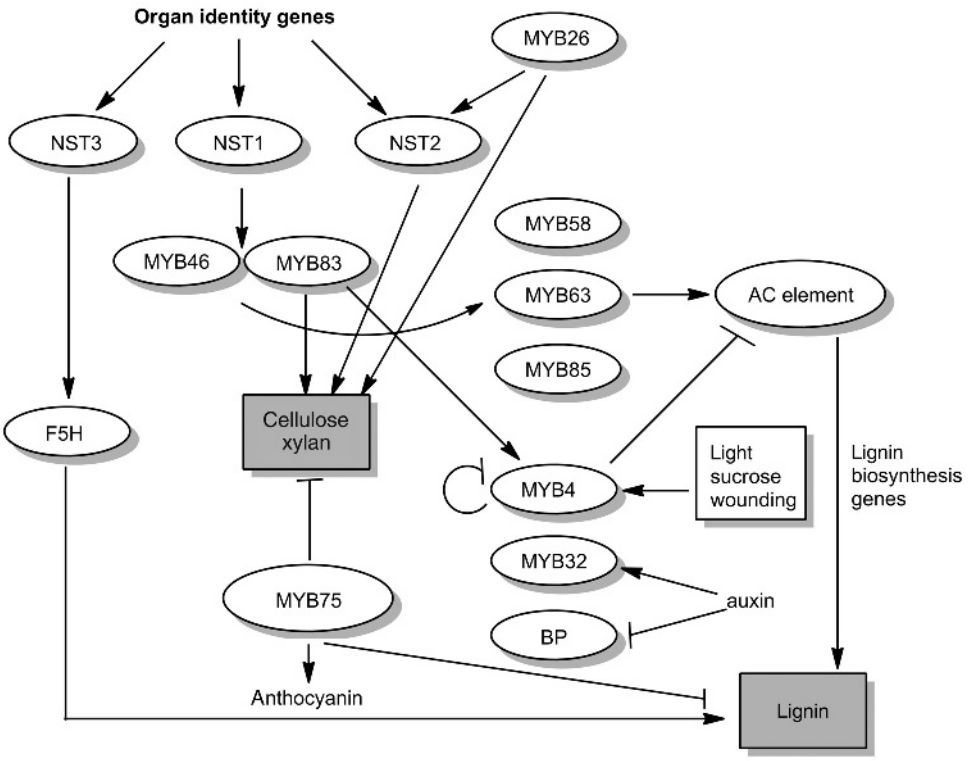


Figure 1.16. Current model of the transcriptional network of secondary cell wall biosynthesis in *Arabidopsis*. Components in oval circles specify regulatory proteins; *arrows* indicate activation or positive influence; vertical bars at the end of a line indicate inhibition or repression. For more details, see the review of Zhao and Dixon (2011).

are presumably all known (see Figure 1.15) and the regulation of their expression has been shown to be governed by several transcription factors of the NAC and MYB gene families (Zhou et al., 2009; Guillaumie et al., 2010; Figure 1.16). There is evidence suggesting that the transcript levels for PAL, C4H, 4CL, HCT, CCoAOMT, and F5H/CAld5H are functionally redundant (Simmons et al., 2010).

Following the biosynthesis of the lignin monomers, they are translocated to the cell wall, where they diffuse until they are oxidized for polymerization within the secondary plant cell wall. This transport is not fully understood, but available evidence suggests that it occurs either as free monolignols or as phenolic glucosides (coniferin, syringin) (Kärkönen and Koutaniemi, 2010). Genes for synthesizing these glycosides have been cloned, but the significance of the "glycoside hypothesis" of monolignol transport is questioned because coniferin and syringin do not accumulate to high levels in angiosperm xylem, and coniferyl and sinapyl alcohols are able to freely diffuse through the plasma membrane (Vanholme et al., 2008).

Once in the cell wall, the lignin precursors are polymerized via radical reaction. The formation of radicals is catalyzed by oxidative enzymes, either H_2O_2-dependent peroxidases or O_2-dependent oxidases/laccases. These enzymes are secreted into the apoplast where they are either soluble or covalently or ironically bound to the cell wall (Kärkönen and Koutaniemi,

2010). The broad substrate specificities and the large gene families of these two classes of oxidases have made it difficult to identify isoforms that are specifically involved in developmental lignification. Coexpression analyses of genes expressed during stages of active lignification or during tension wood formation and gene-family-wide expression have been performed in the attempt to associate individual gene family members with specific processes (Vanholme et al., 2008). Since peroxidases seem to be crucial for lignin biosynthesis, H_2O_2 should be available for peroxidases in the cell wall during lignin formation. Candidate hydrogen peroxide forming apoplastic enzymes are (i) a "cell wall oxidase" that catalyzes the oxidation of NADH to NAD^+, which in turn reduces O_2 to O_2^-, consequently is dismutated to produce O_2 and H_2O_2; (ii) germin-like oxalate oxidases and amine oxidases; (iii) cell membrane NADPH-dependent oxidase (NADPH oxidase); and (iv) cell wall polyamine oxidase (reviewed by Quan et al., 2008).

The question of how monolignols couple during lignifications is still controversially debated (Vanholme et al., 2008). One theory, originally developed by Freudenberg (for details, see Freudenberg and Nash, 1968), contends that lignin monomers are oxidized and then coupled in a combinatorial fashion. Thus, any phenol present in the lignin forming area of the cell wall is capable of becoming subject of the combinatorial free radical-coupling process, depending on its structural compatibility, and its dependency on reaction parameters, such as pH, temperature, ionic strength, monolignol supply, hydrogen peroxide and peroxidase concentrations, and the matrix in general (Vanholme et al., 2008). An alternative hypothesis, that is, the lignin monomers are coupled with absolute structural control by proteins bearing arrays of dirigent sites, has been proposed, but the evidence in favor of this is, at the time of this writing, less convincing than the Freudenberg model (for detailed explanation, see Ralph et al., 2008).

1.5 Strategies for Manipulating Wall Composition

Genetic engineering of crops in order to increase structural carbohydrate content and reduce lignin levels is a promising path that may result in reduced pretreatment severity, facilitate the hydrolysis process, and help recover the maximum amount of sugars. In addition to this, cellulose and hemicellulose degradation enzymes are also being expressed in the cell wall, which decreases the overall cellulase enzyme load during saccharification of biomass (Chandel and Singh, 2011).

1.5.1 Manipulation of Plant Cell Wall Polymer Composition

A major drawback in this area is that, although significant progress has already been obtained in the elucidation of the biochemical mechanisms that ultimately lead to cell wall polymers, the regulation of wall polymer biosynthesis remained largely elusive (Pauly and Keegstra, 2010). As I have explained earlier, UDP-glucose and other hexose-phosphates are the key substrate for biosynthesis of the different nucleotide-sugars required for synthesis of the various wall polymers, and manipulating of the enzymes involved in these steps has therefore been used as one of the earliest targets for altering the cell wall composition of plants. For example, overexpression of the sucrose synthase (for the importance of sucrose in cellulose biosynthesis, see Section 1.4) in poplar leads to a higher proportion and absolute amount (2–6%) of cellulose in fiber cells, whereas growth and biomass remained unaffected (Coleman et al., 2009). The

most obvious strategy, that is, overexpression of the genes encoding the nucleotide-sugar conversion enzymes, has so far not yielded plants with altered wall compositions, possibly because of the redundancy of genes (Pauly and Keegstra, 2010).

Another strategy for modifying the composition of cell walls would be to up- or down-regulate the levels of glycosyltransferases and glycan synthases involved in polysaccharide biosynthesis. Attempts in this direction have so far been made only in model plants like *A. thaliana* and tobacco, which showed that it is principally possible to make dramatic alterations in the hemicellulose composition of a plant without serious consequences for its ability to grow and reproduce (Pauly and Keegstra, 2010).

1.5.2 Manipulation of Plant Lignin Content

As will be described in Chapter 10, lignin is one of the most important negative factors in the conversion of plant material to bioethanol. Analysis of lignin mutants showed that altering the expression of individual genes of its biosynthetic pathway has far-reaching consequences on plant metabolism: reduced lignin contents are typically associated with dramatic changes in the soluble phenolic pools, and different species accumulate various storage and detoxification

Table 1.4. Possible targets for genetic engineering of plants to generate high yielding and less recalcitrant biomass.

Selected Crop Traits	Approach Targeted	Effects Observed
Photosynthesis	Overexpression of phosphoenolpyruvate carboxylase, fructose-1,6-bisphosphatase and sedoheptulose-1,7-bisphosphatase	Increased CO_2 fixation leads to increased dry weight of biomass, enables water resistance
Cell wall composition	Specific cytochrome P450 enzymes, caffeic acid *O*-methyltransferase	Increased cellulose amount, less lignin
Starch composition	Starch enzymes, pullulanase	Alteration in starch structure and increased amount of starch
Stress tolerant	Signal transduction, transcription factors, effector genes	Development of stress-tolerant varieties
Cellulose degrading enzymes	Cellulase expression, β-glucanase expression	Cellulase and β-glucanase production in the cell wall
Grain yield	Enhanced ADP-glucose pyrophosphorylase activity; deregulation of endosperm ADP-glucose pyrophosphorylase activity; stimulation of photosynthesis and carbon metabolism	Improved seed weight and biomass yielding crops
Male sterility and plastid transformation	Engineering cytoplasmic male sterility via chloroplast genome by expression of β-ketothiolase	Impact on the development of routine biofuel crops

Adapted from Chandel and Singh (2011).

products. In addition, alterations in lignin metabolism also influence the biosynthesis of other cell wall polymers. Knowledge about these broader effects is essential to fully comprehend how gene function and cell wall properties are linked, how these cell wall properties are elaborated, and how they relate to the quality of raw material destined for agro-industrial uses (for reviews, see Simmons et al., 2010; Pauly and Keegstra, 2010; Chandel and Singh, 2011).

Manipulation of the genes involved in lignin biosynthesis obvious is the first candidate to try: downregulation of PAL, HCT, C3H, CCoAOMT, CCR, and CAD (for abbreviations, see legend to Figure 1.15); enzymes have been shown to impact the monomeric composition and amount of lignin present. Downregulation of C3H in *Medicago sativa* dramatically shifted lignin composition and structure, and produced a material that was easily digested. Downregulation of 4CL1 in *Populus tremuloides* (poplar) produced a 45% decrease in lignin and which consequently increased the relative cellulose content as well as some of the hemicellulose components (Simmons et al., 2010).

Another approach is the interference with lignin polymerization: the phenolic hydroxyl, *para* to the side chain, is critically important for oxidative coupling and polymerization. Lignin precursor phenolic groups in *ortho* configuration become methylated by the *O*-methyltransferases CCoAOMT and COMT, which results in formation of guaiacyl and syringyl monomers. Marita et al. (2003) have used downregulation of COMT, which is responsible for the methylation of 5-hydroxconiferyldehyde to decrease the flux to sinapyl alcohol. This resulted in the accumulation of 5-hydroxyconiferaldehyde in the cell and its reduction to 5-hydroxyconiferyl alcohol, which becomes incorporated into lignin leading to new types of structures in the lignin polymer. Other means to change lignin structure involved the incorporation of coniferyl ferulate into lignins, which could open the way to obtain easily cleavable ester linkages into its backbone, thus facilitating its depolymerization and solubilization (Grabber et al., 2008).

Torney et al. (2007) reviewed the genetic engineering approaches to improve bioethanol production from maize, which are focused on increasing stress tolerance, the rate of photosynthesis, grain yield, and production of biomass conversion enzymes *in planta* (Table 1.4). These approaches could also be incorporated for the improvement of weedy crops in terms of increased biomass weight, cell wall composition, and biomass conversion assisted by enzyme expression *in planta* (Chandel and Singh, 2011).