

Mathematical modeling of monolignol biosynthesis in *Populus* xylem

Yun Lee, Eberhard O. Voit*

Integrative Biosystems Institute and The Wallace H. Coulter Department of Biomedical Engineering, Georgia Institute of Technology and Emory University, 313 Ferst Drive, Atlanta, GA 30332, USA

ARTICLE INFO

Article history:

Received 25 May 2010

Received in revised form 4 August 2010

Accepted 5 August 2010

Available online 9 September 2010

Keywords:

Biochemical systems theory

Flux balance analysis

Monolignol biosynthesis

Optimization

Parameter estimation

Populus

ABSTRACT

Recalcitrance of lignocellulosic biomass to sugar release is a central issue in the production of biofuel as an economically viable energy source. Among all contributing factors, variations in lignin content and its syringyl–guaiacyl monomer composition have been directly linked with the yield of fermentable sugars. While recent advances in genomics and metabolite profiling have significantly broadened our understanding of lignin biosynthesis, its regulation at the pathway level is yet poorly understood. During the past decade, computational and mathematical methods of systems biology have become effective tools for deciphering the structure and regulation of complex metabolic networks. As increasing amounts of data from various organizational levels are being published, the application of these methods to studying lignin biosynthesis appears to be very beneficial for the future development of genetically engineered crops with reduced recalcitrance. Here, we use techniques from flux balance analysis and nonlinear dynamic modeling to construct a mathematical model of monolignol biosynthesis in *Populus* xylem. Various types of experimental data from the literature are used to identify the statistically most significant parameters and to estimate their values through an ensemble approach. The thus generated ensemble of models yields results that are quantitatively consistent with several transgenic experiments, including two experiments not used in the model construction. Additional model results not only reveal probable substrate saturation at steps leading to the synthesis of sinapyl alcohol, but also suggest that the ratio of syringyl to guaiacyl monomers might not be affected by genetic modulations prior to the reactions involving coniferaldehyde. This latter model prediction is directly supported by data from transgenic experiments. Finally, we demonstrate the applicability of the model in metabolic engineering, where the pathway is to be optimized toward a higher yield of xylose through modification of the relative amounts of the two major monolignols. The results generated by our preliminary model of *in vivo* lignin biosynthesis are encouraging and demonstrate that mathematical modeling is poised to become an effective and predictive complement to traditional biotechnological and transgenic approaches, not just in microorganisms but also in plants.

© 2010 Elsevier Inc. All rights reserved.

1. Background

One of the great obstacles to propelling the cellulosic biofuel industry forward is the natural resistance of plant cell walls to enzymatic and chemical degradation. Overcoming this “recalcitrance” is essential for breaking down cellulose and hemicellulose into fermentable sugars that are readily converted into ethanol, butanol or other biofuels [1]. The plant’s recalcitrance is mainly due to the entanglement of cellulosic microfibrils with other complex cell wall compounds, such as lignin, that inherently serve as the supporting materials of plant structure. While cellulose and lignin play a critical role in providing structural rigidity and defense against pathogen attacks, they are at the same time the major rea-

son for the low efficiency of current methods of biomass conversion, which require the loosening of cellulosic microfibrils before they can be enzymatically catalyzed. This loosening is currently accomplished with an expensive heat and acid pretreatment in the earliest stage of biofuel production [2,3]. Because of the central role of lignin in this process, much attention has been focused on understanding lignin biosynthesis *in situ* and on exploring the potential of introducing transgenic plants with minimal lignin content, which would greatly reduce or even obviate the need for acid pretreatment [4].

Extensive and sustained biochemical and physiological research efforts and, especially, numerous insights from investigations of relevant plant genomes, have shed light on the specific roles of most genes involved in the monolignol biosynthetic pathway, which generates the building blocks of lignin. Importantly, complete genome sequences for two of the prominent model organisms, *Arabidopsis thaliana* [5] and the black cottonwood *Populus*

* Corresponding author.

E-mail addresses: ylee65@mail.gatech.edu (Y. Lee), eberhard.voit@bme.gatech.edu (E.O. Voit).

trichocarpa [6], are available, and many of the relevant genes have been functionally annotated (e.g., [7]). This genome-based information is very valuable but by itself insufficient for explaining or predicting how the monolignol biosynthetic pathway would respond to untested changes in enzyme activities or gene expression, because at least some of the pathway regulation occurs at the metabolic level in a rather complex fashion.

Recently, metabolite (and specifically phenolic) profiling has been used in various transgenic studies to monitor the *in vivo* concentrations of intermediate phenylpropanoid species in the pathway [8,9]. These studies have generated pertinent information that elucidates the lignin monomer biosynthesis from a different perspective and augments the genomic information from earlier studies in a beneficial fashion. Nevertheless, the application of metabolite profiling, for instance, in the characterization of metabolic phenotypes caused by genetic modification [10], is often limited because the levels of some lignin precursors are low and thus difficult to measure.

Concurrent with the advances in genomic and metabolomic analysis, mathematical and computational techniques from systems biology have emerged as an effective tool to help explain the regulation of complex metabolic networks. Examples from yeast demonstrate that sufficient genome annotation, when augmented with biochemical and physiological information, permits the mathematical reconstruction of essentially the entire metabolic network with reasonable fidelity [11,12]. This reconstructed metabolic network can serve as a solid platform from which one may first infer and investigate the metabolic flux distribution and subsequently derive quantitative relationships between genotype, gene expression and phenotype for the pathway of interest.

Two classes of methods are available to achieve these objectives. The first class, including stoichiometric and flux balance analysis, focuses primarily on the connectivity of the metabolic network [13–17]. The methods have been shown to be capable of characterizing possible phenotypes in terms of their steady-state flux distributions based on the pathway stoichiometry and on information regarding the capacities of the involved reaction steps. It has also been shown [18] that stoichiometric models can be converted into linear dynamic models that permit the assessment of small perturbations around the nominal state of the system.

The second class consists of fully kinetic models, which are typically based on differential equations. It includes traditional Michaelis–Menten systems and their generalizations [19] as well as Biochemical Systems Theory [20–24] and lin-log models [25,26] that can be seen as extensions of Metabolic Control Analysis [27–29]. In addition to the connectivity and flux distribution, these models account, at least in principle, for all kinetic, regulatory and dynamic aspects of pathway systems. Fully kinetic models require more extensive amounts of biological information, for instance in the form of metabolite concentrations and enzyme characteristics, or metabolic time series data.

Normally, only one of the two classes is used to model a metabolic network, depending on the questions being asked and information available. Given the limited number of concentration measurements in the monolignol biosynthetic pathway, stoichiometric or flux balance analysis appears to be the model of choice. However, understanding the regulatory mechanisms that are not explicitly taken into account by FBA models constitutes an important step toward applying metabolic engineering techniques to improve biofuel production and seems mandatory before genetic alterations are introduced in natural pathways. Thus, in the spirit of a recent study [30], which proposes a discussion of integrating divergent modeling approaches, we use here a combination of FBA and BST models for analyzing the monolignol biosynthetic pathway at the systems level. This novel combination strategy allows us to harness the regulatory aspects of a kinetic model based

on the metabolic flux distribution obtained from a flux balance model.

Key features of the new strategy are outlined in the following. First, we begin with a minimal amount of experimental information and construct a stoichiometric flux balance model. In the second step, we augment this model using additional biological information, along with various parameter optimization techniques, and morph the static linear model into a dynamic nonlinear model. The ultimate goal of this two-step approach is the establishment of a reliable model that can be used to identify target genes and devise effective strategies for generating modified crops with reduced amounts of lignin. So far, we have not reached the goal of absolute numerical reliability because the currently existing information is still rather scarce. Nonetheless, the resulting model structure appears to be qualitatively adequate and has the capacity to serve as the basis for systematically identifying critical system components (enzymes) whose alterations could improve the yield of fermentable sugars by means of genetic engineering.

2. Methods

2.1. Metabolic mapping

Our main biological target is *Populus* xylem, because a rapidly increasing number of transgenic poplar and aspen varieties within this genus has significantly contributed to our understanding of the enzymes driving the monolignol biosynthetic pathway [31]. Focusing on the metabolic processes occurring in the cytoplasm, we start with the biosynthetic pathway leading to the building blocks of lignin (Fig. 1; also see [Supplementary Method 1](#) for a detailed discussion of how the pathway structure was determined). The pathway generates four alcohols, three of which – *p*-coumaryl, coniferyl, and sinapyl alcohols – are called monolignols. Once synthesized, the monolignols are transported from the cytoplasm to the cell wall, where they are oxidized and polymerized to form lignin. When incorporated into the lignin polymer, these monolignols produce, respectively, *p*-hydroxyphenol (H), guaiacyl (G), and syringyl (S) phenylpropanoid units, which are shown at the periphery of the pathway diagram in Fig. 1. The relative amounts of monolignols, which are affected by a variety of factors [32], determine many of the features of the resulting lignin, such as its structure, toughness and chemical recalcitrance. In dicotyledonous angiosperms, including *Populus*, lignin consists primarily of G and S monomers, whereas the amount of H is negligible. The ratios of lignin monomers and the total lignin content have been closely monitored in transgenic *Populus* variants because of their important role in lignin extractability, forage digestibility [33] and, most importantly, sugar release by enzymatic hydrolysis [4].

In addition to the topology of the network of all enzymatic reactions, it was necessary to account for regulatory mechanisms that are known, alleged, or hypothesized for the monolignol biosynthetic pathway. Correspondingly, we equipped the pathway with regulatory features found in the literature, paying special emphasis to *Populus* (Fig. 1; [Table 1](#)). It should be mentioned that several of the enzymes in the monolignol biosynthetic pathway have multiple isoforms with slightly different kinetics and substrate preferences, and the genes coding for these isoforms are differentially expressed during development and under different environmental cues and stresses. [31]. At this point, this degree of complexity could not be accounted for, due to missing quantitative measurements of the different isozymes in *Populus* xylem, and we focused instead on their collective activity in catalyzing each reaction step. At the same time, if one isoform is known to have a dominant effect over other isoforms, such as Pt4CL1 in aspen xylem [39], the corresponding kinetic constants are assumed to be representative (*cf.* [Supplementary Table 2](#)).

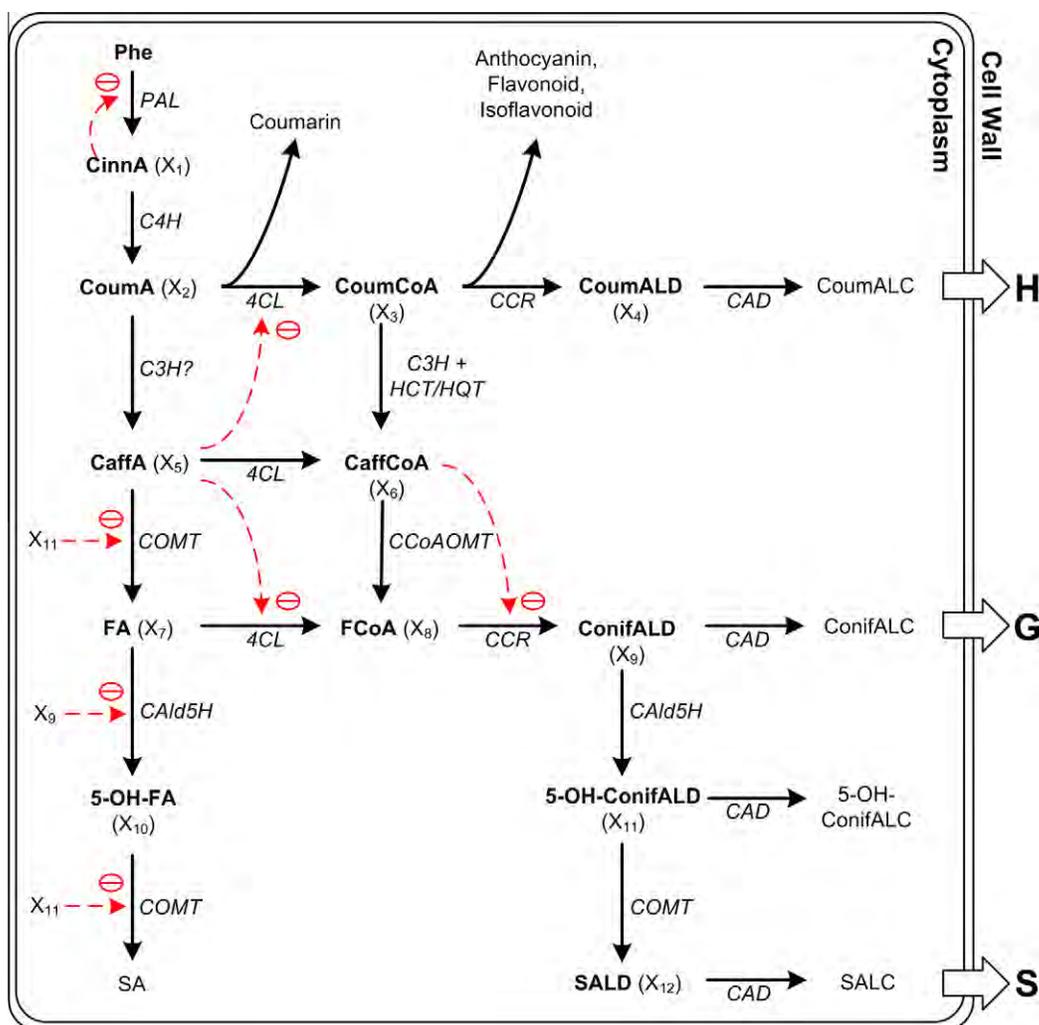


Fig. 1. Generic metabolic map of the monolignol biosynthetic pathway. Metabolites in bold are represented by dependent variables X_i , $i = 1, \dots, 12$, while enzymes are shown in italics. Solid black arrows represent material flow, whereas dashed arrows represent regulatory signals, with negative signs indicating inhibition. Transport processes of monolignols into the cell wall are shown as open arrows. Abbreviations: (1) Metabolites: Phe, phenylalanine; CinnA, cinnamic acid; CoumA, p -coumaric acid; CaffA, caffeic acid; FA, ferulic acid; 5-OH-FA, 5-hydroxyferulic acid; SA, sinapic acid; CoumCoA, p -coumaroyl-CoA; CaffCoA, caffeoyl-CoA; FCoA, feruloyl-CoA; 5-OH-FCoA, 5-hydroxyferuloyl-CoA; SCoA, sinapoyl-CoA; CoumALD, p -coumaraldehyde; ConifALD, coniferaldehyde; 5-OH-ConifALD, 5-hydroxyconiferaldehyde; SALD, sinapaldehyde; CoumALC, p -coumaryl alcohol; ConifALC, coniferyl alcohol; 5-OH-ConifALC, 5-hydroxyconiferyl alcohol; SALC, sinapyl alcohol. (2) Enzymes: PAL, phenylalanine ammonia-lyase; C4H, cinnamate 4-hydroxylase; C3H, p -coumarate 3-hydroxylase; COMT, caffeic acid O -methyltransferase; 4CL, 4-coumarate:CoA ligase; HCT, hydroxycinnamoyl-CoA: shikimate hydroxycinnamoyl transferase; HQT, hydroxycinnamoyl-CoA: quinate hydroxycinnamoyl transferase; CCoAOMT, caffeoyl-CoA O -methyltransferase; CCR, cinnamoyl-CoA reductase; CAD, cinnamyl alcohol dehydrogenase.

Table 1
Documented regulatory signals within the monolignol biosynthetic pathway in *Populus*.

Enzymes	Substrate	Regulator	Kinetics (μM)	Reference
PAL	Phenylalanine	Cinnamic acid	N/A ^c	[72]
4CL	p -coumaric acid Ferulic acid	Caffeic acid ^a	$K_i = 4.37$ $K_i = 4.17$	[39]
CCR	Feruloyl-CoA	Caffeoyl-CoA ^a	$K_i = 15.3$	[73]
COMT	Caffeic acid 5-hydroxyferulic acid	5-hydroxyconiferaldehyde ^a	$K_i = 2.1$ $K_i = 0.26$	[37]
CAld5H	Ferulic acid	Coniferaldehyde ^b	$K_i = 0.59^d$	[35,37]

Abbreviations: PAL, phenylalanine ammonia-lyase; 4CL, 4-coumarate:CoA ligase; CCR, cinnamoyl-CoA reductase; COMT, caffeic acid O -methyltransferase; CAld5H, coniferaldehyde 5-hydroxylase.

^a Competitive inhibitor.

^b Non-competitive inhibitor.

^c No direct evidence has yet been found in *Populus* for this otherwise well-known feedback regulation at the entrance of the pathway.

^d Although this regulation has been experimentally demonstrated in aspen, no quantitative details are known, and the kinetic parameter presented here was measured in the lignifying tissues of sweetgum.

Table 2
Pertinent details of transgenic experiments in Populus.

Enzyme	Enzyme activity (in relation to wild-type) (%)	Lignin composition (S/G; in relation to wild-type) (%)	Species
COMT ^a	32	25	Poplar
CAD ^a	15	100	Poplar
4CL ^b	10	100	Aspen
CAlD5H ^b	280	250	Aspen
CCoAOMT ^c	10 ^d	111	Poplar

The relative proportion of S to G monomers (S/G) was measured by thioacidolysis, which releases the monomers by selectively cleaving the β -O-4 bonds.

^a [41].

^b [42].

^c [43].

^d This particular quantity refers to 10% of wild-type protein amounts.

Since the regulatory signals affect several locations within the pathway, their overall effects are difficult to predict and may even be the cause for counterintuitive observations in transgenic plants. For instance, Fang et al. [40] recently found lignin monomer compositions that cannot be explained solely by the pathway topology in transgenic alfalfa lines with reduced activities of either cinnamate 4-hydroxylase (C4H) or caffeoyl-CoA *O*-methyltransferase (CCoAOMT).

In conclusion, the complexity and multitude of regulatory features that characterize the monolignol biosynthetic pathway render intuitive assessments problematic and highlight the need for mathematical models capable of explaining the functionality of the pathway system.

2.2. Experimental data

The data supporting our modeling effort come in different forms. First, we collected kinetic information and metabolite concentrations from the literature (Supplementary Tables 1 and 2). Secondly, we found pertinent information in five studies of transgenic poplars or aspens, each of which investigated the responses of the pathway to modified protein levels. The investigated proteins were COMT, cinnamyl alcohol dehydrogenase (CAD) [41], 4-coumarate:CoA ligase (4CL), coniferaldehyde 5-hydroxylase (CAlD5H) [42], and CCoAOMT [43] (Table 2). Among these transgenic experiments, three reported an explicit change in the relative proportion of S to G monomers (the so-called S/G ratio), as determined by thioacidolysis. Because lignin content [4] and the S/G ratio [44] are related to the degree of recalcitrance, we will use this ratio as a target indicator of the system's response to genetic manipulations.

Several cautionary notes are in order when we interpret the S/G ratio. First, one should bear in mind that only the fraction of monomers connected by β -O-4 linkages, which accounts for only 20–40% of the lignin by weight, can be extracted by thioacidolysis. Secondly, many of the intervening events, for example, during the transport process or dehydrogenative polymerization, may also contribute to the differences in the observed S/G ratios, but mechanistic details are currently unclear [45]. Third, the composition of lignin monomers is significantly different between two major cell types of xylem tissue, with vessel elements enriched in G monomers and fibers in S monomers [46]. Lastly, genes coding for enzymes like CCoAOMT are expressed in developing vessels but not in fibers, suggesting that different routes to monolignol biosynthesis might be favored in different types of cells¹ [47].

¹ Ideally, a comprehensive analysis of the lignin monomer synthesis in xylem should consist of at least two distinct models, representing the two cell types. The numerical results for any physiological feature of interest, such as the S/G ratio, could then be approximated by combining the two estimates in proportion to their percentage of volume in xylem. While our model could easily be adapted to the two scenarios, currently available data do not allow us to account for such details, and our results therefore reflect averages.

2.3. Mathematical models

We pursued a two-step approach, using complementary methodologies from flux balance and dynamic-kinetic analysis. An overview of the strategy is shown in Fig. 2. First, we converted the pathway (Fig. 1) into a stoichiometric model and used flux balance analysis (FBA) to study phenotypes under different types of constraints [48]. The central concept of FBA is the balanced flux distribution at steady state, which translates mathematically into the equation $\mathbf{N}\cdot\mathbf{v} = \mathbf{0}$, where \mathbf{N} is the stoichiometric matrix of the pathway and \mathbf{v} is a (column) vector of the values of the metabolic fluxes in the network. Normally, there is no unique solution to this problem because one metabolite may be the substrate for more than one reaction, suggesting that the number of fluxes (variables) typically exceeds that of the flux balance constraints at steady state. To determine one best solution, it is customary to apply physico-chemical constraints, as well as an optimization objective like maximal growth. Fast population growth is a reasonable objective for microbial populations, but it is not pertinent here and must be supplanted with different constraints.

Two types of constraints were used here. First, the capacity of each flux v_i must lie within its physiological range $\alpha_i \leq v_i \leq \beta_i$, where we allow $\alpha_i = 0$, and where β_i may be defined as the corresponding steady-state flux in a conventional rate law like the Michaelis–Menten function. Here, all fluxes are assumed to be unbounded (i.e., β_i is defined as $+\infty$), except for the three steps catalyzed by COMT, which are the only reactions for which kinetic constants (K_M and V_{max}) have been characterized for *Populus* protein. While the bounds narrow the range of admissible solutions, they are not stringent enough to identify the optimal solution.

The second constraint is based on the assumption that lignified tissue like xylem has evolved to maximize lignin production in a species- and cell type-specific ratio of monolignols. This assumption is at least partially supported by the observation in poplar xylem that two of the three phenolic glucosides – the storage or detoxification products of hydroxycinnamic acids along the metabolic route to the synthesis of sinapic acid – are barely detectable [43]. This finding suggests that the physiological objective of this pathway is to produce its other end products, namely monolignols. However, this piece of evidence must be handled with caution because the same phenolic glucosides, along with other phenolic compounds like flavonoid and chlorogenic acid, can be abundant in leaves or developing stem tissues [49]. If available, measurements such as the relative amounts of phenolic compounds derived from the pathway are essential for defining the physiological objective within a different context.

The mathematical representation of the physiological objective of monolignol maximization leads to an objective function of the form $\sum_{j \in \Gamma} v_j$, where Γ is the set of fluxes representing the production of the three pertinent monolignols – *p*-coumaryl alcohol, coniferyl alcohol, and sinapyl alcohol. These three fluxes are further constrained by equality constraints such that the corresponding lignin monomer composition reflects the thioacidolysis yields from poplar stem ([50]: Table 3). Mathematically, this modeling approach results in a specific formulation that can be solved with methods of linear programming for which a large number of computational routines exist. In the end, this FBA approach reveals an optimal flux distribution at steady state, and the only inputs needed are the pathway stoichiometry, enzyme capacity measurements, and a predetermined lignin monomer composition.

Various regulatory signals have been identified within the monolignol biosynthetic pathway (Table 1). The mechanisms introduce nonlinearities in the system for which steady-state models like FBA are not sufficient. In the second step of our two-step approach, we therefore use Generalized Mass Action (GMA) models within the framework of Biochemical Systems Theory (BST) to

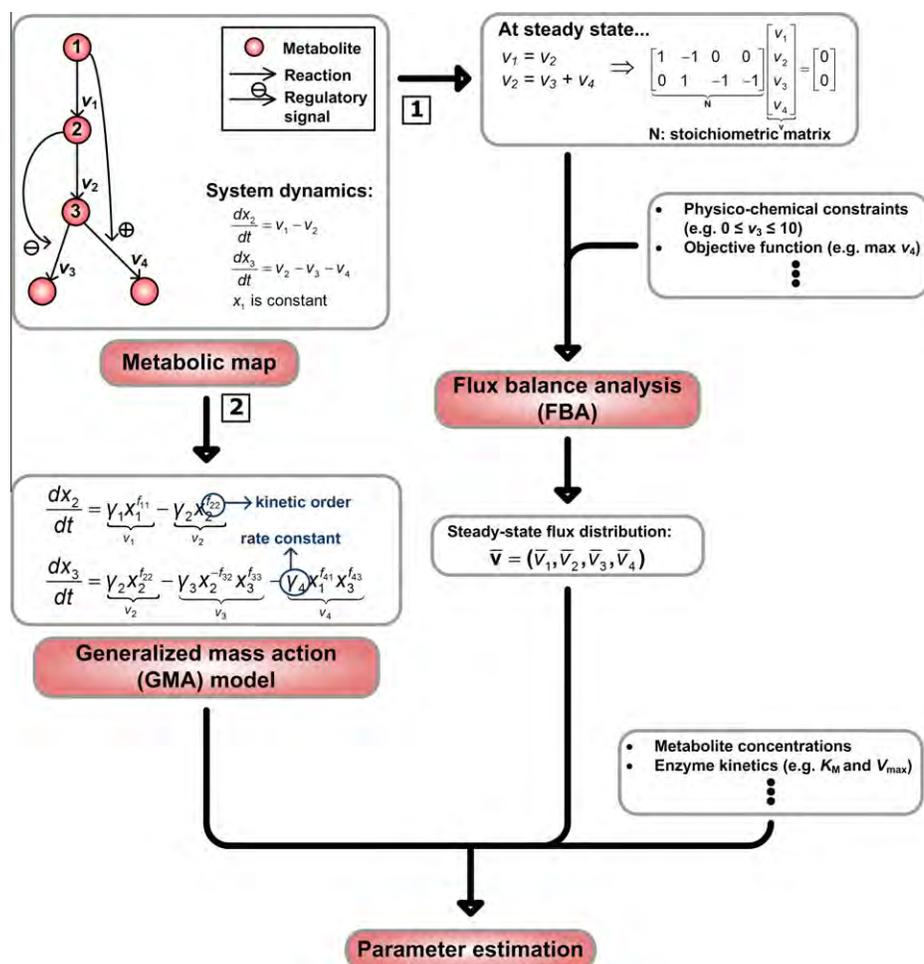


Fig. 2. Overview of the two-step modeling approach. The two-step modeling approach is illustrated here generically by a system with two dependent variables (x_2 and x_3) and one independent variable (x_1). At steady-state, the four fluxes within the system are balanced at two intermediate nodes, resulting in two linear equations with four unknown variables (fluxes). With additional physico-chemical constraints, flux balance analysis (FBA) yields a steady-state flux distribution that satisfies all imposed conditions while optimizing an objective function. Alternately, the same system can be translated into a nonlinear Generalized Mass Action (GMA) model that is characterized by two types of parameters: kinetic orders and rate constants. Collectively, all data, including the steady-state flux distribution, metabolite concentrations, and enzyme kinetic data, are used to estimate the parameters of the ensemble of dynamic models.

Table 3

Minimization of the S/G ratio using the IOM approach (assuming the baseline S/G ratio is 1.8).

No. of enzymes	Modified enzymes ^a	IOM solution ^b (S/G)
1	CAld5H (0.76)	1.3886
2	COMT (0.96) CAld5H (0.71)	1.29
3	C4H (4.31) CAD (1.67) CAld5H (1.34)	1.1133

^a Numbers in parentheses represent the optimized ratio of change in enzyme activities related to the wild-type levels.

^b Average values of the ensemble of model.

account for the documented regulation of the pathway at the metabolic level. Over the past four decades, BST has proven suitable and sufficiently flexible for the analysis of a wide variety of biological phenomena [20–24]. While mostly chosen to model cellular systems in molecular biology, BST has also been applied to plant phenomena such as biomass partitioning in growing trees [51,52] and the so-called 3/2 rule of self-thinning in planted forests [53].

The characteristic feature of BST models is the representation of metabolic fluxes as products of power functions, $\gamma X_1^{f_1} X_2^{f_2} \dots X_n^{f_n}$. Here, each variable X_i describes a metabolite or enzyme in the pathway. The two types of parameters include a non-negative *rate constant* γ that specifies the turn-over rate of the reaction and a set

of exponents f_1, f_2, \dots, f_n , called *kinetic orders*, that characterize the influence of the associated variable on the flux. A real-valued kinetic order f_i is positive (or negative) if an increase in variable X_i is accompanied by an increase (or decrease) in the magnitude of the flux, respectively. The kinetic order of a variable without any effect on the flux is zero, which effectively eliminates the variable from the flux term. If an enzyme-catalyzed reaction had been quantified before as a Michaelis–Menten, Hill, or other similar rate law, it is mathematically easy to convert it into a power-law function [21] (see also [Supplementary Method 2](#)).

2.4. Parameter estimation

For the monolognol biosynthetic pathway, the GMA model consists of 12 dependent variables (X_1, \dots, X_{12} in [Fig. 1](#)) representing the intermediate metabolites involved in the production of monolognols, and one independent variable as the concentration of phenylalanine. As indicated earlier, there are two types of parameters that need to be estimated: kinetic orders f_{ij} and rate constants γ_k . Here, 27 kinetic orders and rate constants are unknown. In general, estimation tasks with such a large number of parameters are computationally intensive and time-consuming. Using the GMA formulation, however, confers two advantages.

First, it is relatively easy to derive parameter values of GMA models, especially for kinetic orders, if information regarding the kinetic features of enzymes and metabolite concentrations is available (*cf.* [Supplementary Method 2](#)). Second, the steady-state flux distribution estimated per FBA helps us circumvent the problem of determining rate constants in the absence of specific flux measurements. As an example, consider a Michaelis–Menten process $V(X) = V_{\max}X/(K_M + X)$ where the maximum rate V_{\max} is unknown. Given a steady-state substrate concentration S and the FBA-predicted steady-state flux V_{FBA} , the rate constant γ for the corresponding power-law term can be represented as:

$$\gamma = V_{FBA}S^{-f},$$

where

$$f = \frac{K_M}{K_M + S},$$

is the kinetic order with respect to the substrate at the steady state. Similar derivations can be applied to conventional rate laws describing competitive or non-competitive inhibition. Details of these types of estimations have been discussed extensively in the literature [21,54] and will not be repeated here.

Once the model is parameterized (in other words, all parameters are assigned values), the first priority is to ensure that no parameter affects the pathway unreasonably strongly. Using sensitivity analysis, we confirmed that the system is indeed robust at the steady state we obtain with FBA (data not shown), indicating that only minor fluctuations in metabolite concentrations and fluxes result from slight changes in parameter values. While a favorable outcome, this robustness is no guarantee that the model is correct. In fact, many parameter values derived from the available data might not be reliable because roughly half of the intermediate metabolites, including the CoA esters, have rather low concentrations *in vivo* (Wout Boerjan, personal communication), and are thus difficult to measure with precision. Computationally, we can explore this uncertainty by systematically changing all parameter values thousands of times and studying how the system responds to such changes. For validation purposes, the observed changes in the S/G ratio from transgenic experiments in poplar or aspen (Table 2) can serve as a quality criterion. To make optimal use of the transgenic experiments for our parameter estimation task, we developed a novel approach consisting of two steps, namely, (1) identification of a subset of significant parameters, and (2) optimization of their values. The steps are summarized in Fig. 3A and discussed in the following.

First, we need an objective criterion to answer the fundamental question of what constitutes a significant parameter. For any transgenic experiment, a parameter is deemed significant if a modest change in its value considerably affects the S/G ratio. To approximate this degree of influence by statistical measures such as Pearson's correlation coefficient or mutual information, we generated a large population of GMA models with different parameter settings, where each parameter (kinetic order) was uniformly sampled from a physiologically realistic range. Given the FBA-derived steady-state flux distribution and the randomly generated values for all kinetic orders, we adjusted each rate constant so that the power-law representation of a flux matched the FBA-derived steady-state value. Typically, the resulting values of kinetic orders are within the range of 0 and 1, if they are associated with substrates, enzymes, and activators, whereas inhibitors are often associated with kinetic orders within the range of -1 and 0 ([21]: Chapter 5). The range of 0 and 1 is also consistent with enzyme-catalyzed reactions following a Michaelis–Menten rate law (Fig. 6A).

With a much reduced number of significant parameters, we gain two important benefits: (1) a reduction – although not total elimina-

tion – of the risk of overfitting; and (2) improved convergence in subsequent parameter optimization tasks, because smaller numbers of parameters are obviously easier to estimate than large numbers. As mentioned earlier, physiological data of the monolignol biosynthetic pathway are available as one-time measurements of the S/G ratio in a number of transgenic experiments. Consequently, our second step – parameter optimization – consists of finding values for those significant parameters that minimize the sum of squared errors (SSE) between the measured and the predicted S/G ratios of all transgenic experiments. Moreover, we characterize an ensemble of GMA models such that all members have comparable training errors in terms of SSE. This notion of finding not just a single best model but an entire class of competent fits is inspired by the argument that inter-individual differences among organisms are reflected in slightly or even moderately different parameter profiles [55]. The search for classes of solutions has also been supported in other scientific domains as diverse as simulations of climate change [56], and models of gene regulatory networks [57] and cell signaling pathways [58].

For readers wishing to explore the modeling approach first-hand, all simulations with our GMA model can be directly performed with the freeware PLAS[®][59]. One simply copies and pastes the system file ([Supplementary Fig. 1](#)) into a blank PLAS file, which is then ready for explorations of the presented or new scenarios. The software permits a variety of dynamic and steady-state analyses of ordinary differential equation models, including sensitivity and gain analysis. Pertinent details about PLAS and the proposed GMA model for the simulation of some transgenic experiment are discussed in [Supplementary information](#) and more generally in [21].

2.5. Pathway optimization

Our monolignol biosynthesis model has the great advantage that it integrates diverse pieces of information from varying experimental conditions. It can be used to address questions like which enzymes should be modified – whether by modulating their expression levels or by improving their turn-over activities through directed evolution [60] – to achieve a higher yield of a desired product. Within the context of biofuel production, genetically engineered crops should of course release significant amounts of fermentable sugars that can be converted into ethanol or other bio-fuel chemicals. In a study on *Populus*, Davison and co-workers [44] indicated that both lignin content and the S/G ratio have significant effects on the yield of xylose, and that a small decrease in S/G ratio alone results in a statistically significant increase in xylose yield. Using our ensemble of GMA models as a framework, we therefore focus on identifying enzymes whose expression levels might allow reductions in the S/G ratio.

GMA models are generally advantageous for modeling the monolignol biosynthetic pathway, but are not trivially optimized with respect to yield because their steady states cannot be computed analytically. This limitation may be overcome with an indirect optimization method (IOM) that permits optimization in an iterative, but much simplified manner [61]. Specifically, IOM allows us to transform the nonlinear problem of minimizing the S/G ratio (or the ratio of fluxes producing coniferyl and sinapyl alcohols), into an iterated linear optimization problem that can be solved with various standard methods, including linear programming. Pertinent details about this approach can be found in [Supplementary Method 6](#). PLAS does not support optimization routines, and we used MATLAB for this particular task.

3. Results

The FBA analysis resulted in an optimal flux distribution within the metabolic pathway system (Fig. 3B) that led to the maximal

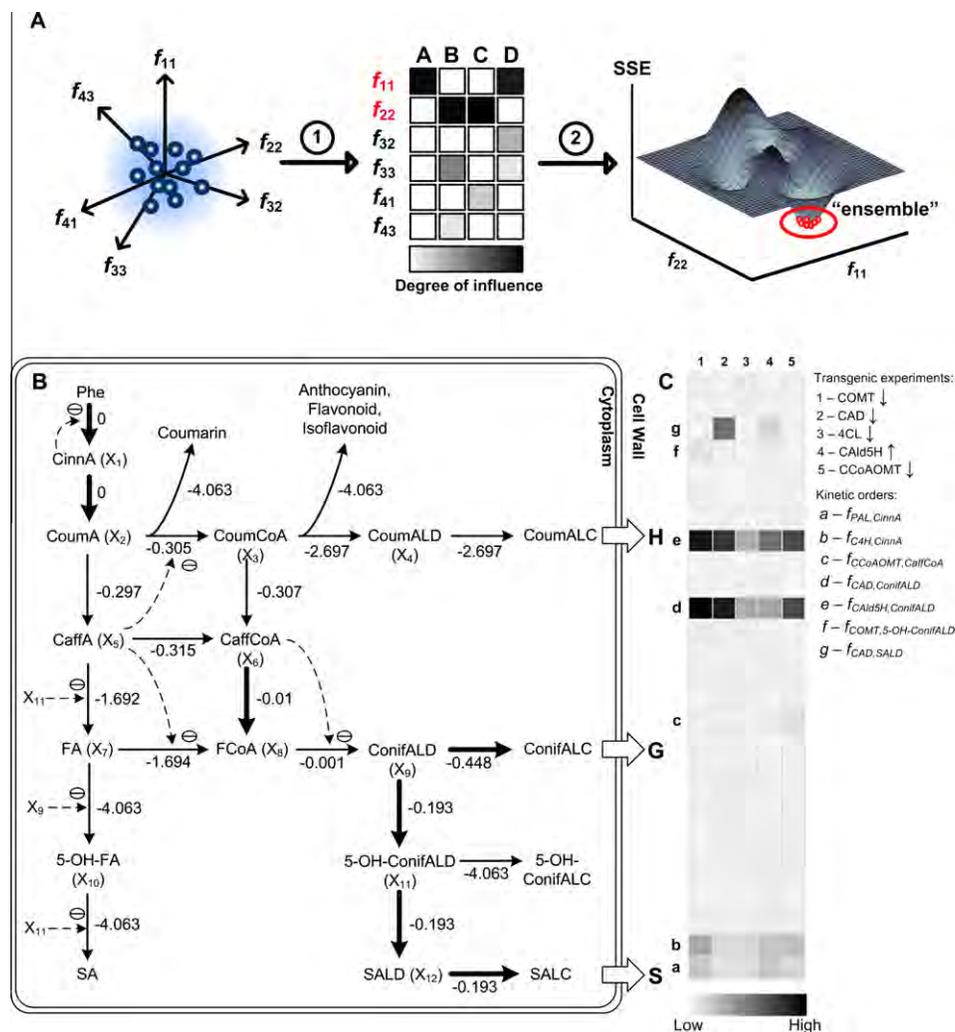


Fig. 3. Steps of parameter estimation. (A) Steps of the parameter estimation process using the system in Fig. 2: (1) constrain each parameter (kinetic order) to a physiologically realistic range and simulate the transgenic experiments in the training set with thousands of randomly sampled parameter profiles; (2) compute a statistical measure (Pearson's correlation coefficient or mutual information) between each parameter and the S/G ratio for all transgenic experiments (A–D) and select statistically significant parameters; (3) values for significant parameters are further optimized to minimize the SSE, and to find an ensemble of models with comparably low SSE. (B) The numerical value next to each reaction represents the magnitude (on a base-10 logarithmic scale) of its steady-state flux, normalized by the input of the pathway, which consists of the reaction converting a constant supply of phenylalanine (Phe) into cinnamic acid (CinnA). (C) The intensity of each box reflects the mutual information (MI) between one parameter and the S/G ratio of one transgenic experiment. Kinetic orders with a statistically significant mutual information score (but not always leading to a large value, such as the kinetic order 'c') are listed on the left of their respective rows. Pathway reactions associated with these significant kinetic orders are represented by heavier arrows.

production of three monolignols in the correct composition. Interestingly, this optimal solution shows that several reactions with relatively high steady-state fluxes dominate the activity of the pathway, whereas other reactions are seemingly inactive. If we connect the dominant fluxes whose steady-state values are within one order of magnitude of the phenylalanine consumption, the resulting route is almost identical to the currently alleged structure of the monolignol biosynthetic pathway in angiosperms [37]. Thus, the purely computational result from the FBA analysis reinforces the point that metabolic pathways are seldom fully connected and indeed use sparse connectivity to bring about specific function. This phenomenon has been widely discussed for microbial metabolic networks [62,63], but our results seem to indicate that the same may be true in plant secondary metabolism as well.

Next, we used the optimal steady-state flux distribution from FBA to construct a dynamic GMA model of the pathway. Converting the metabolic map (Fig. 1) into a symbolic model in GMA format does not take much effort; in fact, this can be done automatically with customized software (e.g., [64]). The much more difficult step, how-

ever, is the numerical identification of parameter values, which is outlined in Fig. 3A and discussed in detail below.

First, by adapting a grid search method used by Alves and collaborators [65], we uniformly sampled every parameter (kinetic order) from a predetermined range of values and generated thousands of GMA models with the same FBA-derived steady-state flux distribution. For each instantiation, we checked local stability (Supplementary Method 4) and discarded unstable parameter profiles. Next, we computed the mutual information (Supplementary Method 5) of each parameter and the output feature of interest, namely the S/G ratio, to evaluate the relative significance of individual parameters (Fig. 3C). Not surprisingly, most parameters are not statistically significant, indicating that only a few parameters have an appreciable influence on the S/G ratio in each transgenic experiment.

Notably, two parameters representing the direct influence of coniferaldehyde on its own consumption, $f_{CAD,ConifALD}$ and $f_{CALd5H,ConifALD}$, are statistically significant in all five transgenic experiments. Although the identification of significant parameters

in our strategy is more or less “biologically blind,” this result can easily be interpreted in terms of the logic of the pathway topology: as shown by FBA and also by thioacidolysis yield, the flux leading to the synthesis of 5-hydroxyconiferyl alcohol is negligible, which means that the formation of 5-hydroxyconiferinaldehyde or coniferyl alcohol from coniferinaldehyde is arguably the principal branch point where the S/G ratio is determined.

In the second half of the parameter estimation process, we generated an ensemble of GMA models that reproduced a training set of experimental results, using a simulated annealing (SA) algorithm (Supplementary Method 7) to find optimal values for the significant parameters. For the five transgenic experiments used as training data (Table 2), the S/G ratios predicted by the ensemble of models are highly consistent with the experimental measurements (Fig. 4). The relative errors in two experiments, where either COMT or CCoAOMT is down-regulated, are slightly greater than the corresponding experimental errors (~3%). Considering that only a handful of transgenic experiments are available for training the models, this level of variance is better than one might have expected.

To assess the reliability of the computed ensemble of models, we used the ensemble to simulate two transgenic experiments not used for training. Specifically, one of the experiments studied a multi-gene co-transformation where the 4CL enzyme activity was reduced by 80% and the CALD5H enzyme activity increased by 2.1-fold [42]. As shown in Fig. 5, the predicted S/G ratio follows the same upward trend and even falls within ~20% of the observed value. In the second transgenic experiment, the CCR transcript levels were severely decreased to < 5% of the wild-type levels [8]. Again, the observed S/G ratio was predicted accurately by the ensemble of models.

Beyond its good agreement with the experimental results, the ensemble of GMA models permits further mechanistic insights. For instance, most of the significant parameters with positive values (which are thus associated with substrates or activators) have optimal values between 0.4 and 0.7, a typical range for kinetic orders estimated from Michaelis–Menten reactions operating close to the K_M (Fig. 6A and B; see also [21]: Chapter 5). By contrast, both $f_{COMT,5-OH-ConifALD}$ and $f_{CAD,SALD}$ take on very small values within the ensemble of models, which according to the theory behind GMA models suggests that both the O-methylation of 5-hydroxyconiferinaldehyde and the reduction of sinapaldehyde to sinapyl alcohol operate at an essentially constant rate that is almost independent of fluctuations in their substrate concentrations. Although there has not yet been direct evidence for this predicted operation close

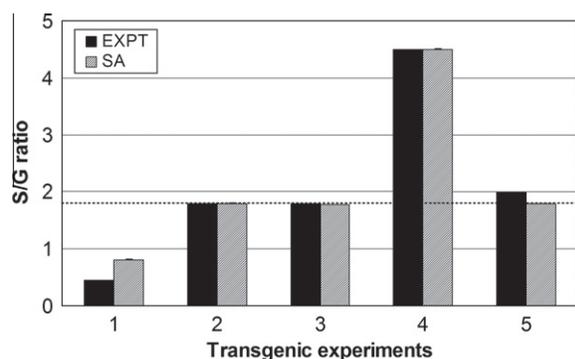


Fig. 4. Simulation results of five transgenic experiments used as training data. Each vertical bar represents either the experimentally observed S/G ratio (EXPT) or the mean of 20 predictions from the ensemble of GMA models fitted by a simulated annealing (SA) algorithm. The transgenic experiments are numbered as in Fig. 3C, with the dashed line featuring the assumed wild-type value (1.8) and the error bar indicating the 95% confidence interval for the mean.

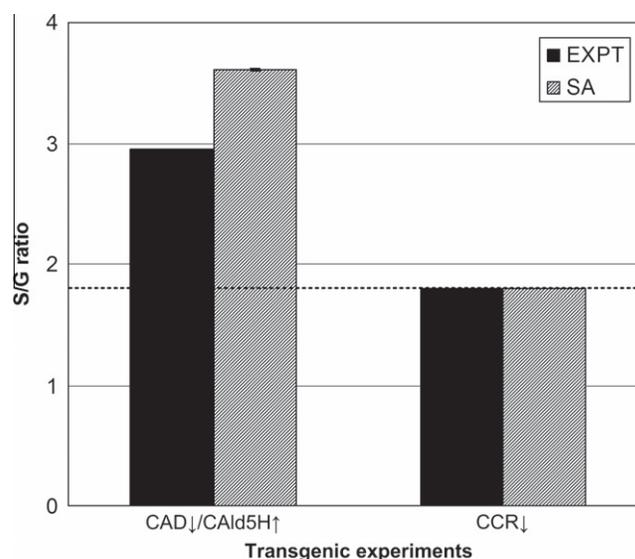


Fig. 5. Simulation results of two novel transgenic experiments. As in Fig. 4, each vertical bar represents either the experimentally observed S/G ratio (EXPT), or the mean of 20 predictions from the ensemble of GMA models fitted by SA. Also, the dashed line features the assumed wide-type value (1.8), and the error bar indicates the 95% confidence interval for the mean. For the CCR down-regulation experiment, the confidence interval is so small ($\sim 10^{-5}$) that it is nearly invisible.

to saturation, one notices that the nominal concentration of sinapaldehyde in wild-type poplar is much greater than the reported Michaelis constant of its CAD-catalyzed reduction to alcohol (see Supplementary Tables 1 and 2 for specific values), which is directly consistent with our model deduction.

Interestingly, the distributions of optimal parameter values reveal a linear relationship between $f_{CAD,ConifALD}$ and $f_{CALD5H,ConifALD}$ (Fig. 7). As discussed in more detail in Supplementary Method 3, this collinearity implies that the ratio between the corresponding fluxes remains unchanged over time and is thus equal to the steady-state value obtained from FBA. More importantly, a constant ratio between these two fluxes suggests that the S/G ratio might be insulated from any genetic modulation prior to the reactions involving coniferinaldehyde, provided that the synthesis of 5-hydroxyconiferyl alcohol is negligible. In fact, this is exactly what happens in transgenic experiments where 4CL (Fig. 4) or CCR (Fig. 5) is down-regulated. Even if the situation is not as expected in the CCoAOMT down-regulation experiment (Fig. 4), the observed S/G ratio is raised only by ~11% despite a 90% decrease in the CCoAOMT protein level.

With an ensemble of models that seems to be qualitatively adequate, we can now apply the IOM approach to minimize the S/G ratio of the monolignol biosynthetic pathway toward a higher yield of xylose. Normally, IOM can be implemented in many different ways. The most common scenario is that all enzymes (genes) involved in the pathway are accessible to manipulations, which unfortunately is not feasible with current biotechnological techniques in plants. [66]. Instead, we mimic the current state of the art (Fang Chen, personal communication) by allowing only one, two, or three enzyme activities to be altered between 5% and 5 times the basal levels. Furthermore, we enforce physiological constraints that are necessary for plant viability and that are discussed in Supplementary Method 6.

The optimization results (Table 3) indicate that by altering the activity levels of three enzymes in prescribed amounts, the S/G ratio predicted by the ensemble of models can be reduced from about 1.8 to about 1.11 – a significant decrease that far exceeds the natural variation observed in poplar [44]. Moreover, by modulating just

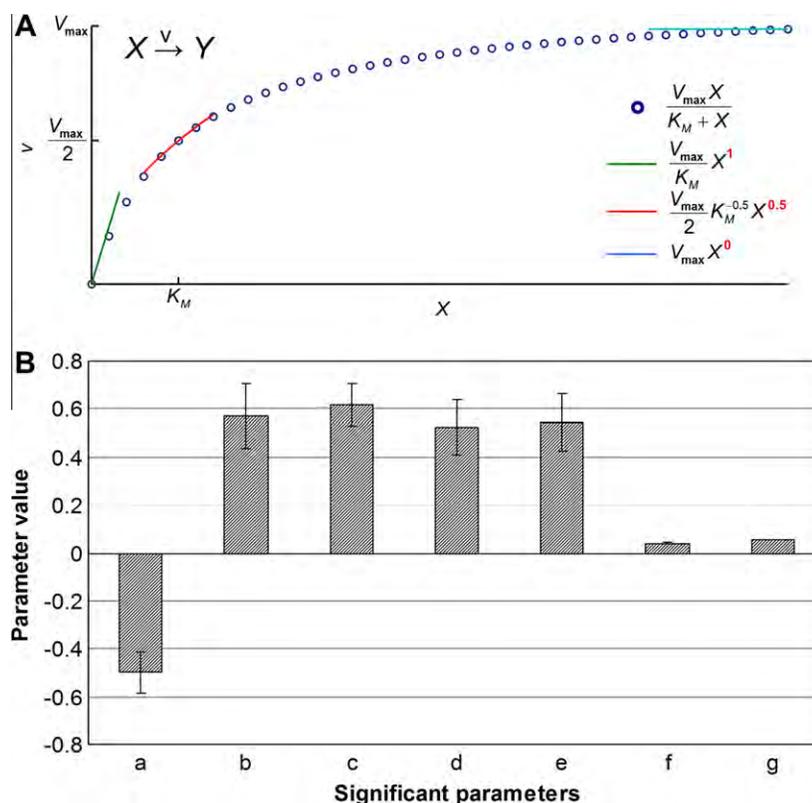


Fig. 6. Illustration of kinetic orders derived from a Michaelis–Menten function and distributions of values for seven significant parameters – the identities of which are described in Fig. 3C – within the ensemble of GMA models. (A) The kinetic order (red number) in each power-law representation of a Michaelis–Menten function is within the range of 0 and 1, with the specific value depending on the assumed *in vivo* concentration of X . If the reaction operates at a point where the concentration of X is much greater than K_M , the corresponding power-law representation has a kinetic order close to zero, implying that the reaction rate is almost unaffected by the concentration of X . (B) As in Fig. 4, the height of a vertical bar is proportional to the mean value of a significant parameter within the ensemble of models fitted by SA, with the error bar representing the 95% confidence interval for the mean. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

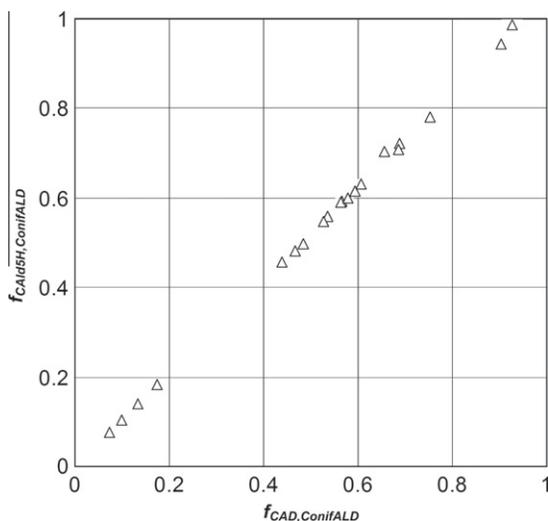


Fig. 7. Plot of $f_{\text{CAD5H,ConifALD}}$ against $f_{\text{CAD,ConifALD}}$. Each point represents the pair $(f_{\text{CAD,ConifALD}}, f_{\text{CAD5H,ConifALD}})$ found in one GMA model within the ensemble obtained through simulated annealing.

one enzyme (Cald5H), we can already achieve $\sim 60\%$ of the maximal reduction that is obtained when three enzymes are manipulated. In other words, the S/G ratio is predicted to decrease from 1.8 to about 1.39 if one down-regulates the enzyme activity of Cald5H by one quarter (the result is easily confirmed in PLAS). Overall, the optimized solutions require only a moderate degree of modulation of

the selected enzymes (from approximately 70% to 4.3 times the wild-type activity levels), which are well within the range of modern recombinant DNA techniques.

4. Discussion and conclusions

The application of mathematical modeling to studies of the monoglignol biosynthetic pathway, or of plant secondary metabolism in general, has not yet attracted much attention, especially when compared with central metabolism in microorganisms. One reason is that the *in vivo* concentrations of secondary metabolites are often low and difficult to measure, which makes quantitative modeling difficult.

In this work, we used diverse types of data to pursue a two-step model analysis of the monoglignol pathway, using both Flux Balance Analysis (FBA) and Biochemical Systems Theory (BST). These two approaches had so far not been combined in the construction of a dynamic model. Thus, we first constructed an initial, coarse FBA model and used it in a second phase as a constraint for developing fully parameterized nonlinear BST models. The result of this dual procedure was an ensemble of models that yield interesting qualitative insights into the topological and regulatory properties of monoglignol biosynthesis. These models also lead to simulation results and predictions that are quantitatively consistent with experimental measurements that were either used for model training or validation. This concordance is quite striking, because the data and information supporting the models are rather scarce and involve a number of assumptions. Two reasons seem to be responsible for the good performance of the model in predicting

the outcomes of validation experiments. The first is the proven robustness of BST models, which is manifest in low model sensitivity with respect to most parameters, as long as the connectivity and regulatory structure of a system is adequately captured by the model equations. The second reason is our strategic and severe model reduction, which effectively eliminated many parameters which we had proven to be relatively inconsequential.

Because we used all available metabolite concentrations and S/G ratios in transgenic experiments, either to estimate unknown parameters or to validate our models, it is presently not feasible to try improving the model further with purely computational means. To construct a “crisper” mathematical model in the future, specific data of the following types will be very helpful. At the metabolic level, intracellular metabolite concentrations, *in vitro* assays of individual enzymes, and perhaps intracellular flux measurements from dynamic labeling experiments [67] are in dire need. As demonstrated in our parameter estimation approach, these data should ideally be accompanied by measurements of lignin monomers from transgenic plants with various genetic modulations of monolignol biosynthesis.

Another source of relevant information will come from gene expression data and specifically from microarray analyses, which have already revealed distinct transcriptional regulation patterns in genes encoding lignin biosynthetic enzymes at different developmental stages [68]. At present, the growth periods in different transgenic experiments span from several months to years, but it is implicitly assumed that enzyme activities are more or less constant. Future experiments and models should account for (slowly) changing levels of enzyme activities over the course of xylem formation during primary and secondary growth. Furthermore, since most reactions within the pathway are catalyzed by several isozymes, changes in gene expression should be confirmed with measurements of changes in enzyme activities. As a first approximation, the number of mRNA copies for each corresponding gene may be an indication of enzyme activity, but direct measurements would eliminate uncertainties associated with different splice variants and posttranslational modifications. Experiments and models should also focus on the dynamics of transcription factors, such as MYB and LIM, that have been found to coordinate the regulation of the expression of genes encoding lignin biosynthetic enzymes [69,70].

The proposed ensemble of models is clearly preliminary. Nevertheless, the models appear to be robust to modest differences in parameter values, are qualitatively consistent with five training experiments, and are even capable of semi-quantitatively reproducing the results of two validation experiments that had not been used for model construction. These initial successes are grounds for cautious optimism that the model might serve as a basis from which future developments may be launched. As an illustration, we demonstrated one of its potential applications in genetic engineering, namely the optimization of the pathway toward a reduced S/G ratio and a higher yield of xylose. The results of this optimization seem to be reasonable in a sense that all proposed changes in enzyme activities are modest and therefore implementable. The estimated improvements in the optimized system are actually very conservative compared with the 75% decrease in the S/G ratio observed in the COMT down-regulation experiment (Table 2). The reason for this discrepancy is that we imposed much more stringent bounds on metabolites than what is observed in the COMT down-regulation experiment. While wider bounds are clearly implementable in optimizations with the computational model and would result in much stronger reductions in the S/G ratio, large metabolite variations *in vivo* might lead to toxicity or reduced viability. Two explanations are possible for the observed 75% decrease in the S/G ratio. First, evidence indicates that metabolites that might be expected to accumulate in the cytoplasm are instead

being transported to the cell wall and incorporated into lignin by so far unknown mechanisms [31], thereby precluding toxicity. Second, the observed variation in the S/G ratio may result from a change in the subcellular structure of pathway enzymes – or alleged “metabolic channeling” [71] – that is currently outside the scope of our GMA models. Taken together, the observed physiological response seems to suggest that our optimization settings might be overly cautious and that the S/G ratio could be reduced even further than predicted.

As new data are being generated in the emerging field of plant systems biology, the next goal will be to integrate a wider variety of “omics” data from different organizational levels into the construction of multi-scale models that will be capable of predicting the physiological consequences of hypothetical transgenic experiments. Models of this capability will be particularly helpful as the corresponding experiments in actual trees are slow and laborious. The need to test model predictions, as well as proposed genetic engineering strategies, will not abate. However, once a model is sufficiently reliable, it may be able to screen out experiments that are unlikely to lead to improved outcomes.

5. Authors' contributions

YL participated in the design of the study, performed the computational and statistical analysis, and drafted the manuscript. EO conceived the study, participated in its design, and edited the manuscript. Both authors read and approved the final manuscript.

Acknowledgements

The authors thank Fang Chen, Wout Boerjan, and Richard Dixon for sharing unpublished technical details. This work was supported by the BioEnergy Science Center (BESC), an US Department of Energy Bioenergy Research Center supported by the Office of Biological and Environmental Research in the DOE Office of Science. BESC is not responsible for the content of this article.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at [doi:10.1016/j.mbs.2010.08.009](https://doi.org/10.1016/j.mbs.2010.08.009).

References

- [1] A.J. Ragauskas, C.K. Williams, B.H. Davison, G. Britovsek, J. Cairney, C.A. Eckert, W.J. Frederick Jr., J.P. Hallett, D.J. Leak, C.L. Liotta, J.R. Mielenz, R. Murphy, R. Templer, T. Tschaplinski, The path forward for biofuels and biomaterials, *Science* 311 (2006) 484.
- [2] C.N. Hamelinck, G. Hooijdonk, A.P.C. Faaij, Ethanol from lignocellulosic biomass: techno-economic performance in short-, middle-, and long-term, *Biomass and Bioenergy* 28 (2005) 384.
- [3] M. Sticklen, Plant genetic engineering to improve biomass characteristics for biofuels, *Current Opinion in Biotechnology* 17 (2006) 315.
- [4] F. Chen, R.A. Dixon, Lignin modification improves fermentable sugar yields for biofuel production, *Nature Biotechnology* 25 (2007) 759.
- [5] The Arabidopsis Genome Initiative, Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*, *Nature*, 408, 2000, pp. 796–815.
- [6] G.A. Tuskan, S. Difazio, S. Jansson, J. Bohlmann, I. Grigoriev, U. Hellsten, N. Putnam, S. Ralph, S. Rombauts, A. Salamov, The genome of black cottonwood, *Populus trichocarpa* (Torr. & Gray), *Science* 313 (2006) 1596.
- [7] C. Somerville, S. Somerville, Plant functional genomics, *Science* 285 (1999) 380.
- [8] J.C. Leplé, R. Dauwe, K. Morreel, V. Storme, C. Lapierre, B. Pollet, A. Naumann, K.Y. Kang, H. Kim, K. Ruel, Downregulation of cinnamoyl-coenzyme a reductase in poplar: multiple-level phenotyping reveals effects on cell wall polymer metabolism and structure, *The Plant Cell* 19 (2007) 3669.
- [9] K. Morreel, J. Ralph, F. Lu, G. Goeminne, R. Busson, P. Herdewijn, J.L. Goeman, J. Van der Eycken, W. Boerjan, E. Messens, Phenolic profiling of caffeic acid O-methyltransferase-deficient poplar reveals novel benzodioxane oligolignols, *Plant Physiology* 136 (2004) 4023.

- [10] F. Chen, A.L. Duran, J.W. Blount, L.W. Sumner, R.A. Dixon, Profiling phenolic metabolites in transgenic alfalfa modified in lignin biosynthesis, *Phytochemistry* 64 (2003) 1013.
- [11] I. Famili, J. Forster, J. Nielsen, B.Ø. Palsson, Saccharomyces cerevisiae phenotypes can be predicted by using constraint-based analysis of a genome-scale reconstructed metabolic network, *Proceedings of the National Academy of Sciences* 100 (2003) 13134.
- [12] J. Förster, I. Famili, P. Fu, B.Ø. Palsson, J. Nielsen, Genome-scale reconstruction of the Saccharomyces cerevisiae metabolic network, *Genome Research* 13 (2003) 244.
- [13] A. Varma, B.Ø. Palsson, Metabolic flux balancing: basic concepts, scientific and practical use, *Biotechnology* 12 (1994) 994.
- [14] K.J. Kauffman, P. Prakash, J.S. Edwards, Advances in flux balance analysis, *Current Opinion in Biotechnology* 14 (2003) 491.
- [15] G.R. Gavalas, *Nonlinear Differential Equations of Chemically Reacting Systems*, Springer-Verlag, Berlin, 1968.
- [16] R. Heinrich, S. Schuster, *The Regulation of Cellular Systems*, Chapman & Hall, New York, 1996.
- [17] G.N. Stephanopoulos, A.A. Aristidou, J. Nielsen, *Metabolic Engineering: Principles and Methodologies*, Academic Press, San Diego, CA, 1998.
- [18] N. Jamshidi, B.Ø. Palsson, Formulating genome-scale kinetic models in the post-genome era, *Molecular Systems Biology* 4 (2008) 171.
- [19] A.R. Schulz, *Enzyme Kinetics: From Diastase to Multi-Enzyme Systems*, Cambridge University Press, 1994.
- [20] N.V. Torres, E.O. Voit, *Pathway Analysis and Optimization in Metabolic Engineering*, Cambridge University Press, 2002.
- [21] E.O. Voit, *Computational Analysis of Biochemical Systems: A Practical Guide for Biochemists and Molecular Biologists*, Cambridge University Press, 2000.
- [22] M.A. Savageau, Biochemical systems analysis. I. Some mathematical properties of the rate law for the component enzymatic reactions, *Journal of Theoretical Biology* 25 (1969) 365.
- [23] M.A. Savageau, Biochemical systems analysis. II. The steady-state solutions for an n -pool system using a power-law approximation, *Journal of Theoretical Biology* 25 (1969) 370.
- [24] M.A. Savageau, *Biochemical Systems Analysis: A Study of Function and Design in Molecular Biology*, Addison Wesley Publishing Company, Reading, MA, 1976.
- [25] V. Hatzimanikatis, J.E. Bailey, MCA has more to say, *Journal of Theoretical Biology* 182 (1996) 233.
- [26] D. Visser, J.J. Heijnen, Dynamic simulation and metabolic re-design of a branched pathway using linlog kinetics, *Metabolic Engineering* 5 (2003) 164.
- [27] H. Kascser, J.A. Burns, The control of flux, in: *Symposia of the Society for Experimental Biology*, 1973, pp. 65–104.
- [28] R. Heinrich, T.A. Rapoport, A linear steady-state treatment of enzymatic chains. Critique of the crossover theorem and a general procedure to identify interaction sites with an effector, *European Journal of Biochemistry* 42 (1974) 97.
- [29] D. Fell, *Understanding the Control of Metabolism*, Portland Press, London, 1997.
- [30] K. Smallbone, E. Simeonidis, D.S. Broomhead, D.B. Kell, Something from nothing-bridging the gap between constraint-based and kinetic modelling, *FEBS Journal* 274 (2007) 5576.
- [31] W. Boerjan, J. Ralph, M. Baucher, Lignin biosynthesis, *Annual Reviews in Plant Biology* 54 (2003) 519.
- [32] M.M. Campbell, R.R. Sederoff, Variation in lignin content and composition (mechanisms of control and implications for the genetic improvement of plants), *Plant Physiology* 110 (1996) 3.
- [33] M. Baucher, B. Chabbert, G. Pilate, J. Van Doorselaere, M.T. Tollier, M. Petit-Conil, D. Cornu, B. Monties, M. Van Montagu, D. Inze, Red xylem and higher lignin extractability by down-regulating a cinnamyl alcohol dehydrogenase in poplar, *Plant Physiology* 112 (1996) 1479.
- [35] K. Osakabe, C.C. Tsao, L. Li, J.L. Popko, T. Umezawa, D.T. Carraway, R.H. Smeltzer, C.P. Joshi, V.L. Chiang, Coniferyl aldehyde 5-hydroxylation and methylation direct syringyl lignin biosynthesis in angiosperms, *Proceedings of the National Academy of Sciences* 96 (1999) 8955.
- [37] L. Li, J.L. Popko, T. Umezawa, V.L. Chiang, 5-hydroxyconiferyl aldehyde modulates enzymatic methylation for syringyl monolignol formation, a new view of monolignol biosynthesis in angiosperms, *Journal of Biological Chemistry* 275 (2000) 6537.
- [39] S.A. Harding, J. Leshkevich, V.L. Chiang, C.-J. Tsai, Differential substrate inhibition couples kinetically distinct 4-coumarate: coenzyme A ligases with spatially distinct metabolic roles in quaking aspen, *Plant Physiology* 128 (2002) 428.
- [40] F. Chen, M.S. Srinivasa Reddy, S. Temple, L. Jackson, G. Shadle, R.A. Dixon, Multi-site genetic modulation of monolignol biosynthesis suggests new routes for formation of syringyl lignin and wall-bound ferulic acid in alfalfa (*Medicago sativa* L.), *The Plant Journal* 48 (2006) 113.
- [41] G. Pilate, E. Guiney, K. Holt, M. Petit-Conil, C. Lapiere, J. Leple, B. Pollet, I. Mila, E.A. Webster, H.G. Marstorp, Field and pulping performances of transgenic trees with altered lignification, *Nature Biotechnology* 20 (2002) 607.
- [42] L. Li, Y. Zhou, X. Cheng, J. Sun, J.M. Marita, J. Ralph, V.L. Chiang, Combinatorial modification of multiple lignin traits in trees through multigene cotransformation, *Proceedings of the National Academy of Sciences* 100 (2003) 4939.
- [43] H. Meyermans, K. Morreel, C. Lapiere, B. Pollet, A. De Bruyn, R. Busson, P. Herdewijn, B. Devreese, J. Van Beeumen, J.M. Marita, Modifications in lignin and accumulation of phenolic glucosides in poplar xylem upon down-regulation of caffeoyl-coenzyme A O-methyltransferase, an enzyme involved in lignin biosynthesis, *Journal of Biological Chemistry* 275 (2000) 36899.
- [44] B.H. Davison, S.R. Drescher, G.A. Tuskan, M.F. Davis, N.P. Nghiem, Variation of S/G ratio and lignin content in a Populus family influences the release of xylose by dilute acid hydrolysis, *Applied Biochemistry and Biotechnology* 130 (2006) 427.
- [45] R. Vanholme, K. Morreel, J. Ralph, W. Boerjan, Lignin engineering, *Current Opinion in Plant Biology* 11 (2008) 278.
- [46] C.C.S. Chapple, T. Vogt, B.E. Ellis, C.R. Somerville, An Arabidopsis mutant defective in the general phenylpropanoid pathway, *The Plant Cell* 4 (1992) 1413.
- [47] C. Chen, H. Meyermans, B. Burggraeve, R.M. De Rycke, K. Inoue, V. De Vleeschauwer, M. Steenackers, M.C. Van Montagu, G.J. Engler, W.A. Boerjan, Cell-specific and conditional expression of caffeoyl-coenzyme A-3-O-methyltransferase in poplar, *Plant Physiology* 123 (2000) 853.
- [48] N.D. Price, J.L. Reed, B.Ø. Palsson, Genome-scale models of microbial cells: evaluating the consequences of constraints, *Nature Reviews Microbiology* 2 (2004) 886.
- [49] R.L. Lindroth, S.Y. Hwang, Diversity, redundancy and multiplicity in chemical defense systems of aspen, in: J.T. Romeo, J.A. Saunders, P. Barbosa (Eds.), *Phytochemical Diversity and Redundancy in Ecological Interactions*, Plenum Press, New York, NY, USA, 1996, p. 25.
- [50] H.D. Coleman, J.-Y. Park, R. Nair, C.C.S. Chapple, S.D. Mansfield, RNAi-mediated suppression of *p*-coumaroyl-CoA 3'-hydroxylase in hybrid poplar impacts lignin deposition and soluble secondary metabolite metabolism, *Proceedings of the National Academy of Sciences* 105 (2008) 4501.
- [51] E.O. Voit, P.J. Sands, Modeling forest growth. I. Canonical approach, *Ecological Modeling* 86 (1996) 51.
- [52] E.O. Voit, P.J. Sands, Modeling forest growth. II. Biomass partitioning in Scots pine, *Ecological Modeling* 86 (1996) 73.
- [53] E.O. Voit, Dynamics of self-thinning plant stands, *Annals of Botany* 62 (1988) 67.
- [54] I.-C. Chou, E.O. Voit, Recent developments in parameter estimation and structure identification of biochemical and genomic systems, *Mathematical Biosciences* 219 (2009) 57.
- [55] M. Vilela, I.-C. Chou, S. Vinga, A.T.R. Vasconcelos, E.O. Voit, J.S. Almeida, Parameter optimization in S-system models, *BMC Systems Biology* 2 (2008) 35.
- [56] J.M. Murphy, D.M.H. Sexton, D.N. Barnett, G.S. Jones, M.J. Webb, M. Collins, D.A. Stainforth, Quantification of modelling uncertainties in a large ensemble of climate change simulations, *Nature* 430 (2004) 768.
- [57] D. Battogtokh, D.K. Asch, M.E. Case, J. Arnold, H.B. Schüttler, An ensemble method for identifying regulatory circuits with special reference to the qa gene cluster of *Neurospora crassa*, *Proceedings of the National Academy of Sciences* 99 (2002) 16904.
- [58] L. Kuepfer, M. Peter, U. Sauer, Ensemble modeling for analysis of cell signaling dynamics, *Nature Biotechnology* 25 (2007) 1001.
- [59] A. Ferreira, *Power Law Analysis and Simulation (PLAS)*. <<http://www.dqb.fc.ul.pt/docentes/aferreira/plas.html>>, 2000.
- [60] J.D. Bloom, M.M. Meyer, P. Meinhold, C.R. Otey, D. MacMillan, F.H. Arnold, Evolving strategies for enzyme engineering, *Current Opinion in Structural Biology* 15 (2005) 447.
- [61] N.V. Torres, E.O. Voit, C. Glez-Alcón, F. Rodriguez, An indirect optimization method for biochemical systems: description of method and application to the maximization of the rate of ethanol, glycerol, and carbohydrate production in *Saccharomyces cerevisiae*, *Biotechnology and Bioengineering* 55 (1997) 758.
- [62] E. Almaas, B. Kovacs, T. Vicsek, Z.N. Oltvai, A.L. Barabasi, Global organization of metabolic fluxes in the bacterium *Escherichia coli*, *Nature* 427 (2004) 839.
- [63] A.L. Barabasi, Z.N. Oltvai, Network biology: understanding the cell's functional organization, *Nature Reviews Genetics* 5 (2004) 101.
- [64] G. Goel, I.-C. Chou, E.O. Voit, Biological systems modeling and analysis: a biomolecular technique of the twenty-first century, *Journal of Biomolecular Techniques* 17 (2006) 252.
- [65] R. Alves, E. Herrero, A. Sorribas, Predictive reconstruction of the mitochondrial iron-sulfur cluster assembly metabolism: I. The role of the protein pair ferredoxin-ferredoxin reductase (Yah1-Arh1), *Proteins: Structure, Function, and Bioinformatics* 56 (2004) 354.
- [66] C. Halpin, Gene stacking in transgenic plants – the challenge for 21st century plant biotechnology, *Plant Biotechnology Journal* 3 (2005) 141.
- [67] R.G. Ratcliffe, Y. Shachar-Hill, Measuring multiple fluxes through plant metabolic networks, *The Plant Journal* 45 (2006) 490.
- [68] M. Hertzberg, H. Aspeborg, J. Schrader, A. Andersson, R. Erlandsson, K. Blomqvist, R. Bhalerao, M. Uhlén, T.T. Teeri, J. Lundeberg, A transcriptional roadmap to wood formation, *Proceedings of the National Academy of Sciences* 98 (2001) 14732.
- [69] B. Karpinska, M. Karlsson, M. Srivastava, A. Stenberg, J. Schrader, F. Sterky, R. Bhalerao, G. Wingsle, MYB transcription factors are differentially expressed and regulated during secondary vascular tissue development in hybrid aspen, *Plant Molecular Biology* 56 (2004) 255.
- [70] A. Kawaoka, P. Kaothien, K. Yoshida, S. Endo, K. Yamada, H. Ebinuma, Functional analysis of tobacco LIM protein Ntlm1 involved in lignin biosynthesis, *The Plant Journal* 22 (2000) 289.

- [71] K. Jørgensen, A.V. Rasmussen, M. Morant, A.H. Nielsen, N. Bjarnholt, M. Zagobelny, S. Bak, B.L. Møller, Metabolon formation and metabolic channeling in the biosynthesis of plant natural products, *Current Opinion in Plant Biology* 8 (2005) 280.
- [72] J.W. Blount, K.L. Korth, S.A. Masoud, S. Rasmussen, C. Lamb, R.A. Dixon, Altering expression of cinnamic acid 4-hydroxylase in transgenic plants provides evidence for a feedback loop at the entry point into the phenylpropanoid pathway, *Plant Physiology* 122 (2000) 107.
- [73] L. Li, X. Cheng, S. Lu, T. Nakatsubo, T. Umezawa, V.L. Chiang, Clarification of cinnamoyl co-enzyme A reductase catalysis in monogenic biosynthesis of aspen, *Plant & Cell Physiology* 46 (2005) 1073.