

Modeling Lignin Polymerization. I. Simulation Model of Dehydrogenation Polymers^{1[OA]}

Frederik R.D. van Parijs, Kris Morreel, John Ralph, Wout Boerjan², and Roeland M.H. Merks^{2*}

Department of Plant Systems Biology, VIB, 9052 Ghent, Belgium (F.R.D.v.P., K.M., W.B.); Department of Plant Biotechnology and Genetics, Ghent University, 9052 Ghent, Belgium (F.R.D.v.P., K.M., W.B.); Centrum Wiskunde en Informatica, 1090 GB Amsterdam, The Netherlands (F.R.D.v.P., R.M.H.M.); Netherlands Consortium for Systems Biology/Netherlands Institute for Systems Biology, 1090 GB Amsterdam, The Netherlands (F.R.D.v.P., R.M.H.M.); and Department of Biochemistry and Department of Energy Great Lakes Bioenergy Research Center, University of Wisconsin, Madison, Wisconsin 53706 (J.R.)

Lignin is a heteropolymer that is thought to form in the cell wall by combinatorial radical coupling of monolignols. Here, we present a simulation model of *in vitro* lignin polymerization, based on the combinatorial coupling theory, which allows us to predict the reaction conditions controlling the primary structure of lignin polymers. Our model predicts two controlling factors for the β -O-4 content of syringyl-guaiacyl lignins: the supply rate of monolignols and the relative amount of supplied sinapyl alcohol monomers. We have analyzed the *in silico* degradability of the resulting lignin polymers by cutting the resulting lignin polymers at β -O-4 bonds. These are cleaved in analytical methods used to study lignin composition, namely thioacidolysis and derivatization followed by reductive cleavage, under pulping conditions, and in some lignocellulosic biomass pretreatments.

Lignins are aromatic polymers that are predominantly present in secondarily thickened cell walls. These polymers make the cell wall rigid and impervious, allowing transport of water and nutrients through the vascular system and protecting plants against microbial invasion. Lignins are heterogeneous polymers derived from phenylpropanoid monomers, mainly the hydroxycinnamyl alcohols coniferyl alcohol (G-monomer) and sinapyl alcohol (S-monomer) and minor amounts of *p*-coumaryl alcohol (H-monomer). These monolignols differ in their degree of aromatic methoxylation (-OCH₃ group; Fig. 1). The

resulting units in the lignin polymer are the guaiacyl (G), syringyl (S), and *p*-hydroxyphenyl (H) units. They are linked by a variety of chemical bonds (Fig. 2) that have different chemical properties (Boerjan et al., 2003; Ralph et al., 2004; Vanholme et al., 2008).

Lignification is the process by which monomers and/or oligomers are polymerized via radical coupling reactions and typically occurs after the polysaccharides have been laid down in the cell wall. Lignin composition varies among cell types and can even be different in individual cell wall layers (Ruel et al., 2009). Lignin composition is also influenced by environmental conditions; for example, lignin in compression wood is enriched in H-units (Timell, 1986). Hence, both developmental and environmental parameters influence the composition and thus the structure of the lignin polymer (Boerjan et al., 2003; Ralph et al., 2004).

Lignin is one of the main negative factors in the conversion of lignocellulosic plant biomass into pulp and bioethanol (Lynd et al., 1991; Hill et al., 2006). In these processes, lignin needs to be degraded by chemical or mechanical processes that are expensive and often environmentally polluting. Hence, major research efforts are devoted toward understanding lignin biosynthesis and structure. It has already been shown that reducing lignin content and modifying its composition in transgenic plants can result in dramatic improvements in pulping efficiency (Pilate et al., 2002; Baucher et al., 2003; Huntley et al., 2003; Leplé et al., 2007) and in the conversion of biomass into bioethanol (Stewart et al., 2006; Chen and Dixon, 2007; Custers, 2009). These altered biomass properties are related to the alterations in lignin composition and structure in terms of the frequencies of the lignin units and the

¹ This work was supported by the Department of Energy Great Lakes Bioenergy Research Center (grant no. DE-FC02-07ER64494), by Stanford's Global Climate and Energy Project (grants to W.B., F.R.D.v.P., and J.R.), by a Marie Curie Intra-European Fellowship (no. MEIF-CT-2005-025084 to R.M.H.M.) and a Marie Curie European Reintegration Grant (no. PERG03-GA-2008-230974 to R.M.H.M.), and by the Netherlands Consortium for Systems Biology, which is part of the Netherlands Genomics Initiative/Netherlands Organization for Scientific Research. Any opinions, findings, and conclusions or recommendations expressed in this publication are those of the authors and do not necessarily reflect the views of Stanford University, the Sponsors of the Global Climate and Energy Project, or others involved with the Global Climate and Energy Project.

² These authors contributed equally to the article.

* Corresponding author; e-mail roeland.merks@cwi.nl.

The author responsible for distribution of materials integral to the findings presented in this article in accordance with the policy described in the Instructions for Authors (www.plantphysiol.org) is: Roeland M.H. Merks (roeland.merks@cwi.nl).

[^{OA}] Open Access articles can be viewed online without a subscription.

www.plantphysiol.org/cgi/doi/10.1104/pp.110.154468

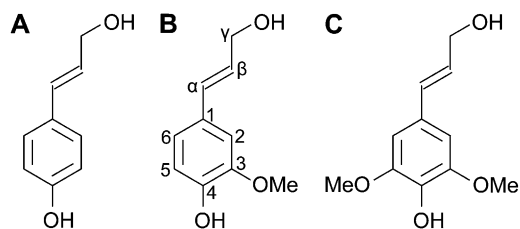


Figure 1. Chemical structures of three monolignols. A, H-monomer (*p*-coumaryl alcohol). B, G-monomer (coniferyl alcohol). C, S-monomer (sinapyl alcohol). G- and S-monomers are considered in our simulations. The G-monomer is methoxylated (-OCH₃ group) on position 3, and the S-monomer is methoxylated on positions 3 and 5.

bond types connecting them and possibly also their interaction with hemicelluloses (Ralph et al., 2004; Ralph, 2006).

To study the parameters that influence lignin structure, lignin polymerization has been mimicked *in vitro* by experiments with dehydrogenation polymers (DHPs; Terashima et al., 1995). Indeed, lignification can be mimicked by oxidizing monolignols using a peroxidase, such as horseradish peroxidase (HRP), and supplying its cofactor hydrogen peroxide, producing synthetic DHP lignins. Monolignol oxidation can also be achieved without enzymes (e.g. by using transition metal one-electron oxidants, such as copper

acetate). Some of these biomimetic DHPs have been suggested to be better models for wood lignins than HRP-generated DHPs (Landucci, 2000).

In DHP experiments, the monolignols are either added in bulk (Zulauf experiment) or dropwise (Zutropf experiment) to the reaction mixture, yielding lignin polymers with very different bond frequencies (Freudenberg, 1956). Zutropf experiments approach the *in vivo* formation of lignin, which depends on the slow introduction of monolignols into the wall matrix via diffusion to the site of incorporation (Hatfield and Vermerris, 2001). Because the exact reaction conditions are known, such *in vitro* experiments have provided insight into the lignification process in plants. In this way, numerous factors were shown to influence lignin structure, including the relative supply of the monolignols, the pH, the presence of polysaccharides, hydrogen peroxide concentrations, and cell wall matrix elements in general (Grabber et al., 2003; Vanholme et al., 2008).

Computer simulations of lignin polymerization can help explain and predict lignin structure from low-level chemical kinetic factors, including subunit-coupling probabilities and monolignol synthesis rates. Such models are helpful in explaining the mechanism behind a range of controlling factors identified in the experimental work, including (1) the ratio of coniferyl versus sinapyl alcohol monolignols, (2) the monolignol supply rate, and (3) the abundance of alternative

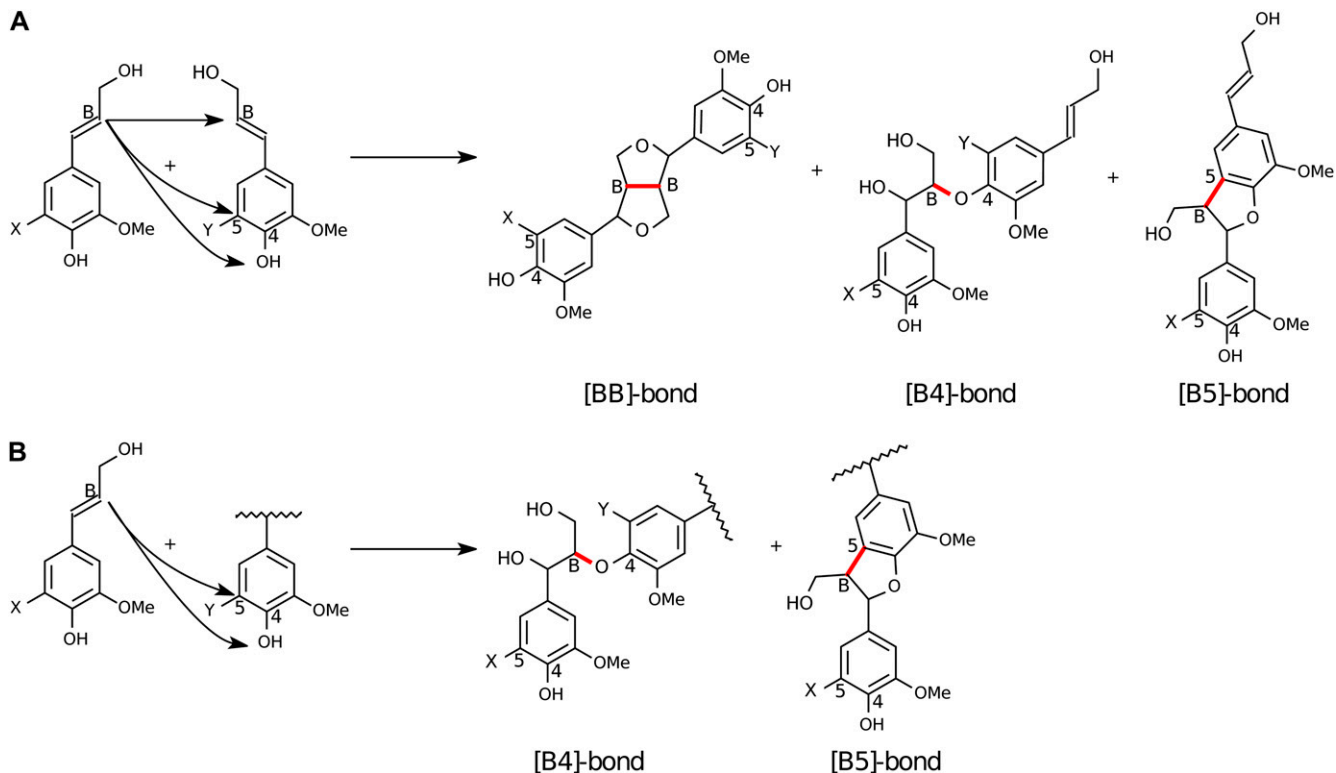


Figure 2. Chemical structures resulting from the possible bonding between two monomers (A) or a monomer and the bindable end of an oligomer (B). X and Y in the monomers denote the absence (for a G-unit) or presence (for an S-unit) of a methoxy group at position 5 (see Fig. 1). The red line indicates the bonds generated by couplings of the B position and B, 4, or 5 position.

monomers present during lignin biosynthesis in mutants and transgenics. Thus, computer models will also help in suggesting new targets for controlled lignin biosynthesis.

Here, we propose a simulation model of synthetic lignin polymerization that is based upon an emerging consensus from a variety of observations and derives from a series of previous models of lignin polymerization (Glasser and Glasser, 1974; Glasser et al., 1976; Jurasek, 1995; Roussel and Lim, 1995). Our model uses a symbolic grammar to describe a constructive dynamical system (Fontana, 1992) or a rule-based system (Feret et al., 2009) in which it is not necessary to define all possible products in advance. We assume that G- and S-monomers and newly formed oligomers couple in a well-mixed medium, depending on coupling rules and experimentally measured coupling probabilities. To develop the model, we have used information from DHP experiments rather than natural lignins, as they are formed in a well-mixed medium and their reaction conditions are well known (e.g. the influx rate of monomers). Using information from natural lignin would have further complicated our model, as the structures of natural lignin polymers are influenced by many factors, including the possible involvement of dirigent proteins (Davin and Lewis, 2005), steric hindrance by polysaccharides, spatiotemporal regulation, and modifications during isolation procedures (Boerjan et al., 2003; Ralph et al., 2004).

Using our simulation models, we study how putative controlling factors of lignin primary structure, including the influx rate of monomers and the relative amount of S-monomers, affect *in silico* lignin synthesis, and we compare our predictions with *in vitro* experiments. To predict the degradability of lignins formed in our simulations, we apply an *in silico* thioacidolysis, which cleaves the polymers at their β -O-4 positions. This simulates the molecular action of two of the most used methods to analyze lignin composition, thioacidolysis (Lapierre, 1993; Baucher et al., 2003) and derivatization followed by reductive cleavage (Lu and Ralph, 1997). The G+S-monomer yield is often taken as a reflection of the fraction of units bound by β -O-4 bonds. Cleavage of β -O-4 bonds is also the most important reaction in kraft pulping of wood (Baucher et al., 2003). The model predicts from first principles (1) that DHP lignins formed under Zutropf conditions have a higher β -O-4 content than those formed under Zulauf conditions, (2) that DHP lignins formed with high S content have a higher β -O-4 content than those formed with high G content, and (3) that a higher β -O-4 content does not necessarily reduce the average length of lignin fragments generated during *in silico* thioacidolysis.

MODEL DESCRIPTION

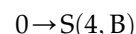
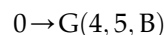
Lignin Polymerization Grammar

Our initial model describes free radical polymerization with a symbolic grammar. This approach de-

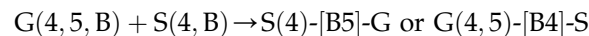
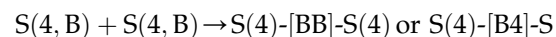
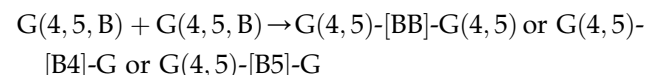
scribes lignin molecules as sequences with symbols representing monomeric units and bonds between units. The symbols G and S represent guaiacyl and syringyl units, the only units that are considered in our model. Square brackets describe chemical bonds; for example, [B5] in S(4)-[B5]-G would denote a bond between position β of an S-unit and position 5 of a G-unit. Round brackets are used to show the bindable positions at the ends of an oligomer. For example, S(4)-[B5]-G has a bindable S-end (phenolic, position 4) and an unbindable G-end. For convenience, the phenolic end is always at the left side in our notation. The corresponding chemical structures can be found in Figure 2.

Monomers and oligomers couple according to a set of rules governed by the chemistry involved (Ralph et al., 2004). For these rules, only the bindable ends of oligomers are considered, by lumping together species with identical reactive sites. This rule-based algorithm (Feret et al., 2009) prevents us from considering an endlessly growing number of reactions; instead, we look at the finite set of reactions between each pair of bindable positions.

A first set of reactions models the influx of monomer radicals into the system. We assume that G-monomers have three free binding positions (4, 5, and β) and that S-monomers have only two (4 and β); position 5 is not bindable because of its substitution by a methoxyl group (Fig. 1). Because of their low frequency in DHP polymers, we did not consider bonds at positions 1 and α . Note that position 4 here strictly refers to the 4-O position. A zero indicates that there are no reactants, allowing us to describe the influx of reactants as a "zero order" reaction:



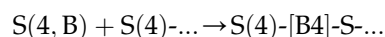
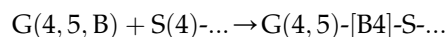
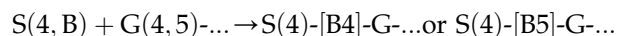
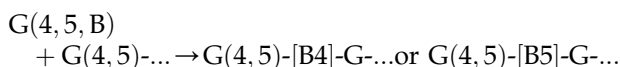
The next set of reactions describes dimerizations (i.e. the coupling of two monomers):



If two monomers bind, each will do so combinatorially at one of its bindable positions, with probabilities based on DHP experiments (see below). At least one monomer always couples at its β position, limiting the number of possible interactions (i.e. [44], [55], and [45] bonds are not possible). This leaves four theoretically possible interactions for heterodimers (G + S), but experimentally only two of those were found in DHPs and poplar (*Populus* spp.) xylem extracts (Morreel et al., 2004). The heterodimers G(4,5)-[BB]-S(4) and S(4)-[B4]-G are below the limit of detection in the DHP reactions from Morreel et al. (2004).

Some bonds are mutually exclusive: if position 4 or 5 is occupied, the unit is etherified so that radical formation is no longer possible and none of the other positions allows further reaction. A dimer always has at least one bindable (phenolic) end, as at least one monomer always couples at its β position. The β - β dimer has two bindable ends.

During monomer-oligomer coupling, the monomer always couples at its β position to the oligomer:



with the ellipses (...) indicating the remainder of the oligomer. As radical formation remains possible from the new phenolic end of the growing oligomer, there is no oligomer termination: oligomers continue to grow as long as monomer radicals are available. Coupling of a monomer exclusively at its β position also means that oligomers cannot have free β positions, so β - β bonds do not form when either a monomer couples with an oligomer or two oligomers couple (i.e. a β - β bond always originates from a dimerization reaction).

Oligomer-oligomer coupling can, in principle, lead to the formation of a branch point. In vitro and in vivo oligomer-oligomer coupling is thought to occur at a lower rate than monomer-monomer and monomer-oligomer coupling, particularly in angiosperms (Jacquet et al., 1997). For that reason, we currently ignore oligomer-oligomer coupling in our model and consider only linear strands of lignin monomers.

Simulation Algorithm

The complete reaction mixture constitutes an ensemble of all unique monomers and oligomer sequences, with their relative abundances, all in a list generated by the model. We also keep a separate list that lumps reagents with identical bindable ends, listing, for example, the amounts of oligomers with one bindable G-end, oligomers with two bindable G-ends, etc. The choice of reaction and the time at which this occurs is determined by a stochastic simulation algorithm (Gillespie, 1977). Figure 3 gives a flow chart of the complete simulation algorithm.

The stochastic simulation algorithm iteratively chooses the next reaction i and the time t at which it takes place. For each reaction, we determine the reaction rate or propensity: $a_i = h_i \times c_i$, with h_i representing an extrinsic reaction rate factor and c_i representing an intrinsic rate constant. The intrinsic rate constant, c_i , depends on the physicochemical features of both reactants, the bond that is formed, the temperature, and the reaction volume (see below). The extrinsic reaction rate, h , is proportional to the collision prob-

ability of the reactants; it is the number of unique combinations between all molecules in the reaction mixture that match both lumped reactants. As the reaction mixture is changed after each reaction, the set of extrinsic values also will change continuously. The algorithm generates a random number, $r_1 \in [0,1]$, from a uniform distribution and selects reaction i for which $\sum_{j=1}^{i-1} a_j < r_1 a_0 \leq \sum_{j=1}^i a_j$, with $a_0 = \sum_j a_j$. The time difference with the previous reaction (τ) is determined by another random number, $r_2 \in [0,1]$, as $\tau = (1/a_0) \ln(1/r_2)$ (Gillespie, 1977).

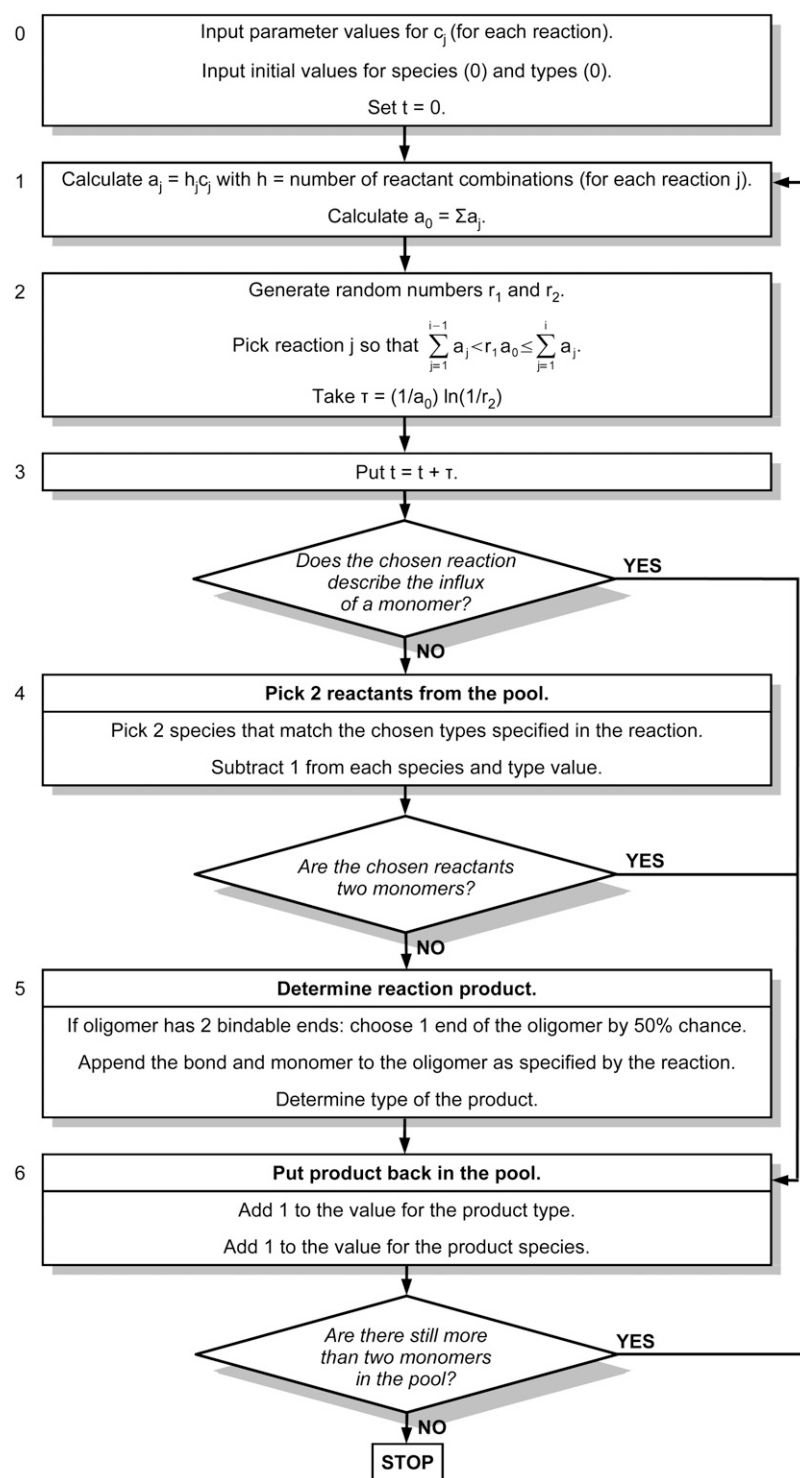
Once a reaction is selected, the algorithm selects the corresponding molecular types from the list of lumped reactants and randomly selects two specific molecules that match these molecular types. Then, the algorithm binds the reactants according to the chosen reaction, which also specifies a bond type, by appending the bond and monomer {e.g. -[4B]-G(4,5)} to the oligomer sequence according to the grammatical rules listed above.

The influx rate for each monomer is the product of the overall monomer influx rate and the G-fraction or S-fraction ($1 - \text{G-fraction}$). The influx is stopped after a maximum number of monomers has been introduced in the reaction mixture (50,000 in the simulations presented here). This approach allows us to separate the effects of molecule density from the effects of influx rates.

Parameter Settings

The reaction constant of a reaction can be considered to be the product of the intrinsic probability that the reactants bind after collision (binding probability), the intrinsic probability that a specific bond is formed (bond probability), and an overall rate constant. The overall reaction rate is unknown, but we used the factor 10^{-6} for practical reasons, to make the simulations less costly. It does not affect the results, as only the relative reaction constants are of importance (data not shown). To a first approximation, we assumed all binding probabilities to be the same for each reactant combination, doubling the propensities for reactions of oligomers with two bindable ends (Bamford and Dewar, 1947). Also, the bond probabilities for monomer-oligomer reactions are unbiased. To assign the bond probabilities for dimerization reactions ([B4], [B5], [BB]), we made an estimate based on data from Zulauf DHP experiments (Morreel et al., 2004) where either only G-monomers or only S-monomers were added to the reaction mixture (Table I). The reaction constants (input) were hand fitted by matching the in silico frequencies of the dimer species (output of a Zulauf simulation) to the in vitro frequencies. Table II lists the resulting reaction constants as used in our simulation. To ascertain that our results did not depend on these specific values of the parameters, our simulations were repeated using unbiased bond prob-

Figure 3. Flow chart of the simulation algorithm. Repeatedly, a reaction is chosen and applied to the reaction mixture. Boxes 0 to 3 show the Gillespie (1977) algorithm, and boxes 4 to 6 show the execution of a reaction. Type includes the lumped description of oligomers (i.e. an oligomer with a bindable G-end or a bindable S-end) as well as the monomer types.



abilities for all monomer-monomer and monomer-oligomer couplings.

Simulation Data Analysis

To analyze the relative frequencies of the reactions and bonds formed during the simulation, we keep track of the frequency by which each reaction occurs. We analyze these data based on the following observ-

ables. The total β -O-4 fraction gives the fraction of β -O-4 bonds from the dimers, and higher order oligomers in the reaction mixture,

$$f_{\beta O4} = \frac{n_{\beta O4}}{n_{\beta O4} + n_{\beta\beta} + n_{\beta 5}}$$

with $n_{\beta O4}$ the number of β -O-4 bonds and $n_{\beta O4} + n_{\beta\beta} + n_{\beta 5}$ the total number of chemical bonds in the reaction

Table I. The abundances of each dimer (relative to all dimers; rows 1–7) and each trimer group (relative to all trimers; rows 8–13) *in vitro* (DHP experiment) and *in silico*

Trimers are grouped according to the monomer-oligomer reaction that produced them (Table II); for example, all trimers in group 8 are the result of coupling a G-monomer radical at its β position with the 4-O position of a G-end radical of dimers 1, 2, 3, or 7. The *in vitro* abundances are deduced from an HPLC chromatogram of a DHP experiment (Morreel et al., 2004). Using the experimental reaction constants from Table II, these *in vitro* abundances match the resulting *in silico* abundances for dimers quite well. A zero (0) denotes that the structure was not found, and a dash (–) denotes that the structure is not possible (e.g. a G-dimer when there is no influx of G-monomers).

Reaction No.	Product	In Vitro Abundances		In Silico Abundances		
		0% G	100% G	0% G	50% G	100% G
		%		%		
1	G(β - β)G	–	11	–	3.6	10.6
2	G(β -O-4)G	–	16	–	4.0	15.6
3	G(β -5)G	–	73	–	20.3	73.8
4	S(β - β)S	92	–	91.2	13.8	–
5	S(β -O-4)S	8	–	8.8	0.8	–
6	S(β -5)S	–	–	–	57.2	–
7	G(β -O-4)S	–	–	–	0.4	–
8	G(β -O-4)G...	–	100	–	6.9	49.1
9	G(β -5)G...	–	0	–	9.0	50.9
10	S(β -O-4)G...	–	–	–	6.3	–
11	S(β -5)G...	–	–	–	6.1	–
12	G(β -O-4)S...	–	–	–	35.7	–
13	S(β -O-4)S...	100	–	100	35.9	–
Ratio dimer:trimer		?	?	51.4:48.6	41.9:58.1	51.1:48.9

mixture. We also keep track of the β -O-4 fraction of bonds formed in dimerization and oligomerization reactions separately. We define the monomer-oligomer coupling fraction,

$$f_{\text{mo}} = \frac{n_{\text{mm}}}{n_{\text{mo}}}$$

with n_{mm} the number of times a dimerization reaction has occurred in the simulation and n_{mo} the number of times a monomer-oligomer coupling has occurred in the simulation. Finally, we analyze the degradability of the reaction mixture after *in silico* degradation (simulated thioacidolysis or derivatization followed by reductive cleavage) by cleaving all β -O-4 bonds in the reaction mixture and comparing the average size of lignin fragments before and after degradation. We then define the monomer yield as $y_{\text{m}} = n_{\text{M,d}}/n_{\text{u}}$, the ratio between the number of monomers after degradation as $n_{\text{M,d}}$ and the number of subunits in oligomers before degradation as n_{u} .

RESULTS

In Silico Lignin Polymerization under Zulauf and Zutropf Conditions

Figure 4 shows the numbers of monomers and oligomers as a function of time under Zulauf and Zutropf conditions for the experimental parameter set based on DHP data. To simulate Zulauf conditions (Fig. 4A), 50,000 monomers (25,000 of each type) were

added to the reaction mixture over a period of 0.1 s. After the formation of dimers and longer oligomers had depleted the monomers, the simulation was stopped. To simulate the Zutropf conditions shown in Figure 4B, the same number of 50,000 monomers was added much more slowly, over a period of 1,000 s. After an initial increase of the number of monomers, the influx of monomers almost equals the consumption by monomer-monomer coupling and monomer-oligomer coupling. The level of monomers continues to decrease slowly as the total level of oligomers rises and more monomers are consumed. The number of oligomers also reaches an almost steady level, as they are both produced and consumed in oligomerization reactions.

Longer oligomers emerge as the reaction proceeds. The total number of oligomers rises and surpasses the number of monomers. Thus, the relative level of monomer-oligomer reactions rises at the expense of dimerization reactions. Dimerization and oligomerization reactions rapidly deplete the monomers after the influx of monomers has been stopped.

Slow Monomer Influx Favors Monomer-Oligomer Coupling and Increases the β -O-4 Fraction

To determine the effect of the influx rate on lignin structure and degradability, simulations were run for seven different influx rates (Fig. 5): the influx rate was gradually decreased by a factor of 10, so that the influx time was increased from 0.1 s (Zulauf) to 100,000 s

Table II. Reaction constants for dimerization reactions (1–7) and monomer-oligomer reactions (8–13) used in our simulations (arbitrary units)

The bond probabilities in dimerizations (i.e. the probability that a certain bond is formed after collision) were estimated by hand fitting the abundances of dimers in a Zulauf simulation experiment (influx time of 0.1 s) to dimer abundances in a Zulauf DHP experiment (Morreel et al., 2004), for G-fractions $[G/(G + S)]$ 0% and 100%. The probabilities for binding and the bond probabilities for monomer-oligomer reactions are unbiased. A separate set of completely unbiased reaction constants (both probability of binding and choice of bond) was used for comparison, as shown in the last column.

Reaction No.	Reaction	Possible Products	Experimental Reaction Constants ($\times 10^{-6}$)	Unbiased Reaction Constants ($\times 10^{-6}$)
1	G(4,5,B) + G(4,5,B) →	G(4,5)-[BB]-G(4,5)	0.20	0.33
2		or G(4,5)-[B4]-G	0.14	0.33
3		or G(4,5)-[B5]-G	0.66	0.33
4	S(4,B) + S(4,B) →	S(4)-[BB]-S(4)	0.96	0.50
5		or S(4)-[B4]-S	0.04	0.50
6	S(4,B) + G(4,5,B) →	S(4)-[B5]-G	0.99	0.50
7		or G(4,5)-[B4]-S	0.01	0.50
8	G(4,5,B) + G(4,5)-... →	G(4,5)-[B4]-G-...	0.50	0.50
9		or G(4,5)-[B5]-G-...	0.50	0.50
10	S(4,B) + G(4,5)-... →	S(4)-[B4]-G-...	0.50	0.50
11		or S(4)-[B5]-G-...	0.50	0.50
12	G(4,5,B) + S(4)-... →	G(4,5)-[B4]-S-...	1.00	1.00
13	S(4,B) + S(4)-... →	S(4)-[B4]-S-...	1.00	1.00
8–13	If oligomer has two bindable ends		$\times 2$	$\times 2$

(Zutropf). For each simulation, 25,000 G-monomers and 25,000 S-monomers were added (S-fraction is 50%). We analyzed the lengths before and after degradation, the β -O-4 fraction (fraction of β -O-4 bonds relative to all bonds), and the monomer-oligomer coupling fraction (fraction of monomer-oligomer coupling reactions relative to all reactions).

Figure 5 shows that DHPs formed at low monomer influx rates (Zutropf; i.e. long influx duration near the

right end of the x axis) are longer (Fig. 5A), have more β -O-4 bonds (Fig. 5B; total β -O-4 fraction), and have a higher monomer yield after degradation compared with those formed at fast monomer influx rates (Zulauf; i.e. short influx duration near the left end of the x axis). Despite the increased β -O-4 fraction at Zutropf conditions, the influx rate hardly affects the average length of degradation fragments (Fig. 5A). This is because the frequency of β -O-4 bonds in Zulauf

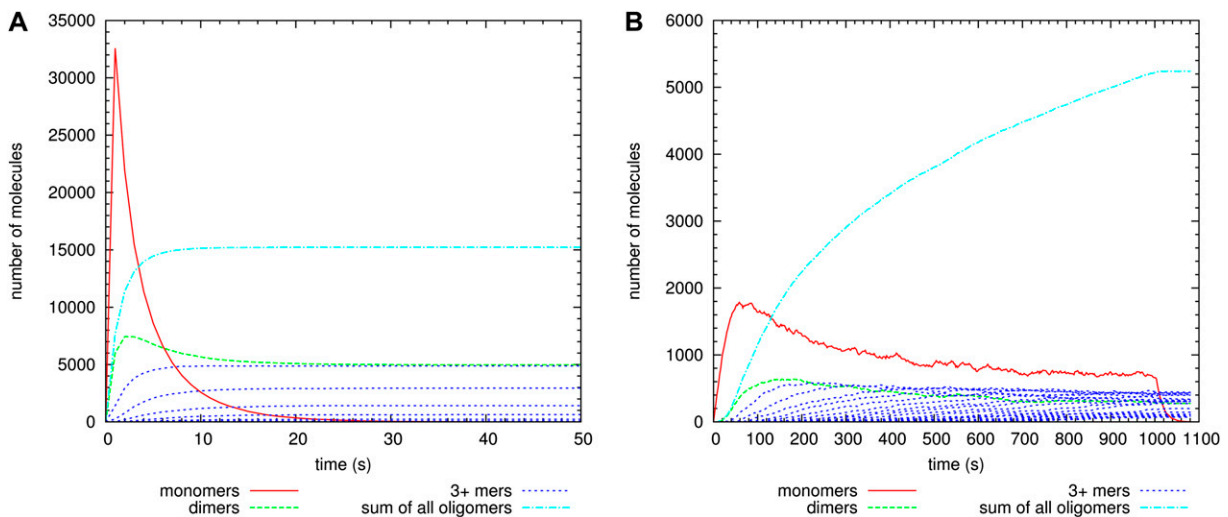


Figure 4. The number of monomers, dimers, trimers, and longer oligomers, as a function of time, and the sum of all oligomers. We stop the simulation when all monomers have reacted. A total of 25,000 G-monomers and 25,000 S-monomers were added in 0.1 s (Zulauf; A) and in 1,000 s (Zutropf; B).

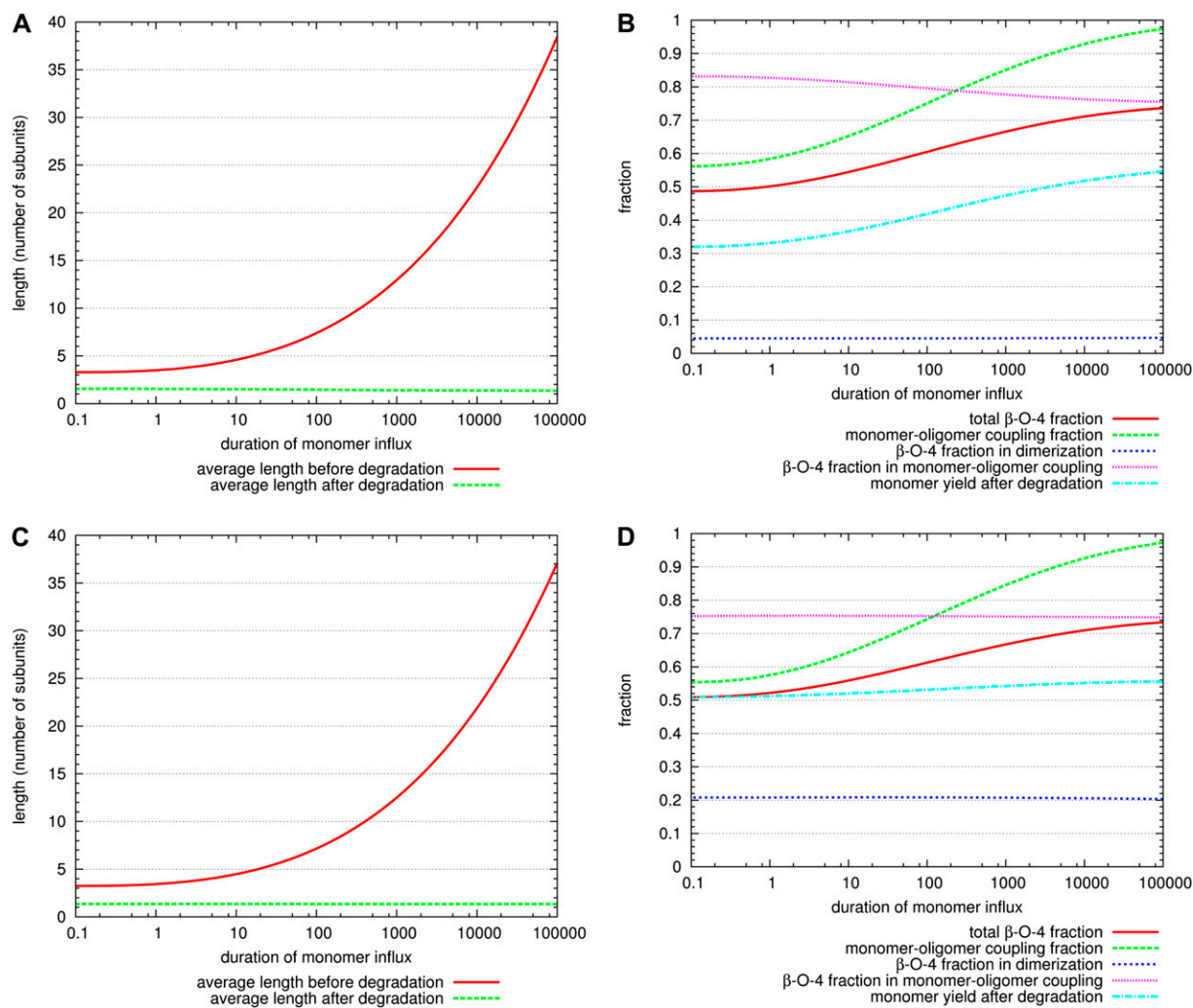


Figure 5. Varying the duration of monomer influx. Influx of 50,000 monomers (25,000 G and 25,000 S) over seven time intervals from 0.1 s (Zulauf) to 100,000 s (Zutropf), using experimental (A and B) and unbiased (C and D) reaction constants. For each data point, the average of 10 simulations was taken (70 simulations in total). The SD values (data not shown) are close to line width. β -O-4 fraction indicates the number of β -O-4 bonds/the number of all bonds; monomer-oligomer coupling fraction indicates the number of times a monomer-oligomer reaction occurred/the number of times any reaction occurred; monomer yield indicates the number of monomers after degradation/the total number of units in oligomers before degradation.

lignins is already rather high (Fig. 5B), so Zutropf reaction conditions cannot reduce the average length of degradation fragments much further. Both under Zulauf and Zutropf conditions, monomer-monomer couplings are initially more frequent than monomer-oligomer reactions, because only a few oligomers are available. However, under Zutropf conditions, oligomers become predominant after an initial buildup of dimers. Thus, under Zulauf conditions, dimerization reactions are most frequent, whereas under Zutropf conditions, monomer-oligomer couplings are more frequent. Thus, lignin chains will be longer on average under Zutropf conditions (Fig. 5A).

Figure 5B shows that the total β -O-4 fraction (full red line) increases with a decreasing influx rate. As the frequency of β -O-4 bonds generated in monomer-oligomer reactions (dotted magenta line) is higher

than in dimerization reactions (spaced dotted blue line), the increase in monomer-oligomer coupling reactions (dashed green line) will also result in an increase of the total β -O-4 fraction (full red line). After *in silico* cleavage of these β -O-4 bonds, the monomer yield will be higher (dashed dotted cyan line) and even seems to be proportional to the total β -O-4 fraction.

In Figure 5B, we also observe that the β -O-4 fraction in monomer-oligomer couplings (dotted magenta line) is slightly higher under Zulauf than under Zutropf conditions (S-fraction is 50%). Subunits binding to an S-ended oligomer always form a β -O-4 bond, while subunits binding to a G-end only form a β -O-4 bond with 50% probability. Thus, under Zulauf conditions, the number of bindable S-ends is apparently higher than under Zutropf conditions. This is due to the

relatively higher number of dimers, which have a higher fraction of bindable S-ends (approximately 72%; Table I) compared with longer oligomers (approximately 48%; Table I).

Higher S-Fraction Increases the β -O-4 Fraction

Next, we varied the relative abundance of G- and S-units to study the effect of monolignol composition on lignin structure and degradability. Here, the S-fraction is defined as the number of S-monomers relative to the total number of monomers [S-fraction = $S/(G + S)$]. Zulauf simulations were run for S-fractions ranging from 0% to 100%, with a 10% interval (Fig. 6), and the lengths of the oligomers before and after degradation, the fraction of β -O-4 bonds, and the monomer-oligomer coupling fraction were determined.

Figure 6 shows that a higher S-fraction results in longer lignin polymers that are easier to degrade (higher β -O-4 fraction). At higher S-fractions, more dimers with two bindable ends are formed, as the reaction constant for S(β - β)S is higher than for G(β - β)G (Table II). Thus, at higher S-fractions, more bindable positions are available and more monomer-oligomer coupling reactions will occur (Fig. 6B).

In a Zutropf model, a similar but more pronounced effect of the S-fraction on lignin is shown (Fig. 6, C and D). Under Zutropf conditions, the average length of lignin (before degradation) is higher than under Zulauf conditions (Fig. 6C). We also observe that the cross-coupling fraction and the β -O-4 fraction are higher under Zutropf conditions (Fig. 6D).

Parameter Sensitivity

Next, we tested to what extent the effects of influx rate and S-fraction depend on the specific values of the reaction constants. We reran our simulations with probabilities for each possible bond set to equal values (Table II). Figure 5 (C and D) shows the effects of the influx rate, and Figure 6 (E and F) shows the effects of the S-fraction for such unbiased bond probabilities. Note that our original parameter settings already assumed unbiased bond probabilities for monomer-oligomer reactions, so the changes can only be attributed to the effect of dimerizations. Although the details differ, the main results hold for such unbiased reaction constants: a slower monomer influx generates lignins with a higher β -O-4 fraction, and lignins synthesized with a high fraction of S-units have more β -O-4 bonds. These results are thus intrinsic to the coupling rules and are only slightly altered by using the experimental set of reaction constants. After all, the fact that one bond cannot be formed (no β -5 bond for S-units and no β - β bond in monomer-oligomer reactions) means the other bonds (β -O-4 in particular) will be more frequent.

Interestingly, the monomer yield following β -ether cleavage from lignins produced under Zulauf condi-

tions is much higher for unbiased parameter settings than for experimental parameter settings (Fig. 5, compare B and D), although the total β -O-4 fractions are nearly identical. The reason for this is that a β -O-4 bond does not always release the same number of monomers, because this depends on the position of this bond in an oligomer. Under unbiased parameter settings, β -O-4 bonds are more likely in dimers, and β -O-4 dimers release two monomers after degradation, whereas an internal unit of an oligomer is only released when it is surrounded by two β -O-4 bonds. Thus, monomer yield is not the perfect measure for the β -O-4 fraction.

DISCUSSION

We have presented a simulation model of in vitro lignin polymerization. The simulation algorithm iteratively chooses a random reaction from a finite set, with each reaction having a certain probability of being selected. The reactions are applied to a reaction mixture, which is symbolized by a list of all monomer and oligomer sequences.

In this model, the β -O-4 fraction can be increased by decreasing the rate of monomer influx. Adding the same number of monomers over a longer period of time gives oligomers the chance to grow longer, even though reactions proceed more slowly because fewer molecules are available in solution. This means monomer-oligomer coupling occurs at a higher rate relative to dimerization, which will increase the β -O-4 fraction of lignin, as monomer-oligomer coupling results predominantly in β -O-4 bonds. This explains why natural lignin and Zutropf DHP lignins are richer in β -O-4 bonds than Zulauf DHP lignin (Hatfield and Vermerris, 2001; Ralph et al., 2008).

Increasing the fraction of S-units also results in a higher β -O-4 fraction and a higher monomer yield after degradation. β -O-4 bonds are relatively more likely with S-units even when all bonds are equally probable, because S-units have one fewer bindable position compared with G-units (C_5 is methoxylated). This result is not completely in accordance with DHP experiments (Kishimoto et al., 2010), which find that the monomer yield is higher in mixed S- and G-monomer experiments than in 100% S-monomer experiments. This discrepancy between our model and the observations of Kishimoto et al. (2010) may be due to our assumption that the monolignol-oligolignol coupling reaction constants were unbiased. If we assumed that S-units have lower binding probability than G-units to bindable S-ends of oligomers (i.e. by setting the reaction constant of reaction 13 to 0.1), our model matches the Kishimoto et al. (2010) observations better (Fig. 7B). For this assumption, the monomer-oligomer coupling fraction decreases at higher S-fractions, because relatively more bonds come from monomer-monomer couplings. So, not only bond probabilities but also binding probabilities affect lignin composition; future work will determine precise

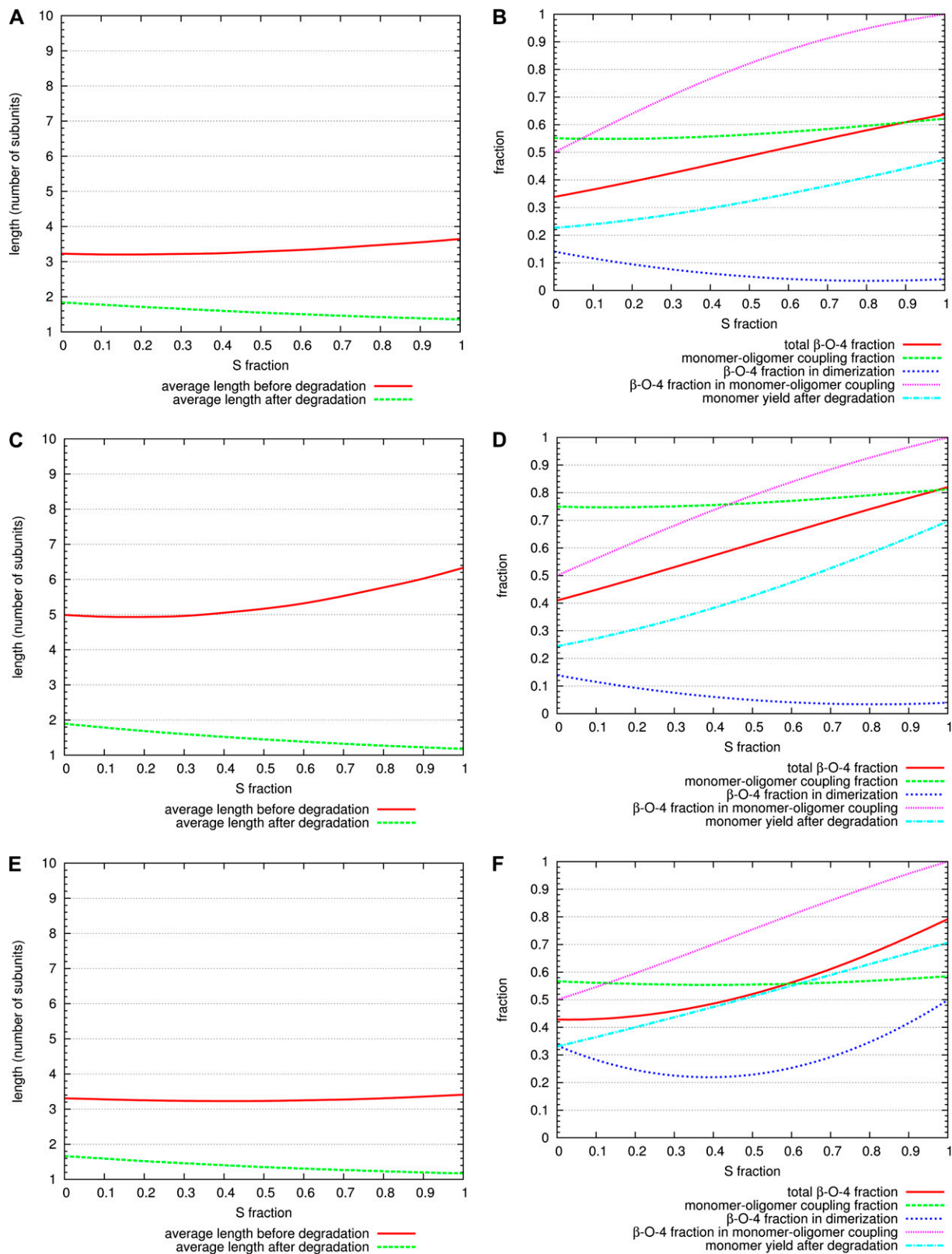


Figure 6. Effect of S-fractions in simulations of a Zulauf (A, B, E, and F) and a Zutropf (C and D) experiment, using experimental (A–D) and unbiased (E and F) reaction constants. In Zulauf conditions, 50,000 monomers are introduced over a time interval of 0.1 s; for Zutropf conditions, the same number of monomers is introduced over 1,000 s. For all eleven S-fractions simulated, the average over 10 simulations was plotted (110 simulations in total). The sd values (data not shown) are close to line width. S-fraction indicates $S/(S + G)$.

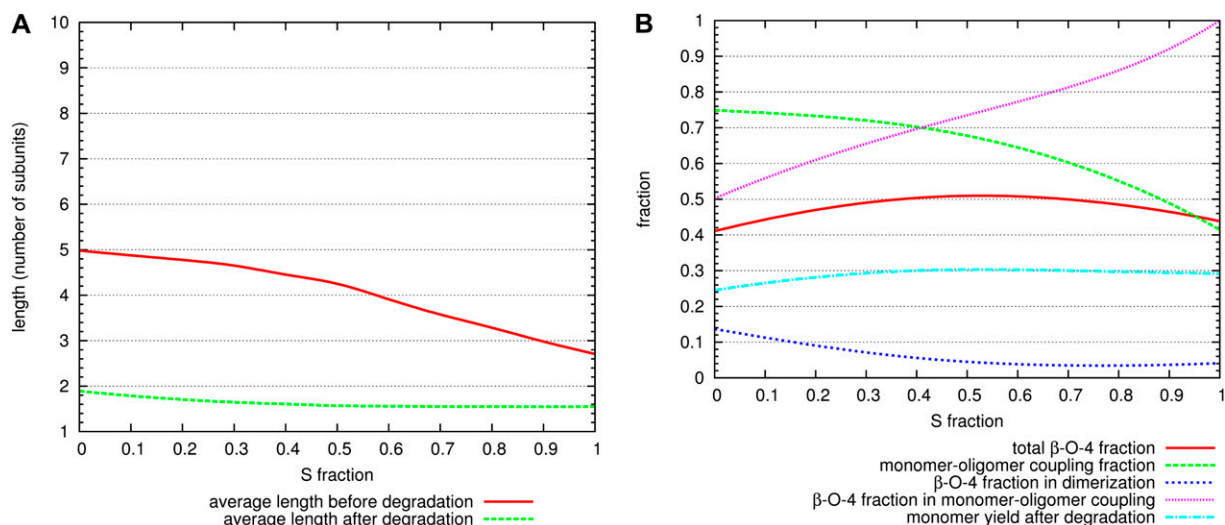


Figure 7. Effect of S-fractions in a simulated Zulauf experiment (influx time of 1,000 s). The reaction constant of reaction 13 (Table II) for coupling between an S-monomer and an S-end of an oligomer $\{S(4,B) + S(4)-\dots \rightarrow S(4)-[B4]-S-\dots\}$ was reduced from 1×10^{-6} to 0.1×10^{-6} . All other reaction constants were as in the experimental parameter set from Table II.

measurements of the binding probabilities between monomer subunits and oligomers. Another observation is that the average lignin chain length increases with the S-fraction (Fig. 6). However, lignin in transgenic plants with a high S content has a lower M_r , as deduced from two-dimensional NMR-based bond frequencies (Stewart et al., 2009). This discrepancy may also be due to our assumption that G- and S-monomers have equal binding affinity to oligomers; after reducing the oligomer-monomer binding probability for S-units, the oligolignols in our simulations shorten at higher S-fractions (Fig. 7A). Our future work will test this model prediction experimentally.

In experiments, the monomer yield is often used as a measure of β -O-4 bond frequency. However, in experiments where the relative number of dimerizations is changed along with the β -O-4 fraction, caution is needed. β -Ether dimers will always release two monomers after cleaving a β -O-4 bond, whereas longer oligomers will release only one; in fact, a monomer can only be released from an internal unit by cleaving the two β -O-4 bonds that may surround it. Thus, if the number of dimerizations is increased (e.g. when the influx rate is increased), the monomer yield will overestimate the β -O-4 fraction.

Unfortunately, little is known about the exact in vitro DHP reaction conditions, in particular for oligomer-oligomer coupling and branching reactions. These were not considered in our model, and neither were some other bond-forming reactions (e.g. the β -1 bond; Ralph et al., 2004). Better estimates need to be made for the binding probabilities by extensive DHP experiments with varying supply rates, varying unit ratios, etc. Nevertheless, the general trends of the reaction constants are considered to be correct. For example, if the β -O-4 fraction in dimerization would

have been higher than in monomer-oligomer reactions, the total β -O-4 fraction would decrease at slower monomer influx, which is contrary to in vitro evidence (Hatfield and Vermerris, 2001).

Another issue is whether the reaction constants of monomer-oligomer coupling reactions are independent of oligomer length, which we implicitly assume by lumping oligomers with identical bindable ends. Experiments have shown that, during polymerization of various monomeric compounds, the reaction rate constants hardly change for oligomers of over four units (Flory, 1953); we are unsure if this observation holds for lignin polymerization. The exclusion of dehydrogenation reactions as a parameter is justified by presuming that the oxidizing agents are available in excess, so that radicalization is not the rate-determining step. However, in vivo, oxidase enzymes (peroxidases and laccases) are not necessarily available in excess and could affect lignin composition. Also note that the oxidizing agent used to make DHPs has a different reactivity compared with the in vivo oxidases. For example, HRP oxidizes S-units less efficiently than G-units (Kishimoto et al., 2010). Our data are based on experiments with copper acetate, which mimics the in vivo oxidation better than HRP (Landucci, 2000).

Our model only considers the in vitro situation (i.e. a well-mixed reaction vessel without any other molecules that could interfere). One hypothesis states that, in vivo, lignification is not a combinatorial process but is tightly orchestrated by proteins harboring arrays of dirigent radical-binding sites (Davin and Lewis, 2005) and that lignification might be a template replication process (Chen and Sarkanen, 2010); a discussion of the likely merits of these hypotheses has been published (Ralph et al., 2008). In our model, lignification is a purely combinatorial process, and it shows that reac-

tion constants will affect these bond frequencies. The in vitro-in vivo differences could thus be the result of different reaction conditions that will affect these reaction constants (e.g. a different pH or the use of a different oxidant). Moreover, the presence of polysaccharides appears to increase the M_n of lignin, favoring the formation of β -O-4 linkages (Houtman and Atalla, 1995).

Several previous studies have modeled lignin polymerization (Glasser and Glasser, 1974; Glasser et al., 1976; Lange et al., 1987; Jurasek 1995, 1998; Roussel and Lim, 1995). These studies predict lignin bond frequencies in reasonable agreement with experimental data but do not predict how they depend on reaction conditions. The SIMREL model (Glasser and Glasser, 1974; Glasser et al., 1976) searches for a reactive position in a polymer and couples a monomer to it, unless the bond distribution in the polymer no longer matches the bond frequencies found in the literature. Thus, this model cannot predict bond frequencies from first principles or reaction conditions. Roussel and Lim (1995) present a two-dimensional cellular automata model of lignification, where lignins are allowed to grow in a two-dimensional space. Monomers are modeled as point particles represented by cellular automata states, and bonds form by joining molecules that diffuse into proximity. The model focuses on the conditions for forming space-filling lignin molecules. Jurasek (1995, 1998) has introduced a model of three-dimensional lignin polymerization, where monomers are represented as ball-and-stick entities and react randomly with existing subunits, selecting a bond type according to probabilities estimated from experimental data. Also interesting is the model by Lange et al. (1987), which simulates the process of lignin degradation by a set of chemical reactions, whereas in our work, β -O-4 bonds are all cleaved at once. Apart from computer models, lignin polymerization has also been approached by quantum chemical and force field methods (Durbeej et al., 2003). These techniques can explain reaction constants, as they investigate electrochemical features of monomers and the energetics of compounds. Molecular modeling techniques have also been used to study the association of lignin and cellulose microfibrils, which exhibit an adsorbing force that may constrain the movement of lignin monomers and oligomers and influence their freedom to participate in random polymerization reactions (Houtman and Atalla, 1995; Besombes and Mazeau, 2005a, 2005b). Thus, such molecular level models are helpful in predicting the details of lignin polymerization and molecular detail and may help to fine-tune models of large-scale lignin polymerization like ours.

Although our model only considers lignin primary structure, the combination of a rule-based, constructive grammar and the Gillespie (1977) model allows us to simulate a well-mixed in vitro reactor and to use quantifiable reaction rates, which enables us to study the effects of a wide range of reaction conditions on

primary structure. The model we used to simulate lignification is still rudimentary; nevertheless, it already offers a better understanding of the process in an in vitro environment. A simple model has the advantage that it makes experimental results easier to explain. Where the results are not in accordance with reality, the explanation has to be sought in different reaction conditions or the simplicity of the model. Thus, our model is a useful tool to predict synthetic lignin structure and should help define strategies to tailor lignin structure in plants to increase degradability. As a future direction, oligomer-oligomer reactions, branching, and radicalization should be included in the model and a better set of reaction constants needs to be established. Future studies will also shed light on the need, or otherwise, for proteins harboring dirigent sites in defining the outcome of lignin polymerization.

ACKNOWLEDGMENTS

Michael Guravage (Centrum Wiskunde en Informatica) is thanked for technical assistance and Milan van Hoek (Centrum Wiskunde en Informatica) for stimulating discussions. Hoon Kim and Fachuang Lu (Department of Biochemistry, University of Wisconsin) are thanked for preparing synthetic lignins upon which the HPLC-mass spectrometry data depended.

Received February 9, 2010; accepted May 13, 2010; published May 14, 2010.

LITERATURE CITED

- Bamford C, Dewar M** (1947) Studies in polymerization. I. A method for determining the velocity constants in polymerization reactions and its application to styrene. *Proc R Soc Lond B Biol Sci* **192**: 309–328
- Baucher M, Halpin C, Petit-Conil M, Boerjan W** (2003) Lignin: genetic engineering and impact on pulping. *Crit Rev Biochem Mol Biol* **38**: 305–350
- Besombes S, Mazeau K** (2005a) The cellulose/lignin assembly assessed by molecular modeling. Part 1. Adsorption of a threo guaiacyl β -O-4 dimer onto a I β cellulose whisker. *Plant Physiol Biochem* **43**: 299–308
- Besombes S, Mazeau K** (2005b) The cellulose/lignin assembly assessed by molecular modeling. Part 2. Seeking for evidence of organization of lignin molecules at the interface with cellulose. *Plant Physiol Biochem* **43**: 277–286
- Boerjan W, Ralph J, Baucher M** (2003) Lignin biosynthesis. *Annu Rev Plant Biol* **54**: 519–546
- Chen F, Dixon RA** (2007) Lignin modification improves fermentable sugar yields for biofuel production. *Nat Biotechnol* **25**: 759–761
- Chen YR, Sarkanen S** (2010) Macromolecular replication during lignin biosynthesis. *Phytochemistry* **71**: 453–462
- Custers R** (2009) First GM trial in Belgium since 2002. *Nat Biotechnol* **27**: 506
- Davin L, Lewis N** (2005) Lignin primary structures and dirigent sites. *Curr Opin Biotechnol* **16**: 407–415
- Durbeej B, Wang YN, Eriksson LA** (2003) Lignin biosynthesis and degradation: a major challenge for computational chemistry. *Lect Notes Comput Sci* **2565**: 137–165
- Feret J, Danos V, Krivine J, Harmer R, Fontana W** (2009) Internal coarse-graining of molecular systems. *Proc Natl Acad Sci USA* **106**: 6453–6458
- Flory PJ** (1953) Molecular size and chemical reactivity: principles of condensation polymerization. In PJ Flory, ed, *Principles of Polymer Chemistry*. Cornell University Press, New York, pp 69–105
- Fontana W** (1992) Algorithmic chemistry. In CG Langton, C Taylor, JD Farmer, S Rasmussen, eds, *Artificial Life*, Vol 2. Addison Wesley, Redwood City, CA, pp 159–210
- Freudenberg K** (1956) Beiträge zur Erforschung des Lignins. *Angew Chem* **68**: 508–512

- Gillespie D (1977) Exact stochastic simulation of coupled chemical reactions. *J Phys Chem* **81**: 2340–2361
- Glasser WG, Glasser HR (1974) Simulation of reactions with lignin by computer (SIMREL). 1. Polymerization of coniferyl alcohol monomers. *Macromolecules* **7**: 17–27
- Glasser WG, Glasser HR, Nimz HH (1976) Simulation of reactions with lignin by computer (SIMREL). 5. Nondehydrogenative polymerization in lignin formation. *Macromolecules* **9**: 866–867
- Grabber JH, Hatfield RD, Ralph J (2003) Apoplastic pH and monolignol addition rate effects on lignin formation and cell wall degradability in maize. *J Agric Food Chem* **51**: 4984–4989
- Hatfield R, Vermerris W (2001) Lignin formation in plants, the dilemma of linkage specificity. *Plant Physiol* **126**: 1351–1357
- Hill J, Nelson E, Tilman D, Polasky S, Tiffany D (2006) Environmental, economic, and energetic costs and benefits of biodiesel and ethanol biofuels. *Proc Natl Acad Sci USA* **103**: 11206–11210
- Houtman C, Atalla R (1995) Cellulose-lignin interactions. *Plant Physiol* **107**: 977–984
- Huntley SK, Ellis D, Gilbert M, Chapple C, Mansfield SD (2003) Significant increases in pulping efficiency in C4H-F5H-transformed poplars: improved chemical savings and reduced environmental toxins. *J Agric Food Chem* **51**: 6178–6183
- Jacquet G, Pollet B, Lapierre C, Francesch C, Rolando C, Faix O (1997) Thioacidolysis of enzymatic dehydrogenation polymers from *p*-hydroxyphenyl, guaiacyl, and syringyl precursors. *Holzforchung* **51**: 349–354
- Jurasek L (1995) Toward a three-dimensional model of lignin structure. *J Pulp Paper Sci* **21**: 274–279
- Jurasek L (1998) Molecular modelling of fibre walls. *J Pulp Paper Sci* **24**: 209–212
- Kishimoto T, Chiba W, Saito K, Fukushima K, Uraki Y, Ubukata M (2010) Influence of syringyl to guaiacyl ratio on the structure of natural and synthetic lignins. *J Agric Food Chem* **58**: 895–901
- Landucci LL (2000) Reaction of *p*-hydroxycinnamylalcohols with transition metal salts. 3. Preparation and NMR characterization of improved DHPs. *J Wood Chem Technol* **20**: 243–264
- Lange H, Wagner B, Yan JF, Kaler EW, McCarthy JL (1987) Computer simulation and experimental studies of alkaline delignification and homogeneous hydrolysis to form omega lignins. In *Fourth International Symposium on Wood and Pulp Chemistry*, Vol 1. Eybens: Imprimerie Jacques Poncet Bresson, Paris, pp 239–244
- Lapierre C (1993) Application of new methods for the investigation of lignin structure. In HG Jung, DR Buxton, RD Hatfield, J Ralph, eds, *Forage Cell Wall Structure and Digestibility*. American Society of Agronomy, Madison, WI, pp 133–166
- Leplé JC, Dauwe R, Morreel K, Storme V, Lapierre C, Pollet B, Naumann A, Kang KY, Kim H, Ruel K, et al (2007) Downregulation of cinnamoyl-coenzyme A reductase in poplar: multiple-level phenotyping reveals effects on cell wall polymer metabolism and structure. *Plant Cell* **19**: 3669–3691
- Lu F, Ralph J (1997) Derivatization followed by reductive cleavage (DFRC method), a new method for lignin analysis: protocol for analysis of DFRC monomers. *J Agric Food Chem* **45**: 2590–2592
- Lynd L, Cushman J, Nichols R, Wyman C (1991) Fuel ethanol from cellulosic biomass. *Science* **251**: 1318–1323
- Morreel K, Ralph J, Kim H, Lu F, Goeminne G, Ralph SA, Messens E, Boerjan W (2004) Profiling of oligolignols reveals monolignol coupling conditions in lignifying poplar xylem. *Plant Physiol* **136**: 3537–3549
- Pilate G, Guiney E, Holt K, Petit-Conil M, Lapierre C, Leplé JC, Pollet B, Mila I, Webster EA, Marstorp HG, et al (2002) Field and pulping performances of transgenic trees with altered lignification. *Nat Biotechnol* **20**: 607–612
- Ralph J (2006) What makes a good monolignol substitute? In T Hayashi, ed, *The Science and Lore of the Plant Cell Wall Biosynthesis, Structure and Function*. Universal Publishers (BrownWalker Press), Boca Raton, FL, pp 285–293
- Ralph J, Brunow G, Harris PJ, Dixon RA, Schatz PF, Boerjan W (2008) Lignification: are lignins biosynthesized via simple combinatorial chemistry or via proteinaceous control and template replication? In F Daayf, A El Hadrami, L Adam, GM Ballance, eds, *Recent Advances in Polyphenol Research*, Vol 1. Wiley-Blackwell Publishing, Oxford, pp 36–66
- Ralph J, Lundquist K, Brunow G, Lu F, Kim H, Schatz P, Marita J, Hatfield R, Ralph S, Christensen J, et al (2004) Lignins: natural polymers from oxidative coupling of 4-hydroxyphenylpropanoids. *Phytochemistry* **3**: 29–60
- Roussel MR, Lim C (1995) Dynamic model of lignin growing in restricted spaces. *Macromolecules* **28**: 370–376
- Ruel K, Berrio-Sierra J, Derikvand MM, Pollet B, Thévenin J, Lapierre C, Jouanin L, Joseleau JP (2009) Impact of CCR1 silencing on the assembly of lignified secondary walls in *Arabidopsis thaliana*. *New Phytol* **184**: 99–113
- Stewart JJ, Akiyama T, Chapple CCS, Ralph J, Mansfield SD (2009) The effects on lignin structure of overexpression of ferulate 5-hydroxylase in hybrid poplar. *Plant Physiol* **150**: 621–635
- Stewart JJ, Kadla JF, Mansfield SD (2006) The influence of lignin chemistry and ultrastructure on the pulping efficiency of clonal aspen (*Populus tremuloides* Michx.). *Holzforchung* **60**: 111–122
- Terashima N, Atalla RH, Ralph SA, Landucci LL, Lapierre C, Monties B (1995) New preparations of lignin polymer models under conditions that approximate cell wall lignification. I. Synthesis of novel lignin polymer models and their structural characterization by ¹³C NMR. *Holzforchung* **49**: 521–527
- Timell TE (1986) Compression wood in gymnosperms. In TE Timell, ed, *Bibliography, Historical Background, Determination, Structure, Chemistry, Topochemistry, Physical Properties, Origin, and Formation of Compression Wood*, Vol 1. Springer-Verlag, Berlin, p 706
- Vanholme R, Morreel K, Ralph J, Boerjan W (2008) Lignin engineering. *Curr Opin Plant Biol* **11**: 278–285